Measuring Behavior 2018

5th-8th June 2018

Manchester Metropolitan University

Conference Proceedings

11th International Conference on Methods and Techniques in Behavioral Research

Volume Editors

Robyn Grant

Manchester Metropolitan University; robyn.grant@mmu.ac.uk

Tom Allen

Manchester Metropolitan University; t.allen@mmu.ac.uk

Andrew Spink

Noldus Information Technology; andrew.spink@noldus.nl

Matthew Sullivan

Manchester Metropolitan University; m.sullivan@mmu.ac.uk

ISBN: 'Measuring Behavior 2018': 978-1-910029-39-8

(c) 2018 Measuring Behavior

Table of Contents

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Preface

Measuring Behavior 2018

Robyn A Grant, Andrew Spink, Matthew Sullivan

These proceedings contain the papers presented at Measuring Behavior 2018, the 11th International Conference on Methods and Techniques in Behavioral Research. The conference was organised by Manchester Metropolitan University, in collaboration with Noldus Information Technology. The conference was held during June 5th – 8th, 2018 in Manchester, UK.

Building on the format that has emerged from previous meetings, we hosted a fascinating program about a wide variety of methodological aspects of the behavioral sciences. We had scientific presentations scheduled into seven general oral sessions and fifteen symposia, which covered a topical spread from rodent to human behavior. We had fourteen demonstrations, in which academics and companies demonstrated their latest prototypes. The scientific program also contained three workshops, one tutorial and a number of scientific discussion sessions. We also had scientific tours of our facilities at Manchester Metropolitan Univeristy, and the nearby British Cycling Velodrome.

We hope this proceedings caters for many of your interests and we look forward to seeing and hearing more of your contributions.

Organisation

Measuring Behavior 2018 was hosted by Manchester Metropolitan University in collaboration with Noldus Information Technology

Conference Chairs

Robyn Grant, Manchester Metropolitan University

Matthew Sullivan, Manchester Metropolitan University

Andrew Spink, Noldus Information Technology

Organising Committee

Tom Allen, Manchester Metropolitan University

Moi Hoon Yap, Manchester Metropolitan University

Yael Benn, Manchester Metropolitan University

Kevin Tan, Manchester Metropolitan University

Gernot Reidel, University of Aberdeen

Natasja Bogers, Noldus Information Technology

Sponsors

Noldus

Manchester Metropolitan University

Los Kaos

isea INTERNATIONAL SPORTS ENGINEERING ASSOCIATION

COMPARATIVE EXERCISE PHYSIOLOGY

WAZA World Association of Zoos and Aquariums

HSBC UK NATIONAL CYCLING CENTRE HOME OF BRITISH CYCLING

NC3Rs National Centre for the Replacement Refinement & Reduction of Animals in Research

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Keynote Speakers

A Window on Welfare - Measuring the expressive qualities of behavior

Prof Françoise Wemelsfelder

Animal and Veterinary Sciences Group, Scotland's Rural College, Edinburgh, UK.

Dynamic models of animal sentience and emotion are gaining momentum, making possible an integrated approach to welfare assessment in which emotion is an expressive aspect of, rather than a separate state from, behavior. Qualitative Behaviour Assessment (QBA) is a 'whole animal' methodology designed to characterise and quantify expressive qualities of animal demeanour, using descriptors such as relaxed, fearful, agitated or content. Such terms are frequently applied in studies of animal temperament and personality, and QBA extends this to include the assessment of animal experience. QBA also builds on qualitative assessment methods used in research with non-verbal human beings. A key factor in developing QBA has been the use of Free Choice Profiling (FCP), a method originally designed for use in consumer and food science. FCP invites groups of observers to generate and then quantify their own descriptive terms and relies on a multivariate statistical technique called Generalised Procrustes Analysis (GPA) to identify common perceived patterns of animal expression. Years of research with a range of animal species support the scientific validity of this approach and suggest it has significant potential, particularly when combined with other measures, to help judge an animal's overall welfare state. QBA has so far not been applied to studies of human welfare; it would be interesting to consider whether and how this might be relevant. In this presentation, I will discuss various key methodological aspects of QBA, their strengths, weaknesses, and potential applications, and illustrate these with examples from QBA research.

Measuring functional morphology and ecological behaviour during the evolution of birds

Karl Bates

Liverpool Univervisty, Liverpool, UK

Birds are one of the most taxonomically and ecologically diverse groups of vertebrates in modern ecosystems. Extant birds also possess a number of highly specialized and in some case unique morphological characteristics, including feathered bodies, wings, a hypermobile neck, and a highly pneumatised skeleton associated with a system of rigid air sacs. These novel morphological features are intrinsically linked to specialized functional traits of birds, notably powered flight, the use of unusual "crouched" bipedal postures during walking, and a unique system of lung ventilation. When and how these morpho-functional specializations evolved, and the extent to which they are interlinked, represent major questions in palaeobiology. In this talk, I will review recent research into the 'evolutionary biomechanics' of bird evolution. This research area is facilitated by a rich fossil record, which details the gradual step-wise acquisition of derived avian morphologies (and by inference mechanics and physiologies) during the Mesozoic Era. Recent work using computer models has demonstrated how temporal and phylogenetic changes in body shape (i.e. mass distribution) and muscle leverage effectively trace the gradual shift from the erect bipedal postures used by basal dinosaurs to the more unusual crouched ('zig-zagged') limb posture seen in extant birds. Quantitative analysis of bone shape reveals unique patterns of morphological regionalization within the necks of living birds, which correlate with variation in the degree of mobility within the neck. As with body shape and limb muscle evolution, evolutionary changes in neck form-function appear to intensify in extinct theropod dinosaurs close to the evolution and diversification of animals with powered flight capability. This strongly suggests that whole-scale changes in morpho-functional anatomy were crucial to the evolution of powered flight and that many novel avian features may well be highly interlinked. Expanding and refining our understanding of this important ecomorphological transition is likely to be challenging, but immediate leaps forward will be realized by further work on extant taxa and the continued development of computer simulation approaches for reconstructing biomechanical performance in extinct animals.

The BioMimetic Approach to Studying the Control and Coordination of Behavior in Robots, Octopuses and other Marine Invertebrates

Frank Grasso

BioMimetic and Cognitive Robotics Laboratory, Brooklyn College, City University of New York.

Soft-bodied animals such as cephalopods (Octopuses, Cuttlefishes, Squids and Nautiluses) are of considerable interests to biomechanicians, neuroscientists and roboticists because our understanding of the motor control in such systems is just in its infancy and offers the possibility of new technologies and understandings of brain function. Their soft bodies mean that they lack the endo-skeletons of vertebrates (e.g., birds, reptiles, mammals) or the exoskeletons of arthropods (e.g., crustaceans like crabs and lobsters or insects) yet cephalopods animals, particularly octopuses are capable of both fine dexterous manipulation and forceful manipulation with the same appendages. Therefore, for roboticists, octopuses provide existence proofs that dexterous and forceful object manipulation are possible in systems lacking hard parts. For neuroscientists, interests lie in uncovering strategies that octopuses must use to make control of hyper-redundant systems manageable. One challenge of studying such high degree of freedom systems involves the quantification of the kinematics of motor behavior. In this talk, I will discuss video methods we have developed for the quantification of fine and forceful manipulation by octopuses. I will discuss the biomimetic approach to understanding control and coordination of behavior with parallel studies in robots.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Meet a polar bear

Kim Kaos

LosKaos

Humans interact with bio-inspired, animal-like entities, in the form of robots, puppets and computer games. We respond to these animals socially and interact with them in a similar way to how we would react to a real animal. In this keynote experience, we will explore this in more detail. You will be able to meet and interact with a “real” animal. Kim will be working with a special (and secret – sorry!) set of mechanisms that allow one person to fully animate a large creature, he will be bringing his polar bear, Bjorn to the conference. Please come along - how often do you get a chance pat a polar bear? After meeting the polar bear, and taking a short break, Kim will talk about his performance experiences, and the psychological and animal behavioural techniques he uses to enhance the live experience of human puppet interactions, and the cultural differences he has observed performing around the world.

Classification of Para swimmers with physical impairments - what should we measure and how?

Carl Payton

Department of Sports Science, Manchester Metropolitan University, Crewe, UK

World Para Sport is the global governing body of the Paralympic movement and Paralympic Games. One of the biggest challenges it faces is to provide a fair classification system for each of the Paralympic sports. Classification is essential to the very existence of sports for athletes with a disability. An effective classification system should provide athletes with a disability with an equitable starting point for competition by minimising the impact that their impairment has on the outcome of the event. The process of classification involves grouping athletes into different classes, using sport-specific assessments designed to evaluate the impact of their impairment on performance.

World Para Sport has decided that the current system used to classify physically impaired swimmers for international competition needs revising. In conjunction with UK Sport, World Para Sport is funding an international research project that will generate a scientific evidence base from which a new system for classifying Para swimmers can be developed. This project will provide World Para Sport with clear recommendations on what biomechanical measures should be taken on swimmers during the classification process and how these should be obtained. This presentation will focus on the measurement techniques employed in the research and on those that might form part of the new classification system.

Workshops

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

How can we turn "Behavioural Tests" into "Behavioural Bioassays"?

R.E. Brown[1]

1. Professor, Department of Psychology and Neuroscience, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4R2

Behavioural assays are essential in neuroscience, but many behavioural studies involve "behavioural testing" rather than "behavioural bioassays". What is the difference? A behavioural test collects data on behaviour in an apparatus such as the open field or elevated plus maze. The apparatus, procedure and methods of data collection are not standardized and the inter-laboratory reliability is often poor. When the results are interpreted, their validity is largely unknown. A behavioural bioassay, on the other hand, is a standardized test procedure which focuses on a single question, such as "What is the visual acuity of this animal?" The purpose of this workshop is to examine and discuss the different behavioural tests used in neuroscience and determine which meet the criteria of "behavioural bioassays" and which do not and to make recommendations for the use of reliable and valid behavioural tests to be used as behavioural bioassays. My plan is to start the workshop with a short lecture on the nature of the problem given by myself (Richard Brown) and then have members of the audience give ad-hoc presentations on their work. I see no reason to have a fixed number of speakers. At the end of the workshop, we should be able to produce a short paper on the issues and give our recommendations.

This workshop would be an extension of my 2016 presentation entitled "The importance of “Behavioural Bioassays” in neuroscience".

References

1. Schellinck, H. M., Cyr. D.P. and Brown, R.E. (2010). How many ways can mouse behavioral experiments go wrong? Confounding variables in mouse models of neurodegenerative diseases and how to control them. Advances in the Study of Behavior, 2010, 41, 255-366
2. Gunn, R.K., Keenan, M-E., and Brown, R.E. Analysis of sensory, motor and cognitive functions of the Coloboma (C3H/HeSnJ-cm) mutant mouse. Genes, Brain and Behavior, 2011, 10, 579-588.
3. O’Leary, T.P. and Brown, R.E. The effects of apparatus design and test procedure on learning and memory performance of C57BL/6J mice on the Barnes maze. Journal of Neuroscience Methods, 2012, 203 (2), 315-324.
4. O’Leary, T.P. and Brown, R.E. Optimization of apparatus design and behavioural measures for the assessment of visuo-spatial learning and memory of mice on the Barnes maze. Learning and Memory, 2013, 20, 85-96.
5. O'Leary, T.P. , Gunn, R.K., and Brown, R.E. What are we measuring when we test strain differences in anxiety in mice? Behavior Genetics, 2013, 43, 34-50.
6. Brown, RE and Bolivar, S. The importance of behavioural bioassays in neuroscience. Journal of Neuroscience Methods, 2017. http://dx.doi.org/10.1016/j.jneumeth.2017.05.022

Can Research Software Engineers help measuring behaviour?

R. Silva[1, 2] and C. Jay[3]

[1] eScience Lab, School of Computer Science, University of Manchester, Manchester, United Kingdom. raniere.silva@manchester.ac.uk
[2] Software Sustainability Institute, Edinburgh, United Kingdom. raniere.silva@software.ac.uk
[3] School of Computer Science, University of Manchester, Manchester, United Kingdom. Caroline.Jay@manchester.ac.uk

Research Software Engineer (RSE) is a term that started to be used in the United Kingdom around 2012 to describe people in a variety of roles who understand and care about both good software and good research. It is an inclusive definition that covers a wide spectrum of people, from a researcher who is primarily focused on getting results for papers but does a lot of programming, to a software engineer who happens to work for a research organisation. Somewhere in the middle lies the RSE who may actually have that as their job title and/or might work for one of the fast-growing Research Software Groups.

The purpose of this discussion is to: examine and discuss the contribution of research software engineers to behavioural science studies; determine the impact of research software engineers in a behavioural researcher group; and to make recommendations for the inclusion of research software engineering positions in grant proposals.

We will start the discussion with a short lecture on the growth of the United Kingdom Research Software Engineer Association and some examples of work done by RSEs that might be interesting to behavioural researchers, and follow this with an interactive discussion on the topic. We will use the information that emerges during the discussion to produce a short paper with our recommendations.

The discussion will cover some or all of the following questions:

1. What research software did you use in your last paper? Make a list.
2. How many of the software applications on your list are open source?
3. How many of the software applications on your list have some documentation?
4. How many of the software applications on your list have any kind of support forum? This can be a email address, a mailing list, web forum, a Q&A website or something else that you can turn for support.
5. Does the software you use do exactly what you want, or could it be better?
6. Do you, or members of your group, develop your own software? This might include anything from scripts for analysing data (e.g. R or Python scripts) to websites.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Measuring Performance at the Great Britain Cycling Team

P.R.Barratt[1]

1 Performance Support Team, Great Britain Cycling Team, UK

At the Great Britain Cycling Team, laboratory-based experimental methods and field-based observational methods are combined to investigate how changes in bicycle setup parameters influence cycling power output. This workshop will demonstrate and showcase the facilities at the National Cycling Center, as part of the conference's scientific tour program. It will then be followed by a workshop discussing how we might measure performance in cycling, with suggestions from the audience.

Tutorials

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

The challenges and opportunities running a behavioral core facility offers

Lior Bikovski

Myers Neuro-Behavioral Core Facility Israel liorbiko@gmail.com

An increasing number of universities, research institutes and companies invest in behavioral core facilities in order to concentrate specific knowledge, assure a high standard of work and provide these specific services to the largest group of people possible. However, running a core facility on a day-to-day basis comes with its own set of challenges and unique opportunities. The quality of research done at these facilities relies on the ability of the facility managers to engage specific challenges, and to bridge the gap between communal and specific needs, as well as the need to identify novel trends and foresee future directions that may be relevant to the core users.

Symposia

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Symposia: Fish as model organisms in behavioural research

The study of fish behaviour has long traditions and goes back to Tinbergen´s studies of stickleback behaviour, and the origin of ethology as a scientific discipline. More than half of all vertebrates are teleost fish, making teleosts the without competition largest vertebrate group. It is also a group displaying enormous diversity in behavioural and physiological adaptations, giving unsurpassed possibilities to find model species to address any biological question. During the last decades zebrafish (*Danio rerio*) has become an increasingly important model species for behavioural research and neuroscience. The symposium will congregate researchers tackling different neurobehavioral questions using fish as models.

Erika Roman and Svante Winberg, Uppsala University

Zebrafish Dravet syndrome models for antiepileptic drug candidates discovery and behavior investigation

Jacmin M.[1,2], Crawford A.D.[1,2]

[1]Theracule S.à.r.l., Esch-sur-Alzette, Luxembourg, maxime.jacmin.001@student.uni.lu; [2]University of Luxembourg, LCSB, Esch-sur-Alzette, Luxembourg, alexander.crawford@uni.lu

Introduction

Dravet syndrome (DS), also known as severe myoclonic epilepsy of infancy (SMEI), is a genetic epileptic encephalopathy with childhood onset [1]. Incidence has been estimated at 1/30,000 [2]. Epilepsy in DS is notably pharmacoresistant, with most children experiencing recurrent seizures for the rest of their lives even after the introduction in Europe of stiripentol, currently the only drug approved for this indication. DS is caused by reduced inhibitory neuron activity, with 80% of Dravet patients having *de novo* mutations in the *SCN1A* gene, which encodes the Nav1.1 sodium channel alpha subunit and is necessary for these neurons [1].

While early-life seizures are perhaps the most striking feature of DS, the most disabling consequences of the disease are often the devastating effects associated with cognitive and behavioral impairment [3]. Indeed, most children with DS develop moderate to severe cognitive delay and some features of autism in the first years of life [4]. When reaching adulthood, cognitive impairment becomes the main feature of their disease as seizures diminish in number, leaving these patients dependent on caregivers for the rest of their lives. Despite this relevance, there is currently no drug in development to treat cognitive impairment associated with epilepsy. Therefore, therapeutics able to treat cognitive impairment in DS have the potential to greatly improve the quality of life of not only DS patients but also those suffering from other forms of severe epilepsy.

Over the last decade, animal models of DS have been established in both rodents and zebrafish. Behavioral analysis of mouse models of DS reveal these mice to also have autistic features and severe cognitive impairments [5,6], including hyperactivity, stereotyped behaviors, social interaction deficits and impaired spatial memory [5,6]. These mouse experiments are very valuable for translational research but also very expensive and time-consuming. Zebrafish is then a more effective animal model to use since they are now a well-established experimental model for epilepsy [7,8,9], and have several advantages as an animal model – including reduced drug dose, greater numbers of animals, reduced costs and increased ease of manipulation – that are useful for screening potentially drug-like compounds, prior to rodent-based assays [10]. Zebrafish larvae with loss-of-function mutations in *scn1a* (*didy* mutants) exhibit hyperactivity, convulsive behavior and spontaneous electrographic seizures, shortened lifespan and a pharmacological profile similar to the one seen in children with DS [9]. However, no behaviour investigation has been performed so far at early-life stage in zebrafish DS models.

The aims of this study are

- To investigate the potential antiepileptic activity of several drug candidates in zebrafish DS models at the larval stage.
- To develop a behavior-based platform in zebrafish larvae in order to first evaluate the similarities between zebrafish and human cognition and second, to establish a high-throughput analysis of possible treatments for cognitive disorder in the context of Dravet syndrome.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Material and methods

Larval locomotor behavior

Zebrafish larvae from either scn1Lab mutants or control groups were tracked using the ViewPoint Zebralab System for Zebrafish™ (Version 3.22, ViewPoint, France) at 5 days post-fertilization (dpf). The system consists of an infrared light source, a digital videocamera to capture larval movements within a defined time period (120 minutes in our experimental set-up) and the software to analyse larval locomotor activity. Zebrafish larvae were placed in a 96-well plate (tissue culture plate, flat bottom, FALCON®, USA); one larva per well. Each well containing a fish was filled with 100 µl of different concentrations of test compounds in embryo medium (1% DMSO) or vehicle (embryo medium only). The larvae thus treated were incubated at 28.5°C in dark and quiet conditions for 1 hour prior to tracking. The movement pattern of the exposed zebrafish larvae was videotracked and assessed for 2 hours. Videotracking of larval movements was started 5 minutes after positioning of the plate in the tracker. The tracker software measured periods of 5 minutes of larvae movement. Results were registered as the average value of the total time of larvae movement for 120 minutes.

Field potential recordings

Open-field recordings were obtained from zebrafish larval forebrain at 5 dpf at 24°C following an incubation with drug candidates or vehicle. A glass electrode, connected to a high-impedance amplifier, was filled with artificial cerebrospinal fluid (124 mM NaCl, 2 mM KCl, 2 mM $MgSO_4$, 2 mM $CaCl_2$, 1.25 mM KH_2PO_4, 26 mM $NaHCO_3$ and 10 mM glucose). A larva was then embedded in 2% low-melting-point agarose and the glass electrode placed into the forebrain of the larvae. Single recordings were performed for ten minutes. Spontaneous epileptiform events were taken into account when the amplitude exceeded three times the background noise. The analysis of spikes was carried out using Clampfit 10.2 software (Molecular Devices Corporation, USA).

Anxiety analysis

Freely swimming 5 dpf scn1Lab mutants and control larvae were pre-incubated in 100 µl of different concentrations of antiepileptic drug candidates or vehicle for 1 hour in individual wells of a 96-well plate at 28°C and then transferred to a 6-well plate filled with embryo medium. The plate was then placed in a videotracking apparatus (Daniovison, Noldus, The Netherlands) and immediately tracked for 5 minutes. Moreover, each well of the 6-well plate was virtually divided in 2 arenas, the centre and the periphery, using Ethovision XT software (Version 11.5, Noldus, The Netherlands) in order to determine (1) the total locomotor activity and, (2) the time spent by the larvae in each arena. It must be noted that the protocol used here was adapted from Grone et al., 2017 [11].

Habituation assay

Freely swimming 5 dpf scn1Lab mutants and control larvae were pre-incubated in 100 µl of different concentrations of antiepileptic drug candidates or vehicle for 1 hour in individual wells of a 24-well plate at 28°C and then placed in a tracking apparatus (Daniovision, Noldus, The Netherlands). After a 10-minutes chamber habituation, larvae were tracked for locomotor behaviour under a repeated sound/vibration stimuli protocol adapted from Wolman et al, 2011 [12]. In a few words, fish were pre-tracked for 30 minutes and then underwent the protocol described in Figure 1.

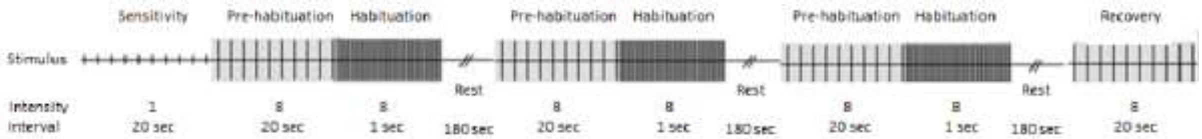


Figure 1. Protocol used for the habituation assessment in zebrafish. Each vertical bar represents a stimulus, interval represents the time between 2 stimuli.

Results

Larval locomotor behaviour

As previously shown by several researches in the field, zebrafish larvae DS models exhibit a strong hyperactivity, most likely due to seizures occurring during the video-recordings [9,13]. We could reproduce these results with our zebrafish line and therefore validate it (see Figure 2A).

Moreover, we made advantage of this feature by assessing the possible anticonvulsant effect of several antiepileptic drug candidates and could identify 6 compounds displaying this outcome (see Figure 2B).

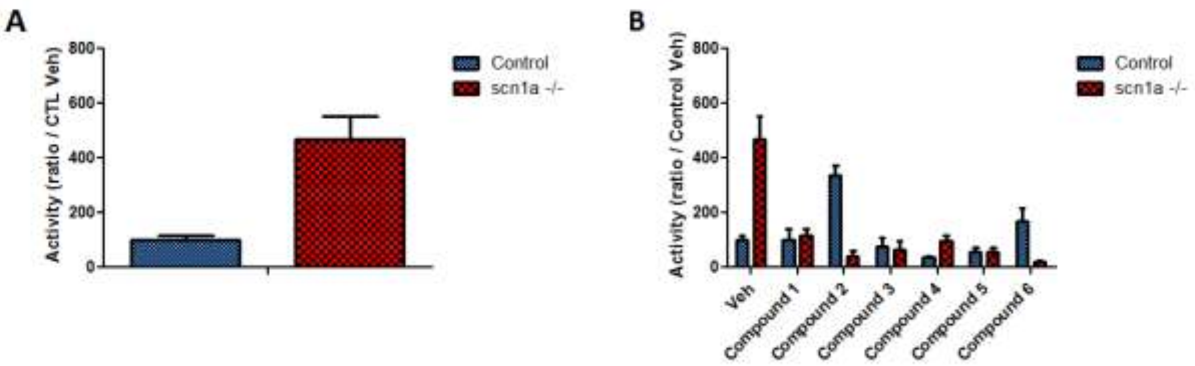


Figure 2. Spontaneous locomotor activity of 5 dpf larvae from control group (Control) or Dravet syndrome group (scn1a -/-) without drugs in panel A and after incubation with several antiepileptic drug candidates in panel B. The activity depicted in the graphs are a ratio to the control group incubated with vehicle (Veh).

Field potential recordings

We validated the results obtained in the locomotor behaviour assay by performing field potential recordings in order to investigate brain activity (and therefore seizures) in these larvae. We could observe a strong seizure phenotype in the mutant larvae compared to controls (see Figure 3A) and a drastic reduction in this brain activity when incubated with our compounds (see Figure 3B as representative example).

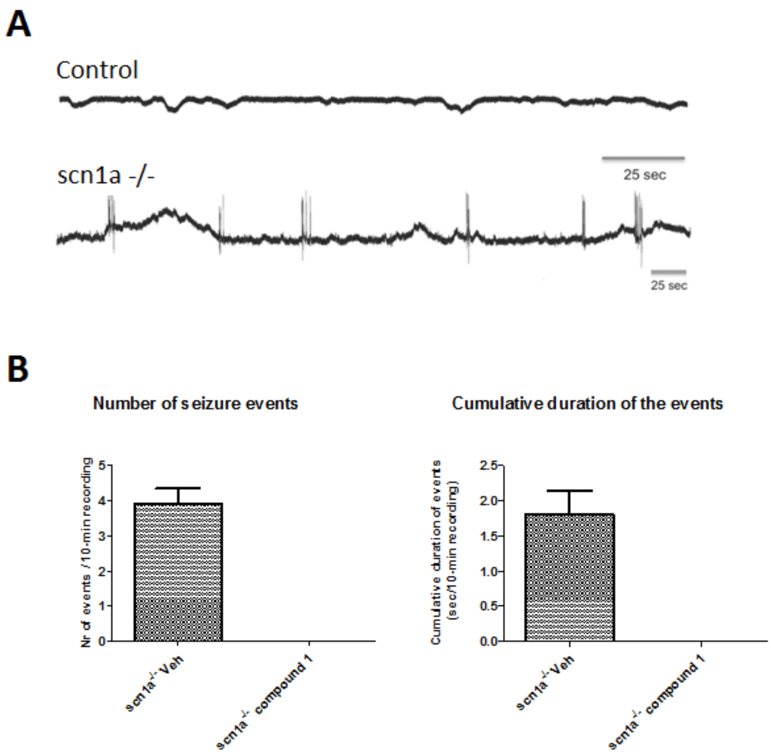


Figure 3. Brain activity analysis using local field potential recordings. (A) Comparison between control larvae and mutant larvae at 5dpf. Each spike represents a seizure. (B) Representation of the number and the cumulative duration of the seizures occurring during the recording after incubation with vehicle (Veh) or compound 1.

Anxiety analysis

Besides the discovery of new molecules for the treatment of Dravet syndrome, our main interest is the establishment of a new platform for behaviour analysis in larvae zebrafish in order to compare it to the human patient phenotype. We therefore investigated the anxiety profile of our mutant larvae since the patients exhibit more stress than normal. We could determine that our Dravet syndrome zebrafish model is significantly more anxious than controls as they were moving less and were staying closer to the wall when placed in a new environment (see Figure 4A and 4B). Unfortunately, so far, we were not able to reverse the phenotype when incubating the fish with our candidate drugs (data not shown).

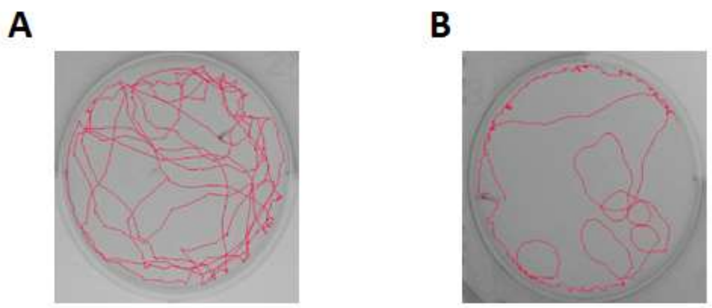


Figure 4. Virtual representation of the locomotor activity of control fish (A) or DS mutant fish (B) when placed in a new environment.

Habituation assay

Finally, the second aspect of Dravet syndrome comorbidities that we wanted to evaluate in zebrafish is the impaired memory formation that patients exhibit. For that matter, we exposed our fish to a continuous stimuli protocol described in the Material and Methods section and could determine that our DS zebrafish model poorly habituate to those stimuli compared to controls (see Figure 5A). Even more, their profile resembles the one of control fish incubated with a known amnesia-inducer compound, MK-801 (see Figure 5A). Interestingly, we were able to reverse this phenotype using one of our compounds (see Figure 5B).

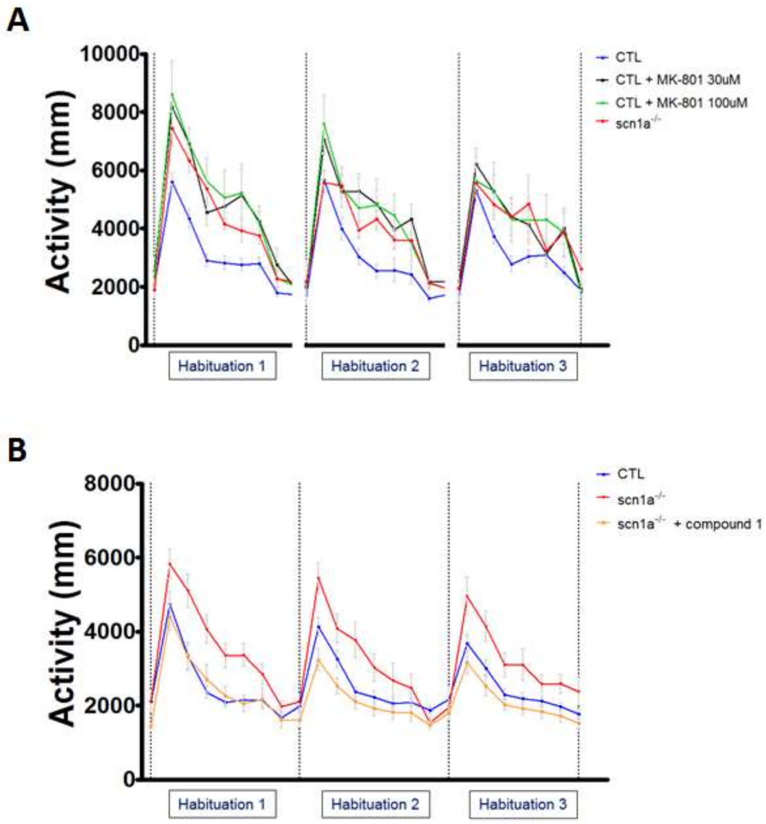


Figure 5. Locomotor activity analysis during the habituation phases of the protocol described above. In (A), we compared the phenotype of control without or with amnesia-inducer drug incubation and in (B), we added an incubation with one of our candidate drug.

Discussion

The results of the present study show that our Dravet syndrome zebrafish mutant line phenocopies human patients and mouse models in terms of locomotor hyperactivity, seizure occurrence and behavioural aspects as anxiety and memory formation. Importantly, the use of zebrafish larvae for these assays had not been extensively tested in the past.

Furthermore, we were able to take advantage of this model for high-throughput screening of anticonvulsant drug candidates in a small amount of time with some possible hits, which would have been difficult to perform in more evolved models as rodents for example.

Ethical statement

The study was conducted in accordance to national and international guidelines (directive 2007/526/EC of the European Commission) for the protection of animal welfare. The Ethical Committee of the University of Luxembourg approved all experimental protocols and animals used in this research.

References

1. Duncan, JS, Sander, JW, Sisodiya, SM, Walker, MC (2006), Adult epilepsy. *Lancet*, 367(9516):1087-100.
2. Hurst, DL (1990), Epidemiology of severe myoclonic epilepsy in infancy. *Epilepsia*, Jul-Aug;31(4):397-400
3. Bender, AC, Morse, RP, Scott, RC, Holmes, GL, Lenck-Santini, PP (2012), SCN1A mutations in Dravet syndrome: Impact of interneuron dysfunction on neural networks and cognitive outcome. *Epilepsy Behav*. 23(3): 177–186.
4. Genton, P, Velizarova, R and Dravet, C (2011), Dravet syndrome: The long-term outcome. *Epilepsia*, 52: 44–49.
5. Ito, S, Ogiwara, I, Yamada, K, Miyamoto, H, Hensch, TK, Osawa, M, Yamakawa, K (2012), Mouse with Nav1.1 haploinsufficiency, a model for Dravet syndrome, exhibits lowered sociability and learning impairment. *Neurobiology of disease* 49, 29-40.
6. Han, S, Tai, C, Westenbroek, RE, Cheah, CS, Potter, GB, Rubenstein, JL, Scheuer, T, de la Iglesia, HO, Catterall, WA (2012), Autistic-like behavior in Scn1a+/- mice and rescue by enhanced GABA-mediated neurotransmission. *Nature*, 489(7416):385-90.
7. Hortopan, GA, Dinday, MT, Baraban, SC (2010), Zebrafish as a model for studying genetic aspects of epilepsy. *Dis. Model. Mech*. 2010 3:144-148.
8. Afrikanova, T, Serruys, AK, Buenafe, O, Clinckers, R, Smolders, I, de Witte, P, Crawford, AD, Esguerra, CV (2013), Validation of the Zebrafish Pentylenetetrazol Seizure Model: Locomotor versus Electrographic Responses to Antiepileptic Drugs. *PLoS One*. 2013; 8(1): e54166.
9. Baraban, SC, Dinday, MT, Hortopan, GA (2013), Drug screening in *Scn1a* zebrafish mutant identifies clemizole as a potential Dravet syndrome treatment. *Nature communications 4*, Article number: 2410.
10. Alfaro, JM, Ripoll-Gomez, J, Burgos, JS (2011), Kainate administrated to adult zebrafish causes seizures similar to those in rodent models. *European Journal of Neuroscience*, Vol 33, pp. 1252-1255.
11. Grone, BP, Qu, T, Baraban, SC (2017), Behavioral Comorbidities and Drug Treatments in a Zebrafish scn1lab Model of Dravet Syndrome. *eNeuro*, Jul-Aug; 4(4): ENEURO.0066-17.2017.
12. Wolman, MA, Jain, RA, Liss, L, Granato, M (2011), Chemical modulation of memory formation in larval zebrafish. *Proc Natl Acad Sci USA*, Sep 13;108(37):15468-73.
13. Zhang, Y, Kecskés, A, Copmans, D, Langlois, M, Crawford, AD, Ceulemans, B, Lagae, L, de Witte, PA, Esguerra, CV (2015), Pharmacological characterization of an antisense knockdown zebrafish model of Dravet syndrome: inhibition of epileptic seizures by the serotonin agonist fenfluramine. *PLoS One*, May 12;10(5):e0125898.

Spiegeldanio: A bold and aggressive fish but what if it loses a fight?

Arshi Mustafa1[,2], Dicle Cetinkaya[1], Xi Cheng[1], Per Ove Thörnqvist[1], Svante Winberg[1]

1 Department of Neuroscience, Uppsala University, Uppsala, Sweden. arshi.mustafa@neuro.uu.se; svante.winberg@neuro.uu.se ; **2 Department of Organismal Biology, Uppsala University, Uppsala, Sweden.** arshi.mustafa@ebc.uu.se

Introduction

Aggression is a competition based survival strategy. The *spiegeldanio* (spd) strain of zebrafish (*Danio rerio*), which has a mutation in the fibroblast growth factor receptor 1a, is bolder and more aggressive than the wild type fish [1]. Usually a socially dominant fish has preferential access to food, mate and shelter, and shows very characteristic postures like erection of the fins. It is also aggressive frequently biting, striking and chasing the subordinate fish as well as threatening its own mirror image in mirror tests [2]. However, what happens when an already known bold and dominant fish like *spiegeldanio* loses a dyadic fight. Spd fish are more aggressive in mirror tests, attacking their mirror image more frequently than wild type conspecifics. However, are they more aggressive in dyadic fights? Do they show an inhibition of aggressive behaviour when losing fights, the typical loser effect? The behavioural inhibition observed in animals losing fights for dominance is at least in part believed to be mediated by an activation of the brain serotonin (5-hydroxytryptamine, 5-HT) system. Do spd fish show a typical increase in brain 5-HT activity in response to social subordination? Dopamine (DA), on the other hand, is associated with aggression and social dominance. What are the effects of winning and losing fights for social dominance in spd fish? In the present study these questions were addressed in an attempt to increase or understanding of the control of agonistic behaviour and social stress.

Animals and Methods

The Spd strain of zebrafish were raised and reared at 27°C in an Aquaneering Zebrafish system at Uppsala University Biomedical Center. The animals were kept at a 14:10 h of light-dark photoperiod. The water used in the fish tanks was Uppsala municipal tap water (pH 7.2-7.6) of which 10% was exchanged daily. Fish were fed twice daily with Tropical energy food (Aquatic Nature, Belgium) and Artemia (Platinum Grade 0, Argentemia, Argent, Aquaculture, Redmond, USA). The use of animals was approved by the Uppsala Animal Ethical Committee (permit Dnr 55/13) and followed the guidelines of the Swedish Legislation on Animal Experimentation (Animal Welfare Act SFS1998:56), and the European Union Directive on the Protection of Animals Used for Scientific Purposes (Directive 2010/63/EU). The fish were transferred to the individual compartments of dimension 29 x 7.5 x 20 cm (length x breadth x height) in experimental tanks used for dyadic interaction and allowed to recover in isolation overnight. These experimental tanks were made from poly methyl methacrylate plastic and each tank was equipped with a submerged pump with filter (Eheim, typ 2006020, pumping capacity 1/h180, made in China), a heater (Sera aquarium, 25W, made in EU) and an air stone, all of which were placed at the back of the tank separated from the fish by a white perforated PVC screen (Figure 1). The setup of the arena was such that the two fish (1 dyadic pair) had an olfactory but not any visual cue of each other before the dyadic interaction. In the mirror test the fish were made to fight against the mirror image that was displayed in the mirror which was pasted on the wall of the arena. Prior to the beginning of the dyadic contest the mirror was covered with a black plexiglas slide cover. The experiment was carried out in the following sequence: The fishes were netted out and placed in the arena in the compartments A and B (Figure 1) and separated from each other by a partition. The cover of the mirror (opaque black PVC partition, Figure 1) was then removed and fish were made to interact with their own mirror image for 10 minutes. Then the slide covering the mirror was pulled down and the middle separating partition was pulled out and the fish were given an opportunity to fight. Dyadic fight was recorded two times, morning and evening on day one with the help of a video filming camera. Then next day in the morning the dyadic fight was again recorded. During the dyadic interaction the two fishes indulged in mutual display of aggressive behaviour which was followed by chasing and biting attacks performed by the dominant fish over the subordinate fish. Then middle partition was introduced again. Fish were given 6 minutes to habituate and the cover from the mirror was removed and fishes were again allowed to interact with their mirror image. Again the mirror was covered and the fish was allowed to get involved in the dyadic fight. Then each fish was taken out from the compartment at the same time and sacrificed for sampling of brain tissue.

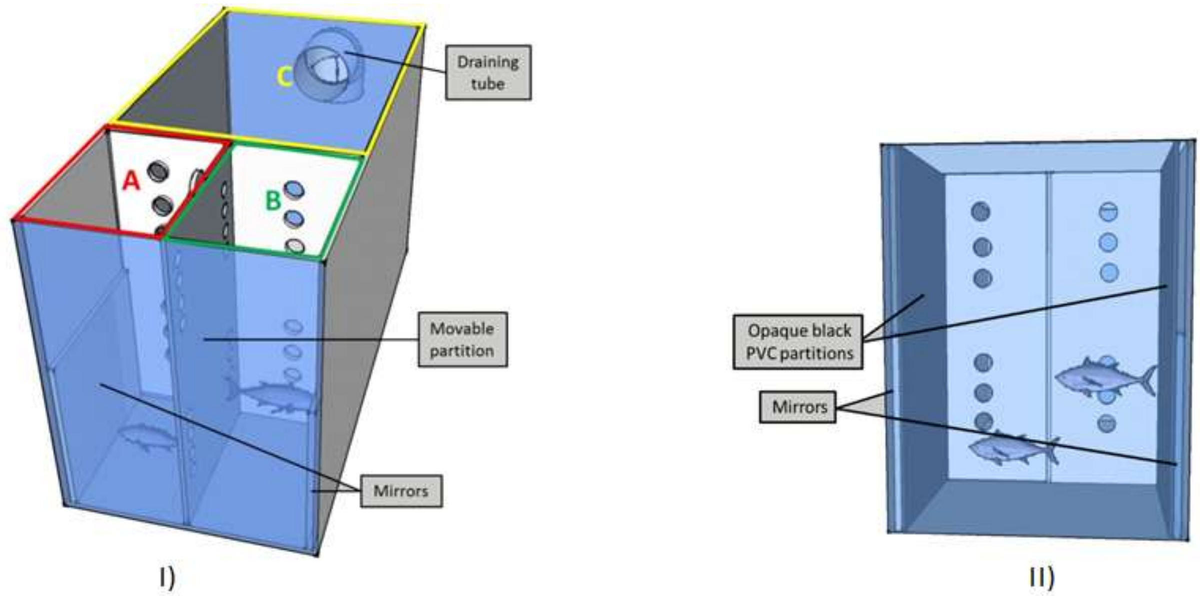


Figure 1 Arena used for Mirror Test and the Dyadic fight

The three dimensional model of tank used in the behavioural tests I) Tank used for mirror test and for dyadic fight later on. It consists of two compartments, A and B. The movable partition separating the two compartments would be removed during the dyadic fight test. Compartment C is located at the back and is separated from the compartment A and B with the help of white coloured opaque perforated partition. It contains an air stone (for diffusion of air bubbles), heater (27°C), water pump (for circulation of water) and a drainage tube to exchange the water. II) Diagram of the settings used for dyadic interactions. The mirrors are covered with the help of a black PVC slide and the middle partition is pulled out. This allows the fish to interact.

Brain dissection and analysis of monaoamines and monoamine metabolites

Brains were divided into forebrain (telencephalon and diencephalon), optic tectum and the rest (here denoted brain stem). The frozen brains were homogenised in 4% (w/v) ice-cold perchloric acid containing 100 ng/ml 3, 4-dihydroxybenzylamine (DHBA, the internal standard) using a Sonifier cell disruptor B-30 (Branson Ultrasonics, Danbury, CT, USA) and were immediately put on dry ice. Subsequently, the homogenised samples were thawed and centrifuged at 15,000 rpm for 10 min at 4^{o} C. The supernatant was used for high performance liquid chromatography with electrochemical detection (HPLC-EC), analysing the monoamines dopamine (DA) and serotonin (5-hydroxytryptamine, 5-HT) as well as the DA metabolite 3, 4-dihydroxyphenylacetic acid (DOPAC) and the 5-HT metabolite 5-hydroxyindoleacetic acid (5-HIAA), as described by Øverli et al. [3]. In short, the HPLC-EC system consisted of a solvent delivery system model 582 (ESA, Bedford, MA, USA), an autoinjector Midas type 830 (Spark Holland, Emmen, the Netherlands), a reverse phase column (Reprosil-Pur C18-AQ 3 µm, 100 mm × 4 mm column, Dr. Maisch HPLC GmbH, Ammerbuch-Entringen, Germany) kept at 40° C and an ESA 5200 Coulochem II EC detector (ESA, Bedford, MA, USA) with two electrodes at reducing and oxidizing potentials of -40 mV and +320 mV. A guarding electrode with a potential of +450 mV was employed before the analytical electrodes to oxidize any contaminants. The mobile phase consisted of 75 mM sodium phosphate, 1.4 mM sodium octyl sulphate and 10 µM EDTA in deionised water containing 7 % acetonitrile brought to pH 3.1 with phosphoric acid. The quantification of samples was done by comparing it with standard solutions of known concentrations. DHBA was used as an internal standard to correct for recovery with the help of HPLC software ClarityTM (Data Apex Ltd, Czech Republic). The serotonergic and dopaminergic activity was measured as the ratio of 5-HIAA/5-HT and DOPAC/DA respectively. The brain monoamines were normalized with respect to brain protein weights which were determined with Bicinchoninic acid protein determination kit (Sigma Aldrich, Sweden). The assay was read at a wavelength of 570 nm with the help of a plate reader (Labsystems multiskan 352, Labsystems Thermo Fisher Scientific).

Results

A clear dominant subordinate hierarchy was established within 30 minutes of dyadic interaction. The number of aggressive acts (bites, strikes and chases) performed by the looser fish decreased significantly from the first dyadic fight to the last (i.e. the fourth) dyadic fight. For the winner fish the number of aggressive acts performed against a mirror during the second mirror test increased or remained same as before after winning a dyadic fight, whereas for the looser fish it decreased significantly. The results from the present study indicate that subordinate fish have higher 5-HIAA/5-HT ratio in the optic tectum as compared to the dominants. More results from this study would be presented at the conference.

References

1. Norton W, Bally-Cuif L (2010) Adult zebrafish as a model organism for behavioural genetics. *BMC Neurosci.* **11**:90.
2. Rowland WJ (1999) Studying visual cues in fish behaviour: a review of ethological techniques. *Env Biol Fishes.* **56**:285-305.
3. Øverli Ø, Harris CA, Winberg S (1999) Short-term effects of fights for social dominance and the establishment of dominant-subordinate relationships on brain monoamines and cortisol in rainbow trout. *Brain Behav Evol.* **54**:263-275.

Behavioral Profiling using a Modified Version of the Zebrafish Multivariate Concentric Square Field™ (zMCSF) Test

Erika Roman, Ronja Brunberg, Arshi Mustafa, Per-Ove Thörnqvist and Svante Winberg

Uppsala University, Sweden; Erika.roman@farmbio.uu.se; svante.winberg@neuro.uu.se

Introduction

Increased understanding of complex psychiatric disorders may demand for more complex test arenas capturing a broader behavioral repertoire [1]. The multivariate concentric square field™ (MCSF), originally developed for rats, is unique in its design by provoking behaviors associated with exploration, risk taking and shelter seeking. Thereby a behavioral profile is generated in a single session [2, 3]. During the last decades, the zebrafish (*Danio rerio*) has become an increasingly important model organism in behavioral neuroscience. Recently, a multivariate test arena for behavioral profiling of zebrafish, i.e. the zebrafish MCSF (zMCSF) test, was described [4]. Despite demonstrating great potential for future use, the arena design was not optimal. For instance, a problem was that the arena was too large resulting in many fish spending approximately 50% or more of the time in the part of the arena that was not a designated zone, which hampered on functional description and behavioral interpretation. Herein, a modified version of the zMCSF test is described.

Animals and method

Male and female domesticated AB fish and wild-caught fish originating from the Calcutta area, India were used. The fish were held in 2.8-liter plastic home tanks (27 ± 1.5°C) in a filtrated recirculating water system (Aquaneering, USA) where 10 % of the water was exchanged daily, and with light/dark cycles of 14/10 hours. The zebrafish were fed twice a day with tropical energy food (Aquatic Nature, Belgium) and brine shrimp eggs that had been hatched in salt water (*Artemia* cysts, Argent Aquaculture, USA). The experimental protocol and use of animals in this study was approved by the Uppsala Animal Ethical Committee, and was consistent with the Swedish Legislation on Animal Experimentation (Animal Welfare Act SFS1998:56) and the European Union Directive on the Protection of Animals Used for Scientific Purposes (2010/63/EU).

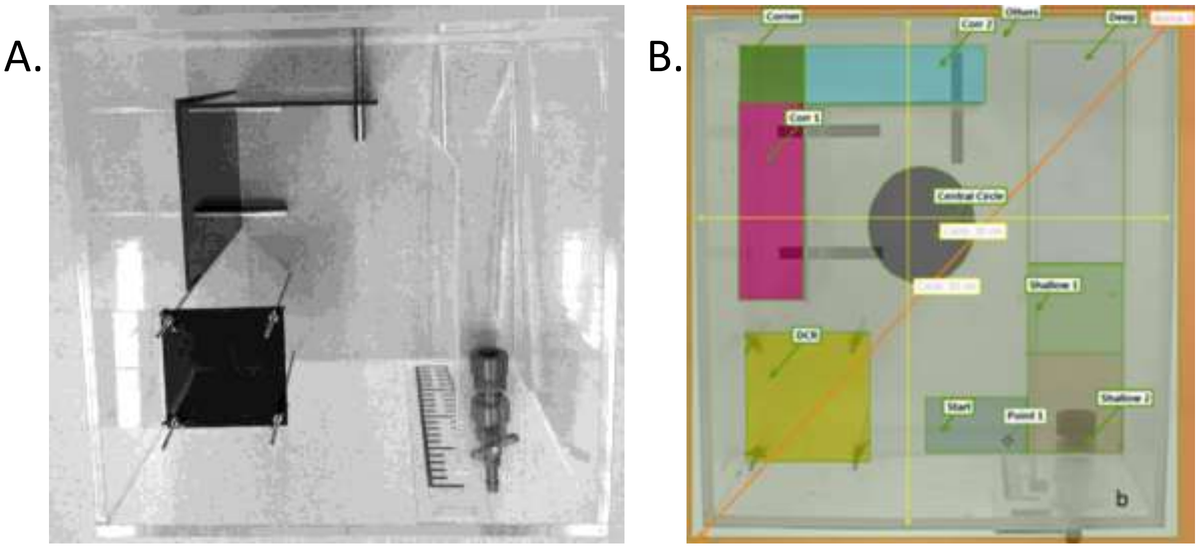


Figure 1. The modified version of the zMCSF (A) with the defined zones (B).

The modified version of the zMCSF arena consists of a tank made of Plexiglas (30 × 30 × 26 cm). The different parts in the zMCSF arena are a Plexiglas ramp with a wall covering approximately 2/3 of the side towards the open arena, a roof made of IR-transparent plastic, and two walls with weights (Figure 1A). These parts are placed in the arena to form the various zones: start, dark corner roof (DCR), central circle, corridor 1, corridor 2, corner and the ramp, which is divided into deep and shallow part 1 and part 2, respectively (Figure 1B). The walls of the arena were sandpapered after the picture in Figure 1A was taken to prevent zebrafish from reflecting in the glass. An IR-light table was placed underneath the arena and an IR-sensitive camera recorded the fish from above. The tank was filled with 24 ± 2°C copper-free water reaching a water depth of 9.5 cm. The fish were caught with a net, released in the arena at the mark in the start zone (Figure 1B), and allowed to freely explore the arena for 30 minutes. Between each fish, the tank was emptied, cleaned with ethanol (96%) and water, and refilled. The fish were tracked using Ethovision® XT 12.0 (Noldus Information Technology, Wageningen, The Netherlands). The number of visits, latency (s) to first visit, total time spent (s), duration per visit (s), distance travelled (cm) and mean velocity (cm/s) in each zone was registered, as well as mean velocity (cm/s) and distance travelled (cm) in the total arena. The duration (% of total trial time) in each zone and the total activity in the arena (sum of all frequencies) were calculated.

Results

In the modified zMCSF, the time spent in the part of the arena that was not a designated zone was decreased relative to what was seen using the previous arena set-up [4]. Moreover, the wall covering parts of the side of the ramp improved detection of the fish in that risk area since the fish had to make an active choice in order to swim up on the ramp. Finally, the inclusion of a start zone enabled detection of fish that immediately after start remained immobile before starting to explore the arena.

Individual differences in explorative strategies were evident in males and females of both strains. When observing the fish, it became evident that some fish swam back and forth on the deep part of the ramp, others moved a bit further up on the shallow part, and some swam all the way up on the shallowest part. Based on this observation the shallow half of the ramp was divided into shallow 1 and shallow 2 in order to be able to detect the most risk-taking fish, i.e. those that swam all the way up into the zone shallow 2.

In a preliminary functional interpretation of the different zones it is evident that the dark corner roof was associated with shelter seeking, while the central circle and the shallow part of the ramp are related to risk-taking behavior. Based on the observation of fish behavior on the deep part of the ramp, this zone is suggested to function as a risk assessment zone. Distance and velocity in the arena, together with the total activity are obvious measures of general activity. Activity in the part of the arena that is not a designated zone together with the corridors and the corner appear related to different explorative strategies.

Discussion

The results of the present study demonstrate that the smaller, modified zMCSF has great advantages compared to the previous version of the arena [4], since the detection of fish in the arena has improved, the fish spend more time in designated zones, and the smaller tank is easier to work with from a practical perspective. The functional interpretation will be further elaborated on using multivariate data analysis approaches. The results from this will set the basis for a trend analysis, which has been useful for interpretation of overall behavioral profiles in rats [3]. The zMCSF needs further validation but it clearly has a great potential in screening behavioral profiles in zebrafish, and in detecting natural variance in phenotypes observed in a population. The need for improved phenotyping strategies has recently been emphasized, and the zMCSF may constitute an important complement to conventional test used in preclinical research

Acknowledgements

Funding from the Facias Foundation (E.R.), the Swedish Research Council (S.W.), and the Gahlins Foundation (E.R., S.W.) is gratefully acknowledged. The study was conducted with support from the Uppsala University Behavioral Facility (UUBF).

References

1. Stewart AM, Kalueff AV (2015). Developing better and more valid animal models of brain disorders. *Behavioural Brain Research* **276**, 28-31.

2. Meyerson BJ, Augustsson H, Berg M, Roman E (2006). The concentric square field: a multivariate test arena for analysis of explorative strategies. *Behavioural Brain Research* **168**, 100-113.

3. Meyerson BJ, Jurek B, Roman E (2013). A rank-order procedure applied to an ethoexperimental behavior model - the multivariate concentric square field™ (MCSF) test. *Journal of Behavioral and Brain Science* **3**, 350-361.

4. Roman E, Tjernström N, Winberg S (2016), Description of a multivariate behavioral test arena for zebrafish – the zebrafish multivariate concentric square field test. *Proceedings of Measuring Behavior 2016*, (Dublin, 25-27 May 2016), 4-6.

Larval zebrafish allows high throughput behavioural analysis: Replacing adults in the study of nociception and drug testing

Lynne U. Sneddon

University of Liverpool, Institute of Integrative Biology. The BioScience Building, Liverpool, L69 7ZB, UK. lsneddon@liverpool.ac.uk

Testing of pharmaceutical drugs and their efficacy is necessary to determine the utility of novel compounds. Zebrafish have many advantages over traditional rodent models in biomedical and fundamental science due to their ease of breeding, short time to maturation (3 months), small size so can be kept in high numbers, rapid development of transparent embryos, ease of genetic modification and visualisation of in internal processes using albino strains. Therefore, the use of experimental fish models and particularly zebrafish is growing on a global scale. Under EU law, zebrafish from first feed are considered protected and thus their use requires licensing since their welfare may be compromised. Zebrafish are not protected in Europe until 5 days post fertilisation (dpf) since they are not fully developed. Therefore, under the principles of the 3Rs larval 5dpf zebrafish are considered as a replacement for using adults making them a more ethical choice and their use is widespread. However, analgesics have never been tested previously nor responses to a variety of noxious thermal and chemical stimuli.

In this project, the value of using 5dpf zebrafish in studies exploring a range of noxious stimuli was tested alongside drugs with pain-relieving properties to determine if 5dpf larval fish could replace adults. A behavioural monitoring system using novel software was developed to monitor the behaviour of 25 larvae at one time exposed to a range of chemical and thermal stimuli that could be potentially painful. Further, analgesic drugs were tested at two doses to determine if these were effective in reducing the behavioural changes. Data files generated by the tracking software were processed with a bespoke algorithm in MATLAB, which can detect various behavioural larvae patterns larvae based upon standard motion features including average velocity (mm s^{-1}), average acceleration or increase in speed (mm s^{-2}), time active (% of total time) and total distance moved (mm). For the analysis of thigmotaxis (the avoidance of a stimulus by moving towards the edge of a well), this was determined as the percentage (%) of time spent active in the outer zone divided by the time spent in both outer and inner zones and as the percentage (%) of the distance swam in the outer zone divided by the distance swam in both the outer and inner zones [1,2]. The advantages of this system is that it was less expensive than commercial systems and allows more detailed behavioural analysis by providing larvae with a larger arena (16.5 x 16.5mm, volume 2191µl) compared with 96 wells (~7mm diameter, volume 320µl) in other systems.

When exposed to noxious stimuli zebrafish larvae significantly reduced their behaviour and this was prevented by the use of an appropriate dose of analgesic drug dissolved in the tank water. A further experiment investigated the impact of stress and fear on the responses to noxious stimuli. The 5dpf larvae were exposed to either a stressor (air emersion), a predatory fear cue (alarm substance) or an anxiogenic (caffeine) alone or prior to noxious stimulation. Both velocity and activity decreased after exposure to the stress and predator cues which were attenuated using etomidate and diazepam, respectively. Noxious stimulation decreased velocity and activity as well, whereas air emersion and alarm substance inhibited these responses consistent with an antinociceptive system seen in other animals [3-6]. We showed there was no effect of time of day on the recordings, therefore, a minimum of eight plates or 200 larvae could be assessed per day which is much higher throughput on studies using adults which in our laboratory would only allow 2 individuals per day. Therefore, this novel tracking system has validated the use of 5dpf zebrafish as replacement for adults. This research was conducted after ethical approval and under Home Office Licensing (UK).

References

1. Al-Juboury, Q., Al-Nuaimy, W., Al-Taee, M.A., Lopez-Luna, J., Sneddon, L.U. (2016). An automatic pattern detection method for behavioral analysis of zebrafish larvae. *13th International Multi-Conference on Systems, Signals & Devices (SSD)* < http://ieeexplore.ieee.org/document/7473748/?reload=true> Accessed 29/03/18.
2. Al-Juboury, Q., Al-Nuaimy, W., Al-Taee, M.A., Lopez-Luna, J., Sneddon, L.U. (2016). Occurrence density index for behavior classification of zebrafish larvae. *13th International Multi-Conference on Systems, Signals & Devices (SSD)* < http://ieeexplore.ieee.org/document/7473764/ > Accessed 29/03/18.

3. Lopez-Luna, J., Al-Jubouri, Q., Al-Nuaimy, W., Sneddon, L.U. (2017). Activity reduced by noxious chemical stimulation is ameliorated by immersion in analgesic drugs in zebrafish. *Journal of Experimental Biology* 220, 1451-1458.
4. Lopez-Luna, J., Al-Jubouri, Q., Al-Nuaimy, W., Sneddon, L.U. (2017b). Impact of analgesic drugs on the behavioural responses of larval zebrafish to potentially noxious temperatures. *Applied Animal Behaviour Science* 188, 97-105.
5. Lopez-Luna, J., Al-Jubouri, Q., Al-Nuaimy, W., Sneddon, L.U. (2017c). Impact of stress, fear and anxiety on the nociceptive responses of larval zebrafish. *PLoS One* 12(8): e0181010.
6. Lopez-Luna, J., Canty, M.N., Al-Jubouri, Q., Al-Nuaimy, W. Sneddon, L.U. (2017d). Behavioural responses of fish larvae modulated by analgesic drugs after a stress exposure. *Applied Animal Behaviour Science* 195, 115–120.

How to Make Use of Individual Variation?

Svante Winberg, Arshi Mustafa, Per-Ove Thörnqvist and Erika Roman

Uppsala University, Sweden; Erika.roman@farmbio.uu.se; svante.winberg@neuro.uu.se

Individuals can be classified in distinct behavioural phenotypes, often referred to as divergent stress coping styles, proactive versus reactive. Proactive animals are bold and aggressive whereas reactive animals show the opposite behavioural profile. Moreover, different populations differ in behaviour, and in sexually mature fish gender differences occur even in non-reproductive behaviour. Individual variation has been seen as a problem since it is a source of variance. However, interspecific divergence in behaviour and physiology may also be used to address questions related to gene-environment interaction, phenotypic plasticity and variable life history traits. Fish displaying divergent behavioural profiles may provide interesting models for studies on affective disorders and drug abuse. We will present the results showing that divergent stress coping styles exists in zebrafish and that these are associated with specific neuroendocrine functions.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Symposia: New tools to analyse motor activity and movement

Increasing age and related neurodegenerative diseases are becoming a significant burden to society. Intrinsic to these are difficulties in movement and injuries related to imbalance and failure of motor programming. Diseases like Parkinsonism, motor neurone disease or amyotrophic latera sclerosis (ALS), but also spinal cord injury or ageing-induced frailties pose a severe risk for the patient and a greater understanding for the underlying mechanisms is needed if we are to develop novel therapies. Multiple experimental models have emerged in which some disease relevant features occur, but our means of testing more detailed anomalies in movement have been lagging behind. In this symposium, we seek to concentrate on the generation of novel analysis tools to extract more and refined data from rodent models of disease. These analysis tools are often based on well-established tests of grabbing, walking or other patterns of movement, but offer a novel approach to detecting and understanding the core principles of behaviour.

Wiktor Niewiadomski and Grazyna Niewiadomska (Polish Academy of Sciences) and Gernot Riedel, Aberdeen University)

Temporal patterns of mice behavior in the horizontal grid test following MPTP

Maurizio Casarrubea[1], Wiktor Niewiadomski[2], Anna Gasiorowska[2], Ewelina Palasz[3], Grazyna Niewiadomska[3], Andrea Santangelo[1,5], Fabiana Faulisi[1], Giuseppe Crescimanno[1] and Gernot Riedel[4]

1. Dept. Bio.Ne.C., Human Physiology Section "Giuseppe Pagano", University of Palermo, Palermo, Italy. 2. Mossakowski Medical Research Centre Polish Academy of Sciences, Warsaw, Poland. 3. Nencki Institute, Warsaw, Poland. 4. Institute of Medical Sciences, University of Aberdeen, UK. 5. Dept. of Neuroscience, Psychology, Drug Research and Child Health, University of Florence, Florence, Italy;
maurizio.casarrubea@unipa.it

Present paper deals with the analysis of behaviour in an experimental model of Parkinson disease obtained following administration of MPTP in mice. Rodents have been tested in an experimental apparatus known as horizontal inverted grid and their behavior analysed by means of a technique known as T-pattern analysis. Preliminary results show clear cut differences between saline administered subjects and MPTP administered ones in terms of behavioral sequencing.

Introduction

The Inverted Horizontal Grid Test (HGT) is a suitable tool to study the behaviour in a mouse model of Parkinson's Disease (PD). On this subject the murine PD model obtained following administration of MPTP is one of the most versatile and translationally relevant. Using this model of PD, we have recently presented an approach able to provide support to the assessment of movement related features once the paws position has been recorded and annotated by an observer (12). In this context it would be of crucial interest to detect possible differences of animal's behavioural sequencing, as indicative of the underlying MPTP-induced breakdown of striatal dopamine. Aim of our study was to utilize an advanced multivariate approach, known as T-pattern analysis (TPA), to study behavioral sequencing in two groups of mice administered with saline or MPTP. TPA is a multivariate technique able to detect the existence of statistically significant temporal relationships among behavioural events in the course of time (2, 9, 10). By means of T-pattern analysis, sequences of behavioural events may be detected. During the last decade T-pattern analysis has been successfully applied to evaluate the structure of behaviour in several experimental contexts concerning both human (6, 8, 13) and non-human behaviour (1, 3-5, 7, 11, 14).

Method

Ten male, three months old, C57BL/6 mice were used. Subjects, bread at the Medical University of Bialystok (Poland), were housed four to six animals per cage. Food and water were provided "ad libitum". Temperature and humidity in housing room were maintained constant at 23 ± 1 °C and 55 ± 5% respectively, under a 12-h light-dark cycle (lights on: 8.00 am). Six C57BL/6 mice were injected four times with MPTP hydrochloride (20 mg/kg in saline, i.p.) at 2-h intervals. Four saline treated C57BL/6 mice served as control. The test was performed on day 10 post-treatment. The HGT apparatus was a square wire grid (12 × 12 cm) surrounded by opaque Perspex walls 9 cm high. Each animal was placed in the center of the grid. After that, the grid was 180° rotated so the mouse was hanging and moving on the underside of the grid. A soft padding below the grid was provided to mitigate possible falling down of the tested subject. A professional digital camera, placed above the apparatus, has been utilized to record all the tests. So collected digital video-files were stored in a PC for following analyses. Annotation of the behaviors was carried out using The Observer software coder (Noldus IT, The Netherlands). The ethogram (Table 1) used to annotate behaviors has four categories: "Paw Displacement", "Paw Suspension", Contact" and "Others", each encompassing various components of the behavioural repertoire.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

CATEGORY “PAW DISPLACEMENT”:
Front-paw Left Displacement (**fld**) - Front left paw, in contact with the grid, is displaced to a new position;
Front-paw Right Displacement (**frd**) - Front right paw, in contact with the grid, is displaced to a new position;
Hind-paw Left Displacement (**hld**) - Hind left paw, in contact with the grid, is displaced to a new position;
Hind-paw Right Displacement (**hrd**) - Hind right paw, in contact with the grid, is displaced to a new position;
CATEGORY “SUSPENSION”:
Front-paw Suspension (**fps**) - Mouse maintains grip with the grid using the front paws. Both the hind paws are not in contact with the grid;
Front-paw Left Suspension (**fls**) - Mouse maintains grip with the grid using its front left paw only. All the remaining paws are not in contact with the grid;
Front-paw Right Suspension (**frs**) - Mouse maintains grip with the grid using its front right paw only. All the remaining paws are not in contact with the grid;
Hind-paw Suspension (**hps**) - Mouse maintains grip with the grid using the hind paws. Both the front paws are not in contact with the grid;
Hind-paw Left Suspension (**hls**) - Mouse maintains grip with the grid using its hind left paw only. All the remaining paws are not in contact with the grid;
Hind-paw Right Suspension (**hrs**) - Mouse maintains grip with the grid using its hind right paw only. All the remaining paws are not in contact with the grid;
CATEGORY “CONTACT”:
Front-paw Left Contact (**flc**) - Mouse displaces its front left paw making contact with one of the four surrounding walls;
Front-paw Right Contact (**frc**) - Mouse displaces its front right paw making contact with one of the four surrounding walls;
Hind-paw Left Contact (**hlc**) - mouse displaces its hind left paw making contact with one of the four surrounding walls;
Hind-paw Right Contact (**hrc**) - Mouse displaces its hind right paw making contact with one of the four surrounding walls;
CATEGORY “OTHERS”:
Wall Nosing (**wan**) - Mouse displaces its head nosing one of the four walls;
Disengaging (**dis**) - Mouse interrupts contact with the grid.

Table 1. Ethogram of mouse behavior in the inverted horizontal grid test

To explore the existence of possible significant relationships among the events in the course of time, multivariate T-pattern analysis has been carried out. This technique can be performed by means of a specific software known as Theme (Noldus IT, The Netherlands; Patternvision ltd, Iceland). T-pattern detection algorithm searches for relationships between events in behavioral data by taking into account, the order, timing, and frequency of these events. For instance, given a hypothetical observational period where several behavioral events do occur, the algorithm compares the distributions of each pair of the behavioral elements A and B searching for an interval so that, more often than chance expectation, A is followed by B within that interval. If such a circumstance does occur, A and B are a T-pattern and indicated as (A B). In a second step, such first level t-patterns are considered as potential A or B terms for the construction of higher-order patterns, e.g. ((A B) C), and so on. When no more patterns are detected, the search is concluded. A more detailed description of concepts, theories and procedures behind T-pattern analysis can be found elsewhere (2, 9, 10).

Ethical statement

All efforts were made to minimize the number of subjects and their suffering. Experimental procedures were conducted in strict accordance with the European Communities Council Directive (2010/63/EU) and with the permission of the First Warsaw Local Ethics Committee for Animal Experimentation (Permission No. 347/2012).

Results

Results are presented in terms of mean occurrences of behavioral components (namely, individual components from the ethogram, not in sequence) and T-patterns. Student’s *t*-test, used to assess possible significant differences between control and MPTP group, revealed no significant differences between groups in terms of mean occurrences of behavioral components (fig. 1). On the other hand, TPA revealed clear-cut differences between control and MPTP group in terms of behavioral structure and number of patterns detected (fig. 2). Notably, 20 different patterns have been detected in control group. On the other hand MPTP administered subjects show 9 different patterns: a reduction of beyond 50%.

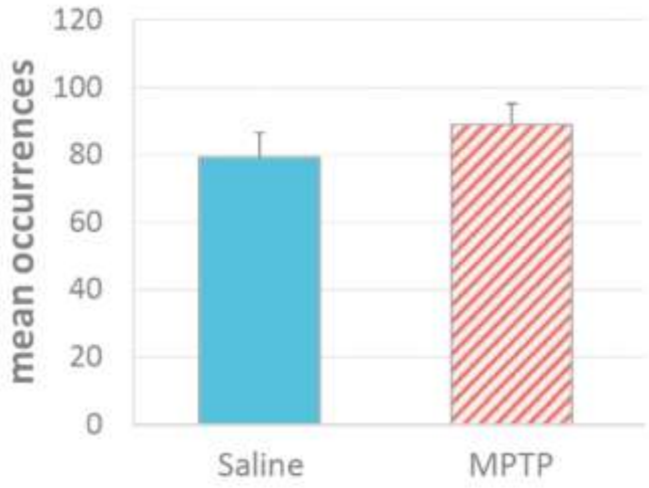


Figure 1. Mean occurrences of behavioural components in Saline (filled bar) and MPTP (dashed bar) groups. Student's *t*-test revealed no significant differences between groups.

Discussion

This study represents the first description of the real time behavioral organization of mice tested in the HGT apparatus. Our preliminary results show that mice behavior in the HGT has complex temporal characteristics and even within the boundaries of a narrow observation window (namely, 30 seconds), several patterns of behavior are present (fig. 2). Figure 1 clearly demonstrates that MPTP did not affect the behavioural repertoire in terms of ability of the rodent to perform individual behavioural elements. On the contrary, important effects of MPTP concern the overall ability of the subject to sequence its behavior. Such a clear-cut impairment appears to be twofold. First, concerning the number of behavioural sequences, MPTP injected mice, in comparison with saline, show a consistent reduction of T-patterns; second, as to the length of detected T-patterns, MPTP group shows, in comparison with saline, significantly shorter sequences. Present results, by showing qualitative changes of behavior following MPTP induced impairment of BG transmission, provide an important contribution in characterizing BG involvement in behavioral sequencing.

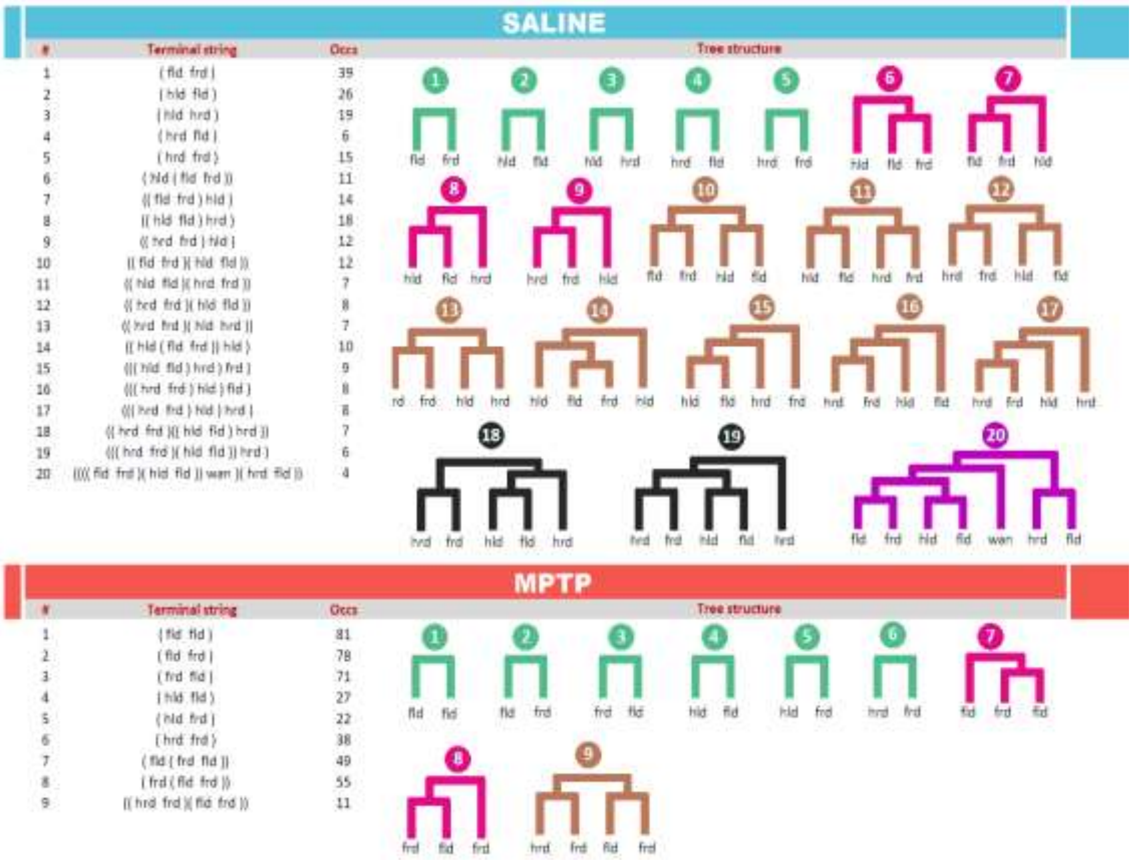


Figure 2. Terminal strings and tree structures of the T-patterns detected in Saline and MPTP groups. Numbers on the right of each string indicate their overall occurrences (Occs). Abbreviations in Tab. 1.

References:

1. Bonasera SJ, Schenk AK, Luxenberg EJ, Tecott LH (2008). A novel method for automatic quantification of psycho stimulant-evoked route-tracing stereotypy: application to Mus musculus. *Psychopharmacology* (Berlin) **196**:591–602.

2. Casarrubea M, Jonsson GK, Faulisi F, Sorbera F, Di Giovanni G, Benigno A, Crescimanno G, Magnusson MS (2015a) T-pattern analysis for the study of temporal structure of animal and human behavior: a comprehensive review. *J Neurosci Methods* **239**:34-46

3. Casarrubea M, Davies C, Faulisi F, Pierucci M, Colangeli R, Partdrige L, Chambers S, Cassar D, Valentino M, Muscat R, Benigno A, Crescimanno G, Di Giovanni G (2015b) Acute nicotine induces anxiety and disrupts temporal pattern organization of rat exploratory behavior in hole-board: a potential role for the lateral habenula. *Front Cell Neurosci* **9**:197 - doi: 10.3389/fncel.2015.00197

4. Casarrubea M, Faulisi F, Caternicchia F, Santangelo A, Di Giovanni G, Benigno A, Magnusson MS, Crescimanno G (2016). Temporal patterns of rat behaviour in the central platform of the elevated plus maze. Comparative analysis between male subjects of strains with different basal levels of emotionality. *J Neurosci Methods* **268**:155-162

5. Casarrubea M, Faulisi F, Cudia A, Cancemi D, Cardaci M, Magnusson MS, Crescimanno G (2017) Discovery of recurring behavioural sequences in Wistar rat social activity: Possible support to studies on Autism Spectrum Disorders. *Neurosci Lett* **653**:58-63

6. Castañer M, Barreira D, Camerino O, Anguera MT, Fernandes T, Hileno R. (2017) Mastery in Goal Scoring, T-Pattern Detection, and Polar Coordinate Analysis of Motor Skills Used by Lionel Messi and Cristiano Ronaldo. *Front Psychol*. **8**:741. doi: 10.3389/fpsyg.2017.00741.

7. Jonsson GK, Thorsteinsson V, Tomasson GG. (2010) Identification of vertical and horizontal movement patterns in cod behavior. In: Proceedings of the 7th international conference on methods and techniques in behavioral research. New York: *ACM*; 2010, http://dx.doi.org/10.1145/1931344.1931354.

8. Kemp AS, Fillmore PT, Lenjavi MR, Lyon M, Chicz-Demet A, Touchette PE, Sandman K (2008). Temporal patterns of self-injurious behavior correlate with stress hormone levels in the developmentally disabled. *Psychiatry Res* **157**:181-189.

9. Magnusson MS (2000) Discovering hidden time patterns in behavior: T-patterns and their detection. *Behav Res Methods Instrum Comput* **32**:93-110

10. Magnusson MS, Burgoon JK, Casarrubea M, Discovering Hidden Temporal Patterns in Behavior and Interaction (2016) *Neuromethods*, vol. 111. Springer New York, NY - DOI: 10.1007/978-1-4939-3249-8 - ISBN: 978-1-4939-3248-1, e-ISBN: 978-1-4939-3249-8

11. Nicol AU, Segonds-Pichon A, Magnusson MS (2015). Complex spike patterns in olfactory bulb neuronal networks. *J Neurosci Methods*. **239**:11-17

12. Niewiadomski W, Palasz E, Skupinska M, Zylinski M, Steczkowska M, Gasiorowska A, Niewiadomska G, Riedel G. (2016) TracMouse: A computer aided movement analysis script for the mouse inverted horizontal grid test. *Sci Rep*. **6**:39331. doi: 10.1038/srep39331

13. Sandman CA, Kemp AS, Mabini C, Pincus D, Magnusson M (2012). The role of self-injury in the organisation of behaviour. *J Intellect Disabil Res*. **56**(5):516-526

14. Santangelo A, Provensi G, Costa A, Blandina P, Ricca V, Crescimanno G, Casarrubea M, Passani Mb (2017) Brain histamine depletion enhances the behavioural sequences complexity of mice tested in the open-field: partial reversal effect of the dopamine D2/D3 antagonist sulpiride. *Neuropharmacology* **113**:533-542

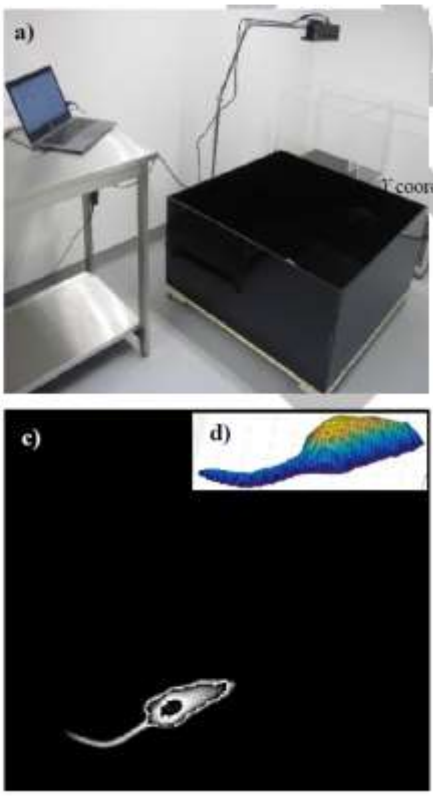
Automatic Classification of Rodents' Behavioral Activity based on Depth Cameras

A. Gerós[1,2], A. Magalhães[1,3] and P. Aguiar[1,2]

1Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal
2INEB – Instituto de Engenharia Biomédica, Universidade do Porto, Portugal.
3IBMC – Instituto de Biologia Molecular e Celular, Universidade do Porto, Portugal. pauloaguiar@ineb.up.pt

The studies on important human diseases such as autism, cancer, Alzheimer's or Parkinson's rely on animal models and behavior analysis, and they play a fundamental role in research (universities, pharmaceutical companies) and also in industry (food production and pharmaceutical companies). Laboratory animals are monitored in terms of motor activity, posture, social interactions, etc., not only to test hypothesis and develop new therapies but also to determine whether a novel therapeutic approach is successful or to further understand complex central nervous system processes [1,2]. With modern computational analysis methods, several approaches have been developed to observe and quantify the behavior of interest directly, instead of relying on a visually behavioral assessment or time-consuming manual annotations [3]. Several studies have addressed the quantification of behavior features by applying automatic systems using video recordings and computer vision methods, to automatically track and characterize the behavior of different animals [4-6]. However, there are still important unsolved challenges in automatic classification/quantification of behavior, namely at the level of correctly and automatically characterizing animal posture, identifying subtle movements of animal's body parts, tracking algorithms in naturalistic environments and characterizing social interactions [7,8]. In fact, the complexity of animals' behaviors together with the lack of precise estimations of their poses in the available two-dimensional (2D) systems impairs a detailed and complete behavior characterization.

The present research addresses specific behavior quantification challenges by developing a methodology to perform three-dimensional (3D) segmentation and tracking of animal's whole-body, and to automatically classify the body parts and simple behaviors of rodents in laboratory environments. In order to perform 3D analysis of



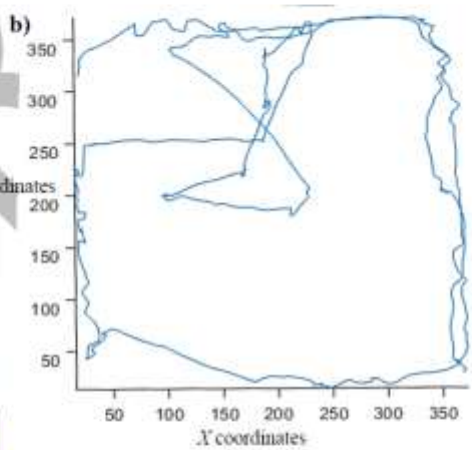


Figure 1. **a)** System's setup: open-field cage, depth camera and computer. **b)** Animal's trajectory inside the open-field cage (x and y coordinates; 30seconds). **c)** Experimental results of whole-body segmentation. **d)** 3D mesh with animal's point cloud.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

laboratory animals using video-based methods, a depth camera was selected for data acquisition. Depth video sequences of animal behavior under normal conditions were recorded using a *Microsoft Kinect v2* camera, which generates depth maps with an operation range from 0.5 to 4.5m and a maximum 30 frames per second acquisition rate. An animal behavior dataset was populated with these short depth video sequences capturing freely-walking movements of *Wistar* rats in an open-field (see Figure 1. **a)**), under different lighting conditions, manually annotated by ethologists. To the best of author's knowledge, such depth datasets are presently unavailable. Using these cameras, the system allows tracking in dark/low contrast conditions: even if background color matches the animal's fur (impairing color segmentation) segmentation in depth (obtained in the infrared spectrum) is possible. Two different background modeling methods were developed to further segment the whole-body of the animal: static median algorithm, followed by background subtraction, and the probabilistic *Gaussian Mixture* model, adapted from Stauffer and Grimson approach [9], to deal with the dynamic background challenge. Both models were successfully implemented for background subtraction and animal's trajectory extraction (see Figure 1. **b)**), using the available dataset, with detailed and precise segmentation of animal's whole-body (see Figure 1. **c)** and **d)**). Six different features were extracted and normalized from the tracked frames, for body parts classification and rodent behavior analysis. A *Support Vector Machines* method was trained to identify three body parts (head, body, and tail) and five types of rodent behaviors (walking, grooming, standstill, rearing, and local exploration), and the optimum kernel type and parameters were applied to train the model, using *5-fold* cross-validation.

The present computational methods, embedded in a fully-functional prototype, are able to perform animal tracking, body-parts segmentation, and behavior recognition, using a *Kinect*-based system to address the challenge of analyzing complex behaviors in 3D. This system is fully automatic and markerless, providing robust and reliable 3D behavioral information. Having this in mind, we believe that this novel system can be used to boost new research and improve the existing methods, in the context of animal behavior analysis in the neurosciences field.

All procedures reported in this research were carried out under personal and project licenses approved by the national authority for animal protection, '*Direção Geral de Alimentação e Veterinária*' (Lisbon, Portugal), and were performed in accordance with the *European Directive 2010/63/EU* on the protection of animals used for scientific purposes.

References:

1. Giancardo, L., et al. (2013). Automatic visual tracking and social behaviour analysis with multiple mice. *PloS one*, 8(9): p. e74557.
2. Raineteau, O. and M.E. Schwab (2001). Plasticity of motor systems after incomplete spinal cord injury. *Nature Reviews Neuroscience*,. 2(4): p. 263-273.
3. Ou-Yang, T.-H., et al. (2011). An infrared range camera-based approach for three-dimensional locomotion tracking and pose reconstruction in a rodent. *Journal of neuroscience methods*,. 201(1): p. 116-123.
4. Kabra, M., et al. (2013). JAABA: interactive machine learning for automatic annotation of animal behavior. *Nature methods*. 10(1): p. 64-67.
5. Pérez-Escudero, A., et al. (2014). idTracker: tracking individuals in a group by automatic identification of unmarked animals. *Nature methods*. 11(7): p. 743-748.
6. De Chaumont, F., et al. (2012). Computerized video analysis of social interactions in mice. *Nature methods*. 9(4): p. 410-417.
7. Egnor, S.R. and K. Branson (2016). Computational Analysis of Behavior. *Annual review of neuroscience*. 39: p. 217-236.
8. Robie, A.A., et al. (2017). Machine vision methods for analyzing social interactions. *Journal of Experimental Biology*,. 220(1): p. 25-34.
9. Stauffer, C. and W.E.L. Grimson. (1999). Adaptive background mixture models for real-time tracking. in Computer Vision and Pattern Recognition. IEEE Computer Society Conference on. 1999. IEEE.

Analysis of movement in Inverted Grid Test: Principles, Confirmation and Automatization

Wiktor Niewiadomski1(*), Ewelina Palasz2, Krzysztof Kodrzycki1,2, Anna Gasiorowska1,2, Gernot Riedel3, Grazyna Niewiadomska1

1Mossakowski Medical Research Centre Polish Academy of Sciences, Warsaw, Poland. wniewiadomski@imdik.pan.pl 2Nencki Institute of Experimental Biology, Polish Academy of Sciences Warsaw, Poland. g.niewiadomska@nencki.gov.pl 3Institute of Medical Sciences, University of Aberdeen, UK. g.riedel@abdn.ac.uk

Complex pathological basis of Parkinson's disease (PD) is a main reason why there are no effective forms of treatment. The supporting therapy could be the usage of physical activity. In many studies trying to test this possibility mouse models are being used, in which the loss of dopaminergic neurons in substantia nigra is caused by neurotoxin MPTP. Detection of motor impairment in MPTP mice is difficult. The inverted grid test, where mouse moves hanging under the mesh is believed to be effective in exposing any motor disorders. Software ***TracMouse*** [2] was designed in order to minimalize the role of rater in judging mouse's behaviour and thus to increase the precision of inverted grid analysis. However, at present the rater spends 20-30 minutes analysing video-recordings of trials lasting 30 seconds. In practical terms this precludes the use of our method in multi-cohort studies. Therefore we have developed a method to automatize detection of grip, its timing, paw position and identification.

Background

The Inverted Grid Test (also termed Traction Test) consists in a mouse or rat being placed onto a grid, which is subsequently turned upside down and in analysing animal movements during its clinging onto grid. It was shown, that this test is sensitive to movement impairments seen in animal models of Parkinson's disease induced by for example the neurotoxin (MPTP), which may cause extensive loss of dopaminergic neurons in the substantia nigra [1]. A systematic review of relevant papers [2] revealed that no agreement in test conditions has been achieved in recent years. In our opinion it was caused, at least in part, by requirement to take subjective decision about some aspects of movement.

Principles

Therefore our first principle was to reduce the need of subjective assessments in animal movements. To this end we limited the role of the rater to the role of detection of simple events: time of beginning and end of gripping the wire with paw, paw position on the grid during grip and identification of that paw. This has to be done in each frame from video-recorded tracks of the Inverted Grid Test. Our second principle was to eliminate the rater from subsequent analysis, by leaving it to a Matlab software script.

Confirmation

Such analysis was applied to a small set of mice, 6 animals acutely treated with MPTP, what destroyed high percentage (76%) of SN dopamine neurons and 4 control mice. It was found that, contrary to expectation based on known PD motor symptoms, no slowing of movement, shortening of its range and reduction of mobility were observed. This finding was confirmed in a study with much larger cohorts (14 animals per group). The neurotoxin MPTP was given chronically over 5 weeks in order to mimic slow onset of this disease and to make room for intervention during the intoxication process. There was no reduction in number of steps per unit of time, their length and paw velocity during step. Physical training was used as therapeutic intervention, its effectiveness was evident as sparing of dopamine neurons in substantia nigra of MPTP treated mice. Our analysis of movement detected some effects of training, and effects of intoxication combined with training, but these could not be explained by MPTP treatment or training alone.

Automatization

At present the rater spends 20 -30 minutes analysing video-recordings of trials lasting 30 seconds. In practical terms this precludes the use of our method in multi-cohort studies; therefore we felt it necessary to automatize detection of grip, its timing, paw position and identification. Although these goals have been achieved, further refinements of the method are possible and required. Each frame of a video recording is analysed separately. The analysing algorithm is based on the YOLO neural network architecture [3]. The use of deep learning methods provided more accurate detection of grip than the standard computer-vision methods. Besides detecting the paw

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

position during a grip, the algorithm detects the nose and therefore auto-identifies between fore and hind paws, and left/right. The current algorithm needs 30 seconds to analyse 30 seconds record of Inverted Grid Test. However, the correctness of algorithm results has to be validated manually and corrections introduced when needed. This amounts to an extra time requirement of 5 minutes for each rater. Of note is that the algorithm has a built-in mechanisms to improve during continuing use. Furthermore, additional rules of inference may improve robustness of paws position detection and paw identification.

Keywords: Inverted Horizontal Grid Test, TracMouse movement analysis, parkinsonian MPTP mouse model, automatization of movement detection

Ethical statement

All efforts were made to minimize the number of subjects and their suffering. Experimental procedures were conducted in strict accordance with the European Communities Council Directive (2010/63/EU) and with the permission of the First Warsaw Local Ethics Committee for Animal Experimentation (Permission No. 347/2012).

Literature

1. Tillerson, J.L., Miller, G.W. (2003). Grid performance test to measure behavioral impairment in the MPTP-treated-mouse model of parkinsonism. *J Neurosci Method* **123**(2), 189-200.

2. Niewiadomski, W., Palasz, E., Skupinska, M., Zylinski, M., Steczkowska, M., Gasiorowska, A., Niewiadomska, G., Riedel, G. (2016). TracMouse: A computer aided movement analysis script for the mouse inverted horizontal grid test. *Sci Rep* **6**:39331. doi: 10.1038/srep39331.

3. https://pjreddie.com/media/files/papers/yolo.pdf

Systematic Data Analysis and Data Mining in Gait Analysis by Heat Mapping

I.K. Timotius[1,5], F. Canneva[2], G. Minakaki3, S. Moceri[2], N. Casadei[4], O. Riess[4], J. Winkler[3], S. von Hörsten[2], B. Eskofier[1] and J. Klucken[3]

1Machine Learning and Data Analytics Lab, Department of Computer Science, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Erlangen, Germany. ivanna.timotius@fau.de 2Department of Experimental Therapy, University Hospital Erlangen (UKEr) and Preclinical Experimental Animal Center, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Erlangen, Germany. stephan.v.hoersten@uk-erlangen.de 3Department of Molecular Neurology, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Erlangen, Germany. jochen.klucken@uk-erlangen.de 4Institute of Medical Genetics and Applied Genomics, University of Tübingen, Germany 5Department of Electronics Engineering, Satya Wacana Christian University, Salatiga, Indonesia.

Introduction

Motor impairment appears as a characteristic of several diseases and injuries [1–4]. Therefore, tests for analyzing motor dysfunction are widely applied across preclinical models and disease stages. Standard analysis addresses an extensive numbers of gait parameters which raises complication and often lead to premature parameter pre-selection. Moreover, cohorts frequently include a limited number of individual animals, which in combination with the extensive number of different parameters analyzed prohibits the application of multivariate data analysis [5]. In most experiments using animal models, it is challenging to collect data from a multitude of subjects due to several reasons, such as the ethical consideration of reducing the number of animals to a minimum [6], the littermate subject requirement and the extensive data collection time needed in a voluntary walking experiment.

The Catwalk XT (Noldus Information Technology, Wageningen, Netherlands [7]) can be used for the sensitive, automated preclinical assessment of rodent gait. This system was initially developed by Hamers and colleagues [8,9] and has since then been widely applied to a range of disorders and screening approaches, including spinal cord injury [10,11], peripheral nerve injury [12], intervertebral disk (IVD) injury [13], Huntington disease [14], Parkinson's disease [15–18], Alzheimer's disease [19], multiple sclerosis [20], osteoarthritis [21], stroke [22], and ataxia [23,24]. Using a high-speed camera, the CatWalk system captures and records rodent silhouette and footprints while the animal transverses an enclosed walkway on glass plate. Figure 1 shows an image captured from a recorded CatWalk video. Based on the dynamic footprint recording and the body silhouette, several gait parameters can be measured. Currently, the CatWalk system can measure and present up to 360 different gait parameters including, speed, stride length, stand time, swing time, duty cycle, cadence, foot angle, regularity, number of body-supports [7]. Normally, the data for pre-selected parameters are then statistically analyzed and interpreted by an expert. For instance in several studies [10,21,23,25], only a selected group of the CatWalk gait parameters are reported and multivariate data analysis has not been performed. The parameter selection process was typically based on the previous research or the expert presumption, which sometimes leaves a question about the importance of the unselected parameters.

Instead of manual parameter selection, here we propose a data mining technique using a systematic approach of initial data analysis (IDA) using a cluster heat map for data visualization. This IDA is useful especially for exploratory research investigating gait abnormality in rodents. Since rodent behavioral studies typically focus on observing the differences between the two or more cohorts of subjects, the IDA demonstrated in this paper is based on the analysis of between-group-differences and/or intra-individual-longitudinal differences. In this paper, we employ the IDA technique in an experiment with endpoint-parameters captured on CatWalk system, exemplifying a standard experimental design and parameters derived from exploratory research. Our proposed IDA could also be used to offer insight on the data for parameter selection in a multivariate data analysis, especially to evade the *curse of dimensionality* [5,26], which arises due to the limited number of subjects within a cohort.

Initial Data Analysis (IDA)

Authors in [27] define IDA as a systematic approach to data inspection and screening. This process is carried out after the data collection but before the formal statistical analysis. The framework of IDA consists of three main steps: (1) Data cleaning, which is a practice of identifying inconsistencies, faulty or outlying data [28,29] (2) Data screening, which aims to describe the data (3) Data reporting, which depicts information for further analysis. Application of those IDA steps on the CatWalk gait parameters reads as follows:

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

1. Data cleaning by first choosing the "compliant" runs from all the recorded runs, then omitting the runs that contain unwanted behaviors, such as sniffing, rearing, sitting, stopping and turning on the alley. The "compliant" runs are selected by the CatWalk software according to the run's duration and speed variation [7]. The unwanted behavior selection is usually completed by an experienced observer.
2. Data screening by calculating differences in gait parameters between cohorts.
3. Data reporting by constructing a cluster heat map.

Data Screening by Analysis of Differences

The between-cohort differences of the studied parameters are then calculated. There are many calculations of differences that can be used in the data screening, including:
1. f-value, which is the ratio of the between-group variability and the within-group variability [30,31].
2. p-value of a t-test, which compares two group means [32].
3. Estimated arithmetic mean, which is a value based on the groups' means and standard deviations. This value is calculated based on the intersection point of the estimated groups' densities [33,34].
4. D-value, which is the ratio of mean difference and standard deviation [6].

Noteworthy, other calculations of differences derived from descriptive statistics are suitable for the data screening. Calculation of difference based on the computation of covariance might not be applicable, especially when there is a limited number of subjects. The choice descriptive calculations could be made based on the consecutive data analysis or requirements derived from the objectives of the study. In cases where there is a need to discern the change direction, signed difference calculation might be particular benefit. A chosen threshold, T_h, is then needed to decide whether a parameter is considered as different.

Data Reporting by Cluster Heat Mapping Approach

Heat maps, two-dimensional color-shaded matrix displays, have been used for visualizing multivariate data in social statistic [35], biology [36], accounting, graphic designing, computer engineering, and others [37]. Data with many rows and columns can be well displayed by a heat map. In designing a heat map, it is important to cluster the rows and columns by their similarities in order to make a heat map easily understandable [38]. This equals the application criteria for clustering of rows and columns.

The data report of CatWalk parameters aims to visualize the differences between cohorts/treatments/genotypes etc. As a report of differences, the colors in a cluster heat map should represent the difference between the cohorts. This color-coding enables easy apprehension of hidden information.

Example

Subjects
As an example of the systematic approach to initial gait analysis, we use an experiment involving 24 C57BL/6N male mice at the age of 7 months, which are grouped into two cohorts: (a) 12 mice of wild-type ("WT") (b) 12 mice of knock-out for the endogenous murine α-synuclein ("KO") [39–41]. Some of the mice in both cohorts received a treatment for 4 weeks. Therefore, after 4 weeks the subjects can be grouped into four cohorts. However for the simplicity of the cluster heat map explanation, we will only use three after-4-week cohorts: (a) 8 WT mice without treatment ("WN") (b) 6 KO mice without treatment ("KN") (c) 6 KO mice with treatment ("KE"). However, the WT mice with treatment ("WE") should not be omitted in the real data analysis. These animals are kept under standard laboratory conditions under Specific-Pathogen-Free (SPF) condition. All research and animal care procedures were performed according to international guidelines and approved by the local animal welfare and ethics committee of Bavaria, Germany (RegUFr#55.2-2532-2-218).

Data Acquisition
The gait parameters were assessed by using CatWalk system before and after the 4-week treatment. During the data acquisition, each mouse could voluntarily walk across the glass plate. Data from 2 - 4 "compliant" runs were collected from each mouse at each time point. The measured gait parameters from all runs for each time point were averaged. Therefore, the experimental unit in this example is the individual animals.

Data Screening
While usually, a non-hypothesis driven approach including all CatWalk parameters is optimal, the hypothesis-driven pre-selection of parameters is often applied by experienced researchers working on well-defined models, which do not require a comprehensive survey. For the purpose of IDA presented in this example, however, a condensation and averaging of certain parameters (body speed calculated from the paws, left/right-paw gait parameters) for the sake of apprehensiveness and ease-of-use was diligently applied, avoiding decreased screening

sensitivity. Therefore, right and left paw measurements were first averaged and the body speed (cm/s) were averaged from four paws. Moreover, the parameters directly connected to paw's intensity and show identical values between the groups, are neglected. The intensity-related parameters were not considered due to their high correlation with the variated software setting in this particular example. Therefore, in this example, there are in total 189 gait parameters of interest. Based on these 189 parameters, we screened:

1. The parameters, which are important in characterizing the disorder, i.e. comparing WT and KO.
2. The parameters, from all the important parameters above, which show differences between WT and KO after four weeks without any treatment, i.e. comparing WN and KN.
3. The parameters, from all the important parameters above, which improve by the four-week treatment on KO mice, i.e. comparing KE and KN.
4. The parameters, from all the important parameters above, which make the treated KO mice differentiable with the untreated WT, i.e. comparing KE and WN.
5. The parameters, from all the important parameters above, which show differences after 4 weeks in both KO and WT mice.

The data screening in this example is done by calculating the signed f-values of each gait parameters between cohorts. A gait parameter is considered different between the cohorts if the f-value is above T_h = 5. The signed f-values in this example were calculated by using Matlab R2015a (8.5.0.).

Data Reporting

A cluster heat map is suitable for the data reporting despite an extensive number of parameters. Since signed f-values are used as the descriptor, we defined graded shades of the color red to depict an increase, and graded shades of blue to describe a decrease.

The clusters in the heat map columns should be organized according to the experimental design. In this example, the columns in the heat map is arranged as follows: The first column depicts the differences between the cohorts at baseline (WT and KO). The next three columns present the differences between the parameters before and after the treatment of both cohorts. The next two columns display the differences after the treatment in the KO mice. Finally, the last column shows the differences between the cohorts after 4 weeks without any treatment. The column arrangement in this example is depicted in Figure 2.

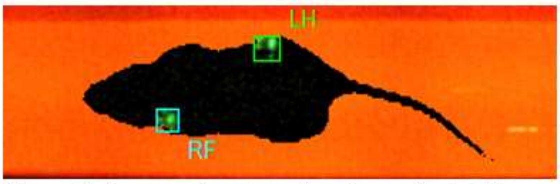


Figure 1. An example of a capture image derived from a recorded CatWalk video, which shows the detected footprints (RF: Right Front, LH: Left Hind).

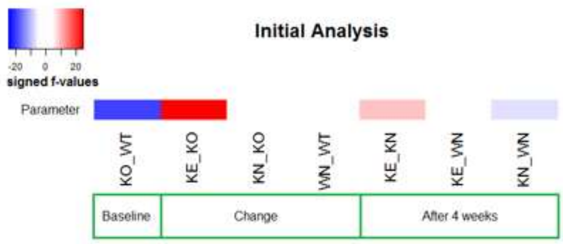


Figure 2. An example of column arrangement in a heat map and an example of a parameter, which is improved by the treatment in the knock-out mice.

The heat map analysis of an example parameter shown in Figure 2 is explained as follows: The color blue in the first column informs us that at baseline the WT mice have a higher parameter value as compared to the KO mice. The color red in the second column informs us that the KO mice with treatment show improvement. The third and the fourth columns explain that there is no improvement in the mice without treatment. The color red in the fifth column informs us that after 4 weeks the with-treatment KO mice perform differently from the without-treatment KO mice. The sixth column explains that after 4 weeks the with-treatment KO mice perform similar with the without-treatment WT mice. The last column informs us that the difference between the KO and WT mice is still noticeable after 4 weeks without any treatment.

The rows in the heat map are ordered according to the parameter similarity. Here the parameters are first clustered according to their importance in differentiating WT and KO using the color code. The parameters, which show differences and higher values for WTs, have negative signed f-values and are clustered together. Conversely, the parameters, which show differences and higher values for KOs, have a positive signed f-values. Next, the parameters are organized according to the between-cohort differences shown at the fourth week without having any treatment (WN and KN). The parameters are afterward organized according to the differences expressed by the treated KO mice (KE) compared with the other cohorts at the fourth week. Subsequently, the parameters are arranged according to their changes before and after 4 weeks of treatment.

The resulted cluster heat map for the 189 parameters of this example is shown in Figure 3 and was generated using R 3.3.0. To make the information in a cluster heat map assessable, it should be accompanied with a table of the important parameter names. In this example, it is listed in Table 1.

Data Analysis

The resulted cluster heat map from the example (Figure 3) suggests the following:

1. According to the results shown in the first column, there are 64 gait parameters that might be important in characterizing KO from the WT. 17 parameters show lower values from the WTs (negative signed f-values), and 47 parameters show higher values compare with the WTs (positive signed f-values). These parameters are highlighted in Figure 2 and are the base for the next analysis.
2. The results in the seventh column present the comparison after 4 weeks without any treatment. From the 64 gait parameters, 13 parameters show differences between the groups without treatment (KN and WN). These 13 parameters are listed in Table 1. These results show that the other 51 parameters have enough improvement even without any treatment.
3. The fifth column shows that from the 64 gait parameters, only two parameters show differences between the KO mice with and without treatment (KE and KN). These two parameters are shown in Table 1. These two parameters are possibly the main advantage of the treatment. However, a further formal statistical analysis is needed to confirm this.
4. The sixth column shows that from the 64 gait parameters, only one parameter, i.e. run maximum speed variation in % (parameter number 54), indicates a difference between KO with treatment and WT without treatment (KE and WN) with a not very high f-value (5.9). Since this parameter indicates a difference between KE and WN, but not between KN and WN, and with the knowledge that WN should have a low-speed variation (%), the speed variation (%) in KE is possibly higher than in KN. However, a deeper analysis is required to certify this.
5. The results shown in the second column through the fourth column indicate that from the 64 gait parameters, some parameters have considerable changes in both cohorts. The parameter changes are observable more in the KO with treatment (KE) compare with in the KO without treatment (KN).

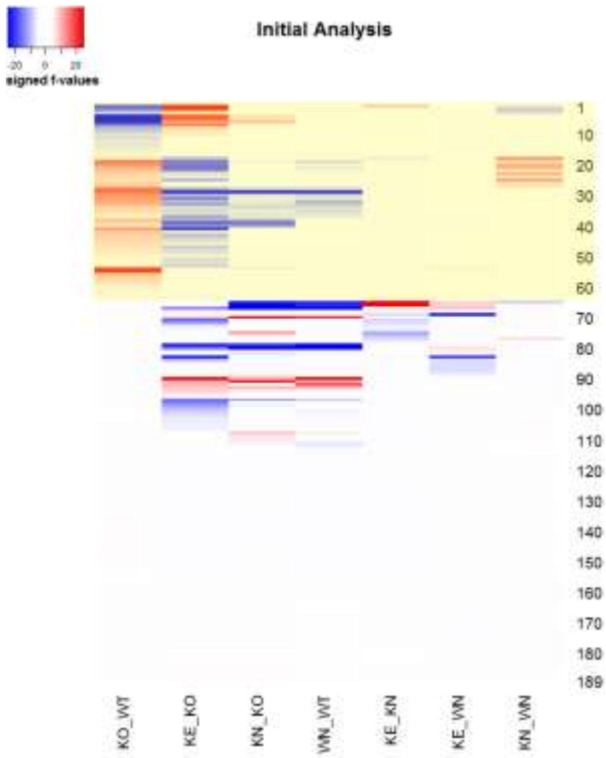


Figure 3. A clustered heat map example taken from a CatWalk experiment. Red depicts an increase, blue depicts a decrease, and yellow shading indicate the 64 important parameters showing differences between cohorts in the baseline.

Table 1. Gait parameters from the 64 important parameters, which indicate differences between cohorts in baseline and after 4 weeks of treatment

Param. Number	**Parameter Names**	**Signed f-value**
	knock-out with and without treatment (KE – KN)	
1	Hind Paws Swing Speed Mean (cm/s)	9.3
18	Front Paws Swing Mean (s)	-6.0
	knock-out and wild-type without treatment (KN – WN)	
18	Front Paws Swing (s) Mean	14.6
25	Front Paws Single Stance (s)	13.2
20	Mean	11.5
19	Front Paws Step Cycle (s) Mean	7.2
23	Front Paws Stand (s) Mean	11.9
21	Hind Paws Swing (s) SD	11.5
1	Hind Paws Step Cycle (s) Mean	-6.9
24	Hind Paws Swing Speed (cm/s)	6.6
27	Mean	7.7
3	Hind Paws Single Stance (s) SD	-7.5
26	Phase Disp. RF->RH SD	7.1
2	Phase Disp. RF->RH CStat R	-9.8
22	Phase Disp. RF->RH CStat SD	5.5
	Cadence	
	Body Speed Variation (%) SD	

SD: Standard Deviation;
R: the strength of directedness, measured the variation of CStat;
CStat: Circular Statistic;
RF: Right Front; RH: Right Hind

Summary: We applied an Initial Data Analysis (IDA) for visualization of CatWalk derived gait parameters by employing a cluster heat map. We exemplify the CatWalk-IDA approach by applying it to a classical before-intervention-after-intervention design in mice and learned that IDA helps the investigator to more easily capture group differences and to condense information. Of note, formal statistical analysis should not be shrunk or omitted and always applied as the next step. This systematic approach has the prospect to be applied in a human gait analysis and in selecting parameters for a multivariate data analysis.

Acknowledgements
Ivanna K. Timotius acknowledges the support from Deutscher Akademischer Austauschdienst (DAAD) by the research grant (number 57129429). Fabio Canneva, Stephan von Hörsten, and Sandra Moceri acknowledge the support by the JPND project CrossSeeds (grant number GRANT#01ED1501C). Jochen Klucken, Georgia Minakaki, and Jürgen Winkler acknowledge the support of the Interdisciplinary Center of Clinical Research (IZKF) of the University Hospital Erlangen, Germany and the Bavarian State Ministry of Education and Culture, Science and Arts in the framework of the Bavarian Research Network for Induced Pluripotent Stem Cells (ForIPS). Moreover, Jochen Klucken acknowledges the Deutsche Forschungsgemeinschaft/German Research Foundation (DFG, no. KL1395/8-1). Nicolas Casadei and Olaf Riess were supported by NEURASYNC: an Academic Industrial Initial Training Network (ITN) through the people program FP7 of the European Union and FCT (PTDC/SAU-NEU/105215/2008) grant. Silke Nuber is acknowledged for her kind help in transferring the mouse model from Tübingen to the study site in Erlangen at Preclinical Experimental Animal Center (PETZ). Bjoern Eskofier gratefully acknowledges the support of the German Research Foundation (DFG) within the framework of the Heisenberg professorship program (grant number ES 434/8-1).

References

1. Jankovic, J. (2008) Parkinson's disease: clinical features and diagnosis. *J. Neurol. Neurosurg. Psychiatry*, **79**,368–376. doi:10.1136/jnnp.2007.131045.
2. Lees, A.J., Hardy, J., and Revesz, T. (2009) Parkinson's disease. *Lancet*, **373**,2055–2066. doi:10.1016/S0140-6736(09)60492-X.
3. Tolosa, E., Wenning, G., and Poewe, W. (2006) The diagnosis of Parkinson's disease. *Lancet Neurol.*, **5**,75–86. doi:10.1016/S1474-4422(05)70285-4.
4. Roos, R.A. (2010) Huntington's disease: a clinical review. *Orphanet J. Rare Dis.*, **5**,40. doi:10.1186/1750-1172-5-40.
5. Salimi, A., Ziaii, M., Amiri, A., Zadeh, M.H., Karimpouli, S., and Moradkhani, M. (2017) Using a Feature Subset Selection method and Support Vector Machine to address curse of dimensionality and redundancy in Hyperion hyperspectral data classification. doi:10.1016/j.ejrs.2017.02.003.
6. Festing, M.F.W. and Altman, D.G. (2002) Guidelines for the Design and Statistical Analysis of Experiments Using Laboratory Animals. *ILAR J.*, **43**,244–258.
7. Noldus Information Technology b.v. (2012) CatWalk XT version 10.0. reference manual. Wageningen, The Netherlands: Author.
8. Hamers, F.P.T., Lankhorst, A.J., Jan, T., Laar, V., Veldhuis, W.B., and Gispen, W.H. (2001) Automated quantitative gait analysis during overground locomotion in the rat: Its application to spinal cord contusion and transection injuries. *J. Neurotrauma*, **18**,187–201.
9. Hamers, F.P.T., Koopmans, G.C., and Joosten, E.A.J. (2006) CatWalk-assisted gait analysis in the assessment of spinal cord injury. *J. Neurotrauma*, **23**,537–548. doi:10.1089/neu.2006.23.537.
10. Hayakawa, K., Uchida, S., Ogata, T., Tanaka, S., Kataoka, K., and Itaka, K. (2015) Intrathecal injection of a therapeutic gene-containing polyplex to treat spinal cord injury. *J. Control. Release*, **197**,1–9. doi:10.1016/j.jconrel.2014.10.027.
11. Hendriks, W.T.J., Eggers, R., Ruitenberg, M.J., Blits, B., Hamers, F.P.T., Verhaagen, J., and Boer, G.J. (2006) Profound Differences in Spontaneous Long-Term Functional Recovery after Defined Spinal Tract Lesions in the Rat. *J. Neurotrauma*, **23**,18–35.
12. Chen, Y.-J., Cheng, F.-C., Sheu, M.-L., Su, H.-L., Chen, C.-J., Sheehan, J., and Pan, H.-C. (2014) Detection of subtle neurological alterations by the Catwalk XT gait analysis system. *J. Neuroeng. Rehabil.*, **11**,62. doi:10.1186/1743-0003-11-62.
13. Miyagi, M., Ishikawa, T., Kamoda, H., et al. (2013) Assessment of Pain Behavior in a Rat Model of Intervertebral Disc Injury Using the Catwalk Gait Analysis System. *Spine (Phila. Pa. 1976).*, **38**,1459–1465. doi:10.1097/BRS.0b013e318299536a.
14. Abada, Y.K., Schreiber, R., and Ellenbroek, B. (2013) Motor, emotional and cognitive deficits in adult BACHD mice: A model for Huntington's disease. *Behav. Brain Res.*, **238**,243–251. doi:10.1016/j.bbr.2012.10.039.
15. Casadei, N., Pöhler, A.-M., Tomás-Zapico, C., et al. (2014) Overexpression of synphilin-1 promotes clearance of soluble and misfolded alpha-synuclein without restoring the motor phenotype in aged A30P transgenic mice. *Hum. Mol. Genet.*, **23**,767–781. doi:10.1093/hmg/ddt467.

16. Saal, K.A., Koch, J.C., Tatenhorst, L., et al. (2015) AAV.shRNA-mediated downregulation of ROCK2 attenuates degeneration of dopaminergic neurons in toxin-induced models of Parkinson's disease in vitro and in vivo. *Neurobiol. Dis.*, **73**,150–162. doi:10.1016/j.nbd.2014.09.013.
17. Wang, X.H., Lu, G., Hu, X., et al. (2012) Quantitative assessment of gait and neurochemical correlation in a classical murine model of Parkinson's disease. *BMC Neurosci.*, **13**,142. doi:10.1186/1471-2202-13-142.
18. Zhou, M., Zhang, W., Chang, J., et al. (2015) Gait analysis in three different 6-hydroxydopamine rat models of Parkinson's disease. *Neurosci. Lett.*, **584**,184–189. doi:10.1016/j.neulet.2014.10.032.
19. Plucińska, K., Crouch, B., Koss, D., Robinson, L., Siebrecht, M., Riedel, G., and Platt, B. (2014) Knock-in of human BACE1 cleaves murine APP and teiterates Alzheimer-like phenotypes. *J. Neurosci.*, **34**,10710–10728. doi:10.1523/JNEUROSCI.0433-14.2014.
20. Bernardes, D., Leite, A., and Oliveira, R. (2017) Comprehensive catwalk gait analysis in a chronic model of multiple sclerosis subjected to treadmill exercise training. *BMC Neurol.*, **17**,160. doi:10.1186/s12883-017-0941-z.
21. Ishikawa, G., Nagakura, Y., Takeshita, N., and Shimizu, Y. (2014) Efficacy of drugs with different mechanisms of action in relieving spontaneous pain at rest and during movement in a rat model of osteoarthritis. *Eur. J. Pharmacol.*, **738**,111–117. doi:10.1016/j.ejphar.2014.05.048.
22. Qin, L., Jing, D., Parauda, S., Carmel, J., Ratan, R.R., Lee, F.S., and Cho, S. (2014) An adaptive role for BDNF Val66Met polymorphism in motor recovery in chronic stroke. *J. Neurosci.*, **34**,2493–2502.
23. Salvi, J., Bertaso, F., Mausset-Bonnefont, A.-L., et al. (2014) RNAi silencing of P/Q-type calcium channels in Purkinje neurons of adult mouse leads to episodic ataxia type 2. *Neurobiol. Dis.*, **68**,47–56. doi:10.1016/j.nbd.2014.04.005.
24. Kyriakou, E.I., van der Kieft, J.G., de Heer, R.C., Spink, A., Nguyen, H.P., Homberg, J.R., and van der Harst, J.E. (2016) Automated quantitative analysis to assess motor function in different rat models of impaired coordination and ataxia. *J. Neurosci. Methods*, **268**,171–181. doi:10.1016/j.jneumeth.2015.12.001.
25. Tatenhorst, L., Eckermann, K., Dambeck, V., et al. (2016) Fasudil attenuates aggregation of α-synuclein in models of Parkinson's disease. *Acta Neuropathol. Commun.*, **4**. doi:10.1186/s40478-016-0310-y.
26. Richard E. Bellman. (1961) Adaptive Control Processes: A Guided Tour. Princenton University Press.
27. Huebner, M., Vach, W., and Le Cessie, S. (2016) A systematic approach to initial data analysis is good research practice. *J. Thorac. Cardiovasc. Surg.*, **151**,25–27. doi:10.1016/j.jtcvs.2015.09.085.
28. Van den Broeck, J., Argeseanu Cunningham, S., Eeckels, R., and Herbst, K. (2005) Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities. *PLoS Med.*, **2**,e267. doi:10.1371/journal.pmed.0020267.
29. Dempster, A.P. (1971) An Overview of Multivariate Data Analysis. *J. Multivar. Anal.*, **1**,316–346.
30. Fisher, R.A. (1921) Studies in crop variation. I. An examination of the yield of dressed grain from broadbalk. *J. Agric. Sci.*, **11**,107–135.
31. Armstrong, R.A., Slade, S. V., and Eperjesi, F. (2000) An introduction to analysis of variance (ANOVA) with special reference to data from clinical experiments in optometry. *Ophthalmic Physiol. Opt.*, **20**,235–241. doi:10.1046/j.1475-1313.2000.00502.x.
32. Mcdonald, J.H. (2008) Handbook of Biological Statistics. Baltimore, Maryland: Sparky House.
33. Timotius, I.K. and Setyawan, I. (2013) Using estimated arithmetic means of accuracies to select features for face-based gender classification. Int.Conf. on Information Technology and Electrical Engineering (ICITEE). Yogjakarta: IEEE. pp. 242–247. doi:10.1109/ICITEED.2013.6676246.
34. Timotius, I.K. (2013) Predicting Receiver Operating Characteristic Curve, Area Under Curve, and Arithmetic Means of Accuracies based on the Distribution of Data Samples. Int. Conf. on Control, Systems and Industrial Informatics (ICCSII). Bandung. pp. 60–64.
35. Loua, T. (1873) Atlas statistique de la population de Paris. Paris: J. Dejey.
36. Weinstein, J.N. (2008) A Postgenomic Visual Icon. *Science*, **319**,1772–1773. doi:10.1126/science.1151888.
37. Wilkinson, L. and Friendly, M. (2009) The History of the Cluster Heat Map. *Am. Stat.*, **63**,179–184. doi:10.1198/tas.2009.0033.
38. Gehlenborg, N. and Wong, B. (2012) Points of View: Heat Maps. *Nat. Methods*, **9**,213. doi:10.1038/nmeth.1902.
39. Abeliovich, A., Schmitz, Y., Fariñas, I., et al. (2000) Mice lacking alpha-synuclein display functional deficits in the nigrostriatal dopamine system. *Neuron*, **25**,239–252.
40. Timotius, I.K., Canneva, F., Minakaki, G., et al. (2018) Dynamic footprint based locomotion sway assessment in α-synucleinopathic mice using Fast Fourier Transform and Low Pass Filter. *J. Neurosci. Methods*, **296**,1–11. doi:10.1016/j.jneumeth.2017.12.004.

41. Timotius, I.K., Canneva, F., Minakaki, G., et al. (2018) Dynamic footprints of α-synucleinopathic mice recorded by CatWalk gait analysis. *Data Br.*, **17**,189–193. doi:10.1016/J.DIB.2017.12.067.

Rodent's Stride Length Depends on Body Size: Implications for CatWalk Assay

I.K. Timotius[1,5], S. Moceri[2], A.C. Plank[2], J. Habermeyer[2], F. Canneva[2], N. Casadei[3], O. Riess[3], J. Winkler[4], J. Klucken[4], B. Eskofier[1] and S. von Hörsten[2]

[1]Machine Learning and Data Analytics Lab, Department of Computer Science, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Erlangen, Germany. ivanna.timotius@fau.de [2]Department of Experimental Therapy, University Hospital Erlangen (UKEr) and Preclinical Experimental Animal Center, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Erlangen, Germany. stephan.v.hoersten@uk-erlangen.de [3]Institute of Medical Genetics and Applied Genomics, University of Tübingen, Germany [4]Department of Molecular Neurology, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Erlangen, Germany. [5]Department of Electronics Engineering, Satya Wacana Christian University, Salatiga, Indonesia.

The analysis of gait is an essential part of rodent phenotyping. Regarding the complex nature of walking, preclinical gait analysis is normally based on several gait parameters (which have been progressively developed [1,2] in recent years) including the animal's stride length, body speed, support, etc. The stride length is altered in several diseases [3–5] and this parameter can be easily translated from rodent models to human patients. However, we hypothesized that the stride length is affected by the body size [6,7], which becomes a challenge in analyzing gait.

Using the gait parameters assessed by CatWalk system [8–10], we examined the correlation between body-length/body-weight with stride length using CatWalk-video-derived body silhouette length. To examine the effect of growth and aging in rodents, we studied wild-type Sprague-Dawley male rats and C57BL/6N mice. Moreover, we examined the body size difference between genotypes in BACSCNA transgenic rats [11,12] and BACHD transgenic mice [13]. The CatWalk data were collected as the rodents walked freely on top of a glass-floored corridor. The rat gait was monitored at 4 different age points (10, 26, 55 and 62 weeks old), whereas the mice were monitored at 3 different age points (20, 32, and 47 weeks old). The rodent numbers included in the experiments were: (a) wild-type male rats (n=16-27/time point), (b) BACSCNA rats (n=19-32/time point), (c) wild-type mice (n=12-13/time point), and (d) BACHD mice (n=9-11/time point). The rodents were maintained under specific-pathogen-free condition. All research and animal care procedures were performed in compliance with international animal welfare standards and approved by the district governments of Lower Franconia, Würzburg, Bavaria, Germany (RegUFr#55.2-2532-2-218).

The correlations between front stride length (averaged from the left and right side) and the body-silhouette-length/body-weight are shown in Figure 1. Significant correlations are shown with both body silhouette lengths and body weight. The correlation between hind stride length and body size showed the similar relationship as the front stride length (data not shown). Besides body size differences due to growth, pathological body size differences were also observed as shown in Figure 2.

Thus, we demonstrated that stride length is highly correlated with body silhouette length (based on the CatWalk video) and body weight. In the near future, we need to subject gait parameters to a scaling process using body weight or silhouette length, raising the possibility of normalization for differences in size as a potential confounding of gait measures.

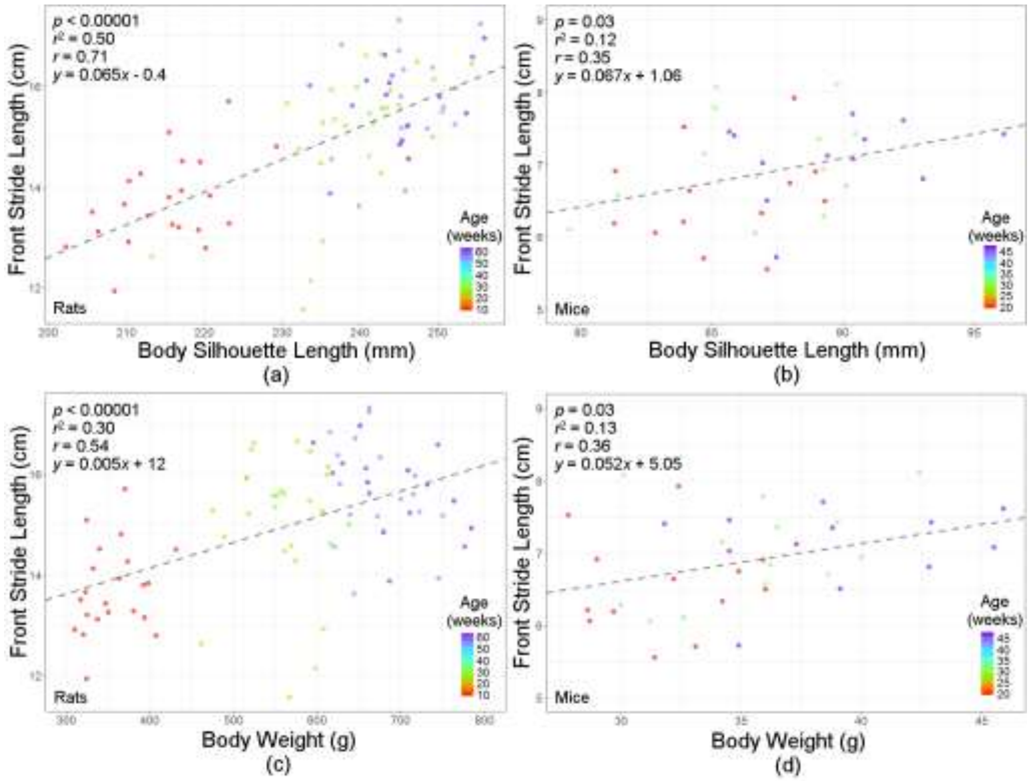
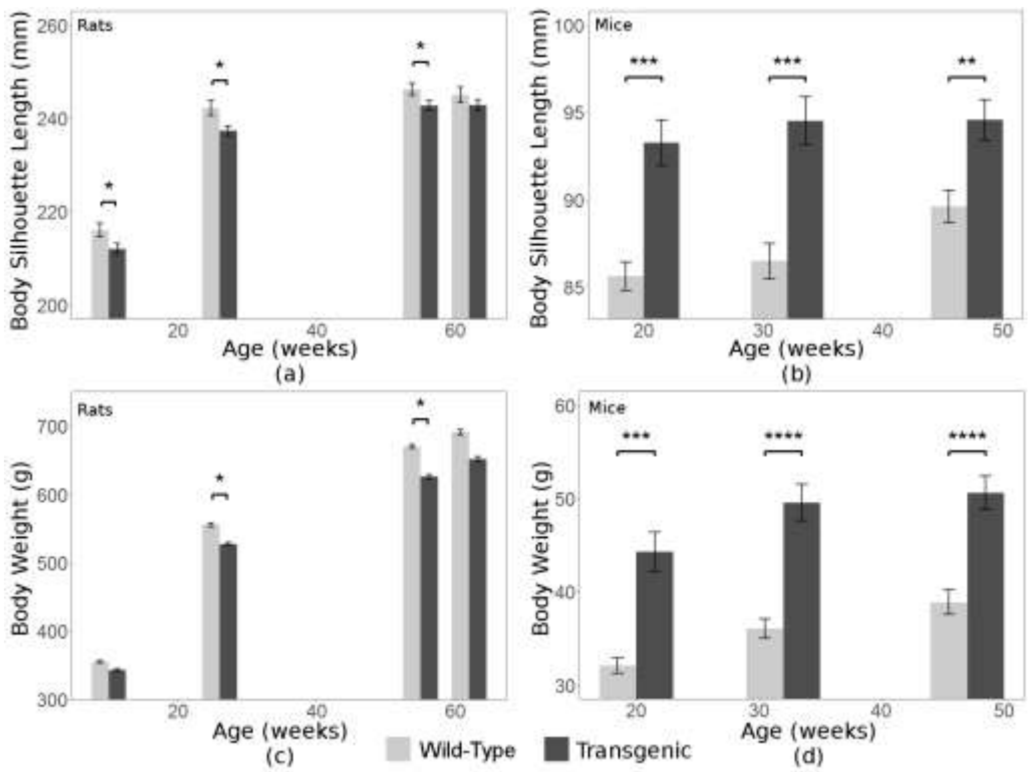


Figure 1. Scatter plot for the correlation of front stride length and body length/weight in (a & c) wild-type rats, body silhouette length and body weight are correlated with $p < 0.00001$ $r^2 = 0.81$, $r = 0.90$ (b & c) wild-type mice, correlation between body silhouette length and body weight is $p < 0.00001$ $r^2 = 0.57$, $r = 0.75$



R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Figure 2. Body silhouette length and body weight of neurodegenerative rodent models compared with their wild-type rodents as a function of age (a & c) BACSCNA rats (b & d) BACHD mice. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$.

Acknowledgements
Ivanna K. Timotius acknowledges the support from Deutscher Akademischer Austauschdienst (DAAD) by the research grant (number 57129429). Stephan von Hörsten, Sandra Moceri, Anne-Christine Plank, Johanna Habermeyer, and Fabio Canneva acknowledge the support by the JPND project CrossSeeds (grant number GRANT#01ED1501C). Jochen Klucken and Jürgen Winkler acknowledge the support of the Interdisciplinary Center of Clinical Research (IZKF) of the University Hospital Erlangen, Germany and the Bavarian State Ministry of Education and Culture, Science and Arts in the framework of the Bavarian Research Network for Induced Pluripotent Stem Cells (ForIPS). Moreover, Jochen Klucken acknowledges the Deutsche Forschungsgemeinschaft/German Research Foundation (DFG, no. KL1395/8-1). Nicolas Casadei and Olaf Riess were supported by NEURASYNC: an Academic Industrial Initial Training Network (ITN) through the people program FP7 of the European Union and FCT (PTDC/SAU-NEU/105215/2008) grant. Silke Nuber is acknowledged for her kind help in transferring the mouse model from Tübingen to the study site in Erlangen at Preclinical Experimental Animal Center (PETZ). Bjoern Eskofier gratefully acknowledges the support of the German Research Foundation (DFG) within the framework of the Heisenberg professorship program (grant number ES 434/8-1).

References

1. Timotius, I.K., Canneva, F., Minakaki, G., et al. (2018) Dynamic footprint based locomotion sway assessment in α-synucleinopathic mice using Fast Fourier Transform and Low Pass Filter. *J. Neurosci. Methods*, **296**,1–11. doi:10.1016/j.jneumeth.2017.12.004.
2. Timotius, I.K., Canneva, F., Minakaki, G., et al. (2018) Dynamic footprints of α-synucleinopathic mice recorded by CatWalk gait analysis. *Data Br.*, **17**,189–193. doi:10.1016/J.DIB.2017.12.067.
3. Hannink, J., Kautz, T., Pasluosta, C.F., Barth, J., Schülein, S., Gaßmann, K.-G., Klucken, J., and Eskofier, B.M. (2017) Mobile Stride Length Estimation with Deep Convolutional Neural Networks. *IEEE J. Biomed. Heal. Informatics*,. doi:10.1109/JBHI.2017.2679486.
4. Danoudis, M. and Iansek, R. (2014) Gait in Huntington's disease and the stride length-cadence relationship. *BMC Neurol.*, **14**. doi:10.1186/s12883-014-0161-8.
5. Amende, I., Kale, A., McCue, S., Glazier, S., Morgan, J.P., and Hampton, T.G. (2005) Gait dynamics in mouse models of Parkinson's disease and Huntington's disease. *J. Neuroeng. Rehabil.*, **2**. doi:10.1186/1743-0003-2-20.
6. Heglund, N.C., Taylor, R.C., and McMahon, T.A. (1974) Scaling Stride Frequency and Gait to Animal Size: Mice to Horses. *Science*, **186**,1112–1113.
7. Neckel, N.D. (2015) Methods to Quantify the Velocity Dependence of Common Gait Measurements from Automated Rodent Gait Analysis Devices. *J. Neurosci. Methods*, **253**,244–253. doi:10.1016/j.jneumeth.2015.06.017.
8. Hamers, F.P.T., Lankhorst, A.J., Jan, T., Laar, V., Veldhuis, W.B., and Gispen, W.H. (2001) Automated quantitative gait analysis during overground locomotion in the rat: Its application to spinal cord contusion and transection injuries. *J. Neurotrauma*, **18**,187–201.
9. Hamers, F.P.T., Koopmans, G.C., and Joosten, E.A.J. (2006) CatWalk-assisted gait analysis in the assessment of spinal cord injury. *J. Neurotrauma*, **23**,537–548. doi:10.1089/neu.2006.23.537.
10. Noldus Information Technology b.v. (2012) CatWalk XT version 10.0. reference manual. Wageningen, The Netherlands: Author.
11. Nuber, S., Harmuth, F., Kohl, Z., et al. (2013) A progressive dopaminergic phenotype associated with neurotoxic conversion of a-synuclein in BAC-transgenic rats. *Brain*, **136**,412–432. doi:10.1093/brain/aws358.
12. Kohl, Z., Abdallah, N. Ben., Vogelgsang, J., et al. (2016) Severely impaired hippocampal neurogenesis associates with an early serotonergic deficit in a BAC α-synuclein transgenic rat model of Parkinson's disease. *Neurobiol. Dis.*, **85**,206–217. doi:10.1016/j.nbd.2015.10.021.
13. Gray, M., Shirasaki, D.I., Cepeda, C., et al. (2008) Full-Length Human Mutant Huntingtin with a Stable Polyglutamine Repeat Can Elicit Progressive and Selective Neuropathogenesis in BACHD Mice. *J. Neurosci.*, **28**,6182–6195. doi:10.1523/JNEUROSCI.0857-08.2008.

Measuring and influencing behavior using biological and neural measures in humans

Modern technology have enabled psychologists to enhance our understanding of and predictive power for human behavior in health and disease, by allowing us to closely link observable behaviors to neural activations (fMRI and EEG), brain structure (MRI) and biological markers (e.g. sweat and heart-rate as a measure of stress). Furthermore, recent developments have enabled techniques of brain stimulation (e.g., tDCS) to be combined with behavioral measures to enhance behaviors such as learning. In this symposium, we will introduce how fNRI, fMRI, EEG, tDCS and biological measures of stress, have been used in our research to measure, understand, and influence human behavior in healthy adults and in patients.

Yale Benn, Manchester Metroplitan University

Using fMRI to Investigate Human Preferences in Information Behaviour

Y. Benn[1] and O. Bergman[2]

1. Manchester Metropolitan University Department of Psychology, Manchester, UK. e-mail: y.benn@mmu.ac.uk
2. Bar Ilan University, Department of Information Science, Israel, e-mail: oferbergman@gmail.com

Introduction

Functional Magnetic Resonance imaging (fMRI), is nowadays a common technique for examining brain activation. The technique relies on measuring Blood Oxygenation Level Dependence (BOLD), assuming that brain activity takes place where oxygen is being consumed. The techniques has allowed for great advances in understanding human cognition and behaviour. In this talk, we will demonstrate how using fMRI, we have been able to address an important long-standing question in Information Sciences. In doing so, we will focus on barriers and difficulties in using the technique for studying human behaviour, as well as difficulties in analysis and interpretation of the data.

Specific information science background

Efficient storage and retrieval of digital data is the focus of much commercial and academic attention. With personal computers, there are two main ways to retrieve files: navigating the folder structure and using the search-box to find a file. In navigation, users move down their virtual folder hierarchy until they reach the target item. When searching, users first generate a query specifying some property of the target file (e.g., a word it contains), and then select the relevant file when the search engine returns a set of results. Despite advances in search technology, users prefer retrieving files using virtual folder navigation [1,2], rather than the more flexible search facility.

The study and results

We used fMRI to investigate the preference for folder-navigation over search. Data was collected from 17 healthy right-handed students (11 females), aged between 19–29 years (mean = 22.8, SD = 2.87), who took part in exchange for a small financial reward. While in the fMRI, participants were asked to either 'search' and 'navigate' for files extracted from their computers the previous day. All MR images were acquired at 3T (Ingenia 3.0T, Philips Healthcare, Best, Holland) using a fifteen-channel radiofrequency receive-only head coil. Cerebral vascular response to the tasks was recorded using the BOLD, T2*-weighted signal time-course. During each functional scan, a time series of 153 dynamic datasets were obtained using a 2-dimensional single-shot, echo-planar imaging (EPI) sequence. The EPI scan parameters were as follows: repetition time (TR) = 3000 ms; echo time (TE) = 35 ms; sensitivity-encoding factor = 1.8; flip angle = 90°; in-plane voxel size = 2.4 mm × 2.4 mm interpolated to 1.8 mm × 1.8 mm; 35 contiguous 2-dimensional transaxial slices each having slice thickness = 4mm. Analysis was done using SPM8 (Wellcome Department of Imaging Neuroscience, London; www.fil.ion.ucl.ac.uk/spm/) implemented in Matlab (The MathWorks Inc., Natick, MA).

Results revealed that navigation resulted in activation of regions similar to that previously observed during real-world navigation in both animals and humans [e.g. 3-5]. In contrast, searching activated 'classic' language regions (Broca's area) [6]. Implication of the results will be discussed.

Ethical statement:

Ethics for the study was granted by the Ethics committee of the University of Sheffield, UK.

References:

1. Bergman, O., Beyth-Marom, R., Nachmias, R., Gradovitch, N. & Whittaker, S. (2008). Improved Search Engines and Navigation Preference in Personal Information Management. *ACM Trans. Inf. Syst.* **26**, 1–24,
2. Whitham, R., & Cruickshank, L. (2017). The function and future of the folder. *Interacting with Computers*, *29*(5), 629-647.

3. Furtak, S. C., Wei, S. M., Agster, K. L., & Burwell, R. D. (2007). Functional neuroanatomy of the parahippocampal region in the rat: the perirhinal and postrhinal cortices. *Hippocampus*, *17*(9), 709-722.
4. Maguire, E. A., Frackowiak, R. S., & Frith, C. D. (1997). Recalling routes around London: activation of the right hippocampus in taxi drivers. *Journal of neuroscience*, *17*(18), 7103-7110.
5. Vann, S. D., Aggleton, J. P., & Maguire, E. A. (2009). What does the retrosplenial cortex do?. *Nature Reviews Neuroscience*, *10*(11), 792.
6. Benn Y., Bergman, O., Glazer, L., Arent, P., Varley, R., & Whitaker, S. (2015). Navigating through digital folders uses the same brain structures as real world navigation. *Scientific Reports*, 14719

Studying stress effects on behaviour by combining subjective, endocrine, physiological and neural measures

Dr Maria I. Cordero[1]

[1]Department of Psychology, Manchester Metropolitan University, Manchester, UK. M.Cordero@mmu.ac.uk

Introduction

Studying the effects of stress on behaviour poses several difficulties for researchers and, unfortunately, overcoming these difficulties (e.g., controlling confounding factors such as time exposed to the stressor or studying a specific type of stressor with well-defined parameters of duration, intensity, predictability, etc.) increase the limitations of the research. Even though the stress response is one of the better preserved biological processes during evolution it varies hugely between species, between individuals of the same species, and even in the same individual not only at different periods of development but within short periods of time, given the huge adaptability of the biological processes triggered. As a biological mechanism whose main function is to preserve survival, stress triggers a multitude of physiological, hormonal and molecular correlates within the organism with multiple effects on brain processes and behaviour. Combining measures of several of these biological processes induced in response to stress is vital in order to understand the effects of stress on behaviour and more importantly on the development of psychopathologies.

In this talk, I will focus on one of our projects, in which we were specifically interested in examining the effects of maternal exposure to interpersonal violence (and related Posttraumatic Stress Disorder (PTSD) and comorbid conditions) on mother-child interaction during a laboratory stressor, and its relationship with children's social-emotional development. The main findings of this project have already been published [1-6].

Methods

Only biological mothers with toddlers between 12 and 48 months-old were asked to participate. Following a screening visit (exclusion criteria: active psychosis and/or substance abuse, physical and/or developmental disabilities that interfere with task completion as well as all contraindications for MRI-scanning), and informed consent, mothers entered a protocol involving two videotaped assessment visits (one with mother only and 2nd with the child), followed by an MRI visit, and an additional videofeedback visit. After each visit, mothers received 50 Swiss francs along with a book or toy for their child.

Violence exposure and other life events were explored via the Traumatic Life Events Interview, PTSD with the Clinician Administered PTSD Scale and Posttraumatic Symptom Checklist-Short Version (PCL-S). Major depressive disorder was assessed using the SCID and the Beck Depression Inventory (BDI). Alexithymia (difficulty in identifying, expressing and describing emotions) was assessed with the Toronto Alexitymia Scale. Infant-Toddler Social-Emotional Assessment (ITSEA) was used to evaluate social-emotional developmental.

The mother-child behavioural observations and saliva sampling were performed during a laboratory session that consisted of a 30-min mother-child interaction procedure known as the "Modified Crowell Procedure"[7], which consisted of free and structured play, separation-reunion, clean-up and exposure to novelty. Toddlers' distress behaviour was scored during mother-child separation episodes. Four salivary samples for the assessment of cortisol were collected from the mother and the child: baseline (before the procedure begin), immediately after the end of the stress situation, and at 30 and 60 min after the end of the experimental condition and during the recovery phase. Each time saliva samples were collected, the mothers were asked to estimate their own and their child's anxiety levels on a Likert scale of 0–5 (0 = almost sleeping, 5 = very anxious). Sampling of cortisol was obtained using the Salivette® system (Salivette®, Sarstedt Inc., Rommelsdorf, Germany; www.sarstedt.com) in which a dry cotton swab is placed in the mouth for 2–3 s to passively absorb saliva. At the end of the aforementioned laboratory session, mothers were provided with a pack containing Salivette® swabs, disposable perforated pacifier, written instructions on saliva collection, self-report forms and asked to collect saliva from themselves and their toddlers 30 min after waking, in the afternoon (between 2 and 3 pm), and prior to bedtime on what the mother considered to be the least stressful day of the week. Saliva samples were frozen at −20 °C until cortisol assay. Upon thawing, samples were centrifuged (2 min at 1000 rpm), the swab was discarded and 100 μl saliva aliquots were collected. Cortisol concentration was measured directly (without extraction) on a cobas e-601 analyzer (Roche) by an electrochemiluminescent immunoassay (ECLIA)

Mother-child interactions of free-play, separation and reunion were videotaped and a research assistant who was blind to case-control status selected the silent excerpts for the fMRI stimulus of play and separation. Mothers viewed 6 silent, 30-second video-excerpts of 3 children, each during the two conditions (separation and play): 1) own child, 2) unfamiliar boy, and 3) unfamiliar girl. The order of presentation was pseudo-randomized across both blocks and participants. Within our fMRI paradigm, this project aimed to test cortico-frontal circuits activation that are known to regulate limbic activity.

Ethics statement: The institutional ethics committee at the Geneva University Hospitals approved this research project which is in accordance with the Helsinki Declaration.

Results

As expected, PTSD mothers and their children showed blunted cortisol reactivity to the laboratory stressor. Maternal cortisol levels were negatively correlated to difficulty in identifying emotions. Our data highlights PTSD-related alterations in cortisol and its relevance to maternal behaviour. Toddlers of PTSD mothers also showed an altered pattern of cortisol reactivity to stress that potentially may predispose them to later psychological disorders.

Furthermore, maternal PTSD severity was significantly associated with decreased ventromedial prefrontal cortex (vmPFC) activation in response to mother-child relational stimuli. Maternal PTSD severity and decreased vmPFC activation were significantly associated to child symptoms/behaviours one year later, and were correlated with emotional dysregulation and risk for child PTSD.

Conclusion

The cascade of biological events in response to a stressor is an important factor affecting behaviour and the development of psychopathologies. Combining accurate psychological, physiological, endocrine and neural measures in response to stress provides key information to help understand individual behaviour and to design tailored intervention programs to efficiently prevent and treat behavioural disorders.

References:

1. Cordero, M.I., et al., *Effects of interpersonal violence-related post-traumatic stress disorder (PTSD) on mother and child diurnal cortisol rhythm and cortisol reactivity to a laboratory stressor involving separation.* Horm Behav, 2017. **90**: p. 15-24.
2. Moser, D.A., et al., *The relation of general socio-emotional processing to parenting specific behavior: a study of mothers with and without posttraumatic stress disorder.* Front Psychol, 2015. **6**: p. 1575.
3. Moser, D.A., et al., *Limbic brain responses in mothers with post-traumatic stress disorder and comorbid dissociation to video clips of their children.* Stress, 2013. **16**(5): p. 493-502.
4. Schechter, D.S., et al., *Maternal PTSD and corresponding neural activity mediate effects of child exposure to violence on child PTSD symptoms.* PLoS One, 2017. **12**(8): p. e0181066.
5. Schechter, D.S., et al., *Methylation of NR3C1 is related to maternal PTSD, parenting stress and maternal medial prefrontal cortical activity in response to child separation among mothers with histories of violence exposure.* Front Psychol, 2015. **6**: p. 690.
6. Schechter, D.S., et al., *The association of serotonin receptor 3A methylation with maternal violence exposure, neural activity, and child aggression.* Behav Brain Res, 2017. **325**(Pt B): p. 268-277.
7. Zeanah, C., et al., *Infant-parent relationship assessment*, in *Handboof of infant mental health*, Press G, Editor. 2000: New York. p. 222-35.

Using Electrophysiology to Provide Novel Insight into the Neural Activity Underlying Ambulatory Behavior

C.E. Craig[1], P.S. Holmes[2] and N.J. Ray[3]

[1.] Research Centre for Musculoskeletal Science and Sports Medicine, Department of Exercise and Sport Science, Manchester Metropolitan University, Crewe, UK. c.craig@mmu.ac.uk [2.] Research Centre for Musculoskeletal Science and Sports Medicine, Department of Exercise and Sport Science, Manchester Metropolitan University, Crewe, UK. p.s.holmes@mmu.ac.uk [3.] Research Centre for Musculoskeletal Science and Sports Medicine, Department of Psychology, Manchester Metropolitan University, Manchester, UK. n.ray@mmu.ac.uk

Introduction

Electroencephalography (EEG) is one of the oldest techniques for measuring brain activity. It is a highly utilized tool within the field of cognitive neuroscience due to its high temporal resolution and provides real-time insight into the neural activity underlying specific human behavior. Historically, these behaviors were limited to sedentary fine motor tasks due to movement artifacts that distort the EEG signal. Recent advances in EEG technology and signal processing algorithms have, however, enabled remote online measurement of neural activity during more gross motor tasks, such as walking. Despite the increasing number of research studies employing mobile EEG [1-6], the field is still in its infancy and no publication to date has utilized this method in a clinical population. This paper will outline how EEG has been used to study ambulatory behavior, and how our lab is using EEG to assess neural activity during adaptive gait and obstacle avoidance in healthy individuals and those with impaired mobility.

Method

A novel experimental method is presented that combines simultaneous mobile 64-channel EEG, wireless electromyography (EMG), motion tracking and mobile eye-tracking, to provide the first comprehensive characterisation of naturalistic gait and obstacle avoidance in older people and people with Parkinson's Disease. These variables were assessed simultaneously during a walking and turning 'on cue' task, in which participants walked either straight towards the end of a 6.5m walkway or turned right or left at an intersection depending on a visual cue presented during the walk. During half of the walking trials, an obstacle was present in one of the pathways. The aim of the study was to provide a detailed examination of the neural, muscular, kinematic and gaze factors that may be associated with falls risk in both populations during naturalistic gait, adaptive gait in which a change of direction is required, and obstacle avoidance. The findings from this research aim to support future studies to develop and measure the impact of interventions to improve mobility and reduce falls risk in older adults and clinical populations. All procedures were approved by the Cheshire faculty ethics committee at Manchester Metropolitan University.

Discussion

This talk will focus on methodological issues that require consideration when designing a study to assess EEG activity during active movement. Using examples from the literature and our experiences during the current project, we will discuss how EEG can provide novel insights into the neural control of ambulatory behavior and the limitations of this technique.

References

1. Gramann, K., Gwin, J.T., Ferris, D.P., Oie, K., Jung, T.P., Lin, C.T., *et al.* (2011). Cognition in action: imaging brain/body dynamics in mobile humans. *Rev Neurosci* **22**, 593-608.
2. Kline, J.E., Huang, H.J., Snyder, K.L., and Ferris, D.P. (2015). Isolating gait-related movement artifacts in electroencephalography during human walking. *J Neural Eng* **12**, 046022.
3. Sipp, A.R., Gwin, J.T., Makeig, S., and Ferris, D.P. (2013). Loss of balance during balance beam walking elicits a multifocal theta band electrocortical response. *J Neurophysiol* **110**, 2050-60.
4. Khanmohammadi, R., Talebian, S., Hadian, M.R., Olyaei, G., and Bagheri, H. (2015). Preparatory postural adjustments during gait initiation in healthy younger and older adults: Neurophysiological and biomechanical aspects. *Brain Res* **1629**, 240-9.
5. Seeber, M., Scherer, R., Wagner, J., Solis-Escalante, T., and Muller-Putz, G.R. (2014). EEG beta suppression and low gamma modulation are different elements of human upright walking. *Front Hum Neurosci* **8**, 485.
6. Knaepen, K., Mierau, A., Tellez, H.F., Lefeber, D., and Meeusen, R. (2015). Temporal and spatial organization of gait-related electrocortical potentials. *Neurosci Lett* **599**, 75-80.

Using Transcranial Magnetic Stimulation to examine the control of manual interception

Joost C. Dessing

Queen's University Belfast, School of Psychology, Belfast, UK. E-mail: j.dessing@qub.ac.uk

Introduction

Transcranial Magnetic Stimulation (TMS) has been used over the past several decades to perturb ongoing brain activity, with a variety of applications. First and foremost, this involved fundamental studies into the role of different brain areas across a range of behaviours. Applications in therapeutic contexts - and evaluations thereof – have also been rife. TMS capitalizes on electromagnetic induction to influence firing patterns of neurons in the brain by creating phasic magnetic fields outside the skull. The best example is an involuntary twitch of a muscle that can be induced by TMS when applied to cells in the primary motor cortex that project directly to this muscle's spinal network. Less straightforward, but consistent, behavioural effects have been reported when other brain areas. While the applications of this technique are widespread, the field in general is hampered by several limitations. In this talk, I will discuss some of these limitations, how they informed the design of a recent study into manual interception, and how they can be overcome in further improvements in research design.

Manual interception

The behavioural regularities of manual interception - catching and hitting - have been studied extensively. The problem for scientists attempting to unravel the principles or mechanisms by which such movements are controlled is that kinematics by definition can be achieved through a variety of control mechanisms [1]. This signals the need for neurophysiological studies. A key question in this field is whether humans predict the interception point in advance of the movement, to be used as the goal position of their arm movement. While this may be the common sense assumption, a range of non-predictive interception strategies are also viable. In a recent preregistered study [2], we used repetitive TMS to examine whether parts of the posterior parietal cortex (PPC) display predictive or non-predictive coding.

The study and Results

We used MRI-based neuronavigation (Brainsight 2; Rogue Research, Montreal, Quebec, Canada) to apply repetitive TMS with a MagStim D70 coil and a MagStim Rapid Stimulator (The Magstim Company Ltd, Whitland, UK). During the inital 500ms of motion of a to-be-intercepted target, 6 TMS pulses were applied (10Hz) at 120% of the individual resting motor threshold to either the left or right PPC during interception of targets shown on a computer screen. Interception involved reaching out to touch the moving target when it reached a line; the target was only visible during the TMS. Right fingertip movements were tracked at 250Hz using an NDI 3D Investigator system (Northern Digital Inc, Waterloo, ON, Canada). Because participants (mean age = 29, age range = 19-54, 13 females, 12 males; all right-handed) fixated centrally and the PPC predominantly employs retinotopic coding for visual targets, diagonal target trajectories (crossing from the left to right visual field, or vice versa) could be used to examine whether rTMS effects reflected the initial or final visual field. We examined both the variability of the horizontal initial movement direction and interception error. All procedures were approved by a local ethics committee.

We found that the superior parietal occipital cortex, SPOC, displayed non-predictive coding: the initial movement direction increased due to TMS when applied within the hemisphere coding the initial target position – this effect disappeared during the movement (no effect for interception error). Follow-up analyses showed that the effect for SPOC occurred predominantly in participants initiating while the TMS was ongoing. No effects were found for the medial intraparietal sulcus.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Discussion

Our study provided clear evidence that the ability to aim ahead when catching, that is, to move directly to the catching position, arises downstream of SPOC. However, I will highlight one of the limitations of TMS to show that these non-predictive nature of the effects does not exclude predictive control of interception altogether. Moreover, I will discuss ways in which TMS research designs can be strengthened further in the future.

References

1. Dessing, J.C., Peper, C.E., Bullock, D., Beek, P.J. (2005). How position, velocity, and temporal information combine in the prospective control of catching: data and model. *Journal of Cognitive Neuroscience* **17**, 668-686.
2. Reid, S.A., Dessing, J.C. (under review). Non-predictive online spatial coding in the posterior parietal cortex when aiming ahead for catching.

Integration of traditional and neuroscientific techniques in the study of consumer behaviour: the contribution of Neuromarketing

The aim of the symposium is to explore and discuss the integration of traditional techniques in the field of marketing research (such as Likert self-report scales, interviews, focus-group, etc.) with the application of neuroscientific techniques as an innovative contribution to the study of consumer behavior. Consumers' choices are often driven by reasons that consumers are not fully aware of. Decision-making is influenced by a complex set of emotions, feelings, attitudes, and values that are impossible to assess simply by asking consumers their opinion. Indeed, traditional techniques such as self-reports or interviews allow measuring mainly conscious and rational reactions toward a product or advertising. In the last decades, there has been a rapidly growing interest in the multidisciplinary field of so-called "neuromarketing" (Lee et al., 2007; Plassmann et al., 2008; Ariely, 2010; Missaglia et al., 2017), which takes advantage of neuroscientific techniques to study consumer reactions. This discipline applies neuroscientific methods and tools that allow measuring consumers' emotional and spontaneous responses in a more objective and observable way.

The aim of this symposium is (a) to describe neuromarketing underlying assumptions, techniques, advantages of this perspective, examining the scientific literature about the use of neuromarketing in consumer studies, and (b) to suggest best practices to apply this novel approach in the marketing domain, with a specific focus on non-invasive methods. Finally, although the perception of packaging has been already explored, nevertheless health content of labels, the presence of specific information and lay-out of the labels and evaluation of the communication conveyed by packaging are other possible elements of interest in future neuromarketing research.

Maurizio Mauri, University of Milan

Emotional Impact played by Art Experience Measured by Means of Neuroscientific Techniques and Traditional Interviews

C. Caldato[1], S. Benedetto[1], E. Bazzan[1], L. Rodighiero[1], M. Mauri[1,2]

[1] TSW s.r.l., User Experience Lab, Treviso, Italy. maurizio.mauri@tsw.it; [2] Department of Food, Environmental and Nutritional Sciences, "Università degli Studi" of Milan, Italy

Neuroaesthetics is a recent research area in cognitive neuroscience [1]. This area focuses the attention on the aesthetic experience about beauty, in particular about visual art. Neuroscientific studies showed how the use of brain imaging techniques such as functional Magnetic Resonance Imaging (fMRI), magnetoencephalography (MEG) and electroencephalography (EEG) can be efficiently applied to study the impact of this kind of experience. Although the scientific debate about explanations and understanding of brain systems involved in aesthetic experience elicited by visual artworks is still open, results show how the possibility to investigate this experience by means of neuroscientific methods is a promising field. In this framework, this research study presents additional empirical evidence supporting the possibility to further developing the application of neuroscientific techniques based on EEG synchronized with portable eye-tracking and galvanic skin responses while subjects are exposed to paintings "in vivo", inside the museum where the paintings are exposed. Preliminary results confirm how portable solutions allow to measure aesthetics experience by means of neuroscientific methods combined with traditional ones based on interviews. The present research project is granted by "Banca Intesa San Paolo" from Italy.

Procedure and Methods

Forty subjects are enrolled in the study and scheduled for a visit. They are invited to visit the "Gallerie d'Italia" Museum, in Milan (Italy). They are all from Italy, distributed according to age (10 subjects for 4 age ranges: 18-28; 29-38, 39-48; 49-58) and gender (50% female). Before the visit, they are welcomed and briefed in a separate room, where they are consequently asked to wear: a) an EEG headset (from Emotive Epoch, 14 channels, 250 Hz); b) a bracelet enabling to measure skin conductance (SC) and heart rate (from Empatica, E4, 4 Hz); c) a portable eye-tracking (from SMI, ETG, 60 Hz). All equipment, after the setting up, is synchronized. A 1 minute of baseline is recorded before starting the test, then subjects are carried to the room where the paintings are exposed to visitors. Each painting is covered by a removable partitioner. Subjects are asked to seat in front of the removable partitioner for a few seconds. At this point, the partitioner is removed to expose the subject to each painting, one by one, in each room where paintings are currently part of the exhibition named "The Last Caravaggio: heirs and new masters", at "Gallerie d'Italia" Museum.

	Michelangelo Merisi, detto Caravaggio (Milano 1571 - Porto Ercole 1610) *Martirio di sant'Orsola* 1610 olio su tela, 143 x 180 cm Collezione Intesa Sanpaolo Gallerie d'Italia - Palazzo Zevallos Stigliano, Napoli
	Bernardo Strozzi (Genova 1581/1582 - Venezia 1644) Martirio di sant'Orsola 1615-1618 olio su tela, 104 x 130 cm Collezione privata. Courtesy Robilant+Voena
	Giulio Cesare Procaccini (Bologna 1574 - Milano 1625) Ultima Cena 1618 olio su tela, 490 x 855 cm (misure complessive) Genova, Basilica della Santissima Annunziata del Vastato
	Giulio Cesare Procaccini (Bologna 1574 – Milano 1625) Martirio di Sant'Orsola 1600-1620 Olio su tela, 130 – 145 Collezione Intesa Sanpaolo Gallerie d'italia

Figure 1. All stimuli presented during the study.

The sequence of painting exposure is randomized for each participant. The paintings are (see Figure 1 – the picture are taken from the web, from links reported in the end of the present paper): a) "Martirio di Sant'Orsola" (Martyrdom of Saint Ursula, oil painting on canvas, 1610), from the painter Michelangelo Merisi, so called "Il Caravaggio"; b) "Martirio di Sant'Orsola" (Martyrdom of Saint Ursula, oil painting on canvas, 1615-1618), from the painter Bernardo Strozzi, so called "Il Cappuccino"; c) "Martirio di Sant'Orsola" (Martyrdom of Saint Ursula, oil painting on canvas, 1610-1620), from the painter Giulio Cesare Procaccini; d) "L'ultima cena" (The last supper, 1616), always from the painter Giulio Cesare Procaccini. During the exposure to all stimuli, eye-tracking data, in sync with EEG data and skin conductance monitoring, are recorded for the first 30 seconds in order to assess emotional reactions by means of biological patterns analyses and comparison within the 4 paintings. At the end of the experiment, subjects are taken in the separate room (the same one where the experiment started), and after removing equipment, a researcher performed a brief interview. As final question of the interview, the researcher asks if the wearing of the EEG headset and the bracelet could affect the exposure and the enjoyment of paintings. All subjects report that after a few seconds they forget about the equipment as they are engaged by the vision of the paintings.

Results

Results about EEG data show significant differences, as resumed by graphs (see Figure 2). Frontal asymmetry is calculated from raw EEG data, according to results and analyses presented in the scientific literature [2,3,4]. According to previous research, alpha rhythms reflect deactivation; assessing the level of activation in each frontal hemisphere, it is possible to detect whether the left or the right frontal lobe is more active, thus revealing "approach" or "withdraw" behavior respectively. The painting reflecting the highest value of frontal asymmetry (approach behavior) is "Sant'Orsola" from Caravaggio. The second one is "Ultima Cena" from Procaccini. The third one is "Sant'Orsola" from Strozzi. The last one (with negative value, reflecting withdraw behavior) is "Sant'Orsola" painting from Procaccini. A t-Student statistical test is performed across all subjects averaged for each painting in terms of frontal asymmetry to assess significant differences. The painting "Sant'Orsola" from Caravaggio elicits a frontal asymmetry that is significantly different from all other paintings. In the same way, the painting "Sant'Orsola" from Procaccini showed significant differences with all other paintings. In other words,

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

EEG data have been able to identify the painting eliciting the highest value of frontal asymmetry reflecting "approach behavior" and the lowest value, always in terms of frontal asymmetry, reflecting, on the opposite, the highest value of "withdraw behavior".

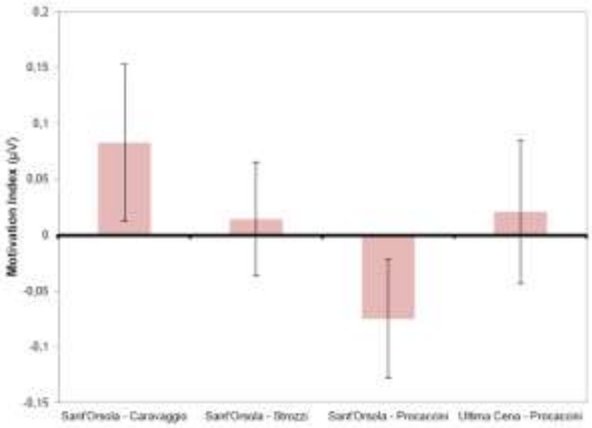


Figure 2. Graphs representing EEG results in terms of Motivation Index (derived from frontal asymmetry of Alpha brain waves) averaged across all subjects for each painting. Values are expressed in microVolts.

Results about Skin conductance data, once again averaged for each subject across each painting, are resumed by graphs (see Figure 3). Skin conductance levels are calculated from raw data [5]. The painting showing the highest value, thus reflecting higher arousal [6,7], is "Sant'Orsola" from Caravaggio. The second one is "Ultima Cena" from Procaccini. The third one is "Sant'Orsola" from Strozzi. The last one (with negative value, reflecting the lowest arousal activation in comparison to baseline) is "Sant'Orsola" painting from Procaccini. A t-Student test is performed across all averages of skin conductance data, calculated for each subject exposed to each painting, to assess significant differences. The painting "Sant'Orsola" from Caravaggio and "Ultima Cena" from Procaccini are significantly different from all others. However, there is not a significant difference between "Sant'Orsola" from Procaccini and "Ultima Cena" from Procaccini. In addition, the painting "Sant'Orsola" from Strozzi does not show significant differences when compared with "Sant'Orsola" from Procaccini and "Ultima Cena" from Procaccini. Skin conductance data analyses are enabling to identify the paintings eliciting the highest values of arousal and the lowest values representing the less arousing painting.

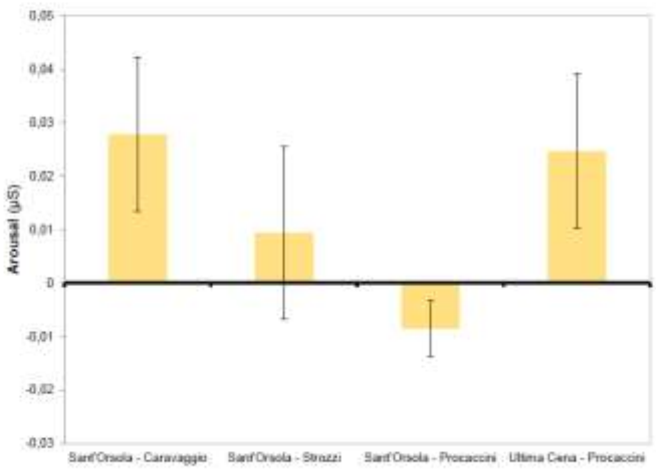


Figure 3. Graphs representing Skin Conductance (SC) results (SC represents the intensity of emotional reaction, in terms of arousal) averaged across all subjects for each painting. Values are expressed in microSiemens.

Results from traditional interviews showed a slight preference for "Sant'Orsola" painting from Caravaggio in comparison to all the other paintings, although all paintings seem to engage all participants in the same way and for different reasons, according to results from interviews. It is worthy to notice that, just for interviews, the limited sample size does not allow any statistical inferential analyses.

Discussion

Considering results from EEG and results from Skin Conductance, it is possible to discern between the different paintings by means of neuro and psychophysiological data. Only combining the two measurements it is possible to assess the emotional impacts played by each painting and distinguishing each painting amongst all the others in terms of biological reactions.

According to most common graphical representations of emotions, where "valence" (the hedonic quality of emotions – in terms of positive vs negative quality) and "arousal" (intensity of emotional reaction) are the most important dimensions enabling to describe the whole panorama of emotional states (See figure 4, readapted from Lang) [8], it might be possible to take advantage of EEG data on the valence dimension, and Skin Conductance data on the arousal dimension, as a potential framework to assess emotional reactions generated by work of art on one side, and the possibility to identify the optimal ways to present works of art in museums (light conditions, position, and so forth) on the other side.

Figure 4. The Valence-Arousal plane (readapted from Lang) [7,8].

In figure 5, data from EEG and Skin Conductance on the valence-arousal plane (data are standardized to the baseline values, that represents the "0" point of the graph) are plotted. In the scientific literature there are already publications addressing the possibility to take advantage of this kind of neuropsychophysiological framework to study the impact of different kind of experience in terms of emotions [7,9]. Looking at the graph in figure 5, the "Martirio di Sant'Orsola" from Caravaggio is located in the quadrant of mid-high approach and mid-high arousal, together with "Ultima Cena" from Procaccini and "Martirio di Sant'Orsola" from Strozzi. However, "Martirio di Sant'Orsola" from Caravaggio shows the highest values and it is possible to claim that elicits the highest approach reaction together with the most intense level of arousal. Amongst the 4 paintings, "Martirio di Sant'Orsola" from Caravaggio is the one that elicits the most optimal emotional reaction. On the opposite, the painting "Martirio di Sant'Orsola" from Procaccini elicits the lowest emotional response, both in terms of arousal and frontal-asymmetry from EEG, showing negative values in comparison to baseline for both measurements. Aside the possibility to assess emotional reactions for each painting, it is worthy to highlight how, in the present research work, emotional reactions are measured "in vivo", while participants are exposed to paintings in the museum where paintings are part of an exhibition.

The application of a neuropsychophysiological rationale in real life situations may increase the opportunities to study human emotions in innovative environments in comparison to traditional laboratory contexts, where subjects' emotional reactions are potentially influenced by the setting of being in a scientific lab. This rationale can open the way to new studies on emotions, enabling to bring higher ecological validity, as it can be applied in any kind of real world situations, from museums to shops, restaurants, etc., where people is much more prone to react spontaneously in terms of emotions. Last but not least, taking into account also results from interviews, it is worthy to notice how it is not possible to take advantage of any statistical analyses performed on results from interviews, due to the limited sample size. On the opposite, results from neuropsychophysiological data about the same sample allow the application of statistical analyzes, thus providing more empirical findings supporting eventual conclusions.

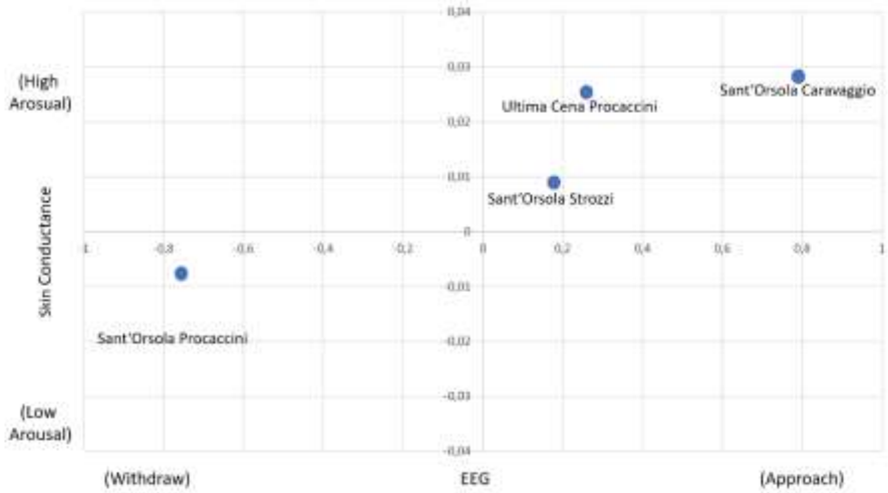


Figure 5. The Valence-Arousal plane representing the emotional reaction for each painting, averaged all subjects

Conclusion

The present research work has been aimed to explore preliminary results enabling to assess emotional reactions played by different Italian paintings from the same period (post-renaissance) in order to evaluate the possibility to assess emotional engagement played by art experience. Taken in the boundary of the present experiment, results are promising, allowing to envisage the application of these methods, based on wearable sensors enabling to monitor neuro and psychophysiological reactions within the museums, to study aesthetic experience with the goal on one hand to identify the optimal ways to present and elicit art experiences, on the other hand also to develop new practices and scientific procedures aimed to enhance the emotional engagement played by this kind of experience in order to promote the diffusion and appreciation of cultural services and sites.

Statement on ethics

The experimental procedures involving human subjects described in this paper are following the ethical principles outlined in the Helsinki Declaration of 1975, as revised in 2000.

Acknowledgements

Financial support for this study was received from the Innovation Centre of “Banca Intesa San Paolo”, one of the most important bank Institutions in Italy.

Reference:

1. Di Dio, C., Gallese, V., (2000). Neuroaesthetics: a review. *Current Opinion in Neurobiology*. **19**, 682-687.
2. Canli, T., Desmond, J.E., Zhao, Z., Glover, G., & Gabrieli, J.D.E. (1998). Hemispheric asymmetry for emotional stimuli detected with fMRI. *NeuroReport*, **9**, 3233–3239.
3. Schmidt, L.A., Trainor, L.J. (2001): *Frontal brain electrical activity* (*EEG*) *distinguishes valence and intensity of musical emotions*. *Cognition and Emotion*, **15(4)**, 487–500.
4. Davidson, R.J. (2004). What does the prefrontal cortex “do” in affect: perspectives on frontal EEG asymmetry research. *Biological Psychology* **67**, 219-233.
5. Bradley, M.M., Codispoti M., Cuthbert B.N., Lang P.J. (2001). Emotion and Motivation I: Defensive and Appetitive Reactions in Picture Processing. *Emotion*, **1(3)**, 276-298.

6. Mauri, M., Magagnin, V., Cipresso, P., Mainardi, L., Brown, E.N., Cerutti, S., Villamira, M., Barbieri, R. (2010). Psychophysiological signals associated with affective states. *In Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pp. 3563-3566.
7. Mauri, M., Cipresso, P., Balgera, A., Villamira. M., Riva, G. (2011). Why is Facebook so successful? Psychophysiological Measures Describe a Core Flow State While Using Facebook. *CyberPsychology, Behavior and Social Networking*, **14(12)**, 723-731.
8. Lang, P.J. (1995). The emotion probe: Studies of motivation and attention. *American psychologist*, ***50*(5)**, 372.
9. Stasi, A., Songa, G., Mauri, M., Ciceri, A., Diotallevi, F., Nardone, G., & Russo, V. (2017). Neuromarketing empirical approaches and food choice: A systematic review. *Food Research International*. https://doi.org/10.1016/j.foodres.2017.11.049

Links to the paintings used as stimuli in the present paper:

1. "Martirio di Sant'Orsola" (Martyrdom of Saint Ursula), from Michelangelo Merisi, so called "Il Caravaggio": https://it.wikipedia.org/wiki/Martirio_di_sant%27Orsola#/media/File:CaravaggioUrsula.jpg .
2. "Martirio di Sant'Orsola" (Martyrdom of Saint Ursula), from Bernardo Strozzi, so called "Il Cappuccino": http://catalogo.fondazionezeri.unibo.it/scheda.v2.jsp?tipo_scheda=OA&id=62793&titolo=Strozzi%20Bernardo,%20Martirio%20di%20sant'Orsola&locale=it&decorator=layout_resp&apply=true
3. "Martirio di Sant'Orsola" (Martyrdom of Saint Ursula), from the painter Giulio Cesare Procaccini: http://www.glistatigenerali.com/arte/un-dipinto-di-caravaggio-non-basta-a-cambiare-il-corso-della-storia/ .

4. "L'ultima cena" (The last supper), from Giulio Cesare Procaccini: https://www.youtube.com/watch?v=b3EwQ1tWutA

Validation of visual attention data comparing 2D and 3D set-up

Jesper Clement

Department of Marketing, Copenhagen Business School; jc.marktg@cbs.dk

In the field of marketing eye-tracking has developed into a common method for the study of consumers' visual attention and response to advertisements, packaging, and in-store experiences. Over the last ten years, many eye-tracking studies have investigated visual attention of such daily decisions, well knowing that there might be a difference between humans' perception of real life exposure and those shown on the screen. New hardware and software have enabled us to use and analyze data from wearable eye-trackers, and by that compare eye-track data from sceneries shown on a screen (2D) with data form real-life experiences (3D). Yet, it also poses a great challenge, as the amount of data and the complexity increases significantly, and what we win on the ecological validity, we might loose on the internal validity. In a set-up we investigated three scenarios of the very same shelf of products in order to compare decision processes in 2D and 3D. In the 2D scenario we used a Tobii 60XL remote eye-tracker and in the 3D scenario we used a Tobii2 mobile eye-tracker (50Hz). The 3D scenario was actually split into two scenarios, as we had the very same shelf display build up in our SenseLab. That way we ended up having three comparable sets of eye-track data related to the same decision process. The outcome showed a more systematic search processes in the 2D scenario and the closer we come to a real-life setting the more chaotic became the search process, including larger head movement and smaller eye-movements.

Another challenge using 3D eye-tracking is, that it easily becomes too resource demanding due to manually interpretation and analysis of the enormous amount of data from a representative sample size. The required workload and the value of the result easily exceeds the cost of the study. A common way for a fast-and-easy analysis of eye-track data is the use of a heatmap, making use of either raw data or number of fixations. Mapping data form a portable eye-tracker recorded in a real-life scenario onto a 2D heatmap is not an easy task. One major challenge is the parallax error due to the natural perspective. In order build a model solving this issue, we took a simple approach using homographic correspondence between a reference image and frames coming from the portable eye-tracker, and to these combined images, we added the gaze-data to 3D-modelled AOIs. The procedure has 3-steps, starting with the design of the 3D AOI. In the second phase, we added the combined frames one by one to this 3D AOI, and then finally we were able to create a 3D heatmap. The proposed model runs automatic with approximately two frames pr. second. We found the 3D heatmap model to work fine for investigating the in-store search process and especially for shelf displays with packaging having distinct image features. However, making a final decision in which consumers might start reading text on the packaging, the 3D heatmap reaches its limit.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Implicit Association Test (IAT) – computer based method to measure implicit attitudes and better predict consumer behaviour

Dominika Maison

Department of Psychology, University of Warsaw.

Even though most consumers think they know why they like or choose the products or brands, very often they do not. Consumers have limited access to their inner processes, and this presents a challenge for practitioners who want to predict consumers' behaviour based on questionnaires. Consumer psychologists have therefore sought reliable and valid ways to overcome this and they have developed indirect measures that rely on non-declarative features of people's responses (Maison, Gregg, 2016). Those indirect measures, often based on physiological (e.g. EEG, GSR, facial muscles movement) or automatic reactions (e.g. reaction time), are called "neuromarketing measures" and they are used to better predict consumer behaviour. There are different examples of such methods based on reaction time (RT), for example the Extrinsic Affective Simon Task (EAST; DeHouwer, 2003), the Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010), as well as various lexical decision tasks where primes and targets are sequentially presented (Klauer & Musch, 2003; Wittenbrink, Judd, & Park, 1997). In all cases, the goal is to detect the presence and quantify the strength of semantic or evaluative associations between objects and attributes.

In the presentation will be discussed potential and limitation of the best known computer-based method of measuring implicit attitudes (IAT – Implicit Association Test) in context of consumer attitudes and behaviour (Maison, Greenwald, Bruin, 2004; Maison, 2016). For many years attitudes were understood as three components constructs (cognitive, affective and behavioural), with assumption that the person has introspective access to his/her attitude, can verbalize it, and express it in a questionnaire. Since 1980s scientists' attention has been drawn to the unconsciousness of attitudes and the automatic character of them. Greenwald and Banaji (1995) introduced concept of *Implicit Attitude* and the method of measuring them – the *Implicit Association Test* (Greenwald, et al. 1998), a computer-based method using reaction-time as an indicator of attitude strength. At the beginning the IAT was used to study racial attitudes, self-concept and self-esteem. Nowadays the IAT is used in many other areas, including consumer attitudes (attitudes toward brands or categories), therefore it can be treated as one of neuromarketing tools.

The ultimate question discussed in the presentation is whether and to what extent indirect measures of consumer attitudes, such as the IAT, can capture implicit content, and provide better understanding of consumer processes or predict consumer behaviour. In particular, how sensitive is the IAT to implicit content, both absolutely and relative to direct measures such as self-report? And to what extent is it capable of delivering unique predictive validity, in virtue of tapping into content that direct measures such as self-report miss?

References

1. Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record*, *60*, 527-542.

2. De Houwer, J. (2003). The extrinsic affective Simon task. *Experimental Psychology, 50*, 77-85.

3. Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.

4. Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27.

5. Klauer, K. C. & Musch, J. (2003). Affective priming: findings and theories. In J. Musch, & K.C. Klauer (Eds.), *The psychology of evaluation* (pp. 7-36). Mahwah, NJ: Lawrence Erlbaum Associates.

6. Maison, D., Gregg, A. (2017). Capturing the consumer's unconscious. Applying the implicit association test in consumer research. C. Jansson-Boyd & M. Zawisza (eds.). *Routledge International Handbook of Consumer Psychology*. London: Routledge. Taylor&Francis Group. pp. 143-164.

7. Maison, D. (2016). Implicit Association Test (IAT): Using Computer-Based Methods to Measure Consumer Implicit Attitudes. [in] K. Nermend & M. Łatuszyńska (eds.). Selected Issues in Experimental Economics. Switzerland: Springer. pp. 107-124.

8. Maison, D., Greenwald, A. G., & Bruin, R. H. (2004). Predictive validity of the Implicit Association Test in studies of brands, consumer attitudes, and behavior. *Journal of Consumer Psychology, 14,* 405–415.

9. Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for radical prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, *72*, 262-274.

Assessment of Storefront Displays with a Multidisciplinary Approach based on Neuroscientific Methods, Self-reports and Anthropological Marketing

M. Mauri[1,2], L. Rodighiero[1], E. Bazzan[1], S. Benedetto[1], C. Caldato[1], L. Armeno[3]

[1] TSW s.r.l., User Experience Lab, Treviso, Italy. maurizio.mauri@tsw.it [2] Department of Food, Environmental and Nutritional Sciences, "Università degli Studi" of Milan, Italy [3]Department of Anthropology, "Università Ca' Foscari" of Venice, Italy

Introduction

Neuromarketing is an interdisciplinary research field that brings together neuroscientific methods and marketing research [1,2]. It focuses the attention on the application of neuroscientific insights, methods and technological solutions into the field of marketing, behavioral economics and consumer psychology, leading to a new academic discipline also known as "consumer neuroscience" [3,4]. Neuroscientific research takes advantage of brain imaging techniques such as functional Magnetic Resonance Imaging (fMRI), magnetoencephalography (MEG) and electroencephalography (EEG) to reveal and understand brain mechanisms. Neuromarketing emphasizes how some of these techniques might be proficiently used to study consumer experience, with the goal to understand how emotional and cognitive reactions can affect purchase intention and preference. Although the scientific debate about implications and understanding of brain processes involved in consumer experience elicited by products and services is still open, pioneering studies show how the possibility to investigate this experience by means of neuroscientific methods is a promising field enabling to improve the design of products and services [5]. In this vision, the present research project presents additional empirical evidence supporting the possibility to further develop the application of neuroscientific techniques based on EEG synchronized with portable eye-tracking and galvanic skin responses while subjects are exposed to a shopping experience "in vivo", in a real shop, where real products, in this case sport shoes, are exposed to real clients. Preliminary results show how portable solutions enable to measure consumers emotional reactions by means of empirical methods while people are exposed to storefront displays, and how this kind of data can be combined with traditional ones based on self-reports and interviews according to Anthropological Marketing perspective [6], leading to the exploration of new insights about consumers' reactions.

Procedure and Methods

Fifteen subjects (9 females, averaged between 18 and 36 years) are enrolled in the study and scheduled for a visit to the shop. They are invited to visit the "Flagship Store" in downtown of Milan (Italy).

Before the visit they are welcomed and briefed in a separate room, where they are consequently asked to wear: a) an EEG headset (from Emotive Epoch, 14 channels, 250 Hz); b) a bracelet enabling to measure skin conductance (SC) and heart rate (from Empatica, E4, 4 Hz); c) a portable eye-tracking (ET) from SMI (Sensory Motor Instruments hardware, equipped with ETG software, 60 Hz). All equipment, after the setting up, is synchronized. A 1-minute baseline is recorded before starting the test, then subjects are carried to the sidewalk outside the shop, in front of the first storefront display amongst five, as the order of storefront display exposure is randomized for each participant. The shop presented 5 storefront displays, numbered from 1 to 5 (see Figure 1): the number 1 is presenting shoes from the brand "Puma", together with a poster showing an American singer, called "Weeknd", well known amongst teenagers; the number 2 is showing shoes from the brand "Adidas"; the storefront display number 3 presents shoes from the brand "Converse"; the number 4 exhibits a big TV monitor, showing continuously audiovisual messages from different brands; finally, the number 5 presents one shoe product only, this time from a new brand quite recently appeared in the Italian market, called "Under Armour", an American company created in 1996 (in comparison to the other three storefront displays, that show brands very well known in the market – as they have been created in the Twenties of last century). Each participant is asked to stay in front of each storefront display (one by one) for 15 seconds. During the 15 seconds, both EEG, SC and ET are monitored and recorded.

After this first phase, the EEG headset is taken out, and subjects are asked to visit the shop for a free walkthrough wearing the eye-tracking glasses and the bracelet enabling to monitor skin conductance. Once completed the free walkthrough (10 minutes maximum), they are asked to accomplish two different tasks: a) find and take a packet of shoelaces and reach the cashier's desk; b) once joined the cashier's desk, they were asked to exit from the shop (as the shop presents 2 different solutions in order to exit, and the project is aimed to explore which one could fit better with the customer experience). At the end of the experiment, subjects are taken again into the separate room (the same where participant started the test), and after removing equipment, a researcher specialized in anthropological marketing [7], performed an interview. Before leaving and getting a coupon of 30 euros as a reimbursement, they are asked to fill in a self-report investigating different factors: memorization of brands presented in the storefront displays, the use of specific labels exposed with specific products, the ability to understand their meaning, the easiness in finding the exit way, how many cashier's desks have been noticed in the shop, etc.

Results

Preliminary results here presented will be mainly focused on the first phase (exposure to storefront displays), as data analyses about tasks mentioned in procedure section is not completed yet.

About eye-tracking results, in Figure 2 the heatmaps for each storefront display are represented. The metrics about average eye-fixation allowed the identification of the visual elements watched at the most during storefront display exposure: they are "price labels", on one side; and "shoe products", on the other side. The storefront display number 4 (see Figure 2) is not reported, as it presents a big TV monitor displaying audiovisual messages changing all the time.

Figure 1. All storefront displays considered presented shoes from specific brands, from left to right: number 1, Puma; number 2, Adidas; Number 3, Converse; Number 4, big TV monitor showing audiovisual messages; number 5, Under Armour.

Figure 2. All storefront displays presented with heatmap performed by eye-tracking SMI ETG software "Begaze" from data collected of the whole sample (15 participants). Looking at the pics, it is possible to notice how shoe products and prizes are

Figure 3. All storefront displays presented with heatmap performed by **analytic** participants: areas of interest (AOI) are focused on price (yellow) and shoe products (blue box)

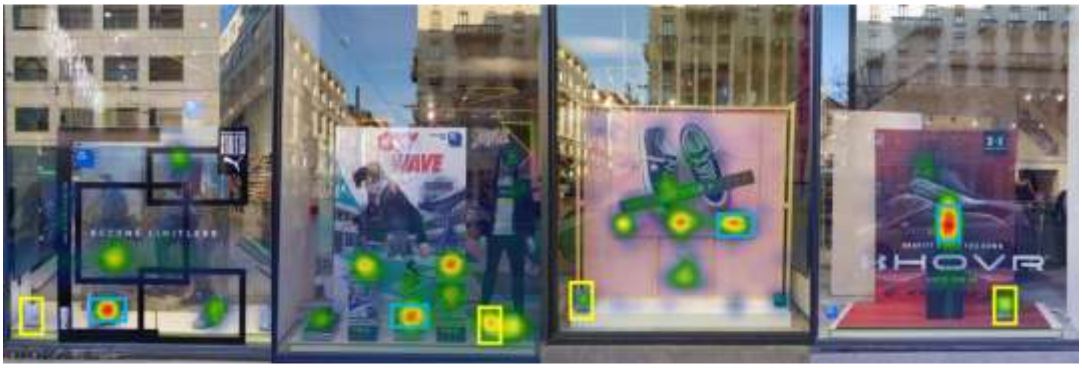


Figure 4. All storefront displays presented with heatmap performed by **creative** participants: areas of interest (AOI) are focused on price (yellow) and shoe products (blue box)

Eye-tracking analyses have been performed also according to results from anthropological marketing interviews, that identified two clusters of consumers: one cluster represented by more analytic people (8 analytic participants over 15), and a second cluster represented by more intuitive people (7 creative people over 15). Two areas of interest focusing the visual attention on price labels and shoes products have been identified for each storefront display. Heatmap from the two clusters are represented in Figure 3 (Analytic participants) and Figure 4 (Creative subjects), where yellow boxes identify AOI about price labels, while blue boxes are identifying AOI about shoe products. Results about dwell time (average time fixation) on a specific element of visual scene, in this case, "prices" and "shoe products", are represented in Figure 5, where AOI about price and shoe products are compared across the two users' clusters (analytic and creative subjects). Statistical analyses do not show any significant difference, although the same patterns, in terms of visual behavior, appear systematically in each storefront display (for 4 times, in each of the 4 storefront displays). The same kind of analyses about eye-fixation time spent on average on prices and shoe products are performed comparing the sample composed by female (8 subjects) vs male (7 subjects), but it is not possible to identify any clear pattern. Also dividing the whole sample for the two age ranges, 18-27 and 28-36, eye-tracking data do not show any pattern appearing in each storefront display.

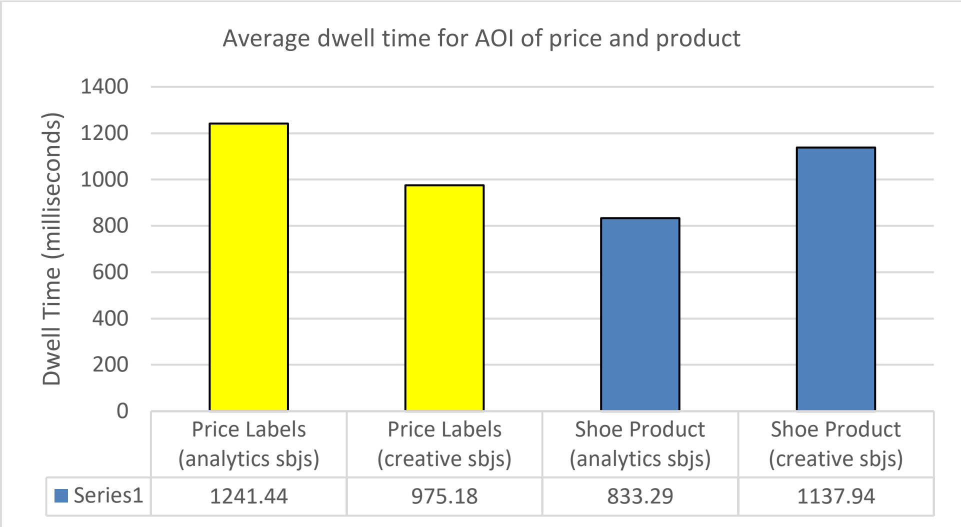


Figure 5. Graphs representing eye-tracking results (average dwell time, that correspond to the average time fixation for each AOI considered – prices and shoe products) across all participants and across all storefront displays: analytic subjects spend more time on prices in comparison to creative ones; creative subjects spend more time watching shoes than prices.

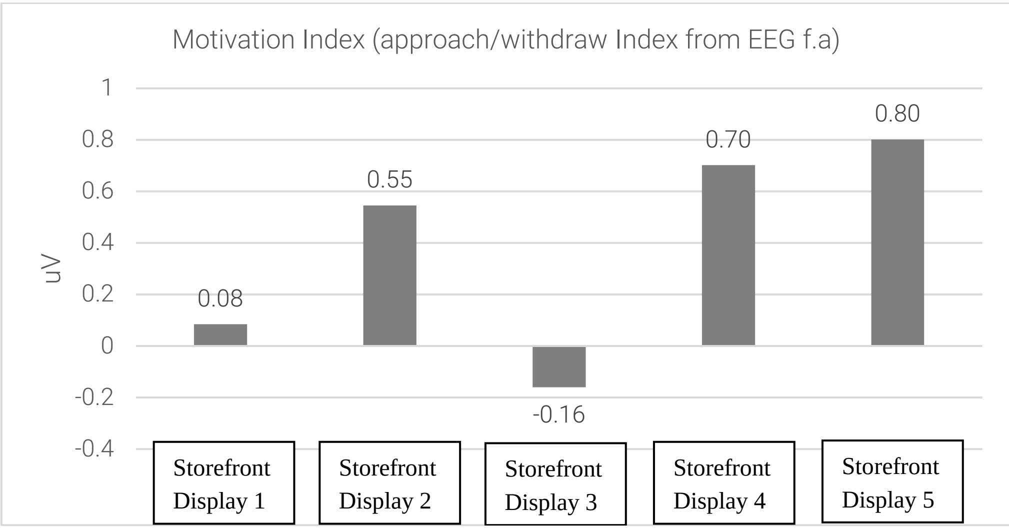


Figure 6. Graphs representing EEG results, in terms of Motivation Index (approach for positive values VS withdraw behavior for negative values, derived from frontal asymmetry – f. a. – of alpha brain waves detected in the frontal lobes) averaged across all subjects, for each storefront display; values are expressed in microVolts.

Considering EEG results, data do not show significant differences, as resumed by graphs (see Figure 6). Frontal asymmetry has been calculated from raw data, according to analyses about alpha rhythms [8,9]. The storefront display reflecting the highest value (approach behavior) is the number 5, Under Armour. The second one is the number 4, equipped with the big TV monitor that shows continuously audiovisual messages. The third one is the number 2, Adidas, followed by the number 1, Puma. The last one (with negative value, reflecting withdraw behavior) is the number 3, Converse. A t-Student statistical test has been performed across all subjects averaged for each storefront display to assess significant differences. No significant differences are emerging from data collected.

Results about Skin conductance data show different trends amongst different storefront displays (see Figure 7). Skin conductance levels are calculated from raw data as the scientific literature shows significant changes according to different stimuli [10,11] and by means of wearable solutions [12]. The stimulus reflecting the highest value (highest arousal) is storefront display number 4, equipped with the big TV monitor. The second one is number 2, Adidas. The third one is number 5, Under Armour, followed by storefront display number 1, Puma. The last one (with negative value, reflecting the lowest arousal activation in comparison to baseline) is the storefront number 3, Converse. A t Student test is performed across all storefront displays to assess significant statistical differences. No significant differences are detected.

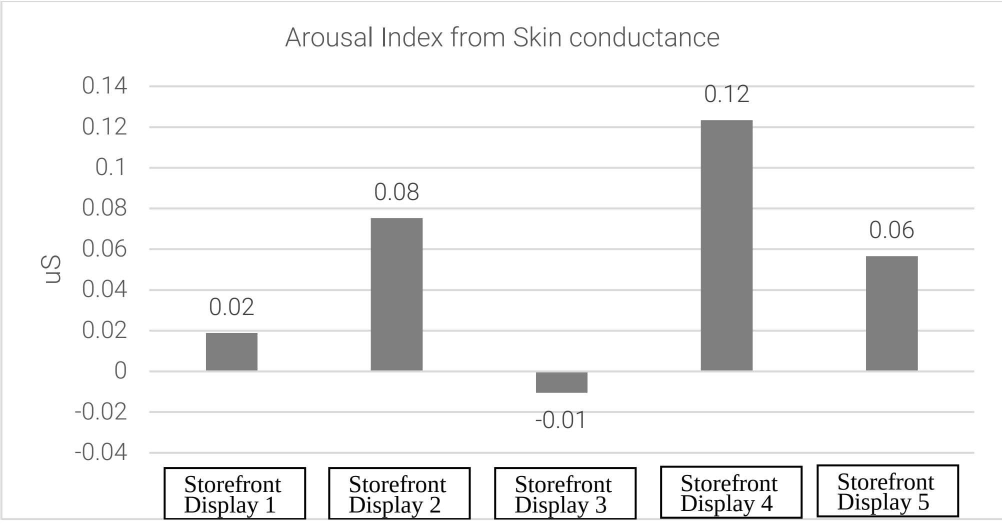


Figure 7. Graphs representing Skin Conductance (SC) results, in terms of intensity of arousal (derived from sweating gland activity from the wrist of the non-dominant hand) averaged across all subjects, for each storefront display; values are expressed in microSiemens.

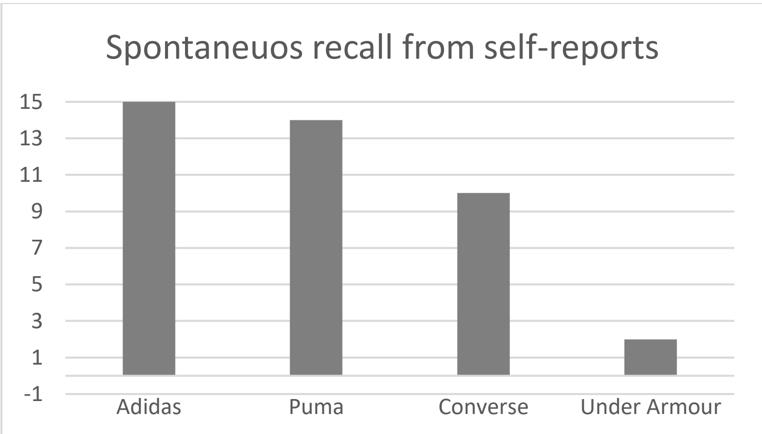


Figure 8. Graphs representing the level of memorization of brands; this kind of data have been collected by means of traditional self-reports.

Results from traditional interviews (see Figure 8) show that the brand “Adidas” exhibits the highest recall performance, as all fifteen subjects reported to spontaneously remember the presence of that brand in the storefront display (15 over 15). The second-best performance about spontaneous recall is provided by Puma (14 over 15). Converse is revealing the third best memorization performance in terms of interviews (10 over 15). The last performance is shown by the brand Under Armour (2 over 15 subjects reported to spontaneously remember this brand presented in the storefront display).

Performing a correlation statistical analysis between data from EEG and data from self-reports, the correlation score has been positive ($p > 0.95$) for the following storefront displays: number 1, Adidas; number 2, Puma, and number 3, Converse. In the correlation analyses, storefront number 4 was not considered, as it was equipped with a monitor, thus it is not comparable with other storefront displays. About storefront display number 5, the correlation analyses is not significant ($p < 0.95$). The same correlation analyses between data from skin conductance and data from spontaneous recall show the same positive correlation ($p > 0.95$) for the same brands: Adidas, Puma and Converse, but not for Under Armour ($p < 0.95$).

Discussion

Considering eye-tracking results from the whole sample of 15 subjects, it is possible to observe how heatmaps highlight the most watched visual elements in storefront display: price labels and shoe products.

On one side, with this method only it is possible to precisely measure the "visual time" spent to process visual elements exposed in storefront displays; interviews, that asked participants to report the most attractive elements in storefront displays, show heterogenous answers, where shoe products are representing one of the several categories emerging from interviews, while price labels are not even mentioned by any participant.

On the other side, there are no significant differences amongst these two kinds of visual elements in terms of eye-fixations. Matching eye-tracking data with anthropological marketing approach that identifies two clusters of consumers, "analytic" and "creative", it is possible to observe different visual behavior: analytic subjects watched longer price labels in comparison to shoe products, while creative participants show the opposite in terms of visual behavior, being more visually attracted by products instead of price labels. It is possible to describe the same visual behavior in each storefront display, even if no significant differences are identified, probably due to the small sample size (8 subjects compared with 7 subjects).

Performing the same analyses on eye-tracking data collected in this study according to other typical factors investigated in marketing research, such as gender and age, no patterns in terms of visual behavior emerge. It seems that anthropological marketing approach may allow to identify more salient clusters in terms of visual behavior even in comparison to gender, being gender one of the most typical factor enabling to profile consumers in marketing [13]. According to the eye-mind hypothesis [14] which states that, when a person is looking at a visual display, his/her gaze point corresponds to the mental representations that is "on top of the stack" of mental operations [15], the attempt to match findings from anthropological marketing with findings from neuromarketing techniques such as wearable eye-tracking methods could lead to a deeper understanding of consumers' behavior. Analytic subjects tend to spend more time on price labels as they rely more on that information when evaluating a new purchase. On the opposite, creative subjects do prefer to watch shoe products as they rely more on product exploration rather than price processing when considering a new purchase. These insights need further exploration and may represent a future research direction to widen the rationale about the role of personality factors as debated in the scientific literature [16], with a more specific focus on the link between personality traits and purchase processes. Considering results from EEG and results from Skin Conductance, it is not possible to discern between the storefront displays by means of neuro and psychophysiological data. Even if there are not significant difference, we tried to plot data collected on a valence-arousal graphical representation as proposed by previous research works (See Figure 9) in the scientific literature [17].

Figure 9. The Valence-Arousal plane (readapted from Lang) [7,8].

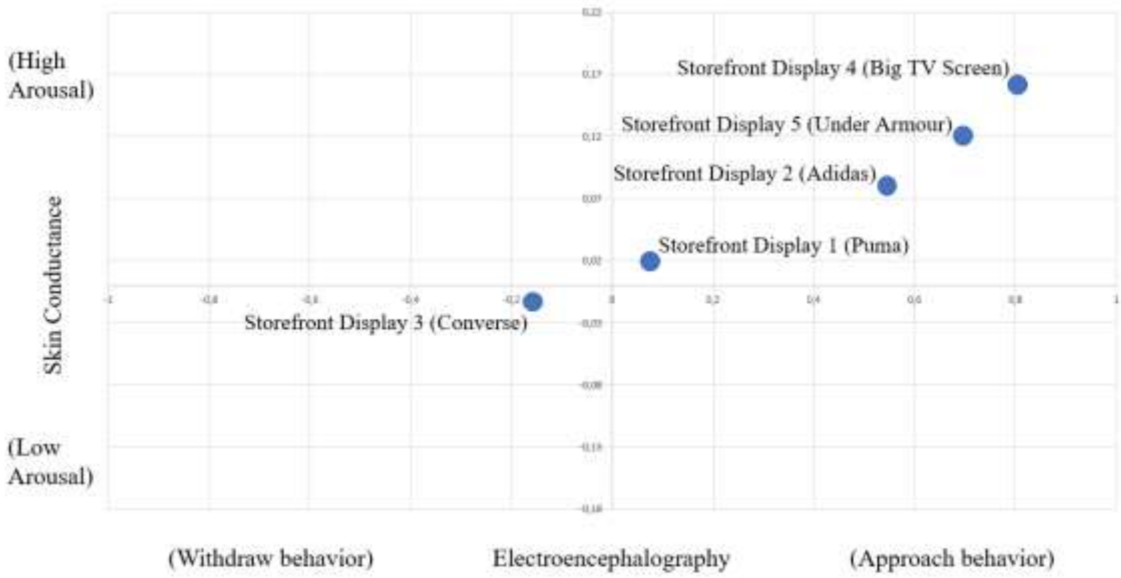


Figure 10. The Valence-Arousal plane where emotional reactions both in terms of EEG data (on the “X” axe, representing the dimension of “valence”) and SC data (on the “Y” axe, representing the “arousal” intensity) are plotted: it is possible to characterize emotional reaction for each storefront display, averaged all subjects.

Looking at the graph where data collected are plotted on the valence-arousal representation (See Figure 10), it is possible to see that data from EEG recordings are represented on the “X” axe, as dimension of valence: it characterizes the hedonic quality of emotion in terms of positive, neutral (the “0” point representing the benchmark according to data collected during the baseline) and negative connotation of felt emotions expressed by the approach / withdraw index emerging in the EEG signals collected. Skin Conductance data are represented on the “Y” axe, referring to the arousal dimension, aimed to represent the intensity of emotions (from low to high arousal), always standardized to the baseline (the “0” point of the graph).

As it is possible to deduce looking at the graph, the Storefront display 4, that was showing a big TV monitor exposing audiovisual messages about many brands and many shoe products, is characterized by the highest values of intensity (highest arousal level) and positive (approach) emotional reactions, according to the average of data collected across all subjects (N. = 15). Storefront display 5, showing one shoe product only, from “Under Armour” brand, is the second most engaging experience in terms of emotional reactions. The third storefront display with highest values both in terms of EEG and skin conductance is the number 2, showing product shoes from Adidas. Finally, it is possible to see also storefront display number 1, showing shoes from brand “Puma”. The only storefront display with negative values in comparison to baseline is “Converse”, as data collected fall, on average, in the quadrant described by mid low arousal levels and mid low negative valence (withdraw behavior). It is worthy to notice that the correlation analyses have been different according to brands. The only brand that does not show significant correlation between neuro- and psychophysiological data and self-reports is the brand “Under Armour”, in storefront display number 5. The main difference between “Under Armour” and other brands is relying on brand equity, as “Under Armour” is recently appeared in the Italian market, while all other brands here considered are in the market since several decades. In this perspective, further investigation about the value of brand equity and its link with the potential to rise an emotional reaction, when exposed to the brand, may allow to better define and understand this interaction. A possible interpretation about the lack of correlation between biological data and memorization for the less known brand may be explained taking into account also eye-tracking data: frontal asymmetry and simultaneous higher level of arousal may be triggered by novelty; however, this positive emotional reaction is not correlating with a more specific memorization of the brand. In terms of eye-tracking data, synchronized with biological data of the same brand, no eye-fixations have been detected on the brand name appearing in the storefront display number 5 representing “Under Armour”, thus, no heatmaps show up in the (small) AOI on brand name. The other storefront displays do not show a different strategy, as the logos and names of the brands are depicted by means of small dimensions as well, however they are much well known and for this immediately recognized/recognizable and remembered. In conclusion, the matching of all source of data analyzed address the need to increase the attention on the way how to convey more visual attention on the brand name, in order to improve memorization performance for those new brands that cannot rely on a long and stable presence in the market. Another consideration is related to the visual display design, as "Under Armour" is the only one that presents one shoe product only, as the other storefront displays present multiple products (three at least). This factor could also explain why Storefront display number 5 (Under Armour) exhibits a "mismatch" in terms of biological reaction and spontaneous recall of the brand. Maybe participants prefer one product only vs multiple options. In summary, new brands could benefit from specific strategy to convey information about their name and logo to improve memorization, not only in terms of dimension of logos and names, but also in matter of lay-out of display, as one product only could enhance the likelihood to spend more eye-fixation on the brand name when it is more visible. Moreover, the present research provides an attempt to show how a marketing strategy based on the assumption that "less is more" may be investigated with a multiple approach. Little is known about shop windows and storefront displays. In this framework as well, there is the need of additional evidence to support this strategy and to further investigate which is the best storefront display design enabling to elicit the highest engagement and memorization in customers.

Conclusion

The present research work has been aimed to explore preliminary results taking advantage of different approaches: one based on innovative neuroscientific methods [1] enabling to measure emotional reactions [2] linked with eye-fixations on specific elements of a real visual scene [14], another relying on traditional self-reports investigating memorization of brands [18], and the last one based on the attempt to make a profile of the participants enrolled, according to anthropological marketing perspective [6]. The mixed-method design, where qualitative (self-reports) and quantitative (Neuro and psychophysiological) data have been collected and merged together, is aimed to assess different storefront displays from the same shop, considering participants as complex agents who require multidisciplinary methods to uncover their mental processes. Taken in the boundary of the present experiment, where no significant difference is emerging mainly due to the small sample size, on one side results do not allow any kind of conclusion in matter of emotional impact played by each storefront display, except for correlation analyses of well-known brands compared with relative new brands; on the other side, it is worthy to notice that anthropological profiling can enable to identify visual patterns systematically emerging in each storefront display, and that the possibility to plot neuro and psychophysiological data on the valence-arousal model might allow to characterize and classify the impact of emotional engagement elicited by different storefront displays. These findings address the need to further investigate this multidisciplinary approach with bigger sample size and with different "in vivo" experiences related to shop experience.

Statement on ethics

The experimental procedures involving human subjects described in this paper are following the ethical principles outlined in the Helsinki Declaration of 1975, as revised in 2000.

Acknowledgements

Financial support for this study was received from the "AW Lab" private company, part of the "Group Bata", one of the most important worldwide corporate in the field of shoe products.

References:

1. Lee, N., Brotheric, A.M., Chamberlain, L. (2007). What is neuromarketing. A discussion and agenda for future research. *International Journal of Psychophysiology*. **63 (2)**, 199-204.

2. Stasi, A., Songa, G., Mauri, M., Ciceri, A., Diotallevi, F., Nardone, G., & Russo, V. (2017). Neuromarketing empirical approaches and food choice: A systematic review. *Food Research International*.

3. Plassmann, H., O'Doherty, J., Shiv, B., Rangel, A. (2007). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences of the United States of America*. **105(3)**, 1050-1054.

4. Plassmann, H., Venkatraman, V., Huettel, S., & Yoon, C. (2015). Consumer neuroscience: applications, challenges, and possible solutions. *Journal of Marketing Research*, **52(4)**, 427-435.
5. Ariely, D., Berns, G.S. (2001). Neuromarketing, the hope and hype of neuroimaging in business. *Nature Review Neuroscience*. **11**, 284-292.
6. Sherry, J.F. (1995). *Contemporary marketing and consumer behavior: an anthropological sourcebook*. Sage Publications, Thousand Oaks, CA.
7. Tian, K., Borges, L. (2011). Cross-cultural issues in marketing communications: an Anthropological perspective of international business. *International Journal of China Marketing*. **2(1),** 110-126.
8. Tomarken, A.J., Davidson, R.J., & Henriques, J.B. (1990). Resting frontal brain asymmetry predicts affective responses to films. *Journal of personality and social psychology*, **59(4)**, 791.
9. Davidson, R.J. (2004). What does the prefrontal cortex "do" in affect: perspectives on frontal EEG asymmetry research. *Biological Psychology* **67**, 219-233.
10. Bradley, M.M., Codispoti, M., Cuthbert, B.N., & Lang, P.J. (2001). Emotion and motivation I: defensive and appetitive reactions in picture processing. *Emotion*, **1(3)**, 276.
11. Mauri, M., Cipresso, P., Balgera, A., Villamira. M., Riva, G. (2011). Why is Facebook so successful? Psychophysiological Measures Describe a Core Flow State While Using Facebook. *CyberPsychology, Behavior and Social Networking*, **14(12)**, 723-731.

12. Picard, R.W., & Healey, J. (1997). Affective wearables. *Personal Technologies*, **1(4)**, 231-240.
13. Darley, W.K., & Smith, R.E. (1995). Gender differences in information processing strategies: An empirical test of the selectivity model in advertising response. *Journal of advertising*, **24(1)**, 41-56.
14. Goldberg, J.H., & Wichansky, A.M. (2003). Eye tracking in usability evaluation: A practitioner's guide. In *The Mind's Eye* (pp. 493-516).
15. Just, M.A., & Carpenter, P.A. (1976). Eye fixations and cognitive processes. *Cognitive psychology*, **8(4)**, 441-480.
16. Feist, G.J. (1998). A meta-analysis of personality in scientific and artistic creativity. *Personality and social psychology review*, **2(4)**, 290-309.
17. Lang, P.J. (1995). The emotion probe: Studies of motivation and attention. *American psychologist*, ***50*(5)**, 372.
18. Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.

Self-serving technology or an employee? A neurological approach

Hendrik Slabbinck[1], Iris Vermeir, Nanouk Verhulst and Maurizio Mauri[2]

1. Department of Marketing, Ghent University; 2. TSW s.r.l., User Experience Lab, Treviso, Italy. maurizio.mauri@tsw.it

Technology transformed the way how services are delivered. Key in this transformation is that human employees are often replaced by self-service technology (SST), but how consumers feel about services delivered by SST or a human service provider remains unclear. SST research to date mainly focused on customers' intentions to use SST, technological readiness, and ease of use of SST. Moreover, empirical studies on consumers' emotional responses during different stages of the service delivery is scarce. Hence, this paper seeks to answer two key questions that remain: (1) Do consumers experience different emotions (arousal and valence) and approach-avoidance behavior during a SST versus a human interaction? and (2) Do consumers experience different emotions and approach-avoidance behavior at different key moments during a service experience (e.g. the moment when a failure happens, a recovery is initiated, …) while using SST or human interaction.

Forty participants (50% men) took part in a between-subject experiment, where they were instructed to make a loyalty card for a book store either by themselves on a computer (SST-condition) or in interaction with an employee (human condition). The storyline of the encounters consisted out of several phases: a welcome, providing personal details to make the loyalty card, a service failure, a service recovery, providing the same personal details, and lastly the confirmation that the loyalty card was created successfully. Both conditions followed the same procedure, except that in the SST-condition the scripted text appeared on a screen and participants had to type in the personal data, whereas in the human condition the employee read out the scripted text and participants vocally gave their personal details.

We investigated the participants' emotions, approach-avoidance behavior, and service satisfaction by combining classic post-experience self-report measures (e.g. service satisfaction, emotions) and neuroscientific tools and measures. This design allows us to study emotions and consumers' reactions during every stage of the customer experience in real-time, instead of the traditional approach of measuring reactions only after interaction has taken place.

We used three different neurophysiological measures. First, changes in GSR by means of EDA electrodes which are used to measure stress reactions and indicate emotional arousal. Secondly, brain activity by means of EEG to assess frontal brain asymmetry. Frontal brain asymmetry gives us an indication of approach-avoidance behavior during the service experience. GSR and brain asymmetry were measured by using the Thought Technology Ltd. Hardware (ProComp Infiniti) and software (BioGraph Infiniti). Lastly, facial muscle activity was measured by means facial recognition software (Facereader 6) which registers facial micro expressions to measure the intensity (occurrence) of 6 basic emotional states (sad, happy, angry, surprise, scared, disgust, contempt + neutral. The Facereader software provides an intensity score between 0-1 for each emotional state, a global valence score of experienced emotions (calculated as intensity of 'happy' minus the intensity of the negative expression with the highest intensity), and a global arousal score based on muscle activation of 20 Action Units (AUs). Next to the neurophysiological measures, we also assessed service experience, emotions, levels of stress and frustration by means of post-experience self-report measures using 7-point Likert scales.

Results show that employee interactions elicit positive emotions, whereas SST fosters negative emotions, but both interaction modes yield similar customer satisfaction levels. Additionally, emotions fluctuated across different service delivery stages.

Measuring the Effects of Visual Cues on Eating Behaviour through a Combined Methodological Platform: Pilot Study

M.A. Vargas[1], J. Brunstrom[2], S. Pérez[1], V. Ciaurriz[1], S. Sobrino[3], B. Martínez de Morentin[1], S. Navas-Carretero[1,4], J.A. Martínez[1,4] and E. Almiron-Roig[1]

[1] Centre for Nutrition Research, University of Navarra, Spain. *ealmiron@unav.es* [2] Department of Experimental Psychology, University of Bristol, United Kingdom [3]Department of Food, Environmental and Nutritional Sciences, "Università degli Studi" of Milan, Italy [4] Spanish Biomedical Research Centre in Physiopathology of Obesity and Nutrition (CIBERobn), Institute of Health Carlos III, Madrid, Spain

The size and design of tableware have been proposed as a potentially effective strategy to modulate how much is eaten at a meal. The mechanisms by which such tools work however are not known, in particular the cognitive processes associated with visual stimuli. To address this gap, 94 lean, overweight and obese men and women will complete a laboratory study where they will self-serve and consume food from a buffet, using a portion control plate with visual stimuli for appropriate amounts of main food groups, or a conventional plate, on two different days in random order. On both sessions participants will complete behavioural and cognitive tests using a novel combined methodological platform that will assess visual attention during the meal, meal microstructure, episodic memory on portion sizes, subjective appetite and, in a sub-sample of subjects, intestinal satiety hormones. In this paper we discuss the setting up and testing of this platform in a pilot sample of 20 participants.

Introduction

Obesity is a recognised global health problem for which effective interventions are needed which can be delivered at a scale for public health impact [1]. There is currently good evidence that people tend to eat more and gain weight when exposed to large portion sizes [2,3] meaning that individuals may need to develop personal strategies to avoid overeating in such environments. In this context, the size and design of tableware have been proposed as a potential practical and effective staratgey to modulate how much is eaten [4] and so appropriately designed tools may help with portion control [5-7]. The particular mechanisms by which portion control tools may work are not known though, in particular the cognitive processes involved and how the effects differ across individuals of different gender and BMI.

To address this research gap we have designed an intervention where 94 lean, overweight and obese men and women will self-serve and consume food from a laboratory buffet, using a portion control (calibrated) plate with visual stimuli for appropriate amounts of main food groups, or a control (conventional) plate in random order, and will complete behavioural and cognitive tests using a novel combined methodological platform. The platform includes an eye tracking device to analyse visual attention, an eating monitor that records eating speed in real time, and a memory reconstruction software to test for memory formation related to portion sizes previously chosen with each plate. At various time points particpants will also complete electronic visual scale (VAS) questionnaires for meal liking, expected satiety and appetite. Portion sizes for all foods chosen and consumed will be covertly measured. In addition, in a sub-sample of participants, the cephalic satiety response will be measured from blood samples drawn at 6 points during the first 90 min after starting the meal. All volunteers will complete three paper-based portion perception questionnaires and a food diary for the remaining of the day from which energy compensation for the energy consumed at lunch will be calculated.

Due to the many variables involved and the tendency for behavioural measures to naturally act as confounding factors amongst themselves [8], collecting accurate data from this study is particularly challenging. In particular, effectively combining and synchronizing visual attention, eating speed and hormonal measures can be complex however this step is cornerstone to relate behavioural, cognitive and physiological responses in real time. In this paper we discuss the setting up and testing of the combined methodological platform for this study in a pilot sample of 20 participants.

Methods

Study design

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

The study follows a within-subjects cross-over design where 47 participants will be randomized to a first session with the calibrated plate, and 47 with the control plate (Figure 1). After a 7-15 day wash out period, participants reverse conditions and repeat the same measures. The study variables and corresponding measuring instruments are described below.

- *Demographic and anthropometric variables***:** age, sex, weight (kg), height (cm), body mass index (BMI) in kg/m^2, weekly hours of moderate to vigorous physical activity; consumption habits and household composition. This information is obtained through published questionnaires and direct measuring in the laboratory (for weight and height).
- *Behavioral variables*: Eating behavior traits (EAT-26 and TFEQ questionnaires, validated versions for the Spanish population) [9,10]; Habitual portion size [11], liking and expected satiety for the meal [12], and hunger, fullness, thirst, nausea before, immediately after and at 3h post meal (100 mm VAS) [13]; Portion-control self-efficacy [14], portion tool acceptance (5-pt Likert scales) [5]; Meal micro-structure measured with the Universal Eating Monitor (UEM) [15], including eating rate (g/min), bite size (g) and deceleration rate (g/sec^2); Gaze direction bias, initial fixation time and gaze dwell time (in ms) [16] for main areas of interest (AOIs) of the meal, measured with a portable eye tracker (Figure 2); Energy compensation (adjustment) at the end of the day (using a food diary) [17].
- *Physiological variables*: glucose, insulin, pancreatic polypeptide and ghrelin in serum and plasma at 0, 5, 10, 60 and 90 min post-consumption, using ELISA [18] in a sub-sample of 34 women.
- *Cognitive variables*: episodic memory for eaten portions analyzed at 3 h post consumption through a computerized task based on a bespoke algorithm [19].

Study plates

The calibrated plate measures 25 cm in diameter and includes demarcations and illustrations (portion size guidelines) for recommended amounts of protein foods, starchy foods and vegetables based on US Department of Agriculture guidelines. It has been specifically designed for this study by Precise Portions LLC based on previous research [5]. The control plate is a white dish of the same size and depth but slightly lighter in weight. Both are ceramic plates, microwave and dishwasher safe, with an enamel finish (Figure 3).

Study foods

The buffet foods include popular foods consumed by the Spanish population as part of a main meal and are: seasoned white rice, boiled vegetables (peas, carrots), meatballs in sauce, wholemeal bread, oil, salt and pepper (Figure 2). The rice, vegetables and meatballs are heated to a temperature of 66°C and presented hot in transparent serving bowls with identical serving spoons. These foods have been chosen because they match the nutritional composition required when using the calibrated plate, plus do not require cutting, i.e. applying strength on the UEM scale (which would alter the readings). Due to their large size, meatballs are presented halved. As this predefines the bite size, the average weight of the cut meat-balls will be recorded and if appropriate, used as covariate in the statistical analyses. Complimentary fruit ad water are provided after the meal. Foods for the memory reconstruction test are exactly the same foods except the bread, fruit and condiments (which are optional), cooked and presented in the same exact format as in the meal.

Equipment

Universal Eating Monitor (UEM). The UEM is designed to analyse meal micro-structural parameters (i.e. bite size, eating rate, deceleration rate and meal duration) when a volunteer consumes a meal sitting at a table [15]. An optimised UEM station was designed and built in-house for this study with particular attention to minimizing background vibration. It was placed in an isolated testing room within the Metabolic Unit with only artificial lightning from above and constant temperature below 30°C. The UEM components are a concealed precision scale (Model MSA5201S-1CE-D0, Sartorius Spain, Alcobendas, Madrid), connected with a serial line to a PC, and located beneath a purpose-built table under a hole, on top of which a place mat is secured to allow positioning of the plate. The PC hosts the Sussex Ingestive Pattern Monitor software (SIPM) [15] supplied by the University of Sussex (UK) programmed to record weight readings from the scale at 2 second intervals (precision 0.1 g). From these readings the average bite size (the difference between each two consecutive weight records), eating rate (grams consumed per minute) and deceleration rate (grams consumed per squared second) are calculated. A dual screen system is used to allow the investigator to programme the software away from the volunteer´s view using a dividing panel (Figure 4). The volunteer screen displays electronic questionnaires and step-by-step instructions for the test. The UEM at the University of Navarra has been purposely built on a bespoke anti-vibratory table containing a steel frame and granite slab measuring 2 x 35 x 35 cm, upon which the Cubis balance rests. The balance includes an auto-calibration function and electronic adjustable levelling legs. It also carries a detachable digital viewer located in a lockable, adjacent drawer.

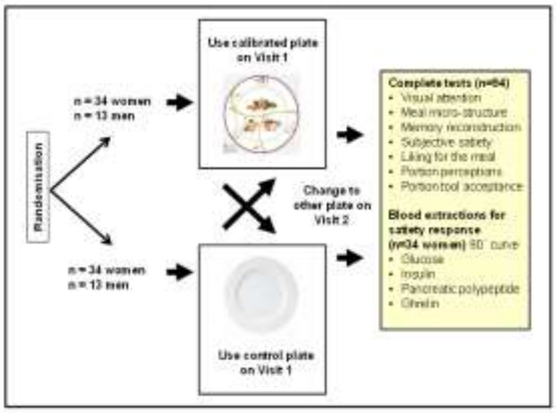


Figure 1. Study design.

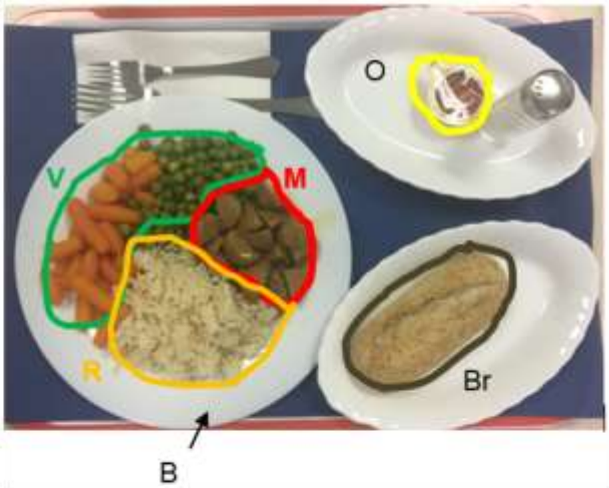


Figure 2. Areas of Interest used for the gaze data analysis in this study. The letters stand for Border, Empty plate (with or without images), Rice (on plate or on fork), Vegetables (both types, on plate or on fork), Meatballs (on plate or on fork), Bread (on plate, hands or side dish), Oil (on hands or side dish). Dwell time on mixed food areas, empty fork and condiments (salt, pepper) will also be recorded for quality control purposes

Figure 3. Plates used for this study. *Left*: calibrated plate (courtesy of Precise Portions LLC, Virginia, USA). *Right*: control plate (purchased from Group Carrefour, France). Both plates measure 25 cm in diameter including a 3.5 cm rim.

Mobile eye tracker. The wearable eye tracking equipment (Tobii Pro Glasses 2) is designed to measure visual attention and pupil dilation during exposure to 3D stimuli. It consists of a head unit (glasses), a recording unit and controller software (Figure 5). The head unit detect eye movement and point of gaze at a frequency of 50 MHz, directed to any 3D space using near-infrared illumination. The glasses are connected to a portable recording unit from which video output can be retrieved and downloaded to calculate dwell times on AOIs. The recording unit is connected to the head unit via an HDMI cable and stores the recorded data on an SD memory card. It is controlled from a tablet or computer running the controller software on a Windows operating system, allowing for managing participants, controlling the eye tracker and viewing both real-time and recorded eye tracking data.

Memory reconstruction software. This software measures episodic memory for portion sizes chosen/consumed in a previous eating occasion. It has been designed by the Nutrition Behaviour Unit at the University of Bristol (UK) using an image batch processing software (XnConvert version 0.84) that allows combining food images from different platforms to generate sequences of increasing or decreasing portions of specific foods (*http://www.xnview.com/en/*). For this study, the images for the starch, protein and vegetable foods come from a digital photographic atlas generated at the University of Navarra and featuring 110 images of the study foods in increasing portion size (5-20 kcal increments), all taken at 90 degrees angle. The software allows the user to select adjustable portions of each meal component to generate a personalized virtual plate, based on a method of adjustment [19] (Figure 6).

Subjects

The subjects for this pilot study were the first 20 volunteers (16 F, 4 M) completing the study. Subjects were recruited from Pamplona, Navarra, Spain, and had a mean ± SD age of 37.4 ± 12.7 years and body mass index of 27.8 ± 4.2 kg/m^2 (overweight), were healthy, with good visual and gastrointestinal (GI) function; absence of eating disorders, regular food habits and liking the study foods. To be eligible for the study all subjects had to pass a screening and familiarization session prior to the study where they consumed 125 g of yogurt on the UEM, while wearing the eye tracker, and completed electronic questionnaires on the UEM. Participants also tasted and rated the rice, meatballs and vegetables. Those assigning a score of 40 or more on a 100 mm VAS for the rice, meatballs and at least one of the vegetables, plus producing valid video and UEM outputs were enrolled. Six of the 20 subjects provided blood measures. Seven subjects used the calibrated plate first and 13 used the control plate first.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Procedures
The daily procedure for the volunteers is shown in Figure 7, and highlights the use of the different technologies (colour-coded by type of measure). This protocol was developed to minimize conflicting measuring times from the various techniques, based on previous research [20]. To decrease confounding due to protocol demands, participants were instructed to look away from the plate while the nurse carried out the drawings and the duration of blood extractions was extracted from the overall meal time for each participant providing blood samples. Video data during extraction were also excluded from analyses.

Figure 4. Universal Eating Monitor station at the University of Navarrra. Diagram courtesy of Borda Laboratorios, Madrid, Spain.

Figure 5. Tobii Pro Glasses 2 eye tracking device. Calibration is performed automatically with the controller software prior to each participant's test using a calibration card provided by the manufacturer. Source: http://www.tobiipro.com/

Figure 6. Screenshot of the portion size reconstruction software. The food photos were taken by a professional photographer using a digital camera with constant lightning and angle, and the same control dish. Portion sizes started at the equivalent of 1 tablespoon and proceeded by 20 kcal increments until the food filled about 3.5 quarters of the plate (assumed to be the maximum volume physically fitting in the plate, based on the study protocol, which required to select at least 3 meal components). The initial 1 tablespoon portion is based on 50% of the average small portion of cooked rice for Spanish consumers. The 20 kcal increments are based on previous research using full plates [19]. For carrots, due to their low energy density, 5 kcal increments were used.

For the setting up of the **combined methodological platform** the following steps were carried out:

1. UEM set up. After initial programming of the SIPM software, two internal tests were carried out with two members of our staff using 125 g of yogurt which they had to consume from a bowl following the UEM standard protocol [15, 20]. This protocol specifies sitting in an upright position without applying any strength on the table from either above or beneath the surface, not moving the container, placing the cutlery on a side dish after finishing or during pauses, not using mobile phones, tablets or other electronic devices except when instructed by the investigator and alerting the investigator using a bell when finished. The same instructions were applied for the screening and intervention sessions. Programming of the UEM software was carried out as separate experiments for the screening and intervention sessions, with particular attention to being able to synchronize the UEM measures with the eye tracker (i.e. being able to start and stop the video recording alongside UEM measures during the food intake task).
2. Eye tracker set up. The Tobii eye tracker was set up in the same environment-controlled room as the UEM. The Pro Glasses controller software was installed on the same PC after resolving initial IT incompatibilities. Calibration and gaze recording tests were carried out on the same internal volunteers

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

while they ate at the UEM station to verify functionality and adjustments to the UEM software were made as necessary. Video data were recorded in MP4 format at 25 fps.

3. Portion size memory reconstruction test trial. Two dry run tests were conducted on an internal volunteer by comparing the recalled and actual portion sizes after food consumption using each plate. To verify that the task was well comprehended by the volunteer, the recalled portions were matched to the actual portions and then checked against the volunteer´s real intentions.
4. Blood extraction protocol feasibility trial. To explore if the set timings for blood extractions were compatible with the UEM and eye tracking measures, we recorded real extraction times against target times for all extractions for the first 6 volunteers who provided blood samples.

Data analysis

To evaluate the feasibility and efficiency of using the combined methodological platform described above the following analyses were carried out in the data arising from the 20 participants:

a) Probability of invalid UEM output (n=40 outputs).
b) Probability of invalid memory test output (n=19, excluding missing data for 1 volunteer)
c) Average time required for blood extractions while participants were eating, computed as absolute and proportional time vs. total meal time (n=12, or 6 subjects x 2 sessions).
d) Average difference between target and real times for blood extractions across all times and conditions (n=60, or 6 subjects x 5 times x 2 sessions), using an independent samples t-test.
e) Quality of video data and identification of AOIs from a sub-sample of videos (n=6). Further video data exploration and the development and piloting of the coding protocol will follow.

Statistical analyses were conducted using STATA v.12. Significance was set at the 0.05 level.

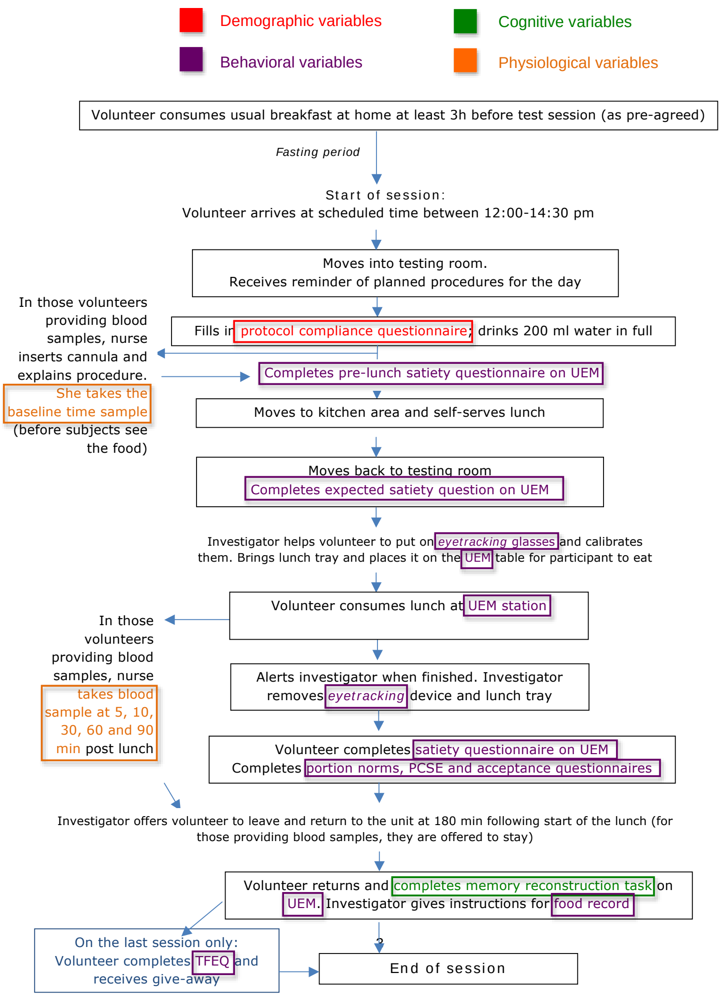


Figure 7. Daily procedure for participants and components of our combined methodological platform.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Results

UEM performance

A total of 40 UEM data outputs were collected for the first 20 volunteers. Of the 40 outputs, only 7 (17.5%) were invalid, including outputs for sessions where participants moved or slid the plate (n=2 outputs); the SIPM software failed due to incompatibility with other equipment (n= 3); the balance had been wrongly tarred by mistake (n=1); and the participant sat incorrectly at the UEM table (with their legs pressed against the table from underneath and so creating negative pressure on the balance) (n=1). Despite that occasional vibration caused by participants unconsciously applying pressure on the table was detected in all the meals, this was not a factor to invalidate the UEM measures since it did not affect the weight recordings.

Memory reconstruction test performance

The memory algorithm showed 100% efficiency across the 20 participants. No participants reported any difficulty with using or understanding the task. Comparison of recalled vs. actual and vs. intended portion size selection by our internal volunteer revealed that all measures were close but larger differences between recalled and selected portion size at the buffet were present for particular foods. This was confirmed in preliminary analyses of differences between eaten vs. recalled portions across the 20 pilot volunteers (Figure 8).

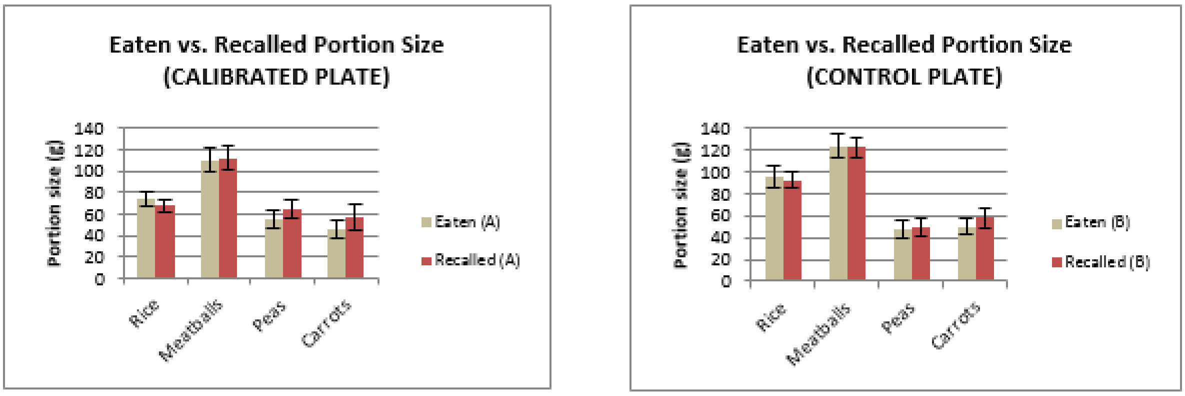


Figure 8. Preliminary results from the memory reconstruction task (n=19-20 subjects after excluding missing data for 1 visit). Comparison of eaten vs. recalled portion sizes after using the calibrated (*left*) or control plate (*right*).

Feasibility of blood extraction protocol

Overall the integration of the blood extraction protocol with the other measurements proved feasible however it required tight attention to measurement timings. It also required good coordination in the team as processing of blood samples for ghrelin analyses requires of the addition of a protease inhibitor immediately after drawing the sample (therefore requiring at least 2 investigators at that point in time).

1.1. Time required for blood extractions while participants were eating

Across the 6 participants who provided blood samples, the mean ± SD time required using the calibrated plate for blood extraction was 201 ± 67 sec and for the control plate it was 272 ± 70 sec. These times represent 23.3 ± 3.1 % and 29.2 ± 8.2 % of the total meal time in each condition. To avoid erroneous calculations in UEM and visual attention parameters, these periods of time will be excluded from the respective analyses. Valid UEM outputs were obtained for all volunteers who provided blood samples in both visit days, except for one volunteer on one visit day (probably due to an IT incompatibility).

1.2. Blood extraction target times

Across the 6 participants providing blood samples, for both sessions and 5 extraction times beyond baseline (5, 10, 30, 60 and 90 min), the average time difference between the target and actual extraction time was 37.0 ± 47.0 seconds when considering absolute differences, and 3.0 ± 59.9 seconds when considering negative and positive differences (which effectively compensated each other). These times did not differ much by plate condition although in absolute terms they were slightly larger for the control plate (Table 1), suggesting that while the overall time diversion from the target times was nearly null (3 seconds), some samples were taken slightly before and some slightly after the target time on average by 1.16 min (Figure 9). Much of the deviation was driven by one outlier measure taken 240 sec away from the target time, while for all other measures the absolute differences ranged from 0 to 0.8 minutes. Excluding this outlier from the analysis did not change the results (absolute differences 33.6 ± 39.0 sec).

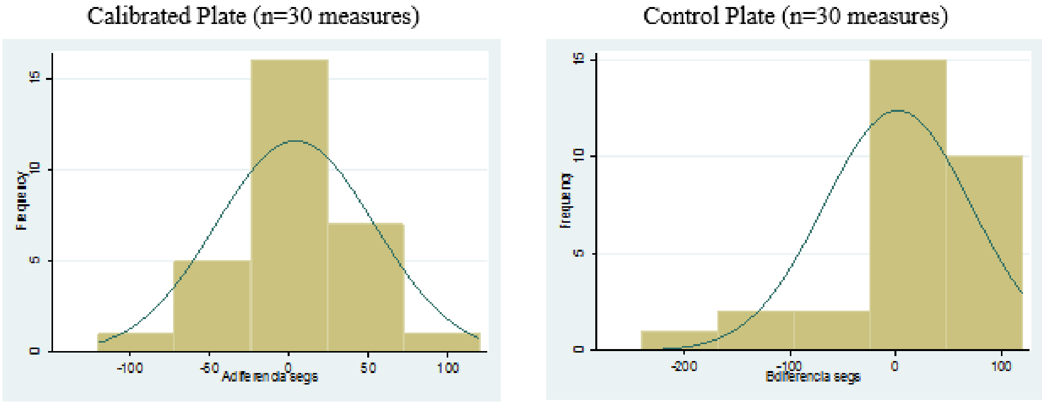


Figure 9. Frequency distribution of the differences between target and actual extraction times for extraction times at 5, 10, 30, 60 and 90 min for 6 volunteers (n=30 measures for each plate). Values on the X axis are seconds.

Table 1. Mean ± SD and 95% confidence interval (C.I.) for the differences between target and actual extraction times (in seconds), for extractions carried out in 6 volunteers at 5, 10, 30, 60 and 90 min after starting the meal, when eating with the calibrated and control plate. Significant differences from zero (zero considered as null deviation) are indicated with an (*).

	Using actual value (>0 or <0)		**Using absolute value**	
	Mean ± SD	**95% C.I.**	**Mean ± SD**	**95% C.I.**
Average of both plates (n=60 measures)	3.0 ± 59.9	(-12.5, 18.5)	37.0 ± 47.0 *	(24.9, 49.1)
Calibrated plate (n=30 measures)	4.0 ± 49.7	(-14.5, 22.5)	32.0 ± 37.7 *	(17.9, 46.1)
Control plate (n=30 measures)	2.0 ± 69.6	(-24.0, 28.0)	42.0 ± 54.9 *	(21.5, 62.5)

* differs from 0 with $p<0.001$.

Eye tracking data

Initially, the eye tracking data collected appear of sufficient quality and value to allow for a detailed analysis of the main AOIs of interest for this study (Figure 2). We are currently developing a coding protocol for manual coding of gaze times using the open access software Lightworks. For the gaze analyses, the main challenge is the presence of gazes on mixed food zones that do not fall under any of the pre-defined AOIs. An initial exploration of the magnitude of fixation time on such mixed food zones across 3 representative participants when they ate with the calibrated plate, indicated that these zones account on average, for less than 5% of the overall fixation time for the meal. Further analyses are under way to confirm these data using 3 additional videos of the same participants when eating with the control plate. This will be followed by protocol finalization and piloting across two trained coders.

Discussion

Taken together, our preliminary observations demonstrate the feasibility of combining eye-tracking, an optimized UEM, and a memory test, to explore underlying processes by which portion-control plates impact energy intake.

Our improved UEM design has resulted in a much lower proportion of erroneous outputs compared with the traditional version, where the balance is directly attached to the surface of the table, therefore allowing vibrations from the participant to directly transfer to the balance. Our previous study using the traditional UEM reported an error rate of 33% due to equipment faults and participants unconsciously unsettling the equipment [20]. This is almost double the error rate observed in this preliminary exploration with an optimized UEM. Another advantage of our methodological platform is that it allows detailed observation of participant´s actual behavior on the UEM thanks to the eye tracker. Often researchers cannot identify the exact cause of erroneous UEM outputs, with these typically being assigned to incorrect seating at the table or moving the food container [20]. With the simultaneous use of the eye tracker we can identify potential failures related to not just the UEM but also other procedures. This can help improving the design of future studies involving combined methodological platforms.

The main challenge in this study was the inclusion of blood extractions as drawing times may interfere with the time recordings for the UEM and with gaze data. While physiological measures have previously been combined with UEM recordings [21], to our knowledge, applying eye tracking measures during a real food intake test has not been attempted before, either alone or in combination with the UEM. A very recent study [22] explored visual attention paid to food during self-serving in a real-life buffet context, however visual attention was not measured during actual eating of the food, and physiological parameters were not monitored. As our study includes a sub-sample of volunteers not providing blood samples, we will be able to isolate any potential effects of the blood sampling procedure itself. In our analysis, blood extraction time was overall similar in both testing conditions, suggesting that the extraction procedure does not condition or cause a change in meal length differentially by testing conditions. This is initially supported by the observed high proportion of valid UEM outputs (11 out of the 13 blood extraction sessions) however this will need confirmation in the final sample. We did detect a slight deviation from target extraction times which is slightly above the recommended 30 sec interval used in a previous study involving similar equipment [18] and so addressing this aspect is important for quality control purposes.

Finally, it is important to consider including a period of training for the staff who will be implementing the protocols, as well as for participants themselves, to ensure everyone is well familiarized with the equipment and the protocol. This is particularly relevant for accurate coding of gaze data and the correct operating of the UEM for which a level of previous experience is required. Ensuring low inter-rater error rates if more than one video data coder is involved is especially important (e.g. through the use of well-developed SOPs), as well as introducing measures to reduce variability in participant's responses due to unfamiliarity with the equipment, foods and the eating environment itself.

In conclusion, this preliminary analysis of a novel methodological platform indicates that combining eye-tracking, an optimized UEM, and a memory test, alongside physiological measures of satiety, is a feasible and promising approach for the detailed analysis of eating behaviour processes in controlled laboratory environments.

Statement on ethics

6. Ethical approval for this study was obtained by the University of Navarra Research Ethics Committee on 27 April 2017 and 17 November 2017 (revised version), with reference number 2017.031mod1.

Acknowledgements

Financial support for this study was received from the Centre for Nutrition Research, University of Navarra, and the Spanish Biomedical Research Centre in Physiopathology of Obesity and Nutrition (CIBERobn). The authors are grateful to Precise Portions LLC and Herba RiceMills SLU for donating study foods and kitchen utensils respectively, as well as to Alejandra Yanes and M. Ysabel Ortega for help with data collection. MAV was supported by a travel grant from MB2018 to present this work.

References

1. Ng M, Fleming T, Robinson M *et al.* (2014). Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384**, 766-781.
2. Zlatevska N, Dubelaar C, Holden SS (2014). Sizing Up the Effect of Portion Size on Consumption: A Meta-Analytic Review. *Journal of Marketing* **78**, 140-154.
3. Rolls BJ. What is the role of portion control in weight management? (2014) *International Journal of Obes*ity **38 (Suppl 1)**,S1-S8.
4. Hollands GJ, Shemilt I, Marteau TM *et al.* (2015) Portion, package or tableware size for changing selection and consumption of food, alcohol and tobacco. *Cochrane Database of Systematic Rev*iews **9**, CD011045.
5. Almiron-Roig E, Domínguez A, Vaughan D, Solis-Trapala I, Jebb SA (2016). Acceptability and potential effectiveness of commercial portion control tools amongst people with obesity. *British Journal of Nutrition* **116**, 1974–83.
6. Kesman RL, Ebbert JO, Harris KI, Schroeder DR (2011). Portion control for the treatment of obesity in the primary care setting. *BMC Research Notes* 4, 346.
7. Pedersen SD, Kang J, Kline GA (2007). Portion control plate for weight loss in obese patients with type 2 diabetes mellitus: a controlled clinical trial. *Archives of Internal Medicine* **167**, 1277-1283.
8. Allison DB, Baskin ML. Handbook of assessment methods for eating behaviours and weight-related problems. Measures, theory, and research. 2nd ed. Thousand Oaks, CA: Sage Publications, Inc; 2009.
9. Sánchez-Carracedo D, Raich I Escursell RM *et al*. (1999). Preliminary adaptation of the Stunkard & Messick Three Factor Eating Questionnaire (TFEQ) with a Spanish university sample [Spanish]. *Psicología*

Conductual **7**, 393–416.
10. Rivas T, Bersabé R, Jiménez M, Berrocal C (2010). The Eating Attitudes Test (EAT-26): reliability and validity in Spanish female samples. *Spanish Journal of Psychology* **13**,1044–1056.
11. Robinson E, Oldham M, Cuckson I, Brunstrom JM, Rogers PJ, Hardman CA (2016). Visual exposure to large and small portion sizes and perceptions of portion size normality: Three experimental studies. *Appetite* 98, 28–34.
12. Forde C, Almiron-Roig E, Brunstrom JM. Expected Satiety: Application to weight management and understanding energy selection in humans (2015). Current Obesity Reports **4**, 131–40.
13. Hill AJ, Blundell JE. Nutrients and behaviour: Research strategies for the investigation of taste characteristics, food preferences, hunger sensations and eating patterns in man (1982). *Journal of Psychiatric Research* **17**, 203–212.
14. Fast LC, Harman JJ, Maertens JA, Burnette JL, Dreith F (2015). Creating a measure of portion control self-efficacy. *Eating Behaviors* **16**, 23–30.
15. Yeomans MR (2000). Rating changes over the course of meals: what do they tell us about motivation to eat? *Neuroscience and Biobehavioral Reviews* **24**:249–59.
16. Werthmann J, Roefs A, Nederkoorn C, Mogg K, Bradley BP, Jansen A (2011). Can(not) Take my Eyes off it: Attention Bias for Food in Overweight Participants. *Health Psychology* **30**, 561–569.
17. Almiron-Roig E, Palla L, Guest K *et al*. (2013). Factors that determine energy compensation: a systematic review of preload studies. *Nutrition Reviews* **71**, 458-473.
18. Yeomans MR, Re R, Wickham M, Lundholm H, Chambers L (2016). Beyond expectations: the physiological basis of sensory enhancement of satiety. *International Journal of Obesity* **40**, 1693-98.
19. Brunstrom JM. Mind over platter: pre-meal planning and the control of meal size in humans (2014). *International Journal of Obesity* **38 (Suppl 1)**, S9-12.
20. Almiron-Roig E, Tsiountsioura M, Lewis HB, Wu J, Solis-Trapala I, Jebb SA (2015). Large portion sizes increase bite size and eating rate in overweight women. *Physiology & Behavior* **139**, 297–302.
21. Näslund E, Gutniak M, Skogar S, Rössner S, Hellström PM (1998). Glucagon-like peptide 1 increases the period of postprandial satiety and slows gastric emptying in obese men. *American Journal of Clinical Nutrition* **68**, 525-530.
22. Wang E, Cakmak YO, Peng M. Eating with eyes - Comparing eye movements and food choices between overweight and lean individuals in a real-life buffet setting (2018). *Appetite* **125**, 152-159.

Symposium: Sequence Analysis Approaches to Discovering Hidden Temporal Patterns in Behaviour

Studying behaviour outside of the laboratory is often difficult as information may be incomplete, idiosyncratic, fragmented, partly qualitative, and rapidly changing. Developing methods to detect and analyse the non-random, syntactic structure of behaviour allows researchers to understand and predict future behaviours. This symposium will provide a series of talks centred around Sequence Analysis methods, which allow the progression of real-world, complex behaviours (both animal and human) to be mapped and understood. The symposium will begin with a brief outline of the basic approach of Lag-Sequence Analysis (LSA), aimed at those with no prior experience or knowledge of Sequence Analysis research. This talk will also focus on Indicator Waves, a novel method of sequencing multiple, concurrent behaviours across time. The second talk will explain the use of proximity coefficients to analyse the interrelationships among sequences of behaviours within and between cases. Using examples from law enforcement interactions, the talk will show how this approach opens up conventional inference testing across variables that distinguish different sequences. The final talks will outline the T-System Approach in Sequence Analysis. The T-System is a formally defined set of probabilistic patterns of relations between behaviours or events on a discrete time scale. This may seem quite similar to LSA; however, the T-System approach allows for much more sophisticated, hierarchical analyses of sequential and concurrent behaviour over time. These complex T-System analyses will be clearly explained to show how hidden behavioural relationships can be detected, which would be impossible with traditional LSA methods. The T-Systems talks will include novel, extended analysis of T-Pattern sets for the new T-prediction and T-retrodiction. In addition, the newly updated THEME software, with increased speed of large dataset analysis, will be outlined, with applications to big data. Finally, drawing from the literature, examples will be provided on how the T-System approach can be applied to research in human and non-human samples. The symposium will end with a discussion and evaluation of which method is best for particular research questions or approaches, and the future of all these methods will be discussed.

David Keatley, University of Lincoln, UK

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Indicator Waves: A new temporal method for measuring multiple behaviours as indicators of future events

Keatley, D. A[1]. and Clarke, D. D.[2]

1Researchers in Behaviour Sequence Analysis (ReBSA) and University of Lincoln, LN6 7TS.
2Researchers in Behaviour Sequence Analysis (ReBSA) and University of Nottingham, NG7 2RD.

The traditional approach to research in Psychology has been to use cross-sectional designs. While temporal methods are not new, they are relatively under-used; however, researchers have been developing methods to analyse temporal dynamics [1 - 3]. Mapping changes in behaviour over time facilitates understanding of causal effects and predicting of future outcomes. To date, there are several main approaches to analysing the influence of variables over time in dynamic systems: Survival Analysis [4 - 5]; Behaviour Sequence Analysis [3], T-Pattern Analysis [6 - 7]; and temporal Social Network Analysis [8 - 9]. All of these methods have proven hugely beneficial in mapping the effects of multiple antecedent factors leading toward different outcomes. The aim of the current presentation is to offer a novel method of temporal measurement, *Indicator Waves*, which allow multiple simultaneous and sequential events to be analysed across varying time-spans. Indicator waves provide easy-to-read statistical wave diagrams, which are interpretable by a wide range of non-specialist end-users. The waves provide quick inference about which risk factors, behaviours, or events (termed 'Indicators') are prevalent across different points in time. The plots at each time point provide a profile of the absence or presence of indicators at that time. The method can be applied to both human and non-human samples.

Understanding the temporal effects of different indicators can provide important insight into the fluctuations of different events across time. For instance, if a researcher were attempting to forecast the next episode of relapse in addictions or self-harm [3, 10, 11], knowing the effects of intrapersonal and interpersonal factors in the preceding minutes, hours, days, weeks, or even months, might allow better prediction of subsequent relapses or episodes. Traditionally, longitudinal studies have attempted to provide this sort of information; however, longitudinal studies incur the limitation of sample drop-out, or having to wait an unknown length of time until relapse (which might never occur!). These limitations are well-known and considered a worthwhile hurdle in order to gain longitudinal data. Some research designs make use of existing, legacy datasets, which removes some of the issues of longitudinal research; however, such datasets can be hard to access or miss key variables that limit the research. A further limitation of temporal analyses, such as Survival Analysis, is that accounting for multiple behaviours across time is not easily possible. A method is required, therefore, that allows for temporal dynamics to be analysed, across multiple indicators.

Behaviour Sequence Analysis (BSA) methods offer researchers the opportunity to measure the relationship between multiple behaviours across time [3]. BSA is based on mathematical Markov models, which analyse the transitions between one event, the *antecedent*, on a following event, the *sequitur*. Complex chains of events are then developed to show how *A* leads to *B*, *B* leads to *C* and so forth. Lag-Sequence Analysis (LSA), which is the most typical form of BSA, typically involves analysing the effect of an antecedent on a sequitur [3, 12 - 13]. The analyses does allow for longer chains to be built, or more distal lags to be measured; however, the underlying focus is on the transition relationships between antecedents and sequiturs. This method has proven useful in a number of domains. The analysis is possible; but, can lead to a combinatorial explosion in analyses [13].

T-Pattern Analysis (TPA) has been proposed, therefore, to overcome the limitations of traditional LSA approaches. TPA allows for complex patterns of sequential and concurrent behaviours to be analysed, and hierarchical patterns to emerge. This method allows for timing of sequential and hierarchical events to be mapped in a dataset, showing fractal dimensions, which can increase prediction of future events [7]. Some of the limitations in Lag-Sequence Analysis are, therefore, overcome with TPA. However, TPA is a much more complex form of analyses, and can require additional training to run and interpret analyses. While the 'pay-off' for understanding such analyses is worthwhile, it can be restrictive to some audiences.

A method is clearly needed, therefore, which brings together the benefits of temporal analyses with multiple events. The method should allow for outputs that can be easily followed and understood, while also allowing multiple concurrent and sequential behaviours to be mapped. Indicator Waves provides all of these strengths, without any of the limitations. Indicator Waves allow multiple behaviours and events to be analysed at varying time scales to be measured at different points in time. The analyses can run on very short timescales (e.g., milliseconds) through to life-history timelines. The output of analyses are waves that outline how likely (or unlikely) the occurrence of indicators are at each time point. End-users can read waves across the graph to see when and where they fluctuate, the Indicator Waves diagram can also be interpreted as ‘profiles’ at particular time points, indicating which events are likely to occur at that time point, and which are less likely to occur.

The Indicator Waves Method

The first stage for researchers is to define the various indicators that they want to measure or map over a time course. This stage is analogous to creating the behaviour list in BSA research [3]. As with BSA, it may be that researchers define the indicators, *a priori*, or *post hoc*; however, ensuring a complete list is developed increases clarity and precision of results. The next step is to decide on the time scale and intervals used to measure the indicator waves across. A coarse time-scale with large intervals may be better suited to life history research; however, smaller intervals can be used for smaller time-scales (e.g., nonverbal communication and micro-gestures). The timeline may also be compressed and elongated depending on research-relevant question that may be of interest. For example, as an outcome event nears, it may be important to investigate the fluctuations of indicators on smaller time scales. It is possible, therefore, to have a timeline that has indicators at monthly ‘check-ups’ or points in time; however, in closer proximity to the outcome event, the timeline may include weekly, daily, or even hourly inputs. The analyses can be ‘zoomed in’ or ‘zoomed out’ of particular time points, to show greater clarity in indicator fluctuations.

Data may be gained from a variety of sources, as with BSA research [3]. Existing, legacy datasets or sources may be used and coded, in written, spoken, or video form. Similarly, researchers may conduct cross-sectional studies, asking participants to recount their previous behaviours and experiences of indicators in the past. Finally, researchers may wish to conduct a longitudinal design with Indicator Waves (or include an Indicator Waves questionnaire in their existing longitudinal research). The Indicator Waves questionnaire is relatively straightforward and simple to design, and can be administered via pen-and-paper, or online questionnaire sites and mobile apps. An example of an Indicator Waves questionnaire is given in Figure 1.

	The previous 1 hour	The previous 24 hours	1 week ago	1 month ago	3 months ago	6 months ago	9 months ago	12 months ago	Over 12 months ago
Felt depressed	☐	☐	☐	☐	☐	☐	☐	☐	☐
Felt angry	☐	☐	☐	☐	☐	☐	☐	☐	☐
Felt despair	☐	☐	☐	☐	☐	☐	☐	☐	☐
Felt pessimistic	☐	☐	☐	☐	☐	☐	☐	☐	☐
Felt anxious	☐	☐	☐	☐	☐	☐	☐	☐	☐
Felt emotionally abused	☐	☐	☐	☐	☐	☐	☐	☐	☐
Felt neglected	☐	☐	☐	☐	☐	☐	☐	☐	☐
Felt frustrated	☐	☐	☐	☐	☐	☐	☐	☐	☐
Felt desperate	☐	☐	☐	☐	☐	☐	☐	☐	☐

Figure 1. Indicator Waves Questionnaire. The Indicators in column 1 are typically much longer lists, involving events, behaviours, or cognitions. The timeline can have greater or shorter intervals, depending on research questions. This example is for an Indicator Waves project predicting an outcome behaviour, it is not always necessary to have an outcome, per se – the method can be used to map trajectories of indicators across time (e.g., across an expedition or treatment programme)

Analyses

Similar to traditional BSA research [3], the first step in Indicator Waves analysis is to gain an insight into the raw frequencies of individual indicators in the dataset. This most closely corresponds to previously conducted research on known risk factors and variables associated with an outcome, it also provides end-users with an initial insight into which indicators occur most or least frequently in the dataset. Frequencies can be globally calculated – across the entire timeline, or can be calculated at each time point. The analysis of frequencies is a simple first step; however, further analyses are required to fully understand the distribution and likelihood of events occurring.

Indicator Waves

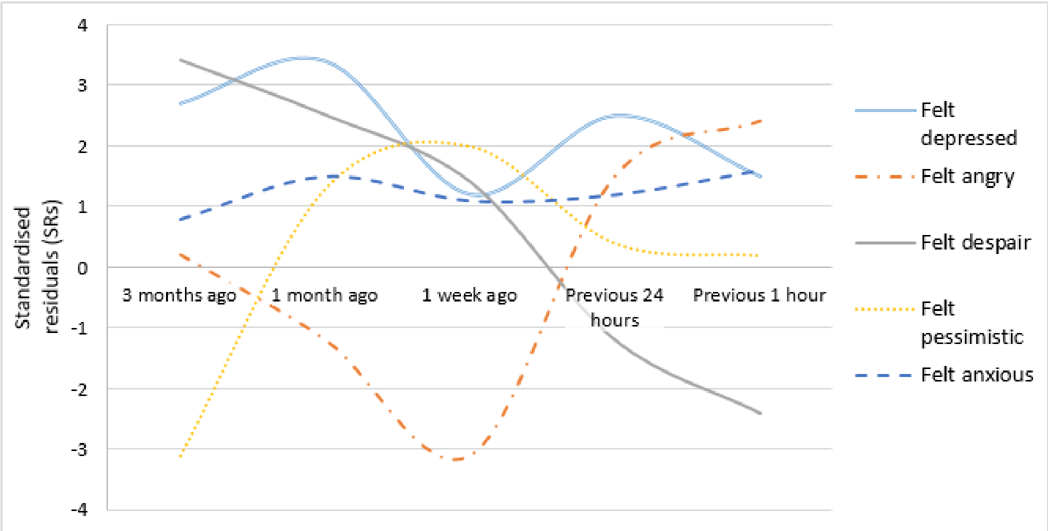
Chi-square analysis of the Indicator Waves frequency matrix shows standardised residuals (SR), which are the preferred analysis approach in Indicator Waves, for the same reasons as given in BSA research [3]. Standardised residuals show which indicators are occurring more times, or fewer times than expected by chance at each time point. A higher SR means that an indicator is occurring more than would be expected by chance, at a particular time point. In contrast, a negative SR shows than an indicator is occurring fewer times than would be expected by chance, at a time point. A table can be developed of SRs to show indicators x time points, which is similar to the transition frequency matrix produced in BSA research [3]

	Previous 1 hour	Previous 24 hours	1 week ago	1 month ago	3 months ago	…
Felt depressed	1.5	2.5	1.2	3.4	2.7	
Felt angry	2.4	1.5	-3.1	-1.2	0.2	
Felt despair	-2.4	-1.2	1.4	2.5	3.4	
Felt pessimistic	0.2	0.4	2.0	1.3	-3.1	
Felt anxious	1.6	1.2	1.1	1.5	0.8	
…						

Table 1. Standardised Residual Matrix (segment from a larger database). Values are standardised residuals (SRs). Positive values show each indicator event occurs more likely than would be expected by chance, at a particular time point. Negative values that the indicator event occurs less likely than would be expected by chance, at that time point. The final column indicates this timeline continued across many more previous time points.

The table provides the complete results for all indicators, at each time point. However, for larger tables, interpreting the table and understanding the combined presence or absence of a number of indicators, across different time points would be extremely difficult. Therefore, the next stage of Indicator Waves is to produce a diagram of the waves, which can be more easily interpreted. The diagram (see Figure 2) shows how different indicators are present or absent at different time points.

The Indicator Waves diagram can be very simply read and understood by individuals without advanced statistical training, making it ideal for applied use. The lines show which indicators (*Felt depressed; Felt angry* etc.) occur at different points in time. Depression, for instance, is likely to occur throughout each time point; however, peaks at around a month prior to the predicted outcome event. Anger (*Felt angry*) peaks an hour



before the outcome; but, dips and is absent a week and a month prior to the outcome event. Each plot along the *x*-axis indicates a specific point in time, and shows how indicators may be present or absent at different times in the build-up to an outcome.

Figure 1. Indicator Waves diagram. Timeline order has been switched from Table 1, to represent passage of time. This is a segment of a longer timeline dataset (spanning many months).

Each point on the *x*-axis is akin to a profile at that point in time. Therefore, an individual experiencing no despair; but, experiencing anger, anxiety, and depression is likely to be in the 'previous 1 hour' point. Similarly, an individual without pessimism; but, increased levels of depression and despair, is likely to b at the '3 months ago' time point. This has a clear clinical application, and advanced statistical approaches are developed to provide an index of matching new cases to known cases in a database.

Discussion and Conclusions

A limitation of previous time-series analysis methods has been the complexity of outcomes when multiple concurrent and sequential events are plotted [3]. A primary aim of applied research should be to develop methods that can be interpreted by lay audiences, without statistical backgrounds. State Transition Diagrams, for instance, are simplified models of complex datasets, which allow individuals to follow chains of behaviour [3, 13]. Although concurrent behaviours can be analysed in BSA, this typically leads to very complex analyses and outputs. For example, from a simple *A→B→C* chain, multiple groupings can be made (i.e., [*A→B; A→C, B→C, AB→C, A→BC* etc.), rendering the simplified State Transition Diagrams prohibitively complex to read. The current Indicator Waves approach, however, allows for complex datasets of sequential and concurrent events to be analysed and plotted in simplified diagrams.

A further benefit of the Indicator Waves approach is the possibility of matching cases to existing datasets. Each time point in the Indicator Waves approach offer a profile of the typical presence or absence of indicators at that point in time. When a new case emerges, further analyses can be conducted to investigate which profile point in time it most closely matches. This offers researchers and practitioners the ability to make inferences about the time point a new case is currently at – and, therefore, indicate how long it is before an outcome event is likely to occur.

Clearly, there are more developments to be made to the Indicator Waves method, and further opportunities to expand the method to analyse wider ranges of data and cases. Already, the Researchers in Behaviour Sequence Analysis (ReBSA) are developing approaches to further clarify the case matching analyses, as well as working on experimental analyses to show the reliability of datasets. This presentation will provide attendees with an overview of the Indicator Waves approach, and highlight some of these further analyses that are currently being developed.

References

[1] P. J. Taylor *et al.*, "Analysing forensic processes : Taking time into account," *Issues Forensic Psychol.*, pp. 43–56, 2008.

[2] J. Ivanouw, "Sequence analysis as a method for psychological research," *Nord. Psychol.*, vol. 59, no. 3, pp. 251–267, Jan. 2007.

[3] D. A. Keatley, *Pathways in crime: an introduction to Behaviour Sequence Analysis*. London: Palgrave Macmillan, 2018.

[4] J. B. Greenhouse, D. Stangl, and J. Bromberg, "An introduction to survival analysis: statistical methods for analysis of clinical trial data.," *J. Consult. Clin. Psychol.*, vol. 57, no. 4, pp. 536–44, 1989.

[5] J. D. Singer and J. B. Willett, "Modeling the days of our lives: Using survival analysis when designing and analyzing longitudinal studies of duration and the timing of events.," *Psychol. Bull.*, vol. 110, no. 2, pp. 268–290, 1991.

[6] M. S. Magnusson, "Discovering hidden time patterns in behavior: T-patterns and their detection.," *Behav. Res. Methods. Instrum. Comput.*, vol. 32, no. 1, pp. 93–110, 2000.

[7] M. S. Magnusson, J. K. Burgoon, and M. Casarrubea, "Time and Self-Similar Structure in Behavior and Interactions: From Sequences to Symmetry and Fractals," in *Discovering Hidden Temporal Patterns in Behavior and Interaction*, New York: Springer, 2016, pp. 3–35.

[8] S. P. Borgatti, M. G. Everett, and J. C. Johnson, *Analyzing Social Networks*. London: Sage, 2013.

[9] G. Robins, *Doing social network research: network based research design for social scientists*. London: Sage, 2015.

[10] R. Wadman *et al.*, "A sequence analysis of patterns in self-harm in young people with and without experience of being looked after in care," *Br. J. Clin. Psychol.*, vol. 56, no. 4, pp. 388–407, 2017.

[11] E. Townsend *et al.*, "Uncovering key patterns in self-harm in adolescents: Sequence analysis using the Card Sort Task for Self-harm (CaTS)," *J. Affect. Disord.*, vol. 206, pp. 161–168, 2016.

[12] R. Bakeman and J. M. Gottman, "Observing behavior: An introduction to sequential analysis," *Cambridge Cambridge Univ.*, 1986.

[13] D. D. Clarke and J. Crossland, *Action systems: An introduction to the analysis of complex behaviour*. London: Methuen, 1985.

An overview of recent research involving the T-pattern model

G.K.Jonsson[1]

[1]Human Behavior Laboratory, University of Iceland, Reykjavik, Iceland. gjonsson@hi.is

The discovery of hidden patterns in behavior is a task frequently faced by numerous researchers across many investigation areas, such as biology, psychology, psychiatry, sport science, robotics, finance etc. Nevertheless, discovering such patterns has proven to be a difficult task due to a lack of three important issues: first of all, adequate formalized models of the kinds of patterns to look for; second, corresponding detection algorithms and, last but not least, their implementation in available software. Over the last decades, these difficulties have been gradually overcome with the mathematical T-pattern model and the continued improvement of a technique known as T-pattern detection and analysis (TPA). Such a technique has been implemented in a software specifically developed for this purpose known as Theme. Using such an approach, recurring synchronic and sequential patterns of behavioral events, otherwise usually hard (or, more often, impossible) to detect, can be discovered and analyzed. TPA has been advantageously used to describe human and animal behavior in various contexts and experimental approaches. Demonstrating this approach we will examine results from several different research projects: temporal patterns in attacking rugby sessions (analyzing play events, location and heart rate of players); temporal pattern detected in data obtained with smart watches; temporal patterns in online gambling records; temporal patterns in financial transactions; temporal patterns in teaching practices, studying if reading instructions are evidenced based; and temporal patterns in behavior of children identified as victims of bullying.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

T-Pattern Detection and Analysis (TPA) with THEME™: New Features

Magnus S Magnusson

University of Iceland, Reykjavik, Iceland

The T-pattern is a multi-categorical hierarchical self-similar (pseudo-fractal) recurrent pattern type occurring on a single real-scale, proposed for the description and detection of (sequential and synchronic) intra- and inter-individual patterns in behavior and other biological phenomena. It is accompanied by an evolution detection algorithm implemented as the THEME™ software dedicated to T-pattern detection and analysis, TPA. Derived and related pattern types have since been gradually defined and added to make up the *T-system* with corresponding detection and analysis algorithms also implemented in Theme (Casarrubea et al 2015; Magnusson, 2016; Magnusson et al. Eds. 2016). From around 1980 Theme has evolved from a 3000-line Fortran IV program for PDP computers to a 300.000-line Delphi Windows program, while data size and speed of processing have increased by orders of magnitude and now the first version using parallel processing will be introduced. An educational (free) and a full version can be obtained from www.patternvision.com. See also http://hbl.hi.is.

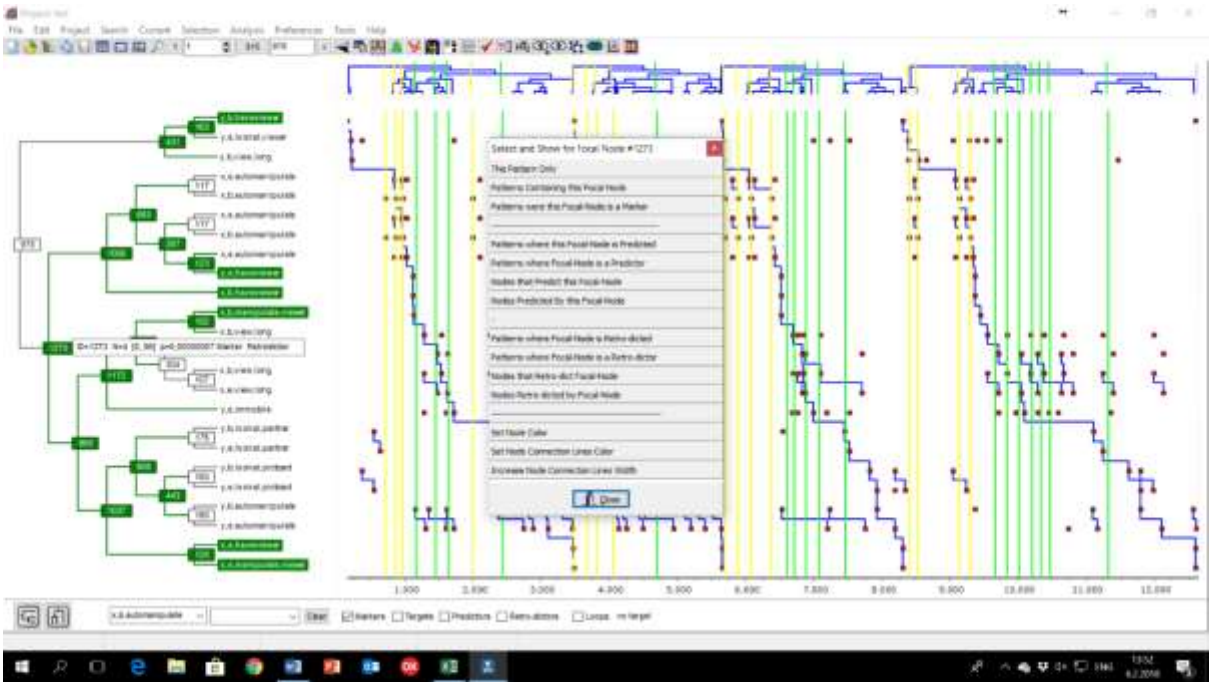


Figure 1. Capture of the screen when studying with Theme the static and diachronic functional structure of T-patterns within a T-pattern set detected within a set of Theme raw data (T-data) files.

Behavioral analysis with Theme results primarily in a set of T-patterns usually with varying frequency and duration, but most importantly, often complex behavioral content and structure. Theme therefore includes many tools for quantitative and structural selection and analysis of individual patterns and pattern subsets, the latter especially useful for the detection of efeccts of independent variables. The most recent tools, already mentioned elsewhere (Magnusson, 2017) allow new kinds of both retro- and prediction, T-prediction & T-retrodiction, based on automatic analysis of relations between pattern all nodes (branches, i.e., terminal and nonterminal nodes) in sets of T-patterns detected in the same data. Corresponding interactive analysis and graphical presentation tools essential for interpretation have also been developed and implemented in Theme and will be presented, see example figure 1.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

References

1. Casarrubea M, Jonsson GK, Faulisi F, Sorbera F, Di Giovanni G, Benigno A, Crescimanno G, Magnusson MS (2015) T-pattern analysis for the Study of Temporal Structure of Animal and Human Behavior: A Comprehensive review. J Neurosci Methods. Vol. 239, pp. 34-46.
2. Magnusson MS, Burgoon JK, Casarrubea M, Eds. (2016) Discovering Hidden Temporal Patterns in Behavior and Interaction: T-pattern detection with Theme. Neuromethods, vol. 111. Springer New York, NY. DOI: 10.1007/978-1-4939-3249-8 - ISBN: 978-1-4939-3248-1, e-ISBN: 978-1-4939-3249-8
3. Magnusson MS (2016) Time and Self-Similar Structure in Behavior and Interactions: From Sequences to Symmetry and Fractals. In Magnusson MS, Burgoon JK, Casarrubea M, Eds. (2016) Discovering Hidden Temporal Patterns in Behavior and Interaction, Neuromethods, vol. 111. Springer New York, NY.
4. Magnusson, MS (2017) Why Search for Hidden Repeated Temporal Behavior Patterns: T-Pattern Analysis with Theme. Int J Clin Pharmacol Pharmacother 2017, 2: 128. https://doi.org/10.15344/2017/2456-3501/128

Using Proximity Coefficients to Examine Behavioural Dependencies Within and Between Cases

P.J. Taylor

Department of Psychology, Lancaster University, Lancaster, UK. p.j.taylor@lancaster.ac.uk; and the University of Twente

The majority of hypotheses about the relationships among behaviours observed over time concern their interdependencies. Researchers look to either affirm predictions about the structure of the relationships—such as, does 'cooperation' tends to occur after 'rationale persuasion?—or identify moderators of this structure—such as, when does speakers' cultural background moderates the existence of the relationship (Giebels & Taylor, 2009)? In this talk I will show how established methods for examining such relationships (e.g., stochastic approaches such as lag analysis, whole-sequence methods such as phase analysis) are usefully subsumed under a general 'proximity' approach in which the standardized average distance between two behaviours provides a bounded measure of their interrelationship. Specifically, I will review a proximity coefficient, first introduced by Taylor (2006), demonstrating how other methods are specific cases of this general coefficient. The *P*-coefficient varies between 0.00 (for behaviours that occur only at the first and last index within a sequence) and 1.00 (for behaviours that immediately proceed another on every occurrence), decreasing monotonically as the average distance between the two behaviours (i.e., units) increases.

I will illustrate the value of the proximity coefficient by presenting three analyses of behaviours that occur in dyadic conversations between police negotiators and hostage takers. The data are transcribed audiotapes of crisis negotiations that have been coded at the utterance level using either Taylor's cylinder model (Taylor, 2002) or Giebels's Table of Ten (Giebels & Noelanders, 2004). These sequences are a useful illustration of what the coefficient offers because crisis negotiations are lengthy interactions that have dynamic relationships among the behaviours (thus, a matrix of proximities among codes shows significant variance) and these interrelationships are driven by various contextual and interpersonal moderators (thus, we can predict differences across explanatory independent variables and across cases). Analyses were conducted on sequences of codes for each negotiation, using the freely available ProxCalc software: https://pauljtaylor.com/data_code/proximity-coefficient-software/

Comparing within a sequence. Often researchers wish to know whether or not there is a dependency between two behaviours within a single case, or across a set of cases where there is no factor of interest discriminating the cases. In this talk I will explore the hypothesis that cooperative-instrumental behaviours by the negotiator are associated with more concessions from the perpetrator than would be expected by chance. The hypothesis is tested using a permutation test wherein the observed *P*-coefficient for the cue-response relationship is compared to a distribution of equivalent coefficients derived after the order of codes within the sequence is randomised. Specifically, by repeatedly generating a 'random' sequence from the same set of codes, it is possible to derive the distribution of possible observed *P*-coefficients given what occurred, and from that we can estimate the likelihood of the observed coefficient occurring by chance. In this case the analysis reveals a significant dependency between cooperative-instrumental behaviours and perpetrator cooperation.

Comparing across sequences. For many researchers, the value of the proximity approach is that it enables comparisons across groups of cases that parallel traditional factorial designs. Although the distribution of *P*-coefficients has been shown to be normal under many conditions, and so researchers might reasonably utilise GLM statistics, this talk will illustrate how to use randomisation tests (Edgington, 1995). Specifically, I will compare the effectiveness of rationale persuasion as a strategy for eliciting cooperation in perpetrators from two cultural groups. The randomization test takes case-based proximities from the original comparison of cultural groups and shuffles them randomly across the conditions, repeatedly, to derive a sampling distribution. This distribution then allows an assessment of the likelihood that the observed average difference in P-coefficient across groups would be observed by chance.

Comparing whole matrices. A matrix of proximity coefficients derived from a sequence serves as a unique measure of the sequence's structure. Accordingly, they provide the basis for several case-focused analyses. I will illustrate how calculating a diversity index (e.g., coefficient of variation, Shannon's H) on each matrix quantifies the degree of dependence among behaviours. Matrices with low variance have little internal structure. Matrices with high variance are quite determined and, thus, the likelihood of a response appearing after a particular cue more predictable. I will also present a comparison of cases similarity that is based on using Pearson's coefficient to derive a simple similarity matrix for each of the original proximity matrices. By submitting the resulting similarity matrix to multidimensional scaling, I derived a visual representation of the similarities and differences among case structures that was interrogated for evidence of a hypothesised typology of cases (Taylor & Donald, 2007). The results show two distinct sub-groups of cases that previous work considered homogenous.

These three example analyses serve to illustrate the kinds of hypotheses that can be explored through the proximity approach. While the focus here is communication behaviours over time, the units within the sequences may represent any form of observation (e.g., states, life events, criminal offences, cf. Taylor et al., 2008) and their recording may be nominal event sequences or timed event sequence data.

References

Giebels, E., & Noelanders, S. (2004). *Crisis negotiations: A multiparty perspective*. Veenendaal, The Netherlands: Universal Press.
Taylor, P. J. (2002). A cylindrical model of communication behavior in crisis negotiations. *Human Communication Research, 28,* 7-48.
Giebels, E., & Taylor, P. J. (2009). Interaction patterns in crisis negotiations: Persuasive arguments and cultural differences. *Journal of Applied Psychology, 94,* 5-19.
Taylor, P. J. (2006). Proximity coefficients as a measure of interrelationships in sequences of behavior. *Behavioral Research Methods, 38,* 42-50.
Taylor, P. J., & Donald, I. J. (2007). Testing the relationship between local cue-response patterns and global dimensions of communication behaviour. *British Journal of Social Psychology, 46,* 273-298.
Taylor, P. J., Jacques, K., Giebels, E., Levine, M., Best, R., Winter, J., & Rossi, G. (2008). Analysing forensic processes: Taking time into account. *Issues in Forensic Psychology, 8,* 45-57.

What do rhythms in behaviour reveal about an animal's status?

Measuring behaviour has always been central to the assessment of an animal’s status, but until recently it has been difficult to precisely measure behaviour over long time-periods. Now that technology allows behaviour to be measured continuously over days, weeks, or even months, rhythms in behaviour can be examined. In humans, rhythms in behaviour alter in poor states of welfare, such as stress, anxiety or pain. Many diseases also change patterns in behaviour. In animals, we are starting to find similar results. Automated and other high-throughput analyses of behaviour mean there are great opportunities to use behavioural rhythms in monitoring or studying disease, affective disorders and animal welfare. This symposium and workshop will explore the latest research in these areas of behavioural rhythms. It will cover:

(i) technology used to capture behaviour;

(ii) applications of studying behavioural rhythms in animals, and

(iii) the theoretical underpinnings of the associations between rhythms and animal welfare.

Christina Umstätter & Marianne Cockburn (Agroscope) and Lucy Asher (Newcastle University)

Rhythms of the night: Circadian rhythms in activity as a potential welfare indicator

JS O'Sullivan [1], C Ladha [1,2] Z Belshaw [3] and L Asher[1]

[1] Centre for Behavior and Evolution, Henry Wellcome Building, Newcastle University, Newcastle, UK [2] VetSens, Newcastle, UK; [3.] Centre for Evidence-based Veterinary Medicine, School of Veterinary Medicine and Science, University of Nottingham, Leicestershire, LE12 5RD; lucy.asher@ncl.ac.uk

Modern sensor technology allows behaviour to be measured continuously over days, weeks, or even months. This means that daily rhythms in behaviour can be examined at resolutions and over durations not previously possible. In humans, circadian rhythms in behaviour alter in poor states of welfare such as stress, anxiety or pain and in response to a wide range of diseases. Here we outline why circadian rhythms in general activity could be a useful indicator of welfare in the domestic dog. We present data from dogs with and without a chronically painful disease to illustrate this point, but also provide a theoretical overview of how circadian rhythms in activity might be associated with welfare. Activity data (measured as the vector magnitude, vm^3, from a collar-mounted tri-axial accelerometer) were collected over 7 days for 20 pet domestic dogs of various breeds with osteoarthritis and 20 healthy controls. We used a cosinor mixed model to explore the degree of fit to a 24 hour circadian cycle and examined the effect of including additional harmonic terms which accounted for shorter cycles (12, 10, 8, 6, 5, 4, 3 and 2 hours added in a stepwise fashion). Overall model fit was judged by R^2 values and individual was included in models as a random effects term. Arthritic dogs displayed heightened nocturnal activity levels and depressed daytime activity levels (Figure 1). The best fitting model for healthy and arthritic dogs included cycles of 24, 12 and 8 hours in duration. Healthy dogs' activity fitted much better to a cosinor model of circadian rhythm (R^2=0.2) than arthritic dogs (R^2=0.04). This suggests that arthritic dogs have a disruption in the normal circadian rhythm of activity. Findings fit with previous owner reports of shifting position and discomfort during sleep in dogs with arthritis. There could be a number of explanations for differences between arthritic and healthy dogs in their activity patterns. The welfare implications of these different explanations is discussed and tentative conclusions presented about the potential for circadian rhythms as a welfare indicator in dogs.

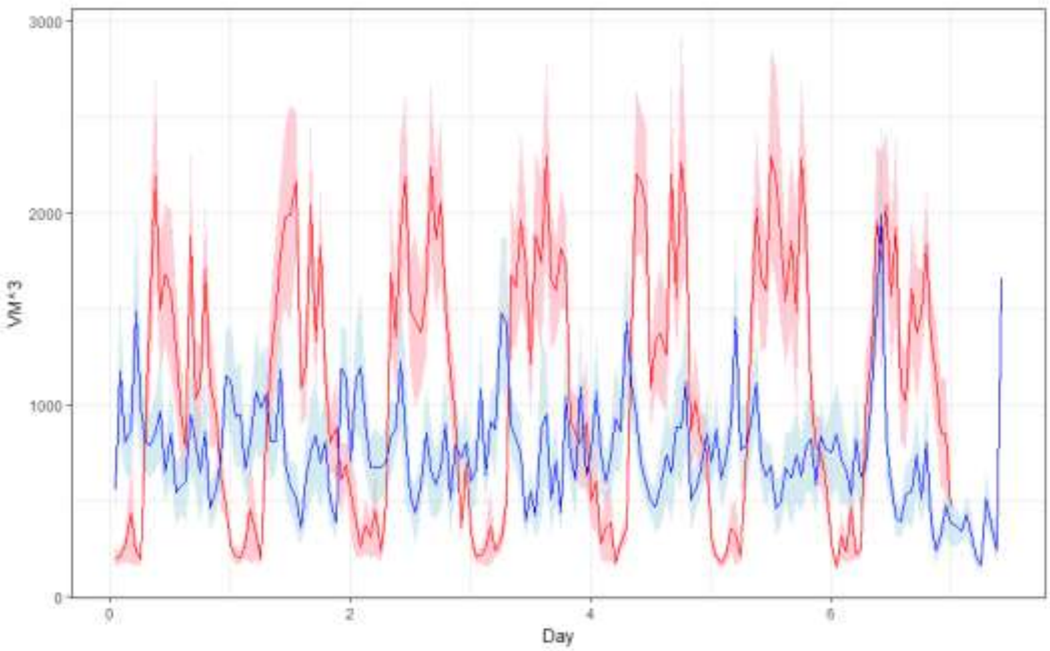


Figure 1. Activity of dogs with (blue) and without osteoarthritis (red) over 7 days. Arthritic dogs have a lack of circadian rhythm to their activity. Solid line is mean vales and shaded area is standard error.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Polysomnography as a Tool to asses equine Welfare

C. Fuchs[1], L.C. Kiefner[1], M. Kalus[1], S. Reese[2], M. Erhard[1], A.-C. Wöhr[1]

1Institute for Animal Welfare, Ethology, Animal Hygiene and Animal Husbandry, Faculty of Veterinary Medicine, Department of Veterinary Science, Ludwig-Maximilian-University of Munich, Germany;
2Institute of Veterinary Anatomy, Histology and Embryology, Faculty of Veterinary Medicine, Department of Veterinary Science, Ludwig-Maximilian-University of Munich, Germany. christine.fuchs@lmu.de

Introduction

Sleep is a naturally recurring state that involves functions of the body and the mind. It is part of every species ethogram and essential for an individual's physical and mental health. Extreme sleep deprivation ultimately leads to death [1, 2]. Nonetheless, when evaluating the well-being of an animal or the welfare aspects of a husbandry system, this functional interaction is rarely considered. Whereas the sleep behaviour of humans has been thoroughly investigated for many years [3], knowledge on animal sleep behaviour is based mostly on visual observations and on electro-encephalic examinations [4]. However, as some animals such as horses sleep while standing, such observations and examinations can be challenging.

Polysomnography (PSG) is a multi-parametric test to determine the depth and quality of sleep. It entails the continuous and synchronous recording of several body functions such as brain wave activity, muscle activity and eye movements and is routinely used in human sleep laboratories [3].

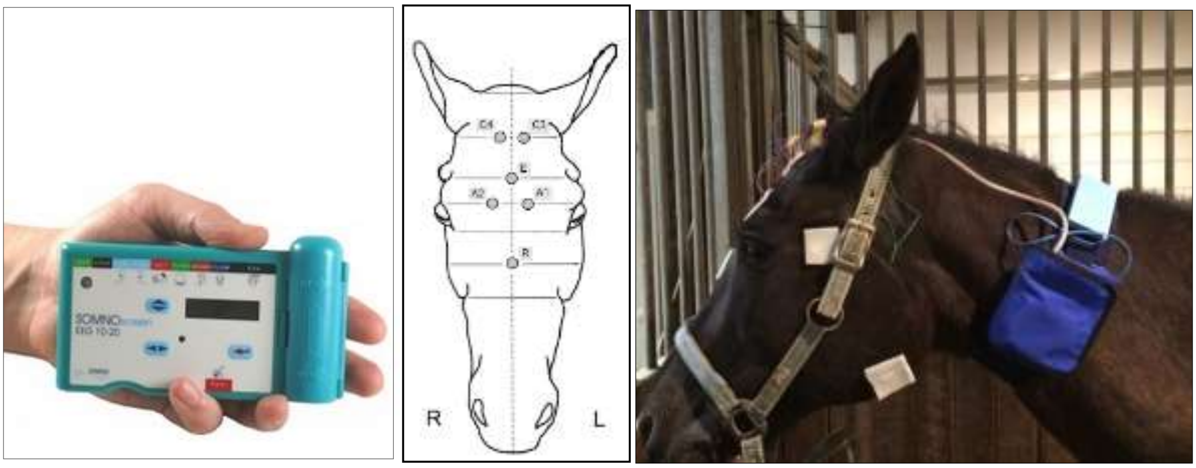
The presented study is a summary of two separate studies. Their objective was to examine the normal sleep behaviour of healthy horses via polysomnography as well as to examine pathological deviations in the equine sleep behaviour.

Materials and Methods

Polysomnographic measurements were made of 7 healthy horses housed in individual stables over four consecutive nights under standardized conditions as a control group. It was compared to the sleeping behaviour of 35 horses which showed collapses during the resting behaviour and associated scars and injuries, but were otherwise healthy.

The presented data were collected on owner's request in the context of veterinary diagnostic measurements and informed client consent was obtained for all animals in the study. The polysomnographic measurements are non-invasive and do not cause anxiety, pain, suffering or damage in the examined animals. All examinations were done in compliance with the German animal welfare law and ethical standards.

The device used was the portable polysomnograph "SOMNOscreen™ plus" (SOMNOmedics GmbH, D-97236 Randersacker) which was fixed at the anterior neck of the horses. Special electrodes with longer cables than usually used for humans were attached with electrode paste, superglue and small paddings (Snøgg Animal Polster) to the head of the horses following a previously established diagram (see Figure 1) [5]. The appropriate skin areas were shaved and degreased with alcohol before the application. For better protection the head and neck of the horses were covered with a Sleezy hood (Horse-friends Sleezy, Loesdau GmbH & Co. KG, D-72402 Bisingen). Data were transferred via a radio module attached to the polysomnography. In addition, recordings from an infrared video camera were transferred synchronously to a computer which was set up next to the stalls.



a) b) c)

d)

Figure 1: a) SOMNOscreen™ plus, **b)** Electrode localization diagram, **c)** SOMNOscreen™ plus in protective cover, connected to the electrodes and fixed to the neck of a horse, **d)** SOMNOscreen™ plus in protective cover, connected to the electrodes and fixed to the neck of a horse covered with a Sleazy hood

To evaluate and analyse the data, we used the polysomnography software DOMINO® (SOMNOmedics GmbH, D-97236 Randersacker) which allows a complete automated analysis immediately after data transfer. However, we aimed to retain measurement segments of low certainty and ensure correct assignment of sleep stages according to horse-specific criteria based on the present evaluations. Thus, we refrained from an automatic analysis. The manual analysis was based on the criteria established by Rechtschaffen and Kales[6] and the observations made on the horses. To classify the equine sleep stages, we combined and modulated our own observations. The human sleep stages N1 und N2 were summarized as light sleep, and stage N3 was assigned to slow-wave sleep (deep sleep). Therefore, the results are based on the 4 stages W = wakefulness, LS = light sleep, SWS = slow-wave sleep, and REM = REM sleep. The video recordings where used to determine the body positions.

Statistical analysis was performed using SPSS Statistics software (IBM Deutschland GmbH, D-71139 Ehningen). To indicate the standardized differences of the different sleep stages between the horses with and without collapses we used the Cohen's d with ≤ 0.2 = small, 0.5 = medium and ≥ 0.8 = large effect size. The Cohen´s d was accompanied by a t-test with $p \leq 0.05$ showing a significant correlation. To demonstrate the correlation between the amount of REM sleep and the number of collapses we used as well the Spearman's rank correlation coefficient (rho).

Results

The control group (n = 7) showed a quite stable total sleep time of 3.5 hours (211,9 ± 29,2 min) per night. The duration of the different sleep stages also appeared to be relatively constant. REM sleep was detected every night in every horse and occurred mainly after midnight. With about 30 min/night and 15 % of the total sleep time REM sleep occupied the smallest amount, slow-wave sleep took up the most time with 65 % of the total sleep time and 40 min/night. The rest of the total sleep time was spent in light sleep. Total and percentage data of sleep duration and sleep stages showed significantly lower intraindividual differences than the times between individuals.

The animals studied mainly lay down after midnight and spent each night a total of about 2.5 hours recumbent in up to 5 phases. REM sleep, which requires complete muscle relaxation, occured only when lying and mostly in sternal recumbency with the head resting on the ground. Slow-wave sleep and light sleep occurred while standing and while lying.

The horses which showed collapses during the resting behaviour (n = 35), suffered from over 60 collapses with up to 199 collapses per day. The individual number of collapses depended significantly on the lying behaviour of each horse. Horses that lay down to sleep showed notably less collapses. Only 2 of the horses that experienced collapses lay down to sleep during the night. On average these 2 horses only slept for 24.5 ± 3.5 minutes in comparison to the horses not experiencing collapses (n = 7), which slept for 134.3 ± 54.5 minutes in a recumbent position. Collapses mainly occured during nighttime hours, especially between 4:00 am and 4:30 am, which corresponds to the period in which horses without collapses spent most of their lying phases and their REM sleep.

The polysomnographic measurements demonstrated that the horses experiencing collapses showed a significantly altered sleeping behaviour compared to horses without collapses. The horses with collapses showed an increased restless sleep profile with a mean value of 0.30 ± 0.09 sleep stage changes per minute. In contrast, this value was almost twice as high as that from the horses with „normal“ sleeping behaviour. Furthermore the horses with collapses spent an increased amount of time in light sleep, less time in slow-wave sleep and significantly less time in REM sleep (see Table 1). The phases of REM sleep of the horses with collapses were significantly shorter with a mean duration of one minute, in comparison to the control horses with a mean duration of more than 4 minutes. These horses also showed more phases of REM sleep with a mean value of 0.10 ± 0.05 per minute which was almost 3 times the value of the horses without collapses (0.03 ± 0.01). The short phases of REM sleep were almost always associated with the collapses and 86,7 % of the collapses occurred in a time frame that was classified as REM sleep. Hence, a strong correlation can be demonstrated between the amount of REM sleep and the number of collapses (rho = 0.56; p = 0.001).

Table 1: Comparison of the mean sleep stage durations between the horses with (n = 35) and without (n = 7) collapses; * = significant correlations, * = large effect size

		Total sleep time (TST)	REM sleep		Light sleep		Slow-wave sleep	
		min	min	% TST	min	% TST	min	% TST
Horses with collapses (n = 35)	Mean	202.1	22.8	10.8	64.6	32.4	114.7	56.8
	Standard deviation	67.9	16.3	6.1	34.2	13.7	44.6	13.2
Horses without collapses (n = 7)	Mean	211.9	31.8	15.2	42.3	18.9	137.8	65.8
	Standard deviation	29.2	5.2	2.6	18.2	6.8	16.4	7.0
Cohen d		-0.2	**-0.8**	**-1.0**	**0.8**	**1.3**	**-0.8**	**-0.9**
Relative difference (%)		0.0	-0.3	-0.3	0.5	0.7	-0.2	-0.1
t-test p =		0.7	**0.013***	**0.005***	0.1	**0.016***	0.2	0.1

90.2 % of the horse owners of the horses with collapses reported on injuries, that were directly related to atonic collapses and 19/35 horses showed present injuries at the time of examination. The injuries ranged from abrasions at the dorsal fetlocks to scars and swollen synovial bursae at the carpal and tarsal joints up to head and tail fractures. Six of the 35 horses showed an altered behaviour since the beginning of the collapses. The examination of the husbandry of these animals revealed a undercut of the recommendations by the German Federal Ministry of Food and Agriculture [7] for the minimal dimension of the lying area in the open stable per horse and accordingly the ground area in a single box in 70.0 % (14/35) of the cases.

Discussion

The results of this work suggest, that a certain level of stability as well as inter-individual and especially intra-individual repeatability of equine sleep behavior exists under relatively standardized conditions similar to humans. Additionally, even different "sleep-types" seem to exist, which could lead to different demands on stabling and handling horses.

Furthermore, the examinations of the horses suffering from collapses, associated with sometimes severe injuries and behavioural problems, demonstrated that horses with atonic collapses suffer from a REM-sleep deficiency or rather a recumbent sleep deprivation and a massively altered sleeping behaviour. The affected horses fall into REM sleep stage while in a standing position, where the characteristically decreased muscle tone during this sleep stage causes the more or less severe collapses that range from slight swaying to sudden complete collapse. It is assumed that the observed altered rest and lying behaviour of these horses is caused by either painful disorders or insecurity and discomfort associated with environmental changes or inadequate husbandry conditions.

The serious consequences of this sleep disorder, diagnosed with polysomnographic measurements, demonstrate, that normal resting behaviour is essential for the physical and mental welfare of the horse. Especially when keeping horses in groups, favoured by an increasing number of owners, emphasis is set on the horses needs for movement, social contact, and species-appropriate feeding. However, the need for normal resting behavior including recumbent rest should not be neglected.

References:

1. Rechtschaffen, A., et al., *Physiological correlates of prolonged sleep deprivation in rats.* Science, 1983. **221**(4606): p. 182-184.
2. Bentivoglio, M. and G. Grassi-Zucconi, *The pioneering experimental studies on sleep deprivation.* Sleep, 1997. **20**(7): p. 570-576.
3. Deboer, T., *Technologies of sleep research.* Cell Mol Life Sci, 2007. **64**(10): p. 1227-1235.
4. Williams, C.D., et al., *Qualitative and quantitative characteristics of the electroencephalogram in normal horses during spontaneous drowsiness and sleep.* J Vet Intern Med, 2008. **22**(3): p. 630-638.
5. Wöhr, A. and M. Erhard, *Polysomnographische Untersuchungen zum Schlafverhalten des Pferdes.* Aktuelle Arbeiten zur artgemäßen Tierhaltung, 2006: p. 127-135.
6. Rechtschaffen, A. and A. Kales, *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. 1969, Washington D.C.: Public Health Service.
7. BMELV, *Leitlinien zur Beurteilung von Pferdehaltungen unter Tierschutzgesichtspunkten*. 2009: BMELV, Referat Tierschutz.

Sensing: Under the covers

Cassim Ladha

Centre for Behaviour and Evolution, Henry Wellcome Building, Newcastle University, Newcastle NE2 4HH, UK; cassim.ladha@ncl.ac.uk

Wearable Sensors have proven themselves useful in many of the sub-disciplines that exist within human and animal movement science. Methods developed on the output from these sensors have already made their way into improving lives and changing welfare practice. Despite their promise, some barriers to using these sorts of sensors still exist and quite often the engineering background needed understand the limitations and realise the potential of these devices is a barrier.

This talk first takes a look under the covers of some well-established consumer sensors (Whistle and Fitbark) in a "dissection style" approach. It then proceeds to evaluate the difference between research grade sensors and highlight the fundamental differences required for longitudinal measurements of raw data; critical for evaluating parameters such as circadian rhythm. This introduction into the session aims to illuminate the strengths and limitations of wearable sensor technology with an aim of aligning expectations with methods and algorithm outputs.

The main part of the talk will focus around the emerging topic of circadian rhythm (or daily activity/in-activity pattern) estimation using wearable sensors. It is understood on a physiological level that the circadian clock is linked to cellular function and regularity is a key indicator of health. It therefor follows that an easy to use measurement of circade signal could be of interest both in clinic and for evaluating new interventions.

By way of approach, the talk will follow a logical progression from the first part of session and describe the fundamentals of how a typical circade detection algorithm works. It will then move onto what parameters can be measured relating to circadian rhythm. The VetSens system will be used as an example sensing platform and by the end of the talk users should feel comfortable taking the first experimental steps collecting and examining data.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Using Skinner boxes to study circadian rhythms in honey bees

Michel B.C. Sokolowski[1] and Coralie Allain

Department of Psychology, INSERM 24 (ERI 24), Université de Picardie

Circadian rhythms have been demonstrated in a great number of animal species including insects and honey bees. To study rhythmic activity in invertebrates, most often, researchers use a locomotion test with an animal inside a small tube with one IR sensor at each end of the tube. We propose an alternative way to measure cyclic activity in bees. Recently, we developed lab Skinner boxes for bees and we suggest here to use conditioning protocols with food reinforcement. With fully automatized protocols, bees can stay in their conditioning chamber during several consecutive days without intervention. Syrup consumption can then be easily measured. In light/darkness alternation, we observe a strong cyclic feeding pattern highly correlated with light alternation. Such pattern is preserved in darkness only, but disappears in light only. We also show that some neonicotinoid pesticides selectively affect circadian rhythms at very low concentrations.

Using activity sensors to characterise behavioural rhythms

Christina Umstaetter[1], Bruna N. Marsiglio-Sarout[2,3], Carol-Anne Duthie, Marie J. Haskell, Anthony Waterhouse and Berger Anne

1. Agroscope; 2. Federal University of Rio Grande do Sul; 3. SRUC; 4. Leibniz-Institute of Zoo and Wildlife Research

With the rapid development of sensor technology it is easy and cost effective to collect monitoring data for behaviour in animals and humans. Therefore, new possibilities are appearing to develop early warning systems to detect health and welfare issues in livestock. However, despite intensive research, reliable and practical solutions are still scarce. One new opportunity is the analysis of behavioural rhythms. A way of understanding this rhythmic structure of behaviour is to study its circadian rhythm. A circadian rhythm could be any biological process that exhibits an endogenous oscillation of approximately 24 hours. Although these biological processes are endogenous, they need a so called 'zeitgeber' as a cue which could be for example light or temperature. To identify such rhythm, Scheibe *et al.* [1] developed a parameter called Degree of Functional Coupling (DFC) to characterize the synchrony between behavioural and environmental rhythms. Therefore, DFC expresses the percentage of cyclic behaviour that is harmonically synchronized with the environmental rhythms, over the 24 hour period. The DFCs can be used to identify welfare issues, with high DFCs in healthy animals and dropping DFCs due to a disturbance [2].

We used this method in sheep and housed cattle to investigate its potential usefulness for farming systems. For sheep, we further investigated how the DFCs for activity behaviour were expressed during the course of a year in extensively kept sheep in the West Highlands of Scotland. A flock of 24 ewes were used wearing three-way accelerometers (IceTag Pro, IceRobotics Ltd., Edinburgh, Scotland) integrated into a collar. DFCs were lower during the harsher winter season and individuals had a different DFC linked to animal performance and responses to weather changes. It was concluded that random regression models based on DFCs effectively identified between-individual variation. These findings create new opportunties for automated phenotyping.

Different parameters were tested for analysis on cattle, from activity to feeding behaviour. Activity was recorded with three-way accelerometer data loggers (IceTag Pro, IceRobotics Ltd., Edinburgh, Scotland) attached to the left hind leg. The feeding behaviour was collected by using an automated feeding system (RIC feed-weigh trough, Hokofarm Group, Marknesse, NL). The experimental design was 2 x 2 factorial comprising 40 crossbred Charolais and 40 purebred Luing assigned to one of two ad libitum diets, a mixed diet and a concentrate based diet. Steers fed with a concentrate based diet had lower activity DFC% and higher variation on its response than steers fed with a mixed diet. In addition, activity DFC% was even lower for Luing fed with concentrate, showing an interaction between breed and diet. In contrast, Luing fed with a mixed diet showed high consistency of activity DFC%. DFCs based on feeding can identify differnet responses compared to DFCs based on activity. Overall, we found that the method provides useful data for livestock farming.

References

1. Scheibe, K. M.; Berger, A., Langbein, J., Streich, W. J., Eichhorn, K. (1999): Comparative analysis of ultradian and circadian behavioural rhythms for diagnosis of biorhythmic state of animals. *Biological Rhythm Research* **30** (2), 216-233.

2. Berger, A., Scheibe, K. M., Michaelis, S., Streich, W. J. (2003): Evaluation of living conditions of free-ranging animals by automated chronobiological analysis of behavior. *Behavior Research Methods, Instruments, and Computers* **35** (3), 458-466.

Use of a precision-livestock-farming technology coupled with time series methods to identify abnormal circadian pattern of activity.

N. Wagner[1,2], V. Antoine[1], J. Koko[1], M.M. Richard[2], R. Lardy[2] and I. Veissier[2]

1Limos, Université Clermont Auvergne, Aubière, France. nicolas.wagner@uca.fr 2 INRA, VetAgro Sup, UMR Herbivores F-63122, Université Clermont Auvergne, Saint-Genès-Champanelle, France

The time-budget of animals and the circadian rhythm of activity are disrupted under stress, disease, or specific physiological states like estrus. Today precision livestock farming technologies (PLF) can automatically monitor several characteristics, including the activity of animals. For instance, with a Real-Time Locating Systems (RTLS) it is possible to infer the behavior of a cow from its location in a barn, and this can be done on a large scale, i.e. many animals for extended periods with a virtually continuous sampling. This opens the way to the use of behavioral indices to finely manage animals. Indeed, [1] designed a model of circadian rhythm of activity of cows using RTLS data. They show that average level of activity of a cow on a given day and its variations during that day differs with specific states (estrus, lameness, mastitis). Moreover, the circadian rhythm fluctuation in activity appear to be particularly sensitive and vary from 1 to 2 days before the disorder is observed by the farmer.

We studied positioning data coming from a UWB-RTLS system (CowView, GEA) installed at INRA experimental unit Herbipole (UE 1414, approved by the French Ministry for Agriculture). From the position and movement of the animal, its gross activity is inferred: resting if in cubicles, feeding if near the trough, ad standing or moving in alleys. The dataset corresponds to the information from cows for two months. It provides for each cow the time spent to rest, to feed, to walk and to stand up per hour. The four values are merged into an activity level by applying a weighted average calculated according to the method proposed in [1]. The weights are rest, -0.34; feed, +0.52; walk, +0.29; stand up, +0.29. We considered the circadian activity of a cow as a time series in order to identify patterns specific to diseases or estrus and to identify the onset of such states. Time series analysis includes powerful tools such as clustering and classification methods. It enables to detect anomalies as well as forecast the future [2].

We split our dataset in two parts. The first part was composed of usual activities and is called the training-set. Here the cows had never been observed by the caretakers as sick or in estrus. The training-set was used to learn the parameters of a regression model that describes the normal activity of a cow. We tested several statistical models (ARIMA) [2] and machine learning models (neural networks e.g. MultiLayer Perceptron (MLP) [3], Support Vector Regression, decision tree for regression, lazy models, etc.).

The second set is called the test-set and contains usual and unusual activities (e.g. days where cows had been noticed as sick or unusual events occurred). The learnt regression model employs the test-set to predict the next hour of a cow's activity as if it was usual. To detect anomalies, we computed the difference between the predicted values and the observed ones (prediction error). If it was significant, we consider this period as abnormal.

We compared the models according to several criteria such as the prediction error and the quality of detection for a particular disorder. We introduced other criteria that are required in practice such as the training time, the robustness of the model to noisy labels and the ability to be deployed on various farms.

Some results are illustrated on Figure 1 for one cow and 8 days. Here an anomaly was detected that coincided with a test of the fire alarm in the barn, which apparently disrupted the activity of the cow.

The first tested methods, such as the MLP, show promising results concerning detection of irregularities. We can obtain a real-time detection for each hour or for each day. At present, the method is linked to the data we collected. In the future, we will extend our work by testing larger datasets and other farms, in order to check the adaptability of the regression methods can be deployed in other settings. If this is successful, we will propose ways to automatize the detection of irregularities and possibly relate the type of irregularity to a specific disorder.

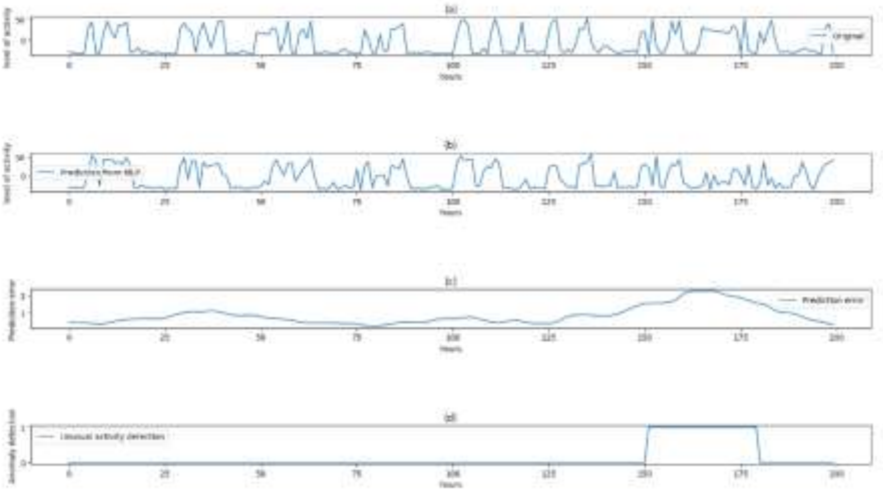


Figure 1: (a), original time series; (b), predicted time series; (c), prediction error; (d), anomaly detected (0, normal; 1, abnormal)

Ethical statement

We did not modify any practice on the farm. All data that we processed were routinely collected without any intervention on animals. There is thus no ethical issues in our study.

References

1. Veissier, I., Mialon, M.M., Sloth K.H. (2017). Early modification of the circadian organization of cow activity in relation to disease or estrus. Journal of dairy science **100**: 3969-3974.

2. Adhikari, R., Agrawal, R.K. (2013). An introductory study on time series modeling and forecasting. LAP LAMBERT Academic Publishing.

3. Gardner, M.W., Dorling, S.R. (1998). Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences. Atmospheric environment **32**: 2627-2636.

New technologies and software to advance the 3Rs and behavioural measurements in animal research

The 3Rs (Replacement, Reduction, Refinement) have long been established as an ethical framework for the use of animals in scientific research, but they are increasingly being recognised for their scientific benefits. This shift in understanding is driving methodological and technological developments that have the potential to significantly impact upon animal welfare and support better science through improved experimental design and outcomes. This symposium will bring together researchers who have developed new technologies and software that not only automate and maximise behavioural measurements across a range of species (zebrafish, mice and macaques), but also demonstrate how implementing the 3Rs can further the progress of scientific research.

Kamar Ameen-Ali, NC3R, UK

Designing robust experiments with the Experimental Design Assistant

N. Percie du Sert

NC3Rs, London, UK. nathalie.perciedusert@nc3rs.org.uk

Recent years have seen growing concern about the reproducibility and reliability of biomedical research using animals. Poor quality experimental design, analysis and reporting have been identified as contributing factors. The NC3Rs has led the development of free resources to support researchers to improve the design, analysis and reporting of animal experiments. This includes the ARRIVE guidelines [1], which provide recommendations on the reporting of the study design, experimental procedures, animal characteristics, housing and husbandry, and statistical analysis in scientific publications, and the Experimental Design Assistant (EDA) [2, 3]. The EDA (https://eda.nc3rs.org.uk) is a web application with a supporting website, which helps researchers design animal experiments, by increasing the transparency of the experimental plan, and providing feedback to improve it.

Features of the EDA include:

- Computer-aided design tool to develop a diagram representing the experimental plan.
- Critical feedback on the experimental plan – using computer-based logical reasoning.
- Statistical analysis suggestions.
- Sample size calculation.
- Randomisation sequence generation.
- Support for allocation concealment and blinding.
- Web-based resources to improve knowledge of experimental design and analysis.

This presentation will introduce the EDA and provide guidance on getting started; it will also include a live demonstration of the system. There is no requirement for any previous knowledge of the EDA.

Ultimately, the use of a tool such as the EDA will lead to carefully designed experiments that yield robust and reproducible data using the minimum number of animals consistent with scientific objectives.

References

1. Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. (2010). Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. PLoS Biol **8**, e1000412, doi:10.1371/journal.pbio.1000412.

2. Percie du Sert, N., Bamsey I., Bate S.T., Berdoy M., Clark R.A., Cuthill I.C., Fry D., Karp N.A., Macleod M., Moon L., Stanford S.C., Lings B. The Experimental Design Assistant. (2017). Nat Methods **14**, 1024-1025, doi:10.1038/nmeth.4462.

3. Percie du Sert, N., Bamsey I., Bate S.T., Berdoy M., Clark R.A., Cuthill I.C., Fry D., Karp N.A., Macleod M., Moon L., Stanford S.C., Lings B. The Experimental Design Assistant. (2017). The Experimental Design Assistant. PLoS Biol **15**, e2003779, doi:10.1371/journal.pbio.2003779 (2017).

3Rs Benefits of Monitoring Group Housed Animals in the Home Cage

Rowland Sillito[1], R Sonia Bains[2], Gareth Banks[2], Adina T Michael-Titus[3], Pat M Nolan[2], Will S Redfern[4], Jordi L Tremoleda[3], Karen Tse[4], Sara Wells[2], Ping K Yip[3] and J Douglas Armstrong[1,5]

(1: Actual Analytics Ltd, UK.; 2: Mary Lyon Centre, MRC Harwell, UK. S.Wells@har.mrc.ac.uk; 3: Blizard Institute, Queen Mary University of London, UK; 4: AstraZeneca Ltd, UK, 5: University of Edinburgh, UK. jdarmstrong@actualanaytics.com)

Automated behavioural monitoring in the home cage has obvious and established advantages in improving the consistency of data captured in behavioural experiments, and with 24/7 monitoring also generates a wealth of data that would otherwise be missed [1]. Both of these features have the potential to reduce the number of animals required to achieve statistical significance in experimental study designs and greatly increase the information gathered from animals in research studies. Here we examine an additional technological avenue that can be exploited to further reduce the burden — recording of individual animal behaviour data in group housed mixed cage experimental designs where control and experimental conditions are mixed in the same social group.

While recording animal behaviour in group housed situations already has significant welfare benefits, the ability to analyse the behaviour of identified individual animals within a group housed setting means that the number of subjects constituting each experimental condition does not need to be augmented merely to meet the criterion of having multiple animals in the cage. A range of technologies exist that permit group housed animals to be monitored using combinations of telemetry, gated experimental chambers attached to a home cage and/or video analytics [e.g. 2-4].

Here we will focus on one of these systems, the Actual Analytics Home Cage Analyser [3,4] which allows individual animal behaviour to be recognised within social groups. A range of experiments conducted using the ActualHCA system in both rats [4] and mice [3], show how it is possible to extract insightful behavioural data tagged to individuals. In terms of 3Rs benefits, the obvious advantage over single housed animals in using such a system is **Refinement** where the stress of isolating the animal from its social group during the study can be effectively removed. This is particularly important in long-term studies, for example in neurodegeneration models [5]. Data in rats also showed that body temperature was lower in isolated animals and that this is prevented by the group situation [4]. There are also additional benefits in that social interactions can be recovered from the data and the potential to observe earlier humane endpoints.

3Rs benefits can also include **Reduction** in animal numbers by multiplexing the datasets extracted from a single cohort of animals. For example, in a recent pharmacological study [6], we extracted data that partially covered the typical functional observational battery alongside detailed locomotion datasets and temperature all from the same single group of animals. In a typical safety pharmacology scenario each of these studies would normally require an additional cohort of animals. Although the study was designed to validate the system we were able to uncover previously unreported effects of well-known pharmacological standards – largely through the ability to capture detailed data through the night phase.

However the mixed cage approach does raise some new challenges. We will present examples from phenotyping studies (mouse) using single and mixed genotype cages as well as data from surgical and pharmacological studies (both in rat). The data on single animals are already large and extremely complex. To extend into mixed cage analysis the accuracy of identity tracking becomes paramount, despite the limited visual tagging opportunities afforded by freely moving animals in a single home cage. The RFID-driven nature of the ActualHCA system implicitly addresses this problem for locomotor (and, for rats, temperature) data, and provides a vital cue for assigning visually detected behaviours to individuals. A more fundamental challenge is posed by the increasing evidence that mixed conditions can have measurable effects on the control animals in the social group. We will present recent evidence (see Figure 1 for a preliminary example) which demonstrates that mixing various behavioural phenotypes in the cage has measurable influences on the control animals within the social group where the controls acquire aspects of an extrinsic behavioural phenotype.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

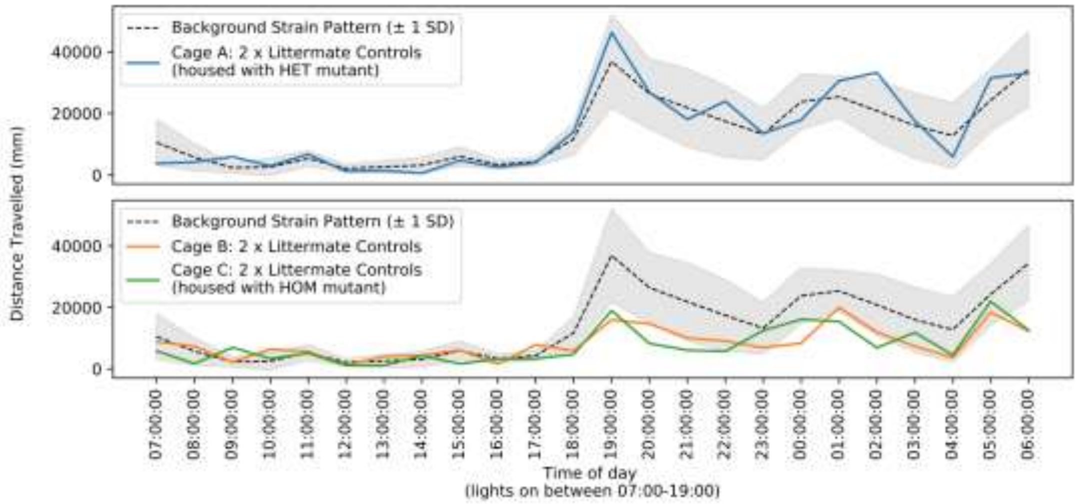


Figure 1: Preliminary data showing the 24 hour activity patterns of littermate controls housed with two variants of an arrhythmic mutant strain. Pairs of control animals can be seen to deviate from the typical activity pattern of the background strain when group housed with a homozygous mutant.

In summary, in-cage monitoring of group housed animals does offer very clear 3Rs benefits but, in doing so, adds a new layer of complexity to the underlying experimental design and data analysis.

Animal Husbandry and Ethical Approval

All studies described here were subject to institutional Animal Welfare and Ethical Review Bodies and all procedures were carried out under UK Home Office Project Licenses.

Availability

Data used in this study including video and annotations will be made publicly available under an open license at the time of full publication. Early access may be requested by contacting the authors. The ActualHCA system (hardware variants for both mice and rats, together with corresponding software – currently version 2.4) is marketed by Actual Analytics Ltd (Edinburgh, UK) http://www.actualanalytics.com

Acknowledgements

This work was supported by the National Council for the 3Rs (UK) under the Crack-It Scheme.

References:

1. Richardson C (2015). The power of automated behavioural homecage technologies in characterizing disease progression in laboratory mice: A review. *Applied Animal Behaviour Science* **163**, 19-27. https://doi.org/10.1016/j.applanim.2014.11.018
2. Vannoni E., Voikar V., Colacicco G., Sanchez M.A., Lipp H-P., Wolfer D.P. (2014). Spontaneous behavior in the social homecage discriminates strains, lesions and mutations in mice. *Journal of Neuroscience Methods.* **234**. 26-37. https://doi.org/10.1016/j.jneumeth.2014.04.026
3. Bains RS., Cater HL., Sillito RR., Chartsias A., Sneddon D., Concas D., Keskivali-Bond P., Lukins TC., Wells S., Acevedo-Arozena A., Nolan PM. And Armstrong JD (2016) Analysis of Individual Mouse Activity in Group Housed Animals of Different Inbred Strains Using a Novel Automated Home Cage Analysis System. *Frontiers in Behavioral Neuroscience* **10(106).** https://doi.org/10.3389/fnbeh.2016.00106
4. Redfern WS., Tse K., Grant C., Keerie A., Simpson DJ., Pedersen JC., Rimmer V., Leslie L., Klein SK., Karp NA., Sillito RR., Chartsias A., Lukins TC., Heward J., Vickers C., Chapman K. and Armstrong JD (2017). Automated recording of home cage activity and temperature of individual rats housed in social groups: The Rodent Big Brother project. *PlosONE* **12(9):** e0181068**.** https://doi.org/10.1371/journal.pone.0181068

5. Bains RS., Wells S., Sillito RR., Armstrong JD., Cater HL., Banks G., Nolan PM (2017). Assessing mouse behaviour throughout the light/dark cycle using automated in-cage analysis tools. *J Neuroscience Methods* **300**, 37-47. https://doi.org/10.1016/j.jneumeth.2017.04.014
6. Tse K., Sillito RR., Keerie A., Collier R., Grant C, Karpe NA., Vickers C., Chapman K., Armstrong JD., Redfern WS. (2018). Pharmacological validation of individual animal locomotion, temperature and behavioural analysis in group-housed rats using a novel automated home cage analysis system: A comparison with the modified Irwin test *Journal of Pharmacological and Toxicological Methods* **94(1),** 1-13. https://doi.org/10.1016/j.vascn.2018.03.008

An intelligent monitoring software tool to measure zebrafish health

Lynne U. Sneddon

University of Liverpool, Institute of Integrative Biology. The BioScience Building, Liverpool, L69 7ZB, UK. lsneddon@liverpool.ac.uk

Improving laboratory animal welfare is an important refinement and a legal requirement in many countries. Fish are used in a variety of experimental contexts including biomedical, pharmaceutical and toxicology research in high numbers. To protect their welfare before, during and after experiments to ensure the validity of results it is vital that researchers can assess alterations in health that may compromise welfare. However, we currently lack the means to do so in a simple cost-effective manner within laboratory fish facilities. Therefore, an automated method, the Fish Welfare Index (FWI), was developed and tested for evaluating welfare by visually tracking and quantifying zebrafish behaviour.

The FWI generates a graduated scale of wellbeing, derived from behavioural measures including activity and distance swum. Experiments involving routine laboratory procedures involved FWI assessment before and after laboratory procedures. Zebrafish movement was individually tracked and videoed in 3D, using two cameras, one frontal and one overhead, before and after relevant treatments. Movement parameters, such as time spent active, distance travelled, space use, were derived from the location coordinates of each tracked subject. The present work consisted of two stages. Firstly, a principal components analysis was performed in order to identify which aspects of behaviour were most involved in the changes in behaviour following the treatments, as shown to be statistically significant. Secondly, based on these main aspects, two derived parameters were combined to construct an automated individual Fish Wellbeing Index (FWI) as a qualitative measure to indicate the probable state of welfare of the subject from its movement to detect whether the subject was experiencing stress or pain.

The FWI was designed to generate a wellbeing indicator in four broad categories, starting at normal and decreasing to an abnormally adverse level of wellbeing using generic labels for ease of reference: 'Healthy', 'Ok', 'Unhealthy' and 'Abnormal'. Results demonstrate the methodology is accurate and reliable since it was based on normal behaviour of zebrafish and no false positive or negatives of blind video analysis were found. The FWI provides a means of gauging welfare in zebrafish and the alteration in normal activity, swimming and space use reflect the impact or severity of the procedures. The FWI results are comparable with statistical analyses of behavioural parameters after treatments, which show that the treatments have significantly affected behaviour with $p \leq 0.05$.

From an ethical perspective based upon the principles of the 3Rs, the FWI represents a significant refinement in the use of fish models in experimentation and could be adopted across a wide range of biological disciplines such as physiology, neurobiology, pharmacology and toxicology. This tool could in principle be extended to other species, for instance aquatic or flying such as birds or drosophila, utilising a 3D (or 2D) environment. This research was conducted after ethical approval and under Home Office Licensing (UK)

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Monitoring Social Relationships in Group-housed Monkeys using Automated Face Recognition and Tracking

C. L Witham[1,2]

[1] Institute of Neuroscience, Newcastle University, Newcastle-upon-Tyne, UK. c.l.witham@ncl.ac.uk, [2] Centre for Macaques, MRC Harwell, Salisbury, UK.

Background

Many laboratory species including mice, rats and monkeys are highly social animals. As social housing becomes more common there is an increasing need to monitor the social relationships between the animals in each group. In species such as Rhesus macaques fight injuries are a major concern; in breeding colonies they are the major cause of veterinary intervention [1]. Understanding the social relationships within a group can help with colony management, may help identify ill or injured animals (especially if they become socially isolated) and could help identify the causes of outbreaks of aggression. However to do these observations in person is time consuming and requires significant training of care-staff. Recent advances have shown it is possible to use automated face recognition to identify individual macaques in real time with high accuracy [2]. We extend these methods to include tracking, which allows more continuous monitoring of groups. This is then used to look at the social relationships within a group (based on proximity between individuals).

Face Recognition and Tracking Framework

Face recognition only works when individuals are facing towards the camera and in combination with face detection is relatively slow. Object tracking methods such as the Kanade-Lucas-Tomasi point tracker can be much quicker and can be used to locate monkeys when they are no longer facing the camera but they cannot be used to identify individuals. Therefore, for the best performance the two methods need to be integrated. We present a framework (in Python) for integrating face detection and recognition with tracking. Face detection and recognition are run every 10 frames to detect and identify individuals. These are then used to start or update a tracker model (kernalized correlation filter; KCF [3]), which is run on every frame. KCF trackers provide good accuracy and speed and are also good at reporting tracking failure. Where face recognition has been used multiple times on a single track the name that appears most frequently is used as the identity for that track. We tested the framework on 16 five-minute video clips filmed from six different groups (each clip contained between 2 and 10 different monkeys). These clips were manually labelled frame by frame for the location of each monkey and the manually scored locations compared to the tracking results.

Social Analysis

To assess the social relationships within the group we use a proximity based association measure based on the number of times two monkeys appear together in the video. These measures are used to form an association matrix for every pair of monkeys within the group. Monte Carlo shuffling is used to find which pairs associate with each other more frequently than by chance alone (preferred associations).

To test the application of these methods we applied the framework to look at the social relationships in one group of 26 monkeys (1 adult male, 10 adult females, 6 juveniles and 9 infants) taken before, during and after a potentially stressful event (annual health screen of group). The results showed an increase in the number of significant preferred associations during the health screen with a specific increase in the number of preferred associations with high ranked animals. This decreased again after the health screen.

We will finish by discussing ongoing work to develop methods for detecting grooming, dominance and aggresson related behaviours.

Ethical statement

The study was conducted at a Home Office licensed UK Rhesus macaque breeding centre and involved non-invasive techniques only. The project proposal was approved by the local Animal Welfare and Ethical Review Board. The housing and husbandry of the animals in the colony complies with the Home Office Code of Practice.

References:

1. McCowan, B., Beisner, B., Hannibal, D. (2017). Social management of laboratory rhesus macaques housed in large groups using a network approach: A review. *Behav Processes*. DOI: 10.1016/j.beproc.2017.11.014.

2. Witham, C.L. (2017). Automated face recognition of rhesus macaques. *J Neurosci Methods.* DOI: 10.1016/j.jneumeth.2017.07.020.

3. Henriques, J.F., Caseiro, R., Martins, P., Batista, J. (2015). High-Speed Tracking with Kernelized Correlation Filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37: 583-596.

Rodent and human studies of social behaviour and its dysfunctions

For the majority of mammalian species, the social environment includes some of the most salient rewarding and aversive stimuli. This is certainly the case in humans, and therefore the psychopathologies of social behaviour that are common in a number of psychiatric disorders are particularly detrimental to daily functioning and life quality. These include excessive or reduced social motivation, reduced reward valuation of social stimuli, heightened motivation to avoid and withdraw from social stimuli, and fear of social stimuli. These can occur in various psychiatric disorders, including autism, social anxiety disorder, depression, schizophrenia, impulse-control disorder and dementia. In all species, the possibility to measure these translational dimensions of social behaviour accurately and unambiguously is dependent on sophisticated automated hardware and software. Examples of the challenges include continuous monitoring of individuals living in social groups, sensitive operant measurement of motivation for social reward in absolute terms and relative to other potential reward stimuli, and responses to social exclusion. The development and availability of automated systems for the measurement of these social dimensions, as presented in this symposium, is allowing for the study and increased understanding of their underlying neurobiology, as well as of the genetic and environmental factors that lead to pathophysiological changes in this neurobiology.

Christopher Pryce, University of Zurich

Social Reward Responsivity in Humans: Insights for and from Autism Spectrum Disorder

I.Dubey & B. Chakrabarti

Centre for Autism, School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK

Preference for social stimuli is evident at a very early age [1] suggesting that social stimuli and interactions are inherently rewarding for humans. Like other rewards, processing of social rewards can be parsed in separable components of liking and wanting. Liking corresponds to the consummatory aspect of reward processing, and refers to the hedonic pleasure derived from a source. Wanting corresponds to the anticipatory aspect of reward processing, and refers to the incentive salience or the reward value that influences an organism to establish contact with the source of that pleasure [2]. In most cases, wanting and liking are temporally associated as one follows the other. In this talk, I will discuss insights from multiple methods used in our and other labs to study these two aspects of social rewards.

Liking has been measured using self-report and observational measures. Chevallier et al [3] used a self-report questionnaire to measure social, physical and other sources of pleasures responses in people with autism spectrum disorder (henceforth autism) and found that they report less pleasure for social situations but not for non-social situations. Observational studies have used eye-tracking tasks to measure preference for social stimuli [4]. It is found that when presented with social and non-social images with matched valence rating, typical adults fixate first at the social images and spend longer duration looking at them than the comparable non-social images. This bias for social rewards is positively associated with trait empathy. These eye-tracking measures have also been used with people with autism and results have consistently shown that people with autism spend longer time looking at non-social stimuli than social [5,6].

Wanting for a stimulus can be quantified by measuring the preference for the cues that have previously been linked to the rewarding stimuli. We developed and tested a simple behavioural paradigm "Choose-A-Movie" (CAM) that evaluates the effort participants make to view social vs non-social stimuli hence estimating the reward value of the stimuli [7]. In this task participants are presented with the cues (coloured boxes) with varies numbers of locks on them (each lock required additional key press to open). These boxes were linked with social and non-social movies. Participants are asked to choose any one box to open on each trial. It was found that typical adults prefer to watch social stimuli more but they trade-off their stimuli preference for effort i.e. they open the box for social stimuli more when it is presented with fewer lock but less when it is presented with more locks. At the same time adults with autism make a similar trade-off but for the non-social stimuli (figure 2).

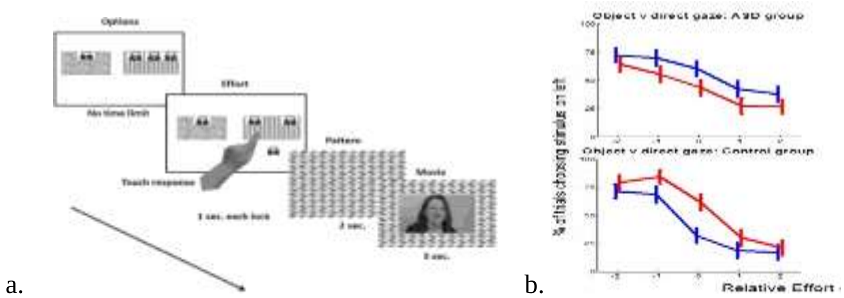


Figure 2: a) Trial structure of the Choose A Movie (CAM) task. b) Results comparing adults with and without autism on the CAM task (figures adapted from [7,8]). The results show logistic regression analysis with relative cost (as addition key presses) on the x axis and percentage of preference for stimuli on left on Y axis. The red line show social stimuli and blue line shows the object stimuli.

Another similar task "button task" targeting a younger group of participants was also developed in our lab to quantify social wanting [9]. In this task two buttons were associated with the images of social or non-social stimuli and children (3-4 years old) made the choices for cued buttons to look at the preferred (social or non-social) images. Children with autism were found to show a reduced relative preference for social rewards, indexed by a lower proportion of touches for the button associated with the social reward image (figure 3).

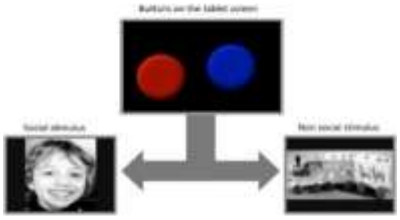
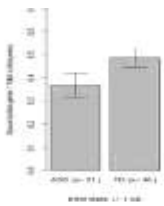
a.  b. 

Figure 3: a) Task structure for the "button task". b) Results on the button task comparing preference for social stimuli vs non-social stimuli between the groups of children with and without autism. Figures are taken from [9]

Comparing social liking and wanting: We are now taking these measures further and have developed tablet version of the "preferential looking" and the "button task" for evaluating both liking and wanting element of social rewards in same set of participants with and without autism. These tasks are being developed as part of a tablet based screening tool of autism risk in low resource settings. In this battery we use the front-camera of regular tablets to track the eye-movements of the participants to quantify the visual preference for social vs non-social stimuli. We also use the button task to measure the social wanting behaviour of the participants. We believe the tablet version of these task will make them accessible to a larger population and evaluation in the natural setup would elicit real life behaviour of the participants.

Ethical statement: Both the reported studies using Choose a movie task [7] and Button Task [9] had received an ethical approval from University of Nottingham and University of Reading respectively.

References

[1] Frank MC, Amso D, Johnson SP. Visual search and attention to faces during early infancy. J Exp Child Psychol 2014;118:13–26. doi:10.1016/j.jecp.2013.08.012.

[2] Berridge KC, Robinson TE, Aldridge JW. Dissecting components of reward: "liking", "wanting", and learning. Curr Opin Pharmacol 2009;9:65–73. doi:10.1016/j.coph.2008.12.014.

[3] Chevallier C, Grèzes J, Molesworth C, Berthoz S, Happé F. Brief report: Selective social anhedonia in high functioning autism. J Autism Dev Disord 2012;42:1504–9. doi:10.1007/s10803-011-1364-0.

[4] Chakrabarti B, Haffey A, Canzano L, Taylor CP, McSorley E. Individual differences in responsivity to social rewards: Insights from two eye-tracking tasks. PLoS One 2017;12:e0185146. doi:10.1371/journal.pone.0185146.

[5] Klin A, Jones W, Schultz R, Volkmar F, Cohen D. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. Arch Gen Psychiatry 2002;59:809–16. doi:10.1001/archpsyc.59.9.809.

[6] Pierce K, Conant D, Hazin R, Stoner R, Desmond J. Preference for geometric patterns early in life as a risk factor for autism. Arch Gen Psychiatry 2011;68:101–9. doi:10.1001/archgenpsychiatry.2010.113.

[7] Dubey I, Ropar D, Hamilton A. Measuring the value of social engagement in adults with and without autism. Mol Autism 2015;6:35. doi:10.1186/s13229-015-0031-2.

[8] Dubey I, Ropar D, Hamilton A. Brief Report: A Comparison of the Preference for Viewing Social and Non-social Movies in Typical and Autistic Adolescents. J Autism Dev Disord 2017;47:514–9. doi:10.1007/s10803-016-2974-3.

[9] Ruta L, Famà FI, Bernava GM, Leonardi E, Tartarisco G, Falzone A, et al. Reduced preference for social rewards in a novel tablet based task in young children with Autism Spectrum Disorders. Sci Rep 2017;7:3329. doi:10.1038/s41598-017-03615-x.

Tracking social hierarchy and interactions of mice in Eco-HAB - a fully automated and ecologically relevant environment for assessment of social behavior

Ewelina Knapskas

Laboratory of Neurobiology of Emotions, The Nencki Institute at Warsaw

Although social interactions between individuals are central to most animals, including humans, we are only beginning to understand the underlying neural processes. The search for the ultimate neural correlates of social behaviors is important not only because it advances our fundamental knowledge but also because it helps to understand the nature of deficits in social interactions, such as impaired social interest or shared affect, observed in autism spectrum disorder. Mouse genetic and idiopathic models of autism spectrum disorder offer the possibility of testing both neural mechanisms underlying specific impairments and therapeutic intervention. However, conventional tests of social phenotyping have repeatedly proven inefficient in differentiating certain genotypes and replicating these differences across laboratories and protocol conditions. In such tests behavior is often studied in pairs under artificial conditions, which ignores the natural tendency of rodents to live in larger groups. The size of the experimental group most likely affects the behavioral outcome of the test. For instance, both affect contagion and social interest can be modulated by social hierarchy. Further, in most studies of sociability in mice, behavioral effects related to anxiety and susceptibility to stress have been overlooked. To overcome irreproducibility and high manpower costs of manual testing as well as confounding effects of anxiety we have developed automated, ethologically relevant behavioral test, which measures spontaneous sociability in group-housed, familiar mice without the presence of a human experimenter.

Eco-HAB is a fully automated system based on RFID technology and inspired by the results of ethological field studies in mice. Group-housed animals equipped with RFID tags live in a spacious, four-compartment apparatus with shadowed areas and narrow tunnels resembling natural burrows. Tube-shaped inter-territorial passages are equipped in radio-frequency identification antennas and two of four housing compartments have impassable, perforated partitions behind which different olfactory stimuli may be presented. Food and water is available in other two housing compartments. Eco-HAB reduces stress by tracking the tendency of animals to voluntarily spend time together in an environment to which they have already been accustomed and utilizes novel sociability measures for group-housed mice. The system is equipped with software for automated data extraction and analysis, enabling quick evaluation of social activity [Fig. 1A, 1].

Social interactions, both in humans and rodents, are modulated by many factors, including the social status of interacting individuals. To study the relationship between social hierarchy and social interactions in the Eco-HAB we have recently developed an add-on to the main system that allows for assessment of social status. The measurement is based on limited access to sweetened water (Fig. 1B). The results of the test correlate well with the commonly used U-tube test of social dominance. The position in social hierarchy is then correlated with in-cohort sociability and following patterns. In-cohort sociability is defined as time spent by each mouse with every other animal within a tested cohort (the time expected to spend together by pairs of mice (based on the random dispersal throughout the system) is subtracted from the actual time that each pair of mice spent together to obtain amount of time related to non-accidental social interactions, Fig. 1C). For analysis of the following patterns we use the data analysis algorhitms to describe the relative positions of mice taking part in the experiment. For each pair of mice, in each experiment phase we can detect following episodes and evasion episodes (Fig. 1D). The data are collected for a mouse model of autism spectrum disorder with impaired social interactions (Fmr1 KO and WT) and C57BL/6J mice.

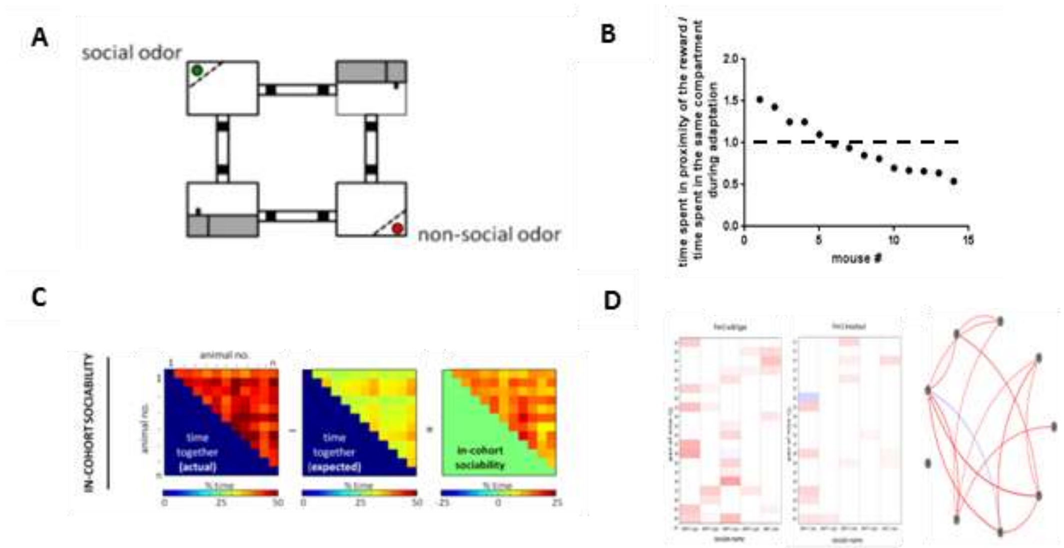


Figure 1. (A) Eco-HAB system. (B) Social hierarchy measured in a cohort of C67bL6 mice. (C) In-cohort sociability. (D) Following patterns, the following episode (red), the evasion episode (blue). The data are shown for a mouse model of autism spectrum disorder with impaired social interactions (fmr1 ko) and control (wt) mice. We can also track relations between subjects – dots with numbers symbolize mice and arrows the directed relations between them, red – following, blue – evasion, wt mice.

Mapping the relationship between social hierarchy and following patterns will allow to study specific deficits in mouse models of autism and test novel therapeutic strategies, as well as study the mechanisms underlying affect contagion and social interest.

1. Puścian, A, Łęski, S, Kasprowicz, G, Winiarski, M, Borowska, J, Nikolaev, T, Boguszewski, PM, Lipp, H-P, Knapska, E (2016). Eco-HAB as a fully automated and ecologically relevant assessment of social impairments in mouse models of autism. *eLife* **5**

Automated analysis of social behavior in groups of mice

T. Peleh[1], L. Canellas-Dols[2], F. Li[3], B. Hengerer[1]

1 Boehringer Ingelheim Pharma GmbH & Co. KG, 88397 Biberach an der Riss, Germany; 2 The University of Manchester, UK; 3 CleverSys, Inc., Reston, VA, USA

Social withdrawal is a key symptom of symptom domain “negative symptoms” of the Research Domain Criteria (RDoC) concept, and is shared among the heterogeneous spectrum of neuropsychiatric disorders (Winograd-Gurvich et al., 2006). Social withdrawal is characterized by the withdrawal of the individual from their social environment and it is one of the first indicators of emerging psychiatric disorders such as schizophrenia and major depression and neurological disorders such as Alzheimer’s disease. There is substantial evidence that social withdrawal is an independent behavioral trait across neuropsychiatric diagnoses with a specific biological basis (Dickerson, 2015). In order to elucidate the biological basis of social withdrawal and develop new targets for treatment, higher-order translational tests in groups of mice need to be developed. Quantitative technologies for assessing behavior in these tests are needed to identify and validate clinically relevant biological substrates and to develop new hypotheses for therapeutic intervention for social withdrawal.

Here, we describe a social arena, which has been developed in the context of IMI PRISM (Psychiatric Ratings using Intermediate Stratified Markers), a consortium to develop quantitative biological measures to facilitate the discovery and development of new treatments for social and cognitive deficits in Alzheimer’s disease, schizophrenia and depression. The higher order mouse social interaction paradigm needed to fulfill the following core requirements:

- Continuous tracing of four individual mice over days
- No visible labeling of the animals
- Large open arena with flexible components for social interaction such as nests or feeding areas
- Automated analysis of the video material for social behaviors such as following another animal or leaving the group
- Flexible software solution to add further analyses when needed

For reliable identification of individual animals, a RFID-based solution has been developed. Each mouse is injected as its back near the tail with a miniature RFID tag (7x1mm), which is detected by an array of RFID sensors evenly distributed over the whole area of the arena (60x80cm). The animals are video-monitored with an infrared camera above the arena, which is evenly illuminated by infrared LED arrays. Animals are detected by contrast between bedding material and fur color and individually identified by synchronizing the video position with RFID identification. Both animals with white and black coat color can reliably be identified., even when sitting nearby in a group. Occasionally misaligned animals are retrospectively re-aligned after unambiguous identification by the RFID signal.

Nests of different sizes, food hoppers, water bottles or other devices such as operant tasks can flexibly be positioned at three sides outside of the arena. RFID sensors monitor the entering of nests by individual mice. To accommodate additional objects such as ramps or walls, individual RFID sensor field can be disabled. The setup allows for maximal flexibility to accommodate future tasks.

For analysis of the various social behaviors which can be observed in groups of mice, a software package has been developed, which continuously monitors the correct identification of the animals by synchronizing the video position with the RFID position. Furthermore, the position, movement speed and direction as well as body shape of all mice visible in the area is assessed. By defining relative position and movements of individual animals, social interactions such as approaching, contact, following and withdrawing of individual mice with other individuals and the whole group can be analyzed in an automatic manner. Furthermore, non-social (rearing, grooming) and locomotor (running, walking, immobility) activities are quantified automatically.

The newly developed social arena is a promising tool for translational studies of social behavior in groups to identify and validate clinically relevant biological substrates of social withdrawal for the development of new therapeutic concepts.

The experiments reported on were conducted under a permit for animal experimentation issued by the Regierungspräsidium Tübingen, Germany.

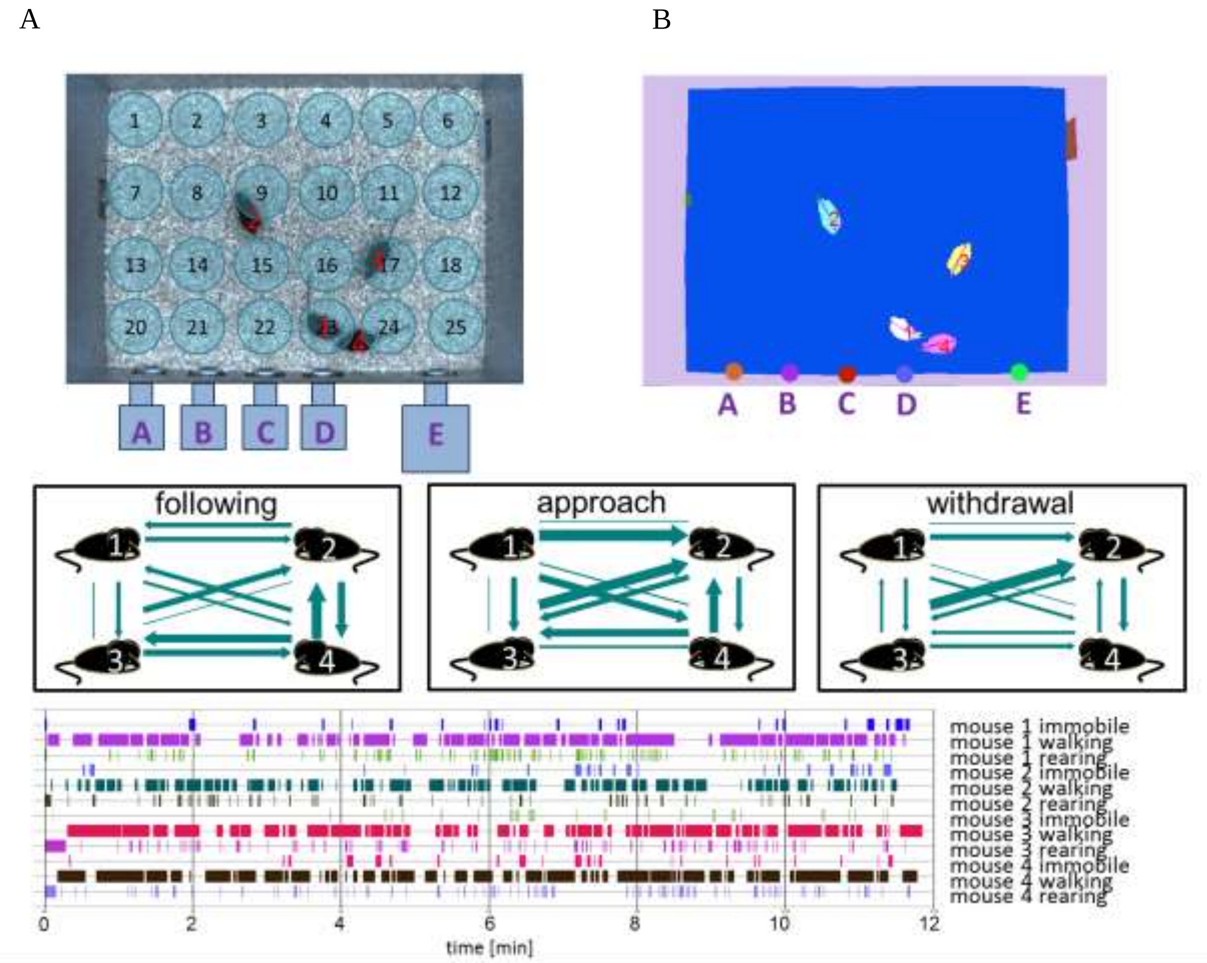


Figure1 A) Social arena with localization of the RFID sensors indicated by numbers 1-25; A-D depict position of four small nests and E of the large nest. B) position and body shape vector of the four mice in the arena C) example of automated behavior analysis

References:

1. Dickerson BC (2015) Dysfunction of social cognition and behavior. *Continuum (Minneap Minn)* **21**:660-677.

2. Winograd-Gurvich C, Fitzgerald PB, Georgiou-Karistianis N, Bradshaw JL, White OB (2006) Negative symptoms: A review of schizophrenia, melancholic depression and Parkinson's disease. *Brain Res Bull* **70**:312-321.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Operant tests for the study of stress effects on gustatory and social reward processing in mice

C.R. Pryce[1], G. Bergamini[1], H. Sigrist[1], F. Odermatt[1] and B. Hengerer[2]

1 Department of Psychiatry, Psychotherapy and Psychosomatics, University of Zurich, Zurich, Switzerland. christopher.pryce@bli.uzh.ch 2 CNS Diseases Research Germany, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany

The Research Domain Criteria (RDoC) project (NIMH, www.nimh.nih.gov/research-priorities/rdoc/index.shtml) provides a framework for translational research into psychological processes that are dysfunctional in mental disorders. Two of the domains are Positive valence systems and Systems for social processes: the former concerns processes that underlie reward-directed behaviour, and the latter processes that underlie social interactions. Of course, social stimuli are also rewarding in many contexts, and impaired social processing is a major pathology in various neuropsychiatric disorders. One major advantage of RDoC is that it promotes animal modelling of specific human psychopathologies, and therefore the detailed, causal study of their underlying neurobiology and pathophysiology.

In adult male C57BL/6 mice we are studying the inhibitory effects on subsequent reward processing of chronic social stress (CSS). CSS comprises 15 days of exposure to aggressive, dominant CD-1 mice, including proximate interaction (somatosensory, attack) for 10 min/day and otherwise separation by a divider through which there is distal interaction (olfactory, visual, auditory), and with a different CD-1 mouse each day [1]. Control mice are maintained in littermate pairs and separated during the light period by a cage divider identical to that used to separate CSS and CD-1 mice. CSS mice exhibit submissive behaviour but are nonetheless attacked and thereby experience a lack of social control; this has analogy with the loss of control, or helplessness, often reported in depression [2]. The overall aim is to increase understanding of CSS-induced cellular and molecular changes in the neurocircuitry that underlies reward processing. Within the RDoC positive valence systems, we focus on reward valuation, i.e. how salient a prospective reward is, and effort valuation, i.e. how much effort will be exerted to obtain a reward. Deficits in either or both of these dimensions could contribute to the core depression symptom of loss of interest as well as to negative symptoms in schizophrenia.

As with the majority of tests of reward processing in rodents, we use sweet-tasting sucrose pellets as a reinforcer i.e. gustatory reward [3, 4]. In a test of gustatory reward valuation using modified operant chambers, a 30-s tone constitutes the conditioned stimulus (CS) during which approaching and responding at a feeder aperture results in delivery of a sucrose pellet and termination of the CS. Successive CSs are separated by inter-trial intervals of 50±30 s. Relative to control mice, CSS mice make fewer aperture responses and acquire the CS-reward association to a lesser extent. In a test of gustatory effort valuation in operant chambers, mice are required to make nosepoke responses to activate delivery of a sucrose pellet, with the number of responses required for each reinforcer increasing on a progressive ratio schedule. Relative to control mice, CSS mice make fewer responses and attain a lower final ratio [3] (Fig. 1A). Interestingly, a similar deficit is observed following pharmacological depletion of dopamine in the nucleus accumbens [5].Therefore, stress results in gustatory reward being less salient under low effort conditions, and such reward becomes even less of an incentive under high effort conditions.

Recently, we have started to investigate CSS effects on interest in mice of the same, C57BL/6 strain. Our CSS experiments are conducted with young-adult males, and females at the oestrous stage of the ovarian cycle are used as socio-sexual stimuli. In a 3-chamber operant apparatus (Fig. 1B), a central chamber is connected by transfer tubes to two side cages. Each side cage contains a divider that is transparent and perforated. A female is placed behind this divider in one side cage whilst the other one remains empty. The central chamber is equipped with 2 operant levers, one on each side wall, pressing of which opens the adjacent vertical door to the transfer tube. To measure social reward valuation, the doors remain open and the proportion of time spent in the female cage is scored: the proportion of total time spent in the two side cages that is spent in the female chamber provides a measure of the salience of the female stimulus, similar to the three-chamber social approach test [6]. To measure

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

social effort valuation, the doors are closed and the amount of time required to reach a fixed-ratio criterion on the lever controlling access to the female cage is scored; the time required to attain criterion is inversely proportional to the effort valuation of the female. It is hypothesized that CSS will lead to a reduced preference to spend time in the female cage and to increased time to attain the operant criterion. The current test apparatus does not allow for detailed analysis of the activity of the mice in the apparatus, and video tracking would provide a very informative addition.

The demonstration that chronic social stress in mice results in motivational deficits with respect to both gustatory and social stimulus modalities would provide a robust and novel model for the trans-diagnostic psychopathologies of low reward interest and apathy. This model could then be applied to increase understanding of the stress-induced pathophysiology in the underlying neurocircuitry and thereby identify novel therapeutic targets.

The experiments reported on were conducted under a permit for animal experimentation issued by the Veterinary Office, Zurich, Switzerland.

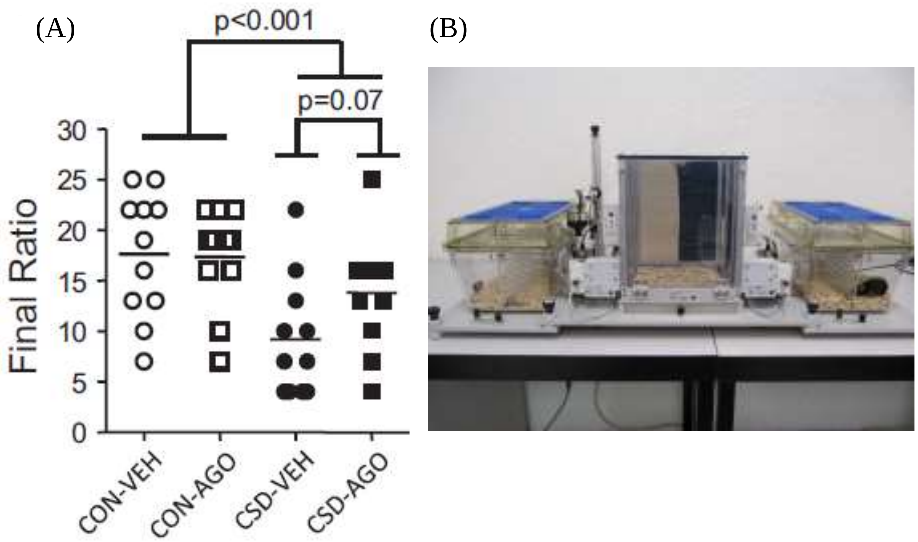


Figure 1. (A) Example of stress-induced reduced effortful motivation on a progressive ratio schedule for sucrose reinforcement in mice. Also shown is the partial reversal of the stress effect by subchronic administration of the antidepressant agomelatine. Taken from ref. 3. (B) Operant apparatus for the study of effortful social motivation in mice.

References:

1. Azzinnari, D., H. Sigrist, S. Staehli, R. Palme, T. Hildebrandt, G. Leparc, B. Hengerer, E. Seifritz, and C.R. Pryce, Mouse social stress induces increased fear conditioning, helplessness and fatigue to physical challenge together with markers of altered immune and dopamine function. *Neuropharmacology*, 2014. **85**: p. 328-341.
2. Pryce, C.R., D. Azzinnari, S. Spinelli, E. Seifritz, M. Tegethoff, and G. Meinlschmidt, Helplessness: a systematic translational review of theory and evidence for its relevance to understanding and treating depression. *Pharmacol Ther*, 2011. **132**: p. 242-267.
3. Bergamini, G., F. Cathomas, S. Auer, H. Sigrist, E. Seifritz, M. Patterson, C. Gabriel, and C.R. Pryce, Mouse psychosocial stress reduces motivation and cognitive function in operant reward tests: a model for reward pathology with effects of agomelatine. *European Neuropsychopharmacology*, 2016. **26**: p. 1448-1464.
4. Bergamini, G., J. Mechtersheimer, D. Azzinnari, H. Sigrist, M. Buerge, R. Dallmann, R. Freije, E. Seifritz, B. Ferger, T. Suter, and C.R. Pryce, Chronic social stress induces peripheral and central immune activation, blunted mesolimbic dopamine function, and reduced reward-directed behaviour. *Neurobiology of Stress*, 2018. **8**: p. 42-56.
5. Bergamini, G., H. Sigrist, B. Ferger, N. Singewald, E. Seifritz, and C.R. Pryce, Depletion of nucleus accumbens dopamine leads to impaired reward and aversion processing in mice: Relevance to motivation pathologies. *Neuropharmacology*, 2016. **109**: p. 306-319.

6. Yang, M., J.L. Silverman, and J.N. Crawley, Automated three-chambered social approach task for mice. *Curr Protooc Neurosci*, 2011. **8.26**: p. 8.26.1-8.26.16.

Autism Spectrum Disorders in Females

David Skuse[1]

1. Behavioural and Brain Sciences, Great Ormond Street Institute of Child Health, University College London.

For many years, our conceptualization of autism was based on a male stereotype. There was a widespread consensus that the sex ratio was at least 4-to-1, males predominating, and epidemiological evidence appeared to support that assumption. An influential theory of autism susceptibility - that it represents an 'extreme male brain' - has subsequently been promulgated with great success.

This 4-to-1 ratio is not consistent across the full range of intelligence quotients (IQ), however, which makes the observation hard to understand in terms of most parameters of genetic risk. No evidence for X-linked susceptibility has been proven. Females with autism and high IQ are rarely clinically identified, yet in those with the lowest intellectual functioning, the sex ratio is no more than 2-to-1. One possible explanation is that our current ascertainment methods for autism are biased toward males. A corollary of this biased ascertainment hypothesis is that so-called 'high-functioning' females with autism are harder to diagnose.

The are several plausible reasons for this biased identification. First, our standardized measures of autism are derived from historical conventions based on the stereotypical symptom profile of boys. Second, boys with autism tend to have co-occurring symptoms and conditions that prompt clinical attention (including attention deficit hyperactivity disorder and disruptive behavior). Girls with autism, on the other hand, tend to have subtler related symptoms (such as social withdrawal, depression and anxiety), so their underlying social communication problems are often overlooked. Third, there is increasing evidence that females with autism show a greater capacity to 'mask' or compensate for their difficulties, and this leads to the development of compensatory behaviors in those who are intellectually able.

This presentation will review empirical evidence to support these hypotheses and will describe the unique qualities of the female phenotype of autism spectrum disorders.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

From behavioral measurements to models that behave

Measuring behavior is not enough to understand it. For instance, the behavior of a fan controls its output, while a thermostat controls its input using its behavior. This distinction is crucial, since one must then acknowledge that behavior in humans, animals and robots, to be adaptive, has to be a closed-loop process. One can then discover the perceptual variables that the organism controls, which is different from how it responds to certain stimuli. To some extent, building is more challenging than intervening or fixing: grounded in (but not restricted to) behavioral measurements, a model that behaves can be more insightful than a model of behavior. Complementary to classification and extrapolation, we seek to discuss relevant methods and techniques to build and test generative models of behavior whose parameters are derived from empirical data.

Warren Mansell, School of Health Sciences, University of Manchester, Manchester, UK and Alex Gomez-Marin, Institute of Neuroscience, CSIC-UMH, Alicante, Spain

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Good-Enough Learning in Rodents and Humans

A. Gomez-Marin[1], S. Ravindranath2, E. Tornese[1], G.E. Loeb[3], R.M. Costa[4]

(1) Behavior of Organisms Laboratory, Instituto de Neurociencias CSIC-UMH, Alicante, Spain; (2) Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey, USA; (3) Dept. of Biomedical Engineering, University of Southern California, Los Angeles, USA; (4) Zuckerman Mind Brain Behavior Institute, Columbia University, New York, USA; agomezmarin@gmail.com

Even when we can confidently claim *that* we taught animals or humans something, we are generally quite ambiguous regarding *what* they actually learned. Namely, it is difficult to specify the content of learning beyond reporting a statistically significant change. As a corollary, two interrelated questions follow: *how?* (implying the dynamics of learning, rather than just paying attention to the end of the learning curve), and by *whom?* (avoiding the tendency to confound individual with average behavior) [4].

Both classical and recent views on learning have proposed that there is an increase in the probability of repeating a specific behavior whose outcome was satisfactory. Some may add that learning leads to optimal solutions in terms of maximizing reinforcers (e.g. food pellets, juice, applause, tenure). Are these two propositions reasonable from the organism perspective? First, if by behavior one means the specific movements or series of motor actions of each individual, then it seems impossible that learning can take place at all. Behavioral variability is a requirement in order to ensure consistency of results achieved in real-life environments, which are inherently noisy and uncertain. Second, can organisms actually optimize their performance in order to reach the global minimum that we experimenters have in our heads? It is more likely that behaving organisms navigate the space of possibilities via trial and error, thus satisfying rather than optimizing [2].

Here we try to articulate a quantitative methodology that merges Perceptual Control Theory [3] with The Good-Enough Principle [4] to study these three aspects of learning (*what*, *how*, by *whom*) that, despite being recognized as essential in theory, defy practice. We address quantitatively the phenomenology of lever-pressing behavior by mice during operant conditioning with different reinforcement schedules. We then present an attempt to transpose that task to humans pressing keys on a computer. Finally, we show a generative model that learns by “running and tumbling” [5] in the space of actions guided by temporal gradients of cumulative reward.

Paradoxically, if animals and humans turn out not to learn what we think we are teaching them, we may end up with the imperative of having to learn what they have to teach us.

References

1. Gallistel, C.R, Fairhurst, S. and Balsam, P. (2004) The learning curve: Implications of a quantitative analysis. *Proceedings of the National Academy of Sciences,* 101 (36) 13124-13131.

2. Simon H.A. (1968) The Sciences of the Artificial. The MIT Press.

3. Powers, W.T. (1978). Quantitative Analysis of Purposive Systems: Some Spadework at the Foundations of Scientific Psychology. *Psychological Review* **85(5)**: 417-435.

4. Loeb G.E. (2012) Optimal isn’t good enough. Biol Cybern. 106(11-12): 757-65.

5. Koshland, D. E. (1980). Behavioral chemotaxis as a model behavioral system. New York: Raven Press.

Deriving Hierarchical Control Structures to Produce Measurable Complex Behaviour

B. Hawker and R. Moore

Department of Computer Science, University of Sheffield, Sheffield, UK, bhawker1@sheffield.ac.uk

Abstract

Controllers, even of a simple nature, are used to produce solutions to real world problems. As problems become more complicated, solutions to produce more refined and robust behaviour are required. As solutions scale up, understanding of the behaviour of the system decreases. This means engineers producing systems are not sure why specific behaviour is the output of the system they produced. Measuring behaviour becomes exponentially difficult as the complexity of the behaviour increases and the judgment of what constitutes the appropriate component of behaviour to measure (frequency, duration, trajectory, angular velocity, etc) becomes unclear. Perceptual hierarchies simplify the problem through focusing on the perceived components of behaviour (and its effects) that matter to the organism in terms of fulfilling its superordinate goals in the hierarchy. Theoretical analysis of systems and their building is required to better understand and thus control the behaviour of a system. Hierarchical control systems, which are compatible with perceptual hierarchies, have long been a strong choice within control theory[1] but with no fundamental understanding how what constitutes or makes said hierarchy. This talk explores the place of hierarchies in control theory and how an understanding of the structure and origin of a hierarchy can make the behaviour of a system more understandable. Furthermore, work conducted on the inverted pendulum problem with hierarchical controllers will show the impressive robustness of simple hierarchical controllers.

Perceptual Control Theory, which combined elements of control theory and behavioural studies, explains how elements of control theory elegantly represent behaviour through the control of perception[2]. Comparing the perceptual signal with a desired value produces an error, which is a measurable quantity representing a magnitude of behavioural response. Complex tasks with complicated behavioural dynamics can be competently performed by agents with hierarchically structured controllers[3]. Thus, hierarchical structures of perceptual control provides an easier method of measuring complex behaviour. Hierarchical structures can be designed differently for the same solution and balanced in different ways, as a result of a lack of understanding of what a hierarchy is. These inconsistencies result in differing performance, robustness and complexity of produced systems. Successful systems cannot always justify the hierarchical structure they chose, citing area specific expertise in the design. Living agents do not require an external agent to explicitly program a hierarchical set of behaviours and responses, meaning that the hierarchical arrangement of simple controllers should be derivable and automatable.

The origin of hierarchical control structures has not been explored, nor has whether they can be developmentally produced autonomously. This talk proposes a method of developmentally producing hierarchical layers of controlled perceptions leading to simple solutions to complex control problems, named 'The Dependency Oriented Structuring Architect' or DOSA. By agents understanding the dependencies of one skill or behaviour on another, an agent can progressively build skills to eventually produce complex behaviour that is robust. This is done by deriving a hierarchy of controllers, beginning with nothing more than the inputs and outputs of the system. This process can be implemented in an artificial agent to produce progressively developed solutions to problems. This aids in reducing the complexity of balancing a large number of control parameters as well as making the solution easier to understand.

This talk will demonstrate how DOSA is able to derive a complex hierarchy to control a robotic arm with multiple complex degrees of freedom through progressive iterations of learning simple skills. Furthermore, use of the underlying methodology of DOSA can be used to optimise hierarchies of behaviour that are already defined. DOSA shall be combined with leading techniques in optimising control hierarchies to improve results. By analysing the perceptual hierarchy when the robot is disturbed, specific perceptual controllers will trigger responses to robustly handle the disruption. Therefore, a perceptual description of behaviour can describe the sophisticated solution of robustly solving a problem.

Reference examples:

1. Brooks, R. (1986). A Robust Layered Control System for a Mobile Robot. *IEEE Journal on Robotics and Automation.* **2(1):** 14-23
2. Powers, W.T. (1978). Quantitative Analysis of Purposive Systems: Some Spadework at the Foundations of Scientific Psychology. *Psychological Review* **85(5)**: 417-435.
3. Powers, W.T. (2008). "Living Control Systems III: The Fact of Control". 127-145.

Acquisition of Control in a Complex Dynamic System Task: Computational Modelling of Spontaneous Improvements in Performance

Vyv Huddy

Div of Psychology & Lang Sciences, UCL

Trial and error behaviour is crucial to the survival of even the most basic forms of life; Koshland (1980) [1] suggested that it is the means by which E.coli navigates nutrient gradients. Marken & Powers (1989) [2] formalised a generative model of *E.coli* behaviour where periodic random changes in direction - or 'tumbles' - allow the rate of nutrient absorption to be controlled. Humans acquire control of apparently much more complex environmental variables. Nonetheless, without prior knowledge of underlying rules, trial and error principles are still crucial for learning. Indeed, they may not only be necessary, but could actually be sufficient to explain performance on some tasks without the need for higher-level strategies. The E.Coli algorithm implements arguably the simplest possible form of trial and error behaviour - our first research question investigated whether simulating the operation of this alone could approximate human performance on a complex dynamic control (CDC) task. This class of tasks are designed so they resemble problem solving in real-world contexts.
The second research question concerned the measurement of behaviour. In the Marken and Powers (1989) simulations, nutrient control emerged abruptly and unpredictably. This is because the algorithm stabilises behaviour as goal achievement occurs and in the absence of contingent environment cues. Similar sudden spontaneous changes in animal behaviour have been observed in conditioning research (see Gallistel, Fairhurst and Balsham, 2004) [3] but have yet to be demonstrated in studies of CDC tasks.

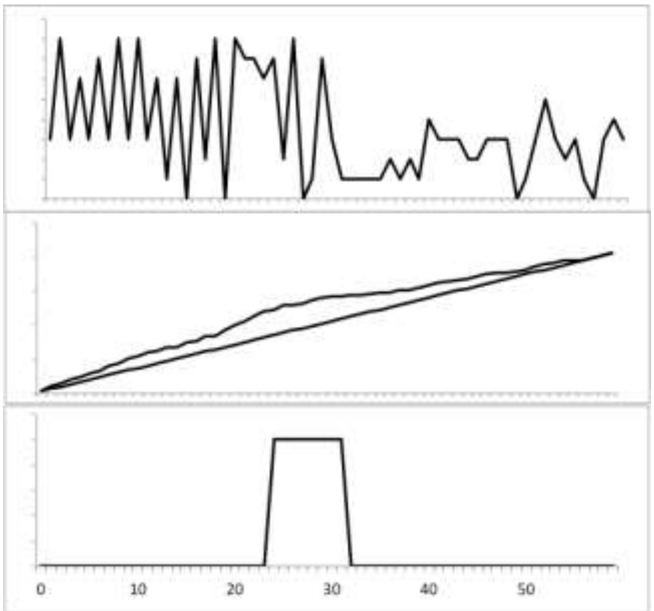
Procedure

We used the classic Broadbent and Berry (1984) [4] 'sugar factory' CDC task to investigate the emergence of control in human and simulated performance. In the task, participants are instructed to control the rate of production of sugar in a simulated factory by changing the size of the work force on each trial. The well-established finding is that participants are able to maintain the sugar production level at a target level.
We built a generative model of performance based on the principle that workforce selections vary in a manner proportional to the current deviation from the goal state. Thus, more stable workforce selections occurred whenever the goal state was closer to being achieved. Twenty-one participants completed the task for comparison with model performance.

Analysis

We implemented two independent algorithms for detecting change points in trial-by-trial time series for both simulated and experimental data. The top panel of the Figure below depicts a single participant's performance, indicated by their error score, on each of the 60 trials of the task.

Visual inspection suggests a change point where more stable control, and generally lower error, occurs after roughly 25 trials (top panel right).



The first algorithm detected a flattening at this point in the cumulative record of error over the course of the trials of the experiment (see middle panel right). The second used sliding window t-test procedure to highlight a region of significance at the same point (see bottom panel right). Change points defined by the two methods were strongly correlated.
The generative model reproduced the pattern of abrupt change points found in experimental data. The simulated and human experimental data also showed a close similarity across range of parameters both at the level of group and individual performance.
These findings have implications for the measurement of behaviour in a range of tasks where adaptive behaviour emerges spontaneously.

References

1. Koshland, D. E. (1980). Behavioral chemotaxis as a model behavioral system. New York: Raven Press.
2. Marken, R. S., & Powers, W. T. (1989) Random-walk chemotaxis: Trial-and-error as a control process. *Behavioral Neuroscience,* 103, 1348–1355.
3. Gallistel, C.R, Fairhurst, S. and Balsam, P. (2004) The learning curve: Implications of a quantitative analysis. *Proceedings of the National Academy of Sciences*, 101 (36) 13124-13131.
4. Berry, D., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology,* 36(A), 209–231

Perceptual Origins of Geometric-Kinematic Constraints in Human Drawing

A. Matic and A. Gomez-Marin

Behavior of Organisms Laboratory, Instituto de Neurociencias CSIC-UMH, Alicante, Spain; agomezmarin@gmail.com

It is known since the 80s that when you write your name on a piece of paper, the instantaneous angular speed is related to the local path curvature according to a power law [1]. Such tight covariance between geometric and kinematic degrees of freedom is not mathematically or physically trivial [2]. Its origin remains debated. Different explanatory frameworks have tried to account for the generation of such constraint, each underscoring (i) kinematics emerging via coupled oscillations of limb segments, (ii) geometrical structure of internal representations related to constant equi-affine speed or (iii) centrally planned optimization via minimum jerk. These competing (and even complementary) views share an emphasis on behavior as production of output. Here, instead, we approach the power law phenomenon thinking of behavior as the *control of perception* [3]. To that end, we used a *closed-loop* perspective to decompose the rich phenomenology exhibited in human drawing. The question then becomes: what are the actively stabilized *invariants* —perceptual variables that humans protect against disturbances while drawing— that give rise to the speed-curvature constraint?

To that end, we designed a set of behavioral experiments where hand movements do not necessarily correspond movements of the cursor on the screen. In this way, we can gain precise experimental control over the feedback function so as to dissect what is controlled visually versus what is controlled kinesthetically and, in each case, identify what perceptual variables (or hierarchies of perceptual variables) are controlled by humans in spontaneously attempting to draw certain shapes. To capture human movement data in real time and at high resolution, we used a Wacom Intous Pro S graphics tablet and a Logitech 3DPro joystick, operating at 200 and 130 Hz, respectively. On the tablet, two-dimensional drawing movements were performed directly with an electronic pen. For the joystick, we used different manipulanda: the two degrees of freedom of the main handle, and the one-dimensional movement of the throttle handle. We wrote custom software in Clojure and C++ to interface with the devices and be able to display on a monitor a whole range of modified curves and trajectories in real time. Our setup then allows to create incongruencies between what the eye sees and what the hand does. For instance, on the visual side, regardless of what humans did with their hand, they would always see movement along an ellipse, but at varying speeds. Or, as they draw an ellipse with a certain eccentricity, they would see an elongated version of it on the monitor. Alternatively, amongst other motor-sensory contingencies, in a less naturalistic situation, the joystick handle could be used as a pen but only controlling geometry while speed being externally preset.

This way of *manipulating and measuring* behavior hopes to make justice to the organism's perspective, namely, to accommodate the fact that movement is not the end of a reflex arc, but a part of a control loop. Perception is in service of action, and action is in service of perception. Or, in the words of the visionary John Dewey: "the so-called response is not merely to the stimulus; it is *into* it" [4].

References

1. Lacquaniti, F., Terzuolo, C., & Viviani, P. (1983). The law relating the kinematic and figural aspects of drawing movements. Acta psychologica, 54(1-3), 115-130.

2. Zago, M., Matic, A., Flash, T., Gomez-Marin, A., & Lacquaniti, F. (2017). The speed-curvature power law of movements: a reappraisal. Experimental brain research, 1-14.

3. Powers, W. T. (1973). Behavior: The control of perception. Chicago: Aldine.

4. Dewey, J. (1896). The reflex arc concept in psychology. Psychological review, 3(4), 357.

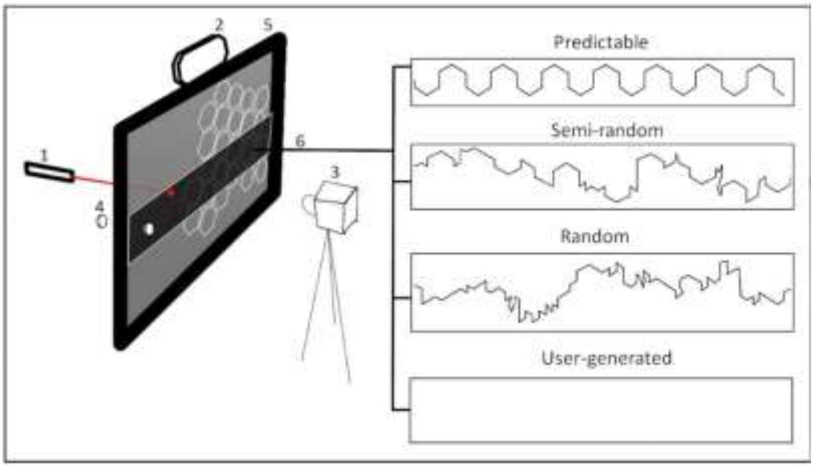
Measuring Behaviour as Perceptual Control Using Human Simulations of Predator and Prey Pursuit

Warren Mansell[1], Maximilian Parker[1], Shaktee Sandhu[1], Tauseef Gulrez[2]

1. University of Manchester, UK; 2. University of Salford, UK

Arguably, predator–prey pursuit provides an analogue of a range of competitive activities in animals and humans such as sports performance, racing and video gaming. We aimed to use this context to examine the extent to which pursuit and evasion are likely to be governed by a the negative feedback control of desired perceptual states in relation to the other agent, versus the use of predictive feedforward models of the other agent. Predictive, feedforward models of control have dominated the last few decades of research. Yet, we anticipated that a negative feedback model would be a better match with the assessed behaviour, because predictive models of the prey's movement by the predator would very soon be outdated through the deliberately unpredictable and random (known as 'protean') behaviour of the prey.
In order to test the above hypothesis, we needed to construct, pilot and modify and detailed set of apparatus, methodology and analysis, which are described in this talk.

Apparatus
A wooden frame (94cm by 61cm) was built with a slot for two transparent Perspex sheets in order to enclose the double-sided hexagonal graph paper placed within the frame (see Figure). This board was placed on top of a table to allow the participants to interact whilst standing. Three strips of clear acetate were constructed with pre-determined patterns of movement printed onto them with varying degrees of predictability: predictable, semi-random and random. Each strip was placed individually, according to their respective condition, upon the hexagonal graph paper at a standardised location, resting between the hexagonal graph paper and Perspex sheet.



A magnetic stylus and magnet were provided for the predator participant and a laser pen was provided for the prey participant. A digital metronome set at 70 beats per minute was used in order to standardise the speed of each participant's movement. A Nikon D5200 DLSR camera was placed at a distance of 110cm perpendicular to the centre of the board and recorded at 25 frames per second the participants' interaction on the board.

Procedure
Each participant was allocated randomly to either the predator or prey role. Then, the predator participant was told to stand outside the room, whilst the prey participant was
goal explained their goal was to avoid getting caught by the predator. The predator participant was then told to enter the room and told that their goal was to catch the prey by staying as close as possible to the prey. Each participant could only move one hexagon space at a time in any direction to the tempo of the metronome. The order of the trials were counterbalanced and randomly assigned.

Analysis
A bespoke MATLAB program was used to analyse the videos in order to code them as x and y coordinates at each strike of the metronome after the start of a trial. This used the intensity of the marker (magnet versus laser)

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

to identify each agent and trace its location on the hexagon grid. Model fitting was also carried out in MATLAB. Competing models of the predator and prey behaviour were compared to the observed data by calculating their absolute distance at each metronome click and averaging across the total number of clicks in a pursuit.
We will report on the results of this analysis with respect to our hypotheses.

From Measurement to Models to Movements: Reproducing Human Tracking Performance with a Model Driven Steering Wheel.

M. G. Parker[1], S. F. Tyson[2], A. P. Weightman[3] & W. Mansell[1]

1Division of Psychology and Mental Health, Faculty of Biology Medicine and Health, University of Manchester; 2School of Nursing and Midwifery, Faculty of Biology, Medicine and Health, University of Manchester; 3Mechanical and Aeronautical Engineering Division, School of Mechanical, Aerospace & Civil Engineering, University of Manchester

Voluntary action is goal oriented. It follows that actions are undertaken in order that the senses perceive a change in the variable of interest until the goal is achieved. Thus, behaviour operates in a closed loop. This is often overlooked in psychological experiments, particularly discrete reaction time tasks. Conversely, continuous tasks such as pursuit tracking enable measurement of the dynamic characteristics of action control. In pursuit manual tracking, participants aim to keep an on-screen cursor aligned with a moving target using a computerised joystick or steering wheel. Computational models of tracking performance can be developed to test theorised motor control mechanisms and elucidate which perceptual variables humans' control during action. By training models on an individual's tracking data, a set of parameter estimates can be derived for each individual. These quantify individual differences in tracking characteristics and strategy. The validity and reliability of this individual model can be assessed by the degree of accuracy by which it simulates that individual's performance when they track new targets. The individual modelling approach is thus a novel method to investigate individual performance.

Perceptual control theory (PCT [1]) provides a basic architecture for a biologically feasible dynamic feedback control unit. The system comprises four parameters which, together, describe a control scheme: Input delay, output gain, damping constant and reference value. This model is evaluated for its fit to human tracking data over a number of experiments with different task conditions. A final study investigates whether the model can be learn naturalistically (in real time) to track targets like human participants.

In the first reported study [2], 20 participants tracked targets that moved in an unpredictable (pseudorandom) pattern. PCT position control models were optimised to the cursor movements of the participants; each participant had a corresponding model. Analysis of model parameters over time revealed that models could capture idiosyncrasies in control strategy, and that these strategies showed temporal consistency. Models were found to be robust because when tested on a different (new) set of pseudorandom tracking data, models accurately simulated each individual's performance. Additionally, the assessment of the contribution to variance of simulation accuracy by each parameter of the model can demonstrate that all parameters are necessary to predict performance, including the unique reference value parameter.

A second study investigated the role of sensory delays in participant and model performance when tracking predictable (sinusoid) and unpredictable (pseudorandom) targets. The position control model was extended to emulate the 'anticipatory' movements observed when participants tracked predictable targets. The extended model compensated for sensory delays by utilizing both position and velocity information in unison, in either a parallel or hierarchical structure. This provides evidence that humans utilise visual velocity information to track predictable targets with zero latency, whilst maintaining a biologically plausible duration of sensory delay (150-200ms).

A final study advances the approach by testing whether a force feedback steering wheel can be driven by these PCT models. Both hierarchical position plus velocity controllers and single-unit position control models are tested. Models will compute cursor positions which are translated to torque outputs which turn the steering wheel to move the cursor to the given position. Models are optimised via live reorganisation; model parameters are updated as the steering wheel tracks in real time. Thus over a number of trials the model develops an accurate tracking strategy which allow it to control the steering wheel in tracking a variety of new targets accurately. We aim to determine whether these models are able to emulate specific individual's control movements on new targets. Individual models will be compared with average models to determine whether models are participant-specific in their predictions, just as in the first study.

We discuss possible applications of such individual model-driven hardware for motor rehabilitation and driving.

References

1. Powers, W.T. (1973) *Behavior: The Control of Perception.* Chicago, IL: Aldine de Gruyter,

2. Parker, M.G., Tyson, S.F., Weightman, A.P., Abbott, B., Emsley, R., Mansell, W. (2017). Perceptual control models of pursuit manual tracking demonstrate individual specificity and parameter consistency. *Attention, Perception,& Psychophysics.*, **79**, (8), 2523–2537.

Facial Expressions Analysis of Human and Animal Behavior

Face and gesture reveal one's health, age, emotion/feelings, intention and is a vital biometric characteristic. Tremendous research is being carried out in the fields of psychology, physiology and computer vision/technology. This symposium focuses on datasets/data sharing, technological challenges in analysis/measurement of human and animal facial expressions and its implication of behaviors to real-world applications. We aim to promote interactions between researchers, scholars, practitioners, engineers and students from across industry and academia on facial expressions analysis. Cross-discipline work is highly encouraged. We welcome original works that address a wide range of issues including, but not limited to:

- Analysis of facial expressions for real-world applications
- Subtle/micro face and gesture movements analysis
- Technology in automated behavior measurement
- Machine learning and deep learning in face and gesture analysis
- Real-time face and motion analysis
- Face and motion recognition on mobile devices
- Pain assessment on veterinary sciences
- The ongoing challenges of subtle motion analysis
- Behavior technology in-the-wild

Moi Hoon Yap (Manchester Metropolitan University, UK, Chair), Adrian Keith Davison (University of Manchester, UK) and Daniel Leightley (King's College London, UK)

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

A new approach to analyse facial behaviour elicited by food products: Temporal Dominance of Facial Emotions (TDFE)

Luis M Cunha[1], Celia Rocha[1,2,3], Rui Costa Lima[2,3]

1. GreenUPorto/DGAOT, Faculdade de Ciências, Universidade do Porto; 2. Sense Test; 3. Lda;

One of the greatest challenges in emotions research is the conception of an efficient tool for their measurement. In the last few years, both explicit and implicit methods had been developed to measure emotions. Within the implicit methods, the FaceReader™ emerged as a tool capable of analysing facial expression patterns whether in recorded videos or real-time capture. From this tool it is possible to automatically analyse facial behaviour expressing emotions intensities in each frame of time, creating big data for analysis, which in turn are difficult to interpret and analyse.

The main goal of this research was to design an algorithm capable of transforming big data of Face Reader into a database for temporal dominance of facial emotions (TDFE).

Measuring Response to Racial Bias Among Preservice Teachers During a Class Intervention

T. A. Drape[1] and S. K. Vincent[2]

1Department of Agricultural, Leadership and Community Education, Virginia Tech, Blacksburg, United States. tdrape@vt.edu; 2Department of Community and Leadership Development, University of Kentucky, Lexington, United States. Stacy.vincent@uky.edu

The demand for multicultural and culturally competent educators is greater today than before. With the changing demographics of the population, the cultural gap between teachers and students must close. Agriculture education is no exception. The role of emotion in teacher education is under-researched and under-theorized. Using facial recognition to determine response during teacher training is one way to gauge student response to help preservice programs understand how students feel subconsciously, even if they are not participating verbally during class discussion. Noldus FaceReader© was used to analyze 14 preservice teachers during an intervention on racism and multicultural education as part of their teacher training program. FaceReader analyzed student emotions throughout the class and the 10-minute discussion that followed the insertion of an emotional stimulus. The class data was used to set a baseline and the 10-minute discussion was sliced out for further analysis. Students' emotions of sad, happy and surprised were the highest discovered. Recommendations include enhancing the quality of multicultural education within the agriculture education preservice training curriculum; assisting students in their own professional growth; and expanding the study to a broader audience.

Introduction

A workforce that is well versed in a variety of cultural working environments is important in order to ensure the United States maintains its global agricultural rank [12]. As the US population continues to diversify, younger generations will be apprised of Hispanics, Asian-Americans, Asians and other races [4]. In the span of 40 years, the non-white population has more than doubled, from 17% to 38% [11]. These trends are not reversible, making demographic diversity in higher education and agriculture concurrent social trends.

Teacher preparation programs can make a difference in increasing multicultural competencies among students by influencing students' course-taking behaviors and ensure that they are exposed to curriculum and activities that help them understand what it means to live in a diverse society [8]. Incorporating content into existing courses is a viable option to help engage faculty who want to contribute to curricular transformation in their department or college.

Need for the Study

The need for multicultural autonomous educators is greater today than ever before. Within the United States the cultural gap between teachers and students continues to broaden [6] and the same is true within school-based agricultural education [5]. The cultural gap doesn't imply that positive impact cannot occur; rather a cultural disconnect is present, which establishes a disconnect the teacher must overcome. Within the cultural disconnect comes a history of traditions, family values and social norms that, through time, develop our psychological norm. Psychological norms effect and create bias, and over a course of experiences called a combinatorial explosion [1], our subconscious emotions welcome and warrant relationships and personality judgments.

Research Objectives

To explore the dynamic of micro expressions among agricultural education pre-service teachers when discussing controversial issues that will affect their treatment of ethnically diverse students, the following objectives were addressed:

- Identify the total emotions exhibited by agricultural education students during a multicultural inclusion lesson.
- Determine the emotions exhibited by the agricultural education students during a 10-minute discussion that followed the insertion of an emotional stimulus.

Methodology

Upon Institutional Review Board approval at the institutions, the researchers designed a lesson to engage the students on the issue of student racial relations. Using webcams from student laptops, the participants were asked to begin recording themselves at the beginning of the class per the instructor's permission. Students who did not consent were not required to record or share their video after the class ended [2]. Of the eighteen student participants, 14 agreed to provide their recorded video for this study. One of the videos failed to provide a substantial number of data points for the FaceReader© software; thus, the study utilized the results of 13 post-secondary agricultural education students.

Data was collected for 50-minutes using student cameras and a stationary camera at the back of the classroom. Data was moved from student computers using an external hard drive or students could opt to upload via Google Drive. Data was analyzed using FaceReader© and analyzed for the duration of each video that each student recorded.

After obtaining consent, students enrolled in an agricultural education course served as the participants to the study. Each student was already admitted into the Teacher Education Program and was 75% completed with their semester of coursework. The students were of junior and senior status. The focus of the course was to provide strategies and approaches for educating culturally different learners.

The facilitator/teacher prompted the students and handed out a news story, which served as an emotional stimulus [9], about a racial hate crime that recently took place on a different university campus. The participants were prompted to utilize skills obtained throughout the semester to find comments of concern. From the comments identified, the students were to think of the origin of each topic. Finally, the students engaged in a 10-minute reflection and discussion of the article, facilitated by the lead instructor. The students were observed consistently from their video recording throughout the 50-minute course.

The instructor served as a moderator during the discussion and assisted with the flow of each topic; progression of topics; engagement of all students; and clarity of understanding when necessary. Once the discussion began, emotions outside of neutral became apparent within the facial expressions of the students.

To assess the students' cultural competence regarding this topic, the instructor probed students with questions that welcomed behaviors of attribution and empathy. For example, at one point in the lesson, the instructor asked, "is there anybody who this article did makes sense too?" Students looked around the room at each other to look for confirmation. At another point, students were asked regarding specific sentences in the article, "How many were angered by the comment from the student?" followed by a "Could you please explain?" These questions triggered emotions as they are asked and as students responded.

Discussion

Objective one sought to identify the emotions exhibited the total data points described in Table 1. Each facial emotion was recorded by 1/3 of a second and reported by total number of seconds. According to Table 1, Neutral facial emotion was identified the most (f = 219,927.24; 79.60%), followed by Sad (f = 22,072.73; 7.99%), Happy (14,310.40; 5.18%), Surprised (f = 7,493.77; 2.71%), Angry (f = 6,198.42; 2.24%), Scared (f = 4,335.49; 1.57%), and Disgusted (f = 1,936.13; 0.70%).

Table 1. Total Seconds of Facial Emotions Data Points Recorded (n = 276,274.20)

Facial Emotion	*f*	%
Neutral	219,927.24	79.60
Sad	22,072.73	7.99
Happy	14,310.40	5.18
Surprised	7,493.77	2.71
Angry	6,198.42	2.24
Scared	4,335.49	1.57
Disgusted	1,936.13	0.70

Table 2 describes the seconds of facial emotions exhibited by the agricultural education students during a class discussion following a silent read (emotional stimulus). Of the facial emotions collected, on average, Neutral was recorded highest (M = 4063.46; SD = 857.80), followed by Sad (M = 446.33; SD = 261.76), Happy (M = 208.22; SD = 162.00), and Surprised (M = 200.97; SD = 157.10). During the 10-minute discussion, 66,995.29 seconds of facial emotions were recorded.

Table 2. Facial Emotions Seconds Recorded During 10-minute Discussion (N = 66,995.29)

Facial Emotion	M	SD	Range
Neutral	4063.46	857.80	1928.21 – 5199.91
Sad	446.33	261.76	83.04 – 890.33
Happy	208.22	162.00	31.08 – 521.33
Surprised	200.97	157.10	25.92 – 243.64
Angry	97.26	92.26	23.38 – 371.61
Scared	91.14	81.10	18.19 – 243.64

Facial Emotion	M	SD	Range
Disgusted	40.26	54.33	09.71 – 204.00

This study is limited to the findings of a single episode within an agricultural teacher education preparatory course. Although the number of participants is minimal, the study provides important exploratory results that emphasize the need to expand the study to a larger audience. During analysis of each student, their prompts after the article from the facilitator turned to disgust, anger, and sadness. The variances of the three emotions remained in that state for .5 to 1 second per expression. The facilitator prompted students with questions as to "how did you feel?" and "what did you think?"

Objective one sought to identify the total number of emotions during the entire 50-minute multicultural inclusion lesson. By capturing video data during the class and pinpointing/targeting the time before and after the stimulus, the total emotional state of each participant was captured and analyzed using FaceReader©. Besides the neutral data point, sad, happy and surprised were the top three emotions discovered in the course, while angry, scared, and disgusted were the lowest. Although, qualitatively, only happy and surprised were identified among the researchers based upon their participation during the lecture; it was clearly discovered that other emotions were present. The micro expression is a reflection of how the student's true behavior. In the future, it is recommended that the students receive an individual analysis of their micro expression breakdown. An individual can work to strengthen bias in order to minimize negative micro expressions. Typical methods for improvement include hours of exposure and immersion exercises [10].

When examining the findings from objective two, the emotions exhibited during the discussion that followed the insertion of an emotional stimulus varied. The students subconsciously responded, even if not actively participating by raising their hand and contributed to the discussion. The instructor used humor at numerous places to ease tension during the discussion, which resulted in a relatively high result of "happy" responses. Taking that into account, students were recorded as sad and surprised during the discussion when the instructor would make connections to the material and their future career trajectory as teachers. The emotion of sadness is often related to empathy [3]; while surprise can be associated with a disconnect, or also referred to as contact, phase in their identity growth [7].

We can conclude: the students did have an emotional response based on their facial analysis outputs. Even if they didn't participate in the class discussion, they were still eliciting responses throughout the 10-minute discussion. We can conclude as preservice teacher educators that they are 1: paying attention, 2: looking for information, and 3: they have feelings, they are responding emotionally in their expressions even if they're not far enough along on the identity development scale, perhaps they're in the pre-encounter stage and this is the first they've really discussed it as a group or as part of their teacher training program.

It would be wise to ask students to come back for qualitative based interviews to get their opinion of how they felt or why they were responding the way they were during the class. Video recording them and analyzing with Noldus Observer would allow the researchers to examine more than the facial expressions during the interviews. Examining change over time from the beginning of a course or teacher prep program and then again more towards the end to see how much change took place over the course of X amount of time.

The facial responses allowed the researchers to see that multicultural issues are necessary and should be woven into the preservice curriculum in more deliberate ways, not a lesson or two throughout the teacher-training program. By incorporating multicultural education, it can help expose students to the issues they may face as future educators. The faculty members also bear responsibility in their need to become multicultural responsive in order to educate and recruit new teachers to the agricultural education profession.

References

Baayen, R. H., Hendrix, P., Ramscar, M. (2013). Sidestepping the combinatorial explosion: An explanation of n-gram frequency effects based on naive discriminative learning. Language and Speech, 56, 329-347.

Den Uyl, M. J ., van Kuilenburg, H. (2005). The FaceReader: Online facial expression recognition. Proceedings of Measuring Behavior 2005, 5th International Conference on Methods and Techniques in Behavrioral Research (Wageningen, 30 August-2 September 2005), 589-590.

Ekman, P. (1982). Emotion in the human face (2nd ed.). Cambridge, MA: Cambridge University Press.

Environmental Systems Research Institute (2011). Major trend revealed in census 2010 data: U.S. race and ethnic diversity increased dramatically in the last decade. <http://www.esri.com/news/arcnews/winter1112articles/major-trend-revealed-in-census-2010-data.html>. ESRI. Accessed 22 August 2016.

Foster, D. D., Lawver, R. G., Smith, A. R. (2015). National agricultural education supply and demand study executive summary. Retrieved from: <http://www.naae.org/teachag/NSD%20ES%20Final%20March%202015%20.pdf.> AAAE. Accessed 4 April 2017.

Gaines, B. (2015). Cultural Gap? Connections and Bridges between Teachers and Students. (Unpublished doctoral dissertation). University of California, San Diego, CA.

Helms, J. E. (1990). Black and White Racial Identity: Theory, Research, and Practice. Contributions in Afro-American and African Studies. New York: Greenwood Press.

Hurtado, S., Milem, J., Clayton-Pedersen, A., Allen, W. (1999). Enacting diverse learning environments: Improving the climate for racial/ethnic diversity in higher education. ASHE-ERIC Higher Education Report, 26, ERIC Clearinghouse on Higher Education, Washington, DC 20036-1181.

Ochsner, K. N., Gross, J. J. (2008). Cognitive emotion regulation: Insights from social cognitive and affective neuroscience. Current Directions in Psychological Science, 17, 153-158.

Petrican, R., Todorov, A., Grady, C. J. (2014). Personality at face value: Facial appearance predicts self and other personality judgments among strangers and spouses. Journal of Nonverbal Behavior, 38, 259-277. doi: 10.1007/s10919-014-0175-3

Tienda, M. (2013). Diversity ≠ inclusion promoting integration in higher education. Educational Researcher, 42, 467-475.

Tindell, S., Young, L., O'Rear, E., Morris, P. (2016). Teaching assistant perspectives on a diversity and social justice education course for collegiate agriculture students. *NACTA Journal*, 60, 158-166.

Can a Machine Learn to See Horse Pain? An Interdisciplinary Approach Towards Automated Decoding of Facial Expressions of Pain in the Horse

P. H. Andersen[1], K.B. Gleerup[2], J. Wathan[3], B. Coles[1], H. Kjellström[4], S.Broomé[4] Y. J.Lee[5], M.Rashid[5], C.Sonder[6], E. Rosenberg[7] and D. Forster[8]

1 Department of Clinical Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden pia.haubro.andersen@slu.se; 2 Department of Large Animal Sciences, Copenhagen University, Copenhagen, Denmark; 3 The Brooke's, London, United Kingdom; 4 Department of Robotics, Perception and Learning, KTH Royal Institute of Technology, Stockholm, Sweden; 5 Department of Computer Science, UC Davis, California, US; 6 Centre for Equine Health, UC San Diego, California, US; 7 Erika Rosenberg Consulting, Berkeley, California, US; 8 Contextual Robotics Institute, UC San Diego, California, US

Because pain is a manifestation of disease, effective pain evaluation is of cardinal importance. Behavioural standardized scales have become key tools for pain evaluation in animals. However, prey animals are challenging as they may display subtle to no visible behaviours during direct observation. Consequently, Machine Learning bears the promise of being a useful clinical and educational tool for many reasons. It can offer pain surveillance 24-7 without disturbing animals; it can generate real-time analysis and warning signals instanteously; and it can transfer valuable expert pain knowledge to novices through illustrative visuals for a better consensus. We have focused on the coding of the equine facial expression of pain - an important spontaneous component of pain behaviour - as the primary basis of Machine Learning input metrics. First, specific anatomy points of key interest are registered as "keypoints." As there are currently no large animal databases that can be used as reference material for training the machines to recognize these key points, members of our team have created an ingenious solution. They have created a method for transferring human database recognition of facial keypoints to the facial expression of horses and sheep, with an impressive low failure rate. The next step is to move from detection of still images to video for better accuracy. The strength of our team is the interdisciplinary approach to address pain, with potentially translational implications for better health and well-being for human and non-human animals.

Introduction

Pain is an important subject of study since it decreases animal welfare. Much controversy exists about the nature of pain and the exact differences between nociception and the experience of pain. The International Association for the Study of Pain [1] defines human pain as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage." Without entering into the details of pain per se, the emotional component, the assessment of pain in animals poses a number of unmet challenges. Animals cannot verbalize their pain as humans can. The subsequent use of standardized behavioural pain scales is challenged by the fact that prey animals, such as horses and cattle, display subtle and less obvious pain behaviours [2].

Moreover, pain is a manifestation of disease. Examples are found in Swedish horse insurance databases where most Swedish horses are registered and disease data have been collected for decades [3]. Such data show that the average lifespan of a Swedish warmblood horse is below half of the natural life expectancy [4]. The most common specific diagnosis is fetlock arthritis, followed by lameness of undefined origin, other locomotor problems, traumatic injuries to the skin, arthritis in several joints, and colic. In Swedish riding school horses, the overall yearly incidence rate was 1584 events of veterinary care and the total mortalities were 790 deaths per 10,000 horse years at risk [5]. These rates are considerably higher than for the general population of insured horses. Nevertheless, there is a large variation in "horse wastage" among riding schools: some riding establishments have strategies that positively effect the risk of injury and death [6]. The large variation in veterinary claims made, in number of days-lost per horse-days-at-risk, for various diseases, indicate that large differences exist in the basic recognition and/or the management of health-related findings, including pain. The number of pain incidents a horse may experience during its life is thus co-dependent on the caretaker's ability to monitor its health and wellbeing.

Pain recognition (as performed by a human observing the animal) is not only challenging but also takes significant training and experience. In addition, certain pain behaviours may only be displayed in the absence of human observers. Consequently, teaching recognition of pain to machines may not only increase the chance of recognizing and monitoring pain: Machine Learning also has the potential of becoming an important and needed two-way educational conduit-learning from expert experience and teaching novices.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Assessment of Pain in Horses

Despite its importance, the identification and management of pain in horses is surprisingly sparsely described in the literature, although different principles of pain quantification have been applied in equine medicine, reviewed by Ashley et al. 2005 [7]. It is generally accepted and proven that no currently known single physiological or biochemical parameter is pathognomonic for pain. Because pain is an emotional experience and the use of pain scales is not yet common, there is little general consensus on: the amount of pain accompanying specific diseases or surgical procedures; whether a particular horse is in pain or not; or whether some pain may be "good" for the horse or not. As an example of lacking consensus, practicing veterinarians scored their assumption of pain in horses with the same diagnoses in a range from non-painful to very painful [8]. Another consequence of the nature of pain is that there is no "gold standard" that can be used in pain studies. Therefore "analgesic testing" may be used where appropriate. If an effective pain medication reverses the abnormal behaviour suspected to be caused by pain towards normal behaviour of the horse, there is reason to believe that the horse was in pain.

Modern horse pain scales are primarily based on behavioural parameters that are more pain-specific than physiological measures [9-14]. These scales were developed for use in hospital environments and the first scales were built on comprehensive registrations of activity budgets for healthy and painful horses.

Equine Pain Behaviour

Horses are flight-fight animals preferring to flee from threatening situations, including pain. If hindered in their natural escape behaviour, they might turn aggressive towards owners and caretakers. Consequently, pain should always be suspected when a horse shows a sudden and/or inexplicable aggressive change in behaviour. In their own surroundings, horses in pain are typically depressed with decreased physical activity, decreased appetite and a diminished interest in socialization. In foreign environments that can be perceived as threatening, this behaviour is often concealed, especially if the horse perceives itself as being observed. Threatening environments could be new stables, hospitals, or the presence of unknown people, as veterinarians or new caretakers. These are the exact same condition for example for a hospitalized, sold, or riding school horse. Horse specific ethograms derived from careful observation of non-painful and painful horses in their own environment is central in the development of a valid pain-scale [7, 15, 16]. Also, the observation of the undisturbed horse is a prerequisite for the specificity and sensitivity of the scale.

Composite Measure Pain Scales

The current most valid pain scales contain scorings of combinations of behavioural and physiological indicators of pain, so called composite measure pain-scales (CMP-scales) [14]. CMP-scales for the evaluation of abdominal and/or orthopaedic pain in horses, with subsequent initial attempts at validation and determination of intra- and/or inter-observer agreement have been proposed by a number of research groups [12, 15-17]. These CMP-scales have been used successfully in experimental pain models. One CMP-scale has been applied in a clinical setting where it discriminated well between horses with and without pain and had good inter-observer reliability [17]. Recently, the performance and robustness of a CMP scale in horses with and without pain, developed by our group, was evaluated among veterinary students with little clinical experience. The scale was robust and had good inter-observer agreement for horses with and without pain [18].

The CMP-scale used in this study contained a range of simple descriptive scales which score pre-selected behavioural and/or physiological features. Scale categories that are simple to understand and score would be, for example: the assignment of the location and position of the horse in box and the scoring of the height of the head of the horse in relation to the withers. Also, the interaction between the horse and the caretaker can be scored quite objectively. However, the CMP-scale also contains subjective scorings derived from human medicine, the Visual Analogue Scale of Pain Intensity (VASPI) and the Numerical Rating Scale (NRS). The VASPI scale is defined only by its two endpoints - "no pain" and "worst imaginable pain." There are no marks or definitions between these endpoints. While the human patient scores his/her own perceived pain intensity, the subjective elements in the CMP scale are scored by the observers of the horse. The VASPI scale accounts for information that potentially could fall outside the rigidly defined and very simple scoring criteria in a painful horse. This VASPI continued to score high in painful horses: we later defined that the facial expressions were key to high VASPI scores. Therefore, the concept of facial expressions of pain were developed to be included in the CMP. Horse owners and horse practitioners have over the years intuitively used the "worried" look as an un-specific indicator of disease. Yet no studies exist on the reliability of these signs [19].

Facial Expressions of Pain

Facial expressions of pain are known to be part of a communication system in humans. They are considered necessary to interact with other people [20]. Spontaneous facial expressions of pain are believed to be innate responses that reflect activation of the nociceptive system. Contrary to this innate response, "stoicism" is considered an active suppression of the pain expression.

During 2014, two independent research groups published, for the first time, investigations on the facial expressions of pain in horses [21, 22], showing that horses exhibit a range of facial expressions when experiencing episodes of acute pain. In [21], pain was induced in otherwise healthy horses using known pain models, whereas the horses in [22] were postsurgical hospitalized patients, where the effects of residual anaesthetic drugs and fatigue were present together with the postoperative pain from castration. Nevertheless, a range of facial cues appeared to be similar, namely "low or asymmetrical ears," "angled appearance of the eyes," "withdrawn/tense stare of the eyes," "medio-laterally dilated nostrils," and "visibly increased tension of the muscles of the mouth, lips and chin." Interestingly, these features correspond to the more formalized ontology [23] described below.

Can People "See" Pain in Horses?

It is widely accepted that humans have a neural apparatus for processing facial cues to recognize emotions, including pain [24, 25]. This has proven useful as a tool in pain assessment in non-verbalizing humans such as infants [26]. Facial expression of durations less than 0.5 seconds may be interpreted and training may further improve this decoding ability [27]. A pilot study was conducted to investigate if persons of different background could be trained to assess clinical pain in video films of horses [28]. The study showed that at least some people can be trained to recognize moderate and severe pain with moderate rater agreement in video clips of horses when video clips were selected and trimmed by the research team. However, movement, stress, coat colour and nervous behaviour of the horse interfered with the correct interpretation. Other and less time consuming objective and automated methods are therefore necessary for correct decoding of the facial cues in a large number of horses. Sensitivity and specificity could not be calculated due to the pilot nature of the study and the lack of gold standard for pain.

Computer Assessment of Equine Pain Expressions

As discussed above, there is significant evidence that horses communicate their experiences of pain in a rich manner to other horses and humans, not least through their repertoire of facial expressions. Moreover, it has been shown that humans can learn to recognize these expressions of pain. However, as mentioned above, horses may hide these expression under conditions they perceive as threatening. Therefore, we argue that there are great benefits of training computers to recognize horse pain, using different types of Machine Learning methods.

Facial Action Coding System (FACS)

The Facial Action Coding System (FACS) provides a method for identifying and recording facial expressions based on the movement of the underlying facial muscles [29]. FACS exhaustively describes all observable facial behaviour in terms of the muscular actions that comprise it, using elements called action units (AUs). Each AU, designated by an arbitrary numeric code, denotes the movements of an underlying facial muscle group. FACS coders rely on direct observation of facial muscle movement as well as changes in facial morphology (e.g., the position of the eyebrows, size/shape of mouth, lips, or eyelids, the appearance of various furrows, creases, bulges of skin, etc.) to determine which AU(s) occurred. Any interpretations made about the emotional meaning of the observed AUs occur post-coding. After appropriate training, human observers are able to use the FACS-system with high agreement between coders. FACS has been adapted to several animal species, for example orangutans [30], cats [31], mice [32] and dogs [33].

Although FACS may not always be sufficient to capture the emotional state of some animals [31], the objective nature of this coding scheme makes it highly suitable for automation. We will now review the adaption to horses.

EquiFACS

The adaption to horses, EquiFACS, is based on a thorough dissection of underlying facial musculature and filming of naturally occurring facial expressions of the horse [23]. EquiFACS is strictly based on muscle anatomy while other ethograms may use concepts such as grimaces that may reflect changes in other adjacent tissues. This difference may be of importance when evaluating pain score scales based on different ethograms.

Facial expressions in horses are dynamic and often complex signals that can change rapidly in response to a range of environmental stimuli and internal affective states [34]. The facial expressions of ridden horses may reflect responses to a range of stimuli, including responses to the signals from the rider, pressure from saddle and tack [35, 36], and interactions with the surroundings, in addition to pain. It is therefore imperative to take these complications into account when designing studies evaluating facial expressions in horses. A carefully prepared protocol must include control of exposure to external stimuli and factors so that demonstrated differences really can be ascribed to pain. The lack of a gold standard for pain is an issue that must be addressed when implementing computer recognition of pain.

One important consequence of the rapidly changing facial expressions across mammals is the use of still photographs which may pose limitations to the interpretation of facial expressions. Selected stills images are prone to selection bias. In addition, the use of a single frame carries an inevitably loss of information in temporal distribution, as shown previously in the work by Wathan [23]. Wathan argues that certain facial movements can only be distinguished accurately from sequences. This is particularly true of facial movements around the muzzle and corresponds with Mullard et al.'s [37] results that show that indicators around the muzzle were poorly interpreted. Preparation of un-validated and "case-based" ethograms based on still images out of context – a situation known from for example blinded scorings - is therefore not advisable.

Can computers be trained to "see" pain in horses?

Computer Vision science has now advanced to the point where automatic facial expression recognition systems can be used in the investigation of behavioural research [38-41]. During the past decade, fully-automated systems have been developed for recognition of basic human emotions (neutral, anger, disgust, fear, joy, sadness, and surprise) in video streams of human faces. For example, Littlewort et al. [39] achieve 100% accuracy on emotion classification for four of the seven expressions and are additionally able to code videos and images with action unit (AU) activations and intensities in real-time. Performance on pain identification in humans is similarly high, with Rodriguez et al. [42] reaching 91.3% accuracy. Figure 1[43] show a visualization of what a Machine Learning model has learnt to detect in face images.

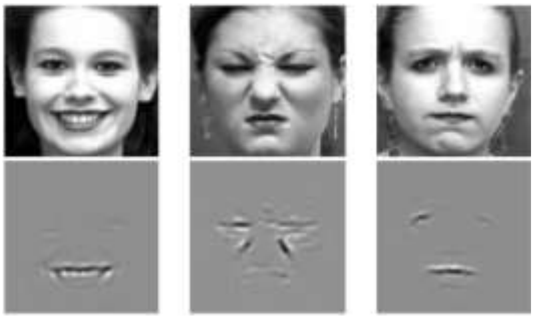


Figure 1. Top row: Images with happiness, disgust, and anger expressions. Bottom row: facial patterns that a Machine Learning model (CNN) can learn to detect (Selvaraju et al. 2017).

These systems learn to identify patterns of facial features that distinguish one expression from another, using a large set of "ground truth" examples provided by expert human observers. They not only give comparable performance to human experts but are also often able to produce real-time predictions for video data. Meanwhile, human experts can take up to 3 hours to code a 1-minute video by hand. There are no existing analogous systems for recognizing facial expressions or pain in videos of animals as there are for humans.

A rudimentary issue with animals is the need for registration of the facial image before expression recognition can be attempted. An approach to this is to detect facial keypoints, i.e. points in an image or video that indicate the location of an important part of the face. For example, Figure 2 [44] shows the facial keypoints that indicate the location of the horse's nose, mouth and eye corners. The registration of the facial image, "face alignment," uses the keypoints to rotate, scale, or otherwise transform images so that the location of the keypoints are approximately the same across all images in the training data. This helps in extracting useful features during both training and testing. Keypoints are also used to extract features around parts that visually change with AU activations. Therefore, it is important to reliably detect facial keypoints so that high performing systems can be developed.

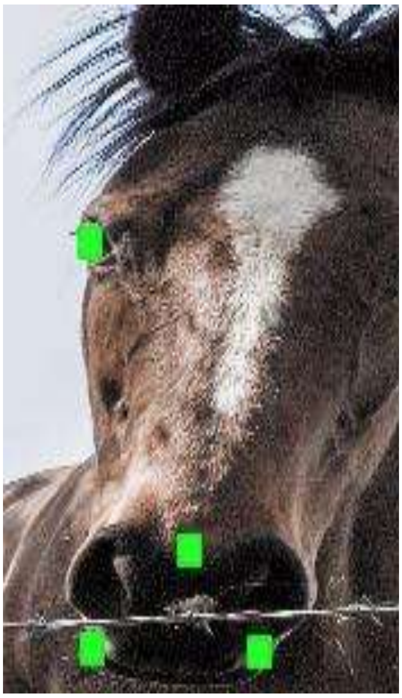


Figure 2. Example of facial keypoints on a horse face, from Rashid (2017).

Keypoint detection, like other Machine Learning tasks, needs training data. While there are large datasets with human facial keypoint annotations, e.g. [45], there are unfortunately no large datasets of animal facial keypoints that could be used to train a Machine Learning method from scratch – for instance, the sheep dataset from [46] has only ~600 images. To address this limitation, Rashid et al. [44] developed a method to transfer information from human datasets for animal keypoint detection by training a Convolutional Neural Network (CNN) to reduce the structural face shape differences between human and animal faces. The method achieved state of the art performance on both horse and sheep facial keypoint detection with 8.36% and 0.87% failure rates, respectively.

In parallel, Bhatti [47] has developed a preliminary horse pain detection method based on feature extraction with Gabor filters and linear classification. Even with this standard method, they reached a classification accuracy of up to 78% accuracy in horses. Experiences with the performance improvement in human pain expression recognition with the introduction of Deep Neural Networks, e.g. [42], lead us to hope for massive improvements in automatic recognition of equine pain expressions using these kinds of methods. We are exploring two different traits, one where we focus on the face and register facial images using keypoint detection [44] and capsule networks [48, 49] and another where the method learns more holistic cues about the entire horse pose, weight distribution, and spatial behaviour.

Moreover, the current methods for horse pain detection are based on still images. This means that temporal information has been ignored. As shown with both humans [42] and horses [21], temporal information is important when a human interprets signs of pain. We therefore focus the holistic behaviour study on temporal information, using a combination of Long-Short Term Memory (LSTM), a temporal Deep Neural Network, and CNN.

Concluding comments

The work described in this paper is part of a larger effort aimed at improving standards for recognizing and assessing pain in horses and other mammals. Also, we wanted to investigate the potential development of technology platforms that can monitor and detect pain automatically in real-time. During this process, we have learned that an interdisciplinary approach is needed from the outset, for example, to avoid differences in descriptions of AUs or collection of data that inadvertently teaches the machine inaccurate categories. The clinical implications of valid and reliable pain recognition, assessment and monitoring are critically significant for both animal welfare as well as human health and wellbeing. The technology platforms could eventually also support a productive research agenda to further our understanding of pain and related states such as fear, fatigue and stress.

References

1. *IASP Taxonomy*. 2016 [cited 2016 July 7th]; Available from: http://www.iasp-pain.org/Education/Content.aspx?ItemNumber=1698.
2. Taylor, P.M., P.J. Pascoe, and K.R. Mama. (2002). Diagnosing and treating pain in the horse.*Veterinary Clinics of North America: Equine Practice*. **18**, 1-19.
3. Egenvall, A., J.C. Penell, B.N. Bonnett, P. Olson, and J. Pringle. (2006). Mortality of Swedish horses with complete life insurance between 1997 and 2000: variations with sex, age, breed and diagnosis.*Veterinary Record*. **158**, 397-406.
4. Wallin, L., E. Strandberg, J. Philipsson, and G. Dalin. (2000). Estimates of longevity and causes of culling and death in Swedish warmblood and coldblood horses.*Livestock Production Science*. **63**, 275-289.
5. Penell, J.C., A. Egenvall, B.N. Bonnett, P. Olson, and J. Pringle. (2005). Specific causes of morbidity among Swedish horses insured for veterinary care between 1997 and 2000.*Veterinary Record*. **157**, 470-477.
6. Egenvall, A., C. Lonnell, C. Johnston, and L. Roepstorff. (2010). Orthopaedic health status of horses from 8 riding schools - a pilot study.*Acta Veterinaria Scandinavica*. **52**.
7. Ashley, F.H., A.E. Waterman-Pearson, and H.R. Whay. (2005). Behavioural assessment of pain in horses and donkeys: application to clinical practice and future studies.*Equine Veterinary Journal*. **37**, 565-575.
8. Price, J., J.M. Marques, E.M. Welsh, and N.K. Waran. (2002). Pilot epidemiological study of attitudes towards pain in horses.*Veterinary Record*. **151**, 570-575.
9. Raekallio, M., P.M. Taylor, and M. Bloomfield. (1997). A comparison of methods for evaluation of pain and distress after orthopaedic surgery in horses.*Journal of Veterinary Anaesthesia*. **24**, 17-20.
10. Price, J., S. Catriona, E.M. Welsh, and N.K. Waran. (2003). Preliminary evaluation of a behaviour-based system for assessment of post-operative pain in horses following arthroscopic surgery.*Veterinary Anaesthesia and Analgesia*. **30**, 124-137.
11. Sellon, D.C., M.C. Roberts, A.T. Blikslager, C. Ulibarri, and M.G. Papich. (2004). Effects of continuous rate intravenous infusion of butorphanol on physiologic and outcome variables in horses after celiotomy.*Journal of Veterinary Internal Medicine*. **18**, 555-563.
12. Graubner, C., V. Gerber, M. Doherr, and C. Spadavecchia. (2011). Clinical application and reliability of a post abdominal surgery pain assessment scale (PASPAS) in horses.*Veterinary Journal*. **188**, 178-183.
13. van Loon, J. and M.C. Van Dierendonck. (2015). Monitoring acute equine visceral pain with the Equine Utrecht University Scale for Composite Pain Assessment (EQUUS-COMPASS) and the Equine Utrecht University Scale for Facial Assessment of Pain (EQUUS-FAP): A scale-construction study.*Veterinary Journal*. **206**, 356-364.
14. Gleerup, K.B. and C. Lindegaard. (2016). Recognition and quantification of pain in horses: A tutorial review.*Equine Veterinary Education*. **28**, 47-57.

15. Lindegaard, C., M.H. Thomsen, S. Larsen, and P.H. Andersen. (2010). Analgesic efficacy of intra-articular morphine in experimentally induced radiocarpal synovitis in horses.*Veterinary Anaesthesia and Analgesia*. **37**, 171-185.
16. Bussieres, G., C. Jacques, O. Lainay, G. Beauchamp, A. Leblond, J.L. Cadore, L.M. Desmaizieres, S.G. Cuvelliez, and E. Troncy. (2008). Development of a composite orthopaedic pain scale in horses.*Res Vet Sci*. **85**, 294-306.
17. van Loon, J.P.A.M., W. Back, L.J. Hellebrekers, and P.R. van Weeren. (2010). Application of a Composite Pain Scale to Objectively Monitor Horses with Somatic and Visceral Pain under Hospital Conditions.*Journal of Equine Veterinary Science*. **30**, 641-649.
18. Lindegaard, C., K.B. Gleerup, and P.H. Andersen. *Evaluation of pain in hospitalized horses – robustness of a composite pain scale.* . in *27th NKVet Pain Symposium*. 2012. Uppsala, Sweden.
19. Fraser, J.A. (1969). SOME OBSERVATIONS ON BEHAVIOUR OF HORSE IN PAIN.*British Veterinary Journal*. **125**, 150-&.
20. Haxby, J.V., E.A. Hoffman, and M.I. Gobbini. (2000). The distributed human neural system for face perception.*Trends in Cognitive Sciences*. **4**, 223-233.
21. Gleerup, K.B., B. Forkman, C. Lindegaard, and P.H. Andersen. (2015). An equine pain face.*Veterinary Anaesthesia and Analgesia*. **42**, 103-114.
22. Dalla Costa, E., M. Minero, D. Lebelt, D. Stucke, E. Canali, and M.C. Leach.(2014). *Development of the Horse Grimace Scale (HGS) as a Pain Assessment Tool in Horses Undergoing Routine Castration*. **9**, DOI: 10.1371/journal.pone.0092281.
23. Wathan, J., A.M. Burrows, B.M. Waller, and K. McComb.(2015). *EquiFACS: The Equine Facial Action Coding System*. **10**, 35 DOI: 10.1371/journal.pone.0131738.
24. Kadosh, K.C. and M.H. Johnson. (2007). Developing a cortex specialized for face perception.*Trends in Cognitive Sciences*. **11**, 367-369.
25. Deyo, K.S., K.M. Prkachin, and S.R. Mercer. (2004). Development of sensitivity to facial expression of pain.*Pain*. **107**, 16-21.
26. Poole, G.D. and K.D. Craig. (1992). JUDGMENTS OF GENUINE, SUPPRESSED, AND FAKED FACIAL EXPRESSIONS OF PAIN.*Journal of Personality and Social Psychology*. **63**, 797-805.
27. Matsumoto, D. and H.S. Hwang. (2011). Evidence for training the ability to read microexpressions of emotion.*Motivation and Emotion*. **35**, 181-191.
28. Gleerup, K.B., B. Forkman, C. Lindegaard, and P.H. Andersen. *Facial expressions as a tool for pain recognition in horses*. in *The 10th International Equitation Science Conference*. 2014. Denmark.
29. Ekman, P., W. Friesen, and J. Hagar, *Facial Action Coding System*. Research Nexus. 2002, Salt Lake City.
30. Caeiro, C.C., B.M. Waller, E. Zimmermann, A.M. Burrows, and M. Davila-Ross. (2013). OrangFACS: A Muscle-Based Facial Movement Coding System for Orangutans (Pongo spp.).*International Journal of Primatology*. **34**, 115-129.
31. Caeiro, C.C., A.M. Burrows, and B.M. Waller. (2017). Development and application of CatFACS: Are human cat adopters influenced by cat facial expressions?*Applied Animal Behaviour Science*. **189**, 66-78.
32. Langford, D.J., A.L. Bailey, M.L. Chanda, S.E. Clarke, T.E. Drummond, S. Echols, S. Glick, J. Ingrao, T. Klassen-Ross, M.L. LaCroix-Fralish, L. Matsumiya, R.E. Sorge, S.G. Sotocinal, J.M. Tabaka, D. Wong, A.M.J.M. van den Maagdenberg, M.D. Ferrari, K.D. Craig, and J.S. Mogil. (2010). Coding of facial expressions of pain in the laboratory mouse.*Nature Methods*. **7**, 447-U452.
33. Waller, B.M., K. Peirce, C.C. Caeiro, L. Scheider, A.M. Burrows, S. McCune, and J. Kaminski.(2013). *Paedomorphic Facial Expressions Give Dogs a Selective Advantage*. **8**, 6 DOI: 10.1371/journal.pone.0082686.
34. Hintze, S., S. Smith, A. Patt, I. Bachmann, and H. Wurbel.(2016). *Are Eyes a Mirror of the Soul? What Eye Wrinkles Reveal about a Horse's Emotional State*. **11**, 15 DOI: 10.1371/journal.pone.0164017.
35. Doherty, O., V. Casey, P. McGreevy, and S. Arkins.(2017). *Noseband Use in Equestrian Sports - An International Study*. **12**, 18 DOI: 10.1371/journal.pone.0169060.
36. Fenner, K., S. Yoon, P. White, M. Starling, and P. McGreevy.(2016). *The Effect of Noseband Tightening on Horses' Behavior, Eye Temperature, and Cardiac Responses*. **11**, 20 DOI: 10.1371/journal.pone.0154179.
37. Mullard, J., J.M. Berger, A.D. Ellis, and S. Dyson. (2017). Development of an ethogram to describe facial expressions in ridden horses (FEReq).*Journal of Veterinary Behavior-Clinical Applications and Research*. **18**, 7-12.

38. Huang, J., K. Craig, D. Diaz, K. Sikka, A. Ahmed, L. Terrones, G. Littlewort, M. Goodwin, and M. Bartlett. (2014). Automated facial expression analysis can detect clinical pain in youth in the post-operative setting.*Journal of Pain*. **15**, S3-S3.
39. Littlewort, G.C., M.S. Bartlett, K. Lee, and Acm, *Faces of Pain: Automated Measurement of Spontaneous Facial Expressions of Genuine and Posed Pain*. Icmi'07: Proceedings of the Ninth International Conference on Multimodal Interfaces. 2007. 15-21.
40. Bartlett, M.S., G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, J. Movellan, and I.C. Soc, *Fully automatic facial action recognition in spontaneous behavior*. Proceedings of the Seventh International Conference on Automatic Face and Gesture Recognition - Proceedings of the Seventh International Conference. 2006. 223-228.
41. Bartlett, M.S., G.C. Littlewort, M.G. Frank, and K. Lee. (2014). Automatic Decoding of Facial Movements Reveals Deceptive Pain Expressions.*Current Biology*. **24**, 738-743.
42. Rodriguez, P., G. Cucurull, J. Gonalez, J.M. Gonfaus, K. Nasrollahi, T.B. Moeslund, and F.X. Roca. (2017). Deep Pain: Exploiting Long Short-Term Memory Networks for Facial Expression Classification.*IEEE transactions on cybernetics*.
43. Selvaraju, R.R., M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, and IEEE, *Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization*, in *2017 IEEE International Conference on Computer Vision*. 2017, IEEE: New York. 618-626.
44. Rashid, M., X. Gu, Y.J. Lee, and IEEE, *Interspecies Knowledge Transfer for Facial Keypoint Detection*, in *30th IEEE Conference on Computer Vision and Pattern Recognition*. 2017. 1600-1609.
45. Kostinger, M., P. Wohlhart, P.M. Roth, H. Bischof, and IEEE, *Annotated Facial Landmarks in the Wild: A Large-scale, Real-world Database for Facial Landmark Localization*. 2011 IEEE International Conference on Computer Vision Workshops. 2011, New York: IEEE.
46. Yang, H., R. Zhang, P. Robinson, and IEEE, *Human and Sheep Facial Landmarks Localisation by Triplet Interpolated Features*, in *2016 IEEE Winter Conference on Applications of Computer Vision*. 2016.
47. Bhatti, K., (2015). Automated detection of pain in horses through facial expression analysis. Chalmers University of Technology: Gotenburg, Sweden.
48. Sabour, S., N. Frosst, and G. Hinton. (2017). Dynamic Routing Between Capsules.*arXiv*. https://arxiv.org/pdf/1710.09829.pdf accessed April 2018.
49. Hinton, G., S. Sabour, and A.B. Frost, (2018). Matrix capsules with EM routing. *International Conference on Learning Representations*. Vancouver Canada. https://openreview.net/pdf?id=HJWLfGWRb accessed April 2018.

Measuring recognition of emotional facial expressions by people with Parkinson's disease using an eye-tracking methodology

Karen Lander, Judith Bek and Ellen Poliakoff

The University of Manchester, UK; karen.lander; judith.bek; ellen.poliakoff@manchester.ac.uk

Parkinson's disease can cause a reduction in the facial expression of emotion (e.g., Simons et al., 2004). This reduced expressiveness is likely to impact on communication and social interaction in people with Parkinson's, resulting in feelings of isolation and reducing quality of life.

Research in healthy populations has established that motion is an important cue when making face emotional expression judgements. Indeed, moving facial expressions are known to be recognised more accurately (Trautmann, Fehr & Herrmann, 2009) and quickly (Recio, Sommer & Schacht, 2011) than static expressions (see Krumhuber, Kappas & Manstead, 2013 for review). Importantly for people with Parkinson's, the reduction in their own facial expressions may influence their ability to use motion as a cue to recognition of others' expressions. Recent meta-analyses (Tickle-Degnen & Doyle Lyons, 2004; Péron et al., 2012) suggest that individuals with Parkinson's are more impaired in the recognition of negative emotions (anger, disgust, fear, sadness) than positive emotions (happiness), although some studies have found no deficit (Gray et al., 2010). This previous work with Parkinson's patients has only used static stimuli, with no consideration of the information added by seeing the face in motion. In addition, healthy people move their eyes to informative areas of the face when identifying emotional expressions whereas people with Parkinson's disease may be less effective in their exploration of the face.

The presented study uses eye-tracking methodology to measure how people with Parkinson's disease move their eyes during a facial emotion recognition task. In our pilot work, two groups of participants were recruited: (1) people with Parkinson's disease (n = 9) and (2) neurologically healthy controls of a similar age range (n = 10). All participants provided written informed consent prior to taking part in the study, which was approved by an NHS Research Ethics Committee. Participants viewed short video clips or a single static image of male and female actors (3 male & 3 female actors portraying each expression x static/dynamic; total 84 trials). Static and dynamic expressions were presented in a randomized order. expressing each of the six basic emotions (happiness, sadness, anger, disgust, fear and surprise) and neutral expressions, and then selected an emotion word that matched the expression. The 6 emotions represent universal innate human responses that are independent of culture and learning. Eye movements while viewing static and dynamic expressions were recorded using an Eyelink 1000+ eye tracker with remote pupil capture. No differences between groups were found for basic eye movement metrics (number and duration of fixations, number of saccades and average saccade amplitude), or for accuracy of emotion recognition. Consistent with findings from young healthy participants, accuracy was higher for moving than static faces ($p = .043$). Fewer fixations ($p = .002$) and saccades ($p = .002$) were made when viewing moving than static stimuli, and fixations were longer ($p = .014$), suggesting that attention may have been more focused on specific features of moving faces. This will be explored in further analysis of interest areas.

The findings of this pilot study demonstrate that people with Parkinson's can utilise motion information as a cue for emotion recognition. The study also highlighted challenges in measuring the eye movements of people with Parkinson's, in whom oculomotor abnormalities are common.

References

Gray, H.M., Tickle-Degnen, L. (2010). A meta-analysis of performance on emotion recognition tasks in Parkinson's disease. *Neuropsychology, 24,* 176–191.

Krumhuber, E.G., Kappas, A. & Manstead, A.S.R. (2013). Effects of dynamic aspects of facial expressions: a review. *Emotion Review, 5,* 41–46.

Péron, J., Dondaine, T., Le Jeune, F., Grandjean, D., & Vérin, M. (2012). Emotional processing in Parkinson's Disease: A systematic review. *Movement Disorders, 27,* 186-199.

Recio, G., Sommer, W., & Schacht, A. (2011). Electrophysiological correlates of perceiving and evaluating static and dynamic facial emotional expressions. *Brain Research, 1376,* 66–75.

Simons, G, Pasqualini, M.C.S., Reddy, V. & Wood, J. (2004). Emotional and nonemotional facial expressions in people with Parkinson's disease. *Journal of the International Neuropsychological Society, 10(4),* 521-535.

Tickle-Degnen, L. & Doyle Lyons, K. (2004). Practitioners' impressions of patients with Parkinson's disease: the social ecology of the expressive mask. *Social Science & Medicine, 58,*603–614.

Trautmann, S.A., Fehr, T. & Herrmann, M. (2009). Emotions in motion: dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain Research, 1284*, 100–115.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Modelling attributes of common mental health disorders in adolescents using audio-visual data

Daniel Leightley

King's Centre for Military Health Research, King's College London, London, United Kingdom.; daniel.leightley@kcl.ac.uk

Background

Society has a moral obligation and duty of care to service personnel and their families, as enshrined in the Armed Forces (AF) Covenant [1]. However, stigma associated with mental ill health is a significant barrier to accessing help [2], with the prevalence of Common Mental Health (CMH) problems and alcohol misuse higher in the military than the general population that is great cause for concern. The House of Commons Defence Select Committee [3] has stated that more needs to be done to support the AF community, particularly in mental health of families [4]. To date, no studies have looked at the impact upon an adolescent of a parent serving in the AF. Research has shown that adolescents of military families are exposed to a range of stressors that are not experienced by their civilian counterparts. Digital technology shows great promise in giving the public, patients, healthcare professionals and researchers' new and innovative means of accessing help, collecting data, communicating, processing, monitoring and intervening in mental health.

Aim

The aim of this work is to explore the feasability of using audio and visual data obtained from adolescents being in conversation with military parents, using big data analytics and machine learning to imprve identification of the adolescents' mental health.

Data

The King's Centre for Military Health Research has, over the last 2 years, collected a wide range of data types from adoleents of military families; the study is called Service Parents' & Adolescents' Challenges & Experiences Study. Fivehundred gigabities of self-reported questionnaire, audio and video data has been collected and clinically evaluated and coded. Of great interest is the audio-video data which includes hours of conversations between the parent and adolescent. These have been coded to identify key expressed emotions such as warmth, negativity and sensitivity, which in turn can be used to infer mental health wellbeing.

Methods

An offline framework has been developed using Graphical Processing Units, and consists of two phases; collaborative filering and audio models;

- <u>Collaborative Filtering</u> models are used to analyse the expressed behavior of the parent and adolescent, from their behaviour captured on video. A schema has been developed to identify warm and negative body language and basic identification of facial expressions.
- <u>Audio models</u> are used to analyse speech by frequency, pitch and tone to identify warm and negativity in the speech.

Each stage model attributes are extracted and combined to provide a holistic overview of the adolescents' mental health state by modelling attributes extracted from the audio-visual data.

Results

Initial exploration of the results demonstrates that the proposed framework is feasible in detection common mental health disorders in adolescents. However, further work is required to improve the sensitivity, automation and data availability.

References:

[1] House of Commons Defence Committee, "The Armed Forces Covenant in Action Part 5: Military Casualties , a review of progress," London, 2015.

[2] M. L. Sharp *et al.*, "Stigma as a barrier to seeking health care among military personnel with mental health problems," *Epidemiol. Rev.*, vol. 37, no. 1, pp. 144–162, 2015.

[3] The Ministry of Defence, "THE ARMED FORCES COVENANT," Ministry of Defense, 2015.

[4] B. Farmer, "Prince Harry to help tackle mental health in the Armed Forces." [Online]. Available:

http://www.telegraph.co.uk/news/2017/10/09/prince-harry-help-tackle-mental-health-armed-forces/. [Accessed: 13-Oct-2017].

Baby FaceReader AU classification for Infant Facial Expression Configurations

Andreas Maroulis

VicarVision, Amsterdam, the Netherlands, andreas@vicarvision.nl

Intro

Baby FaceReader (BFR) is a novel computer vision solution used to automatically code facial expressions in infants based on the Noldus FaceReader [1] Using the Viola-Jones algorithm [2], BFR detects an infant face in a video or an image, models the detected face using a 3D Active Appearance Model (AAM) [3], and then uses neural network classifiers to determine facial Action Unit (AU) activation based on Oster's Baby Facial Action Coding System (Baby FACS) [4]. BFR applies a 3D AAM on 2D images / videos that is more robust to movement than 2D AAMs. The 3D AAM is design by mapping 3D models of faces on the equivalent 2D models of the same faces, thus backpropagading a function that can infer a 3D model for any new 2D image of face. While Baby FACS AUs objectively quantify individual movements in the face and seemingly any combinations of those movements, it has 2 main limitations:

1. Studying co-occurrence of individual action units to determine holistic global expressions remains a tedious, time-consuming tasks
2. Understanding the affective and cognitive meaning of each action unit or co-occurrence of action units is not within the scope of the coding manual

As a result, researchers often code facial expressions in more holistic terms such as positive, negative, pain, smiles and cry faces [5, 6, 7]. While these studies show high reliability scores in coding of these facial expressions, they do not investigate the morphological differences that differentiate one facial expression from another (i.e. what makes a negative facial expression negative, and is that different from a pain or cry expression?). For such a facial expressions codes to be truly reliable they must be expressed in objective terms, i.e. using combinations of AUs to define the morphology of each defined expression. This study will build on previous work [8] and will have 2 main goals:

1. To improve BFR's AU classifiers. In [8] BFR exhibited a 0.60 agreement score against manually annotated images. The gold standard for an agreement score in facial action coding is 0.70 [9].
2. To use the improved BFR classifiers to objectively define a variety of facial expressions such as positive, negative, pain, smiles and cry faces in terms of AUs.

Method

To evaluate the BFR's improved performance we will follow the same procedure as [8]. Using the Baby FACS dataset [4] of 74 images. Contrary to FACS, Baby FACS defines AU3 as Brow Knitting (the movement caused by a contraction of the corrugator supercili muscle that causes the eyebrows to move towards each other), and AU4 as Brown Knotting (the movement of the procerus muscle that causes the eyebrows to lower). The adult FACS manual collapses those two movements into one (AU4). Similarly, we will collapse AUs 3 and 4 into one category (3+4) to fit the current framework of FaceReader. We will do the same of AUs 26 and 27 (26+27). Finally, we will evaluate BFR AU classification results for AUs 1, 3+4, 5, 6, 7, 9, 10, 12, 15, 17, 18, 20, 25, 26+27.

To investigate the AU characteristic of holistic facial expressions we will run BFR's improved classifiers on 3 different datasets:

- The City Infant Faces Dataset [5]: A collection of 154 images of infants (age: 0 - 12 months) annotated as positive, negative and neutral facial expressions.
- The COPE (Classification of Pain Expressions) Database [6]. A collection of 204 facial images of neonates annotated as rest, cry, air stimulus, friction and pain expressions.
- Messinger Databrary Dataset [7]. A collection of 10 videos of 6-month olds in a Face-to-Face-Still-Face Paradigm coded for AUs 6+12 (smile faces), and 6+20 (cry faces)

We will evaluate AU activation for all the aforementioned images / videos in the datasets and provide a suggested AU coding scheme for coding each expression.

Results

Preliminary results are the following:

Using BFR, we managed to model 72% of the face in the City Infant FaceDataset. This created a dataset of model infant faces with 43 Positive facial expressions, 34 Negative Facial Expressions, and 32 Neutral Facial expressions.

Figure 1 shows box plots of the AU intensity for each of the 3 categories of expressions. Activations of of AUs 6, 12, and 25 are often associated with positive facial expressions. Any intensity of AUs 25 can also be associated to negative expressions. Any intensity of AU43 is also most often associated to negative facial expressions. Low to medium intensities of AUs 1, 4, 6, 7, 10 are also associated to negative facial expressions. Finally, low to medium intensities of AUs 1, 5, 17, 25, and 26 are also associated to neutral facial expressions. It should be noted that Baby FaceReader has reported F1 scores of at least 0.60 in most of the aforementioned AUs [8].
We also executed a BFR analysis on the COPE infant databases. Due to occlusions on most of the infant faces from hats, blankets and hands, as well as the very young age of the infants in the dataset (3-6 weeks), the current version of BFR could not model this dataset.
Finally, a preliminary analysis of the Messinger Dataset showed that BFR could model at least 70% the faces in video frames of the dataset. Detailed AU classification will be presented during MB2018.
Furthermore, improvement of BFR's AU classifiers is ongoing and final results will presented at MB2018, together with an updated AU classification of the CITY infant faces databases. We do expect to reach an average agreement score of at least 0.7. We will also train the BFR AAM on younger infant faces with some occlusions to account for the difficulties presented in the COPE infant dataset. We expect our updated AAM and classifiers to indicate expressions categories that may overlap (e.g. cry, negative, pain) and will thus suggest more objective classification categories based on AU activation.

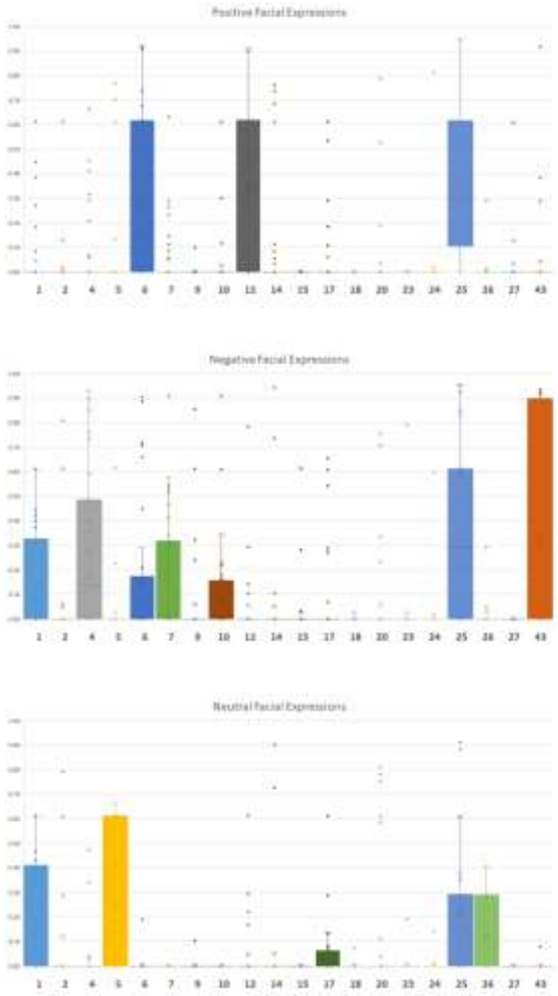


Figure 1: AU intensity box plots for Positive, Negative, and Neutral Facial Expressions of City Infant Faces Databases

Discussion
While coding infant expressions in terms of AUs may be objective measure to quantify facial expressions, creating categories of facial expressions objectively defined in terms of AUs will lead to faster coding procedure of facial

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

expressions that is easier to understand. Similarly, Ekman and Friesen have created EMFACS [10], a simplified version of the FACS manual [9] to quickly code emotional facial expressions for adults.

Acknowledgements
The research received funding from the Marie Sklodowska-Curie Actions of the Horizon 2020 Framework Program H2020-MSCA-ITN-2015. The program name is "Brainview" and the REA grant agreement n°A140857. We would also like to thank Rebecca Webb, Susan Brahnam, and Daniel Messinger for permission to use the City Infant, COPE, and Messinger Databrary Dataset respectively for such research.

References:

1. Noldus (2016). FaceReader: Tool for automatic analysis of facial expression: Version 7.0. Wageningen, the Netherlands: Noldus Information Technology B.V.
2. Viola, P., & Jones, M. J. (2004). Robust real-time face detection. International journal of computer vision, 57(2), 137-154. Chicago.
3. Cootes, T. F., & Taylor, C. J. (2004). Statistical models of appearance for computer vision. Imaging Science and Biomedical Engineering (University of Manchester, Manchester, UK).
4. Oster, H. (2016). Baby FACS: Facial Action Coding System for infants and young children. Unpublished monograph and coding manual. New York University.
5. Webb, R., Ayers, S., & Endress, A. (2017). The City Infant Faces Database: A validated set of infant facial expressions. Behavior research methods, 1-9.
6. Nanni, L., Lumini, A., & Brahnam, S. (2010). Local binary patterns variants as texture descriptors for medical image analysis. Artificial intelligence in medicine, 49(2), 117-125.
7. Messinger, D. (2014). Facial expressions in 6-month old infants and their parents in the still face paradigm and attachment at 15 months in the Strange Situation. Databrary. Retrieved February 7, 2018 from http://doi.org/10.17910/B7059D
8. Maroulis, A. Spink, A.J., Theuws, J.J.M., Oster, H., Buitelaar, J. (2017). Sweet or Sour: Validating Baby FaceReader to Analyse Infant Responses to Food. Poster to be presented in 12th Pangborn Sensory Science Symposium, 20-24 August 2017, Providence, Rhode Island
9. Ekman, P., Friesen, W. V., & Hager, J. C. (2002). Facial action coding system: The manual on CD-ROM. Instructor's Guide. Network Information Research Co, Salt Lake City.

Friesen, W. V., & Ekman, P. (1983). EMFACS-7: Emotional facial action coding system. Unpublished manuscript, University of California at San Francisco, 2(36), 1.

Automatic mutual gaze detection in face-to-face dyadic interaction videos

Cristina Palmero[1,2], Elsbeth A. van Dam[2], Sergio Escalera[1], Mike Kelia[2], Guido F. Lichtert[3], Lucas P.J.J. Noldus[2], Andrew J. Spink[2], Astrid van Wieringen[3]

1 Department of Mathematics and Informatics, University of Barcelona and Computer Vision Center, Barcelona, Spain. c.palmero.cantarino@gmail.com, sergio@maia.ub.es; 2 Noldus Information Technology, Wageningen, the Netherlands. {elsbeth.vandam, andrew.spink, mike.kelia, lucas.noldus}@noldus.nl ; 3 ExpORL, Department of Neurosciences, KU Leuven, Leuven, Belgium. {astrid.vanwieringen, guido.lichtert}@kuleuven.be

Introduction

Mutual gaze, also known as eye contact, is an important non-verbal cue in social signal processing and a promising measure of quality of interaction, particularly playing a crucial role in parent-child interactions [1]. Eye gaze behavior, as an indicator of human visual attention, is also analyzed to assess communication skills, such as in turn-taking or during group meetings, and to detect possible behavioral disorders such as autism and attention deficit hyperactivity disorder (ADHD) [2].

Much effort has been made for automatic mutual gaze detection using computer vision and machine learning techniques. The existing approaches can be mainly divided in two categories: head-mounted or remote camera-based systems. Head-mounted eye trackers are employed for automatic gaze detection in a variety of settings, as they can track the iris with high accuracy and model the eye using geometric-based methods after a user-specific calibration stage [3]. For instance, Ye et al. [4] combined wearable eye-tracking glasses and a First Person View (FPV) camera to determine whether the person in front of the camera was looking at the eye-tracker wearer. To do so, they combined the point of gaze estimated by the eye tracker and the extracted features of the captured FPV image and fitted them into a Random Forest. Egocentric vision was also used by Chong et al. [5], who used a Convolutional Neural Network (CNN) to predict eye contact while implicitly estimating head pose from FPV images.

Despite the high accuracy of head-mounted systems, especially those based on eye trackers, they are not suitable for assessing gaze behavior in naturalistic social contexts, where a non-obtrusive system is preferred. In that case, remote camera based systems offer a trade-off between accuracy and usability. With remote off-the-shelf cameras, eyes may not be visible, so a common alternative is to approximate gaze direction with head pose. For example, the method from Marín-Jiménez et al. [6], which detects if two people are looking at each other in TV videos. They use the scale of the subjects' heads to estimate the depth positioning of the actors, and combines it with the estimated head pose to derive their gaze volumes in 3D. If the volumes intersect, a *'Looking at each other'* event is detected. More recently, Massé et al. [7] proposed a Bayesian switching dynamic model to detect the visual focus of attention in multi-party interactions, using head pose as the observed variable, and the visual focus and gaze direction as the latent variables.

Other remote camera based methods are mainly focused on gaze estimation, such as the work from Zhang et al. [8], which proposes a full-face appearance-based CNN model to estimate the 3D line of gaze in human-computer interaction tasks. Contrary to geometric-based methods, appearance-based approaches learn the mapping between gaze directions and face appearances [3]. Therefore, they can be applied to low resolution images, but need a big training dataset to generalize among different subjects. Surprisingly, most of existing approaches only deal with human-computer interaction tasks, where users look straight to the screen or mobile phone. In naturalistic interactions, however, the range of gaze directions and head movements is not constrained. Therefore, generic, yet still accurate, remote third-person gaze estimation for in-the-wild, free-moving-head settings, remains an open issue.

In this work, we automatically detect mutual gaze instances in naturalistic face-to-face dyadic interactions using 2 calibrated RGB monocular cameras. In our setting, each camera is located in front of a participant. This way other behaviors, such as facial expressions, can be easily observed. We propose a novel system that combines geometric and appearance-based methods to detect *'Looking at other's face'* events. Such events can then be combined to find mutual gaze instances. The system is independent of subject's appearance and orientation, does not need user-specific calibration, and works in any camera configuration, as long as there is an overlapping visible zone between both cameras.

Our contribution is two-fold: first, we propose a new deep full-face appearance-based 3D person-independent gaze estimation model that achieves state-of-the-art performance in a free-moving-head setting; second, we present a end-to-end system, built on top of the gaze estimation model, that combines the geometry of the scene and the estimated features of each participant to perform the mutual gaze detection task.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Methodology

Generally, gazing to a specific target is achieved by a combination of eye and head movements. To determine if one person is looking at a specific target, we therefore need to know his head position and orientation in the scene, the eyeballs orientation and the target location. Following a recent line of gaze estimation work [8], we believe that full face images implicitly contain the necessary information to estimate gaze directions. We take the initial point of the line of gaze as the center point between the eyes. Assuming calibrated cameras and scene, our system proceeds in a frame-based fashion in two steps. First, the 3D line of gaze is estimated for each of the participants, and second, the line of gaze is used, along with the participants location in the 3D scene, to detect *'Looking at other's face'* events. An overview of the system is shown in Figure 1.

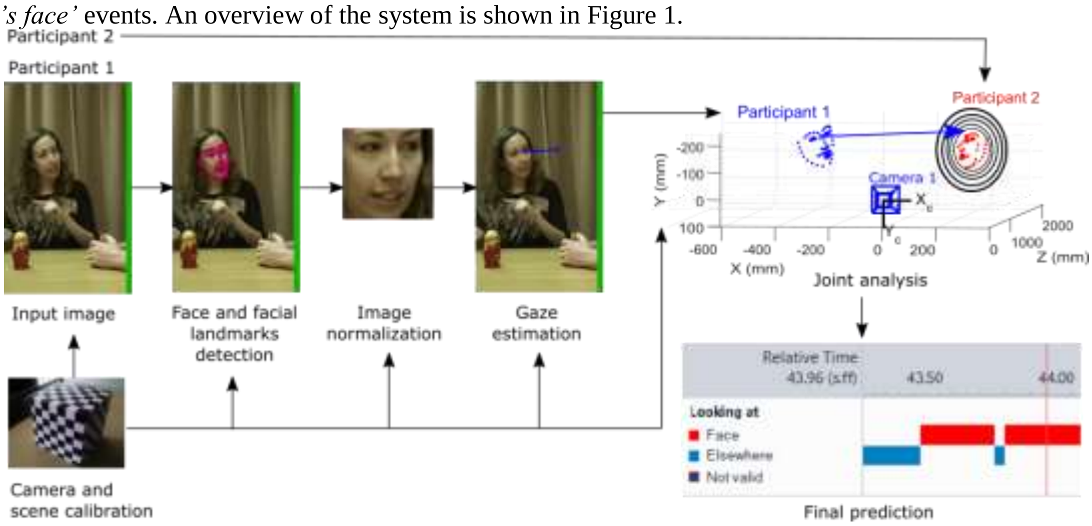


Figure 1. Overview of our approach. Assuming a calibrated 2-camera setting, our approach estimates the facial landmarks, head pose and line of gaze for each of the participants in the 3D scene. The estimated data is then combined along with the geometric 3D information of the scene to detect *'Looking at other's face'* or *'Looking elsewhere'* events.

For each participant, the system first performs face and 2D facial landmark detection using Constrained Local Neural Fields from Baltrušaitis et al. [9][10], which trains a model to perform the detection of 68 face landmarks together, along with separate sets of Point Distribution Models and patch expert models for eyes, lips and eyebrows. Using their 3D face model, we estimate the head position and rotation in the 3D scene using the PnP algorithm based on direct least-squares [11], and compute the 3D face landmarks with respect to the camera coordinate system.

Prior to the gaze estimation stage, a normalization in the 3D space is carried out by scaling and rotating the camera, similar to [8]. A virtual camera is created so that gaze estimation can be performed in a normalized space with fixed intrinsic camera parameters and reference point location. This normalization translates into the image space as a centered cropped face patch where head roll rotation has been removed – the horizontal axis of the camera is parallel to the horizontal axis of the head, and the camera looks at the midpoint of the face landmarks from a fixed distance. Likewise, facial landmarks and ground-truth 3D gaze vectors are also transformed to the normalized space. This normalization allows the gaze estimation model to be applied regardless of the original camera configuration and reduces the appearance variability. Finally, ground-truth gaze vectors are converted to 2D gaze angle representation.

The normalized face image is the input of the gaze estimation model (see Figure 2). It consists in a novel full-face, appearance-based deep CNN model that performs 3D gaze regression. It is based on the VGG-16 deep network architecture [12] which consists in 13 convolutional layers, 4 max pooling layers, and 3 fully connected (FC) layers. We added two dropout layers between the FCs as regularization. All layers use Rectified Linear Unit (ReLU) activations except the last fully connected layer, which performs a linear regression. The output is the normalized 2D gaze angle (horizontal, φ, and vertical, θ) in the camera coordinate system. As a loss function we use the average Euclidean distance between the predicted and ground-truth 3D gaze vectors. After inference, the predicted line of gaze is denormalized back to the original 3D space.

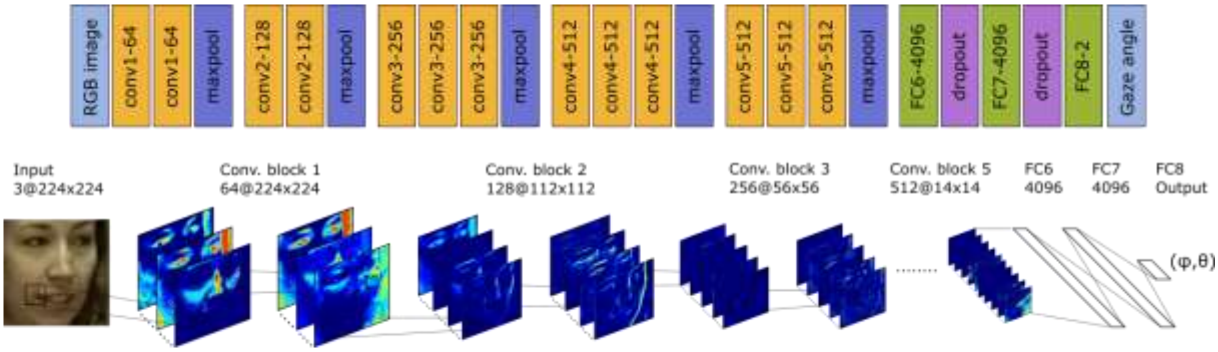


Figure 2. Overview of our gaze estimation model, based on the VGG-16 architecture.

Once the participants' location and their estimated lines of gaze are computed, the system combines this 3D information alongside the scene calibration data to perform the joint analysis in a geometric fashion. Generally, if the predicted 3D line of gaze of participant 1 intersects participant 2's head, a *'Looking at other's face'* event is detected, and *'Looking elsewhere'* otherwise. If no face is detected, or the estimated landmarks are not certain enough, the frame is classified as *'Not valid'*. This approach assumes accurate knowledge of the target head location and precise predicted gaze vectors. However, directly using the obtained estimated lines of gaze to detect the event of interest is challenging for variable face appearances and arbitrary camera-target configurations, especially when using more than one camera. To mitigate these possible errors in head location and gaze estimation, or inaccuracies in the calibration model, the system includes a threshold selector, in which the user can select the maximum distance from the participant 1's line of gaze to the participant 2's head allowed to classify the event as *'Looking at other's face'* with at least 50% certainty. To do so, the head of each of the participants is surrounded by a number of equidistant concentric ellipsoids, which can be interpreted as the certainty factor. The first ellipsoid is fitted to the head, and the user-defined thresholds determine the radii of the middle ellipsoid in its three axis (horizontal, vertical and depth). The gaze-head intersection point is computed performing ray-ellipsoid intersection, using binary search to find the smallest intersected ellipsoid.
The final predictions are represented by numbers:-1 for *'Not valid'* frames, 0 for *'Looking elsewhere'*, and from 1 to *N* for *'Looking at other's face'*, *N* being the number of used ellipsoids (1 for the lowest certainty, *N* the highest). Since the system uses a frame-based gaze estimation model, where one frame's estimation is independent of the previous one, the predictions can fluctuate among frames. A median filter can therefore be applied to smooth the predictions.

Data and Experimental Design

In this section, we discuss the performance of the person-independent gaze estimation model and perform a pilot test to validate the effectiveness of the proposed mutual gaze detection system.
There are few publicly available datasets devoted to in-the-wild 3D gaze estimation, as most of them focus on human-screen interaction with a limited head and gaze directions range. We therefore decided to use the EYEDIAP dataset [13], which consists of 3-minute videos of 16 young adult subjects (12 male, 4 female, mostly Caucasian, age not specified), looking at a specific target (moving floating ball or moving point on screen), with 2 different lighting conditions for some of the subjects. The videos are further divided in static and moving head pose for each of the subjects. Note that this dataset does not contain participants wearing glasses or other face accessories.
Our model, pre-trained with the VGG-Face dataset [12] (a dataset of faces consisting of 2622 identities), was fine-tuned with the floating target, static and moving head pose EYEDIAP videos of all subjects. We filtered out those frames where the face or landmarks were not properly detected by [9] or where the ground-truth gaze vectors were not correct, which resulted in a total of 88770 frames (see Figure 3). We fine-tuned all the layers except the first convolutional block.

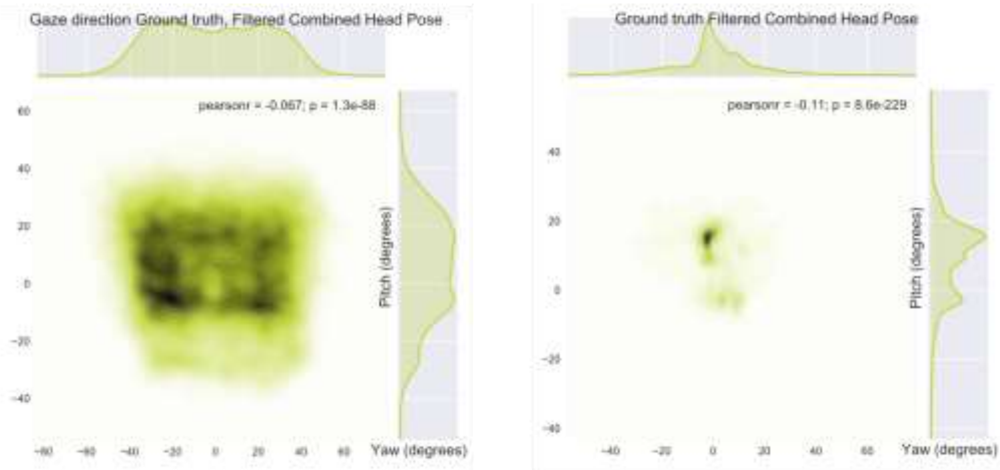


Figure 3. Distribution of ground-truth gaze directions and head poses in terms of yaw (horizontal axis) and pitch (vertical axis) angles, with respect to the camera coordinate system, for the filtered floating-target EYEDIAP subset.

To validate the accuracy of the person-independent gaze model, we performed leave-four-subjects-out cross-validation. We trained the model using the ADAM optimizer [14] with an initial learning rate of 0.0001, dropout of 0.3, and batch size of 64 frames. The number of training epochs, which is the number of passes of the full training set through the network, was set to 10. To extend the filtered dataset we applied on-the-fly data augmentation to the training set, with the following random transformations: horizontal and vertical shifts, zoom, illumination changes, horizontal flip, and additive Gaussian noise. The final model used in the mutual gaze detection system was trained using all the subjects.

Furthermore, to assess the performance of the mutual gaze detection system, we conducted a pilot validation at KOCA (Koninklijk Orthopedagogisch Centrum Antwerpen) in Antwerp, Belgium. KOCA is a home guidance service that supports parents in raising their deaf child and has a standardized laboratory living room to perform video recordings of face-to-face interaction sessions [1]. We recorded a 4-minute video with two participants, one female inside the age range of the training dataset (*Participant 1*), and a male outside the age range (*Participant 2*). *Participant 1* wore glasses the last 2 minutes of the interaction. The interaction was recorded using two calibrated monocular RGB cameras placed about 2 meters in front of each participant, who were seated around a table, placed at 90º to one another. The video signals from both cameras were merged in the control room to one split-screen image by a mixer (see Figure 4). The final video was recorded at 25 frames per second, with a resolution of 720x576 pixels. The interaction session was manually annotated with frame-level accuracy using The Observer XT 12.5 [15] with the following behaviors: *'Looking at other's face'*, *'Looking at object'*, *'Looking at hand'*, *'Looking elsewhere'*. We set the head thresholds to X = 20 cm, Y = 40 cm, and Z = 20 cm for both participants, and the window filter size to 7 frames.

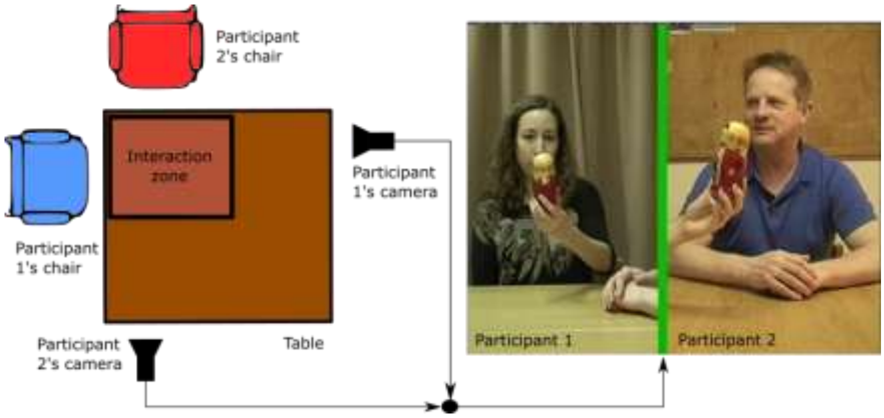


Figure 4. Setup of the pilot validation at KOCA. The interaction was recorded using 2 calibrated monocular RGB cameras placed about 2 meters in front of each participant, who were seated around a table, placed at 90º to one another. The video signals from both cameras were merged to one split-screen image by a mixer.

Results

The trained gaze estimation model achieved an average error of 5.5 degrees after 10 epochs in the leave-four-subjects-out cross validation experiment (see Figure 5). It is difficult to compare this result to the current state-of-the-art person-independent, calibration-free, gaze estimation methods, since they are generally validated with the screen-target EYEDIAP subset. Notwithstanding, the lowest obtained error for that subset has been 6 degrees [8]. Taking into account that the screen-target subset has less variability in head and eye movements than the floating-target subset, we demonstrate that our model is competitive against state-of-the-art approaches, obtaining better results in a less constrained setting.

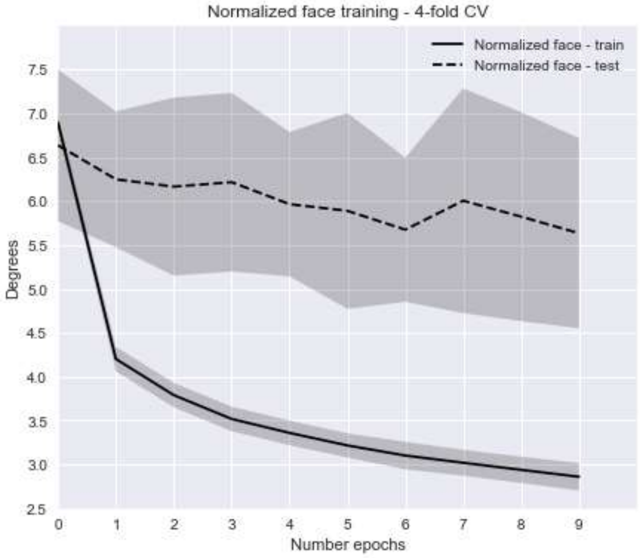


Figure 5. Epoch progression for leave-four-subjects-out cross validation, aggregating the results of each cross-validation run.

Figure 6 shows some examples of qualitative results of the mutual gaze detection system for both participants in the second evaluation scenario. For *Participant 1*, in general, facial landmarks and the line of gaze are visibly well estimated. Problems usually arise with near-to-profile head positions, as face landmarks cannot be properly located and the gaze model has not been trained for that angle range. We also see a clear difference when wearing glasses. The gaze predictions are smoother when the participant is not wearing them, whereas they are less precise and have greater fluctuations with glasses, especially when the glasses frame hides more part of the eyes region. For *Participant 2* we observe a particular effect. While the X component of the gaze vector is usually correct, the Y component is looking upwards most of the time, highly likely due to the difference in appearance of the eye region of the participant with respect to the gaze model training subjects. This can also be observed in some frames where the eyes landmarks are detected below the eyes.

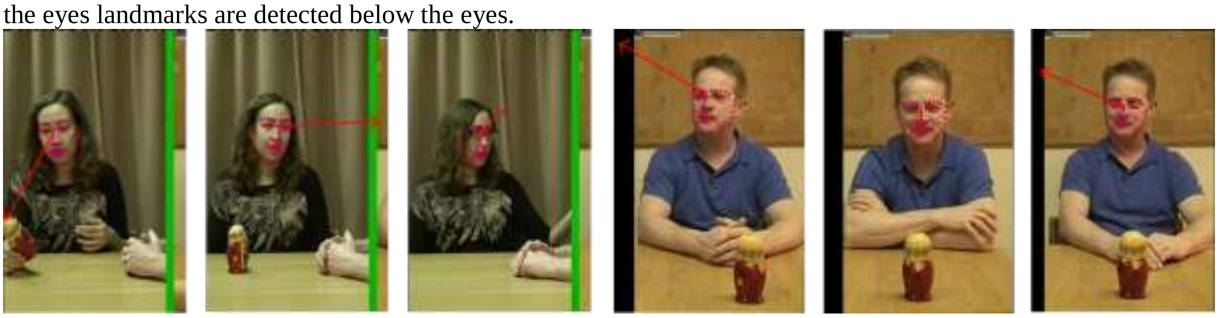


Figure 6. Examples of qualitative results for *Participant 1* (left) and *Participant 2* (right): (1) looking at an object; (2) looking at the other's face; (3) looking at an object, but landmarks and gaze are not correctly estimated, either because of the glasses or the more near-to-profile position; (4) looking at the other's face; (5) looking at the object; (6) looking at the other's face, but eye landmarks are not correctly estimated.

Finally, the event predictions were exported to The Observer XT to compare them to the annotated ground truth. The certainty values were converted to behavior modifiers. To make both coding schemes comparable, '*Looking at object'* and *'Looking at hand'* behaviors used in the manual annotation were treated as *'Looking elsewhere'*. We observed that, even though the Y component of *Participant 2*'s line of gaze was not correctly estimated, the event predictions were mostly right thanks to the head threshold parameter, which mitigates this deviation. Reliability analysis was carried out omitting the modifiers. A percentage of agreement of 70.2% was found using the duration-sequence method with a Kappa value of 0.62, which can be regarded as a good level of agreement. Manual examination of the List of Comparisons generated by the software revealed that the majority of the occasions when the predictions were not the same as the ground truth was when the subjects very briefly looked elsewhere, and when one of the participants was looking at an object located in between of them.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Discussion and Conclusions

We proposed a system, built on top of a novel state-of-the-art appearance-based gaze estimation model, that combines the 3D information of the scene and the participants' location and estimated line of gaze to detect *'Looking at other's face'* events in a calibrated 2-camera setting. The system can be used by observational researchers to automatically detect where a person is looking at and, more specifically, if there is mutual gaze in a dyadic face-to-face interaction.

The gaze estimation model has been validated in the same age range of the training set and with no face accessories, obtaining state-of-the-art results. The pilot validation has also shown promising results, even though is yet unclear how good the system will work with other age ranges, especially children and the elderly, as they have different face features. This is a crucial aspect to take into account, as the model is based on appearance. Furthermore, the model was not trained with people wearing glasses or special accessories. In the future, the model will be further trained with bigger age range and appearance variability in order to ensure it works in other settings, or create specific models for each age range. Note that the performance of the system also depends on the accuracy of the detected face landmarks, so more accurate face landmark estimation methods should be studied.

The joint analysis approach has also further room for improvement, since the final accuracy currently depends on the distance between participants. For instance, if the two subject locations are fixed during the interaction, the system could extract CNN features for known *'Looking at other's face'* events in an active learning scenario.

Acknowledgements

We would like to thank Eric Coenen, KOCA, for letting us use their recording room and equipment to perform tests and record the video sample. This work has been supported by the FP7 people program (Marie Curie Actions), REA grant agreement no FP7-607139 (iCARE — improving Children Auditory REhabilitation), and partially supported by the Spanish project TIN2016-74946-P (MINECO/FEDER, UE) and CERCA Programme / Generalitat de Catalunya. We gratefully acknowledge the support of NVIDIA Corporation for the donation of the Titan Xp GPU used for this research. Portions of the research in this paper used the EYEDIAP dataset made available by the Idiap Research Institute, Martigny, Switzerland.

References

1. Lichtert, G.F. (2016). Observing and classifying mutual gaze patterns. Measuring Behavior 2016.
2. Leekam, S. R., Hunnisett, E., Moore, C. (1998). Targets and cues: Gaze-following in children with autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines* ***39*****(7)**, 951-962.
3. Hansen, D. W., Ji, Q. (2010). In the eye of the beholder: A survey of models for eyes and gaze. *IEEE transactions on pattern analysis and machine intelligence* **32(3),** 478-500.
4. Ye, Z., Li, Y., Fathi, A., Han, Y., Rozga, A., Abowd, G. D., Rehg, J. M. (2012). Detecting eye contact using wearable eye-tracking glasses. In *Proceedings of the 2012 ACM conference on ubiquitous computing*, 699-704.
5. Chong, E., Chanda, K., Ye, Z., Southerland, A., Ruiz, N., Jones, R. M., ..., Rehg, J. M. (2017). Detecting Gaze Towards Eyes in Natural Social Interactions and Its Use in Child Assessment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **1(3)**, 43.
6. Marín-Jiménez, M. J., Zisserman, A., Eichner, M., Ferrari, V. (2014). Detecting people looking at each other in videos. *International Journal of Computer Vision* **106(3),** 282-296.
7. Massé, B., Ba, S., Horaud, R. (2017). Tracking Gaze and Visual Focus of Attention of People Involved in Social Interaction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
8. Zhang, X., Sugano, Y., Fritz, M., Bulling, A. (2017). It's written all over your face: Full-face appearance-based gaze estimation. *Proc. IEEE International Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*.
9. Baltrušaitis, T., Robinson, P., Morency, L. P. (2016). Openface: an open source facial behavior analysis toolkit. *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, 1-10.
10. Baltrušaitis, T., Robinson, P., Morency, L. P. (2013). Constrained local neural fields for robust facial landmark detection in the wild. *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*, 354-361.
11. Hesch, J. A., Roumeliotis, S. I. (2011). A direct least-squares (DLS) method for PnP. *Computer Vision (ICCV), 2011 IEEE International Conference on*, 383-390.
12. Parkhi, O. M., Vedaldi, A., Zisserman, A. (2015). Deep Face Recognition. *BMVC* **1(3)**, 6.
13. Funes Mora, K. A., Monay, F., Odobez, J. M. (2014). Eyediap: A database for the development and evaluation of gaze estimation algorithms from rgb and rgb-d cameras. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 255-258.
14. Kingma, D. P., Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

15. The Observer XT, http://www.noldus.com/observer. Accessed 7 February 2018.

What should I annotate? An automatic tool for finding video segments for EquiFACS annotation

M. Rashid[1], S. Broome[2], P. H. Andersen[3], K. B. Gleerup[4], Y. J. Lee[1]

1 Department of Computer Science, University of California at Davis, Davis, US. mhnrashid@ucdavis.edu; 2 Department of Robotics, Perception and Learning, KTH Royal Institute of Technology, Stockholm, Sweden; 3 Department of Clinical Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden; 4 Department of Large Animal Sciences, Copenhagen University, Copenhagen, Denmark

Facial Action Coding System (FACS) is a system to define and name facial movements by their appearance on the face. Originally developed for humans, various mammal specific FACS have been developed (1-3), including Equine FACS (4) for describing horse facial action movements.

FACS coding of videos is a laborious but necessary process for the scientifically robust description and identification of facial expressions in mammals. Additionally, datasets with FACS labels are a necessary prerequisite for training machine learning methods for the automatic identification of action units as well as expressions, and have been successfully developed for human beings (6-8). While it is relatively easy to ensure that each frame in a human video dataset is easy to annotate with FACS coding - human subjects can be asked to face the camera during filming - the same is not easy, or in most settings, possible, to ensure for horses.

Therefore, EquiFACS annotators face an added challenge when annotating: finding video segments that are most suitable for annotation. The process of finding these segments is cumbersome, and presents extra time overhead to actual FACS coding and verification. We present a solution to this challenge by introducing a tool that can automatically find and mark segments in videos where horse faces are visible and in a pose ideal for annotation. The tool can reduce time spent on finding annotatable segments from half a day to 10 minutes for a 24 hour long video after frame extraction. Furthermore, it has a high recall rate of 94.92%.

In horses, video segments with the full horse face visible from a side or 45º angle are ideal for EquiFACS annotation since the movement of facial muscles is clearly observable. Experienced annotators can identify such video segments while viewing a video at four times its real-time speed, with additional time spent to verify candidate segments. Overall, the process of finding video segments suitable for annotation can take an annotator the same amount of time as a quarter to half the length of a recorded video. This time overhead limits the number and length of footage that is collected as well as annotated.

Our proposed annotation tool can suggest candidate segments automatically. The software takes as input a video, and outputs a continuous confidence value between 0 and 1 for each time step in the video; 0 indicates that the time step is not usable at all, and 1 indicates high confidence in its usability. This confidence value allows users to prioritize their annotation efforts and improve productivity by going from most usable to least usable video segments. In addition, the tool also finds and localizes the horse head in each frame. The tool is extremely fast, and takes less than 13 seconds on a 25 minute video after frame extraction. See Figure 1 for an example output.

The tool works by first extracting a video frame every *n* seconds and running a deep convolutional neural network (CNN) that is trained to find front and side view horse faces on each extracted frame. The frequency of frame extraction, or *n*, can be set by the user and is set to one frame every 5 seconds for experiments in this paper. The confidence value of detections in each frame are recorded. Finally, detections with confidence less than 0.2 are ignored, averaged for every minute window, and plotted against video time.

The backbone of our tool is YOLO v.2 (9). YOLO is a CNN-based real-time object detection system (10). We adapt YOLO to detect two types of 'objects' - horse faces in side view, and horse faces in front view. Figure 2 shows an example of each of these categories. Training the detector to distinguish and identify these two types of horse faces correctly requires annotated training data: images of horses with front and side view faces marked. We used the dataset from (11) that had horse face bounding box annotation. We further marked each annotated horse face as either side view, front view, or neither if the face was mostly self-occluded. In addition we manually annotated and added frames from two twenty minute surveillance videos of horses. The addition of these frames was important to correct the domain difference between the dataset from (11), which comprised of images collected from the internet, and surveillance footage used by our collaborators for EquiFACS annotation. The addition of video frames improved performance from 87.01% recall to 94.92% recall.

While the tool has been developed for EquiFACS annotations, it can be easily used by researchers in other areas to find and label parts of videos most suitable for further analysis. It is currently freely available for EquiFACS annotators on our GitHub page (https://github.com/menorashid/darknet). Future plans include integrating the tool with previous work (11) to find and zoom in on parts of the face that are important for the identification of action units (eg. eyes) so that users can more easily observe and annotate action units.

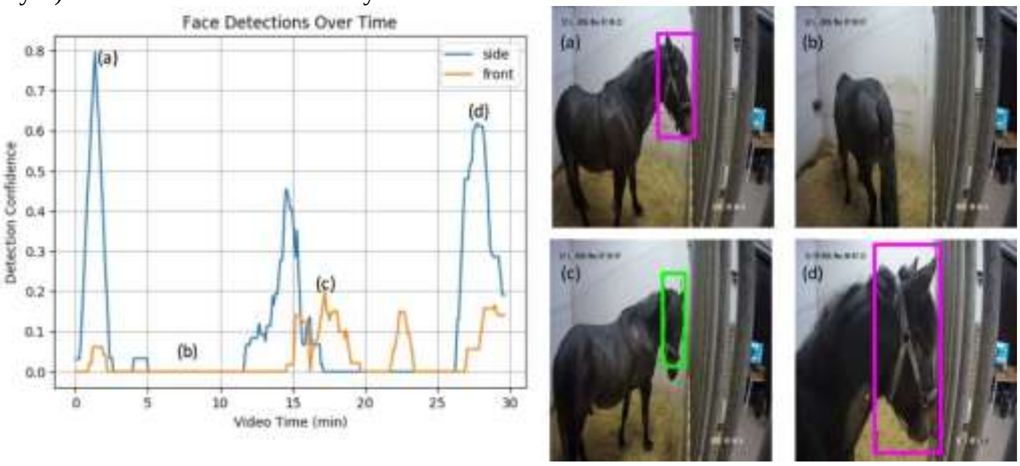


Figure 2 The output of our tool over a 30 minute video with sample frames with no, side and front view horse head detections.

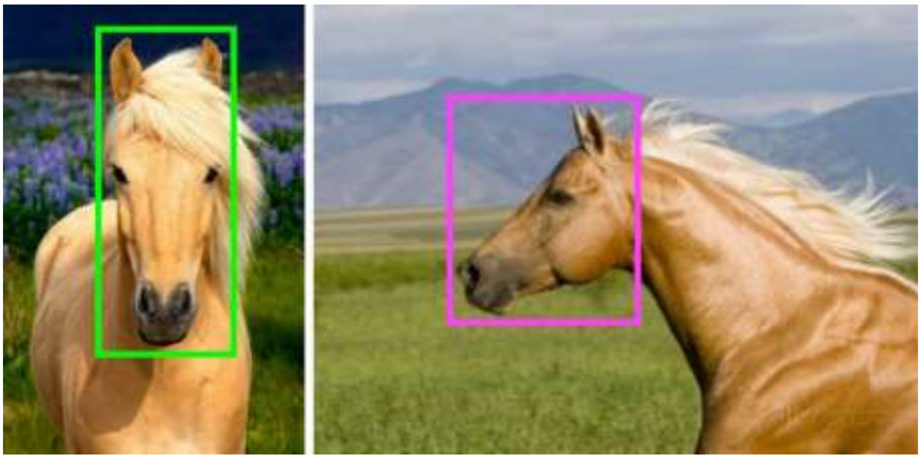


Figure 3 Examples of front and side view images used for training the horse head detector.

Acknowledgements

This research was supported in part by a Hellman Fellowship to Yong Jae Lee.

References

[1] Caeiro, Cátia C., et al. "OrangFACS: A Muscle-based facial movement coding system for orangutans (Pongo spp.)." International Journal of Primatology 34.1 (2013): 115-129.
[2] Caeiro, C. C., A. M. Burrows, and B. M. Waller. "Development and application of CatFACS: Are human cat adopters influenced by cat facial expressions?" Applied Animal Behaviour Science 189 (2017): 66-78.
[3] Langford, Dale J., et al. "Coding of facial expressions of pain in the laboratory mouse." Nature methods 7.6 (2010): 447.
[4] Wathan, Jen, et al. "EquiFACS: the equine facial action coding system." PLoS one 10.8 (2015): e0131738.
[6] Lucey, Patrick, et al. "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression." Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on. IEEE, 2010.
[7] Mavadati, S. Mohammad, et al. "DISFA: A spontaneous facial action intensity database." IEEE Transactions on Affective Computing 4.2 (2013): 151-160.
[8] Zhang, Xing, et al. "BP4D-spontaneous: a high-resolution spontaneous 3D dynamic facial expression database." Image and Vision Computing 32.10 (2014): 692-706.
[9] Redmon, Joseph, and Ali Farhadi. "YOLO9000: Better, Faster, Stronger." The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017.

[10] Everingham, Mark, et al. "The PASCAL visual object classes (VOC) Challenge." International Journal of Computer Vision 88.2 (2010): 303-338.
[11] Rashid, Maheen, Xiuye Gu, and Yong Jae Lee. "Interspecies knowledge transfer for facial keypoint detection." The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017.

Taking Research to the Community: An Initial Protocol for the Experimental Measurement of Emotional Expression in People with Dementia

Gemma Stringer[1, 2], Laura Brown[2] and Moi Hoon Yap[1]

[1] Manchester Metropolitan University, M1 5GD, UK. M.Yap@mmu.ac.uk
[2] University of Manchester, UK.

Increasing evidence shows that people with dementia display impairments in the recognition and production of facial expressions of emotion. There have been a number of studies regarding the recognition of facial expressions in people with dementia, findings show that people with dementia display deficits in recognising facial expression of emotions. However, these results are inconsistent and conflicting. Research concerning production of facial expressions is limited, employing a range of data collection methods, such as separate stimulus display and recording equipment, resulting in limited accuracy and consistency and making it difficult to compare results across studies. In addition, research is often constrained to laboratory-based experiments and research in the community is restricted to nursing homes, limiting the types of participants involved.

This study combines methods from experimental psychology and computer science in order to investigate the ability of people with dementia to recognise emotional facial expression and to accurately characterise facial expressions of emotion in people with dementia in community settings. Three groups of participants were recruited for this study: (1) people with dementia (n=17), (2) cognitively healthy controls over the age of 65 (n=22), (3) cognitively healthy controls under the age of 35 (n=20). All participants provided written informed consent prior to taking part in the study, which was approved by the Manchester Metropolitan University Research Ethics Committee. The present study is the first to simultaneously assess facial expression production in relation to cognitively healthy controls and people with dementia, and thus these data offer a unique point of comparison. For the expression recognition task, participants viewed short video clips of the six basic emotions (happiness, sadness, anger, disgust, fear and surprise) performed by a trained Facial Action Coding System (FACS) [1] coder. These emotions represent universal innate human responses that are independent of culture and learning. Participants then viewed the same expressions for a second time, after each expression was displayed the participant was presented with a choice of six different emotions on the screen and were asked to respond by pressing the button corresponding to the emotion they thought was being expressed in each case.

The results showed people with dementia were less able to recognise five out of six emotional expressions: happy, sad, fear, anger and surprise compared to healthy controls under and over the age of 65. There was no difference between healthy controls and people with dementia in the recognition of disgust. Findings from previous research support but also contrast with these results. Support comes from [2] who found a deficit in the recognition of happy, sad, anger, fear and surprise but a preservation of the recognition of disgust, suggesting this might be attributable to the relative sparing of the basal ganglia in AD. The findings are also somewhat supporting by Maki *et al.* [3] who found deficits in people with AD in the recognition of sadness and anger. However, they also found that recognition of happy facial expressions was relatively preserved in AD patients, which contrasts with the findings of the present study.

References

1. P. Ekman and W. V. Friesen. (1978) Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologies Press.
2. Henry, J. D., Ruffman, T., McDonald, S., O'Leary, M. A., Phillips, L. H., Brodaty, H., & Rendell, P. G. (2008). Recognition of disgust is selectively preserved in Alzheimer's disease. [Research Support, Non-U S

Gov't]. Neuropsychologia, 46(5), 1363-1370.
3. Maki, Y., Yoshida, H., Yamaguchi, T., & Yamaguchi, H. (2013). Relative preservation of the recognition of positive facial expression "happiness" in Alzheimer disease. [Research Support, Non-U S Gov't]. Int Psychogeriatr, 25(1), 105-110.

Measuring Zoo Animal Welfare

Understanding how wild animals housed in zoos and aquariums are faring is one of the responsibilities of zoo professionals today. Are animals having a good quality of life? Do they seem happy, engaged, curious, or are they fearful or anxious? Do they have extensive behavioural repertoires usually found for this species, or does it seem to be a reduced range? And how does this vary between individuals? Measuring different aspects can aid to gauge what an animal might be experiencing. Considering behavioural categories as well as how, reflecting not only physical behaviour recorded in classic ethograms, but also the expressive qualities emerging from the many different kinds of physical behaviours. Research is necessary to make evidence-based decisions regarding animal welfare efforts, with the aim to promote, as much as possible, positive animal welfare. This symposium brings together zoo animal welfare experts offering their perspectives on measuring wild animal welfare.

Sabrina Brando, World Association of Zoos and Aquariums, Switzerland

A Proactive Monitoring Programme for Ageing Zoo Animals

S. Chapman[1], P.Dobbs[2]

1. Chapman Zoo Consultancy, chapmanzooconsultancy@gmail.com; 2. Twycross Zoo, Atherstone, United Kingdom, phillipa.dobbs@twycrosszoo.org

Many species kept in zoological collections live longer than they would do naturally in the wild (4). These older animals can be described as being elderly, ageing or geriatric and will present with a variety of clinical conditions which require treatment and management (1,2,3). Clinical conditions such as osteoarthritis, spondylitis and dental disease are painful and debilitating which may not become apparent to keepers/caregivers until significantly progressed. This presentation will outline the model of a proactive ageing zoo animal care programme which can be applied across all taxa.

The programme was designed to include monitoring of both behaviour and clinical health. Animals were classified as 'ageing' by being 75% through their natural longevity. Longevity data from the wild was used but where this information was not available, data pertaining to the longevity in captivity was used. Once identified, the individual's behavioural, reproductive and clinical histories were reviewed. Two methods of assessment were used: 'hands off' and 'hands on' methods to create a database of information to feed into an overall welfare assessment. Initially, animal's behaviour was discussed with the keepers and any specific health concerns identified e.g. stiff movement, poor appetite. Training needs were identified e.g. crate training, weigh scales training, injection training and plans put in place.

Methods of 'hands off' included behavioural observations; the use of positive reinforcement training; collection of biological samples; weight and body condition scoring and thermal imaging. Behavioural observations included keeper daily records and veterinary observations. Basic behavioural ethograms were used to assess various aspects including social interactions, enclosure use, gait analysis and response to medication. These ethograms were repeated after periods of time to monitor change or following a course of medication. Video footage was also used especially for gait analysis and night time data. Goals of training included the ability to weigh individuals, closely inspect or examine the animal and the collection of biological samples e.g. blood, urine. Training was also used to ensure individual dosing of medications. Diagnostic tests were performed through training and clinical follow up achieved through training in many cases. Visual body condition scoring was used in conjunction with body weights to monitor health. Thermal imaging was a useful tool for identifying areas of inflammation, monitoring progress of conditions and response to treatment.

A 'hands on' health assessment included a weight and body condition score; clinical examination including detailed dental examination; collection of biological samples e.g. blood, urine, faeces; diagnostic tests e.g. radiography, ultrasonography and treatment as appropriate e.g. dental extractions.

Following health assessments, the clinical findings and results of tests were discussed with the appropriate animal managers and treatment options outlined and implemented. This information was fed into an overall welfare assessment which was carried out in conjunction with veterinary and animal staff. At this point, time lines for clinical follow up were agreed and criteria for euthanasia were developed in some cases. Training needs were also discussed as were any enclosure modifications or dietary alterations which could be beneficial for the animal. Animals were not isolated for treatment and always managed within their social groups.

Creation of clear, written criteria for euthanasia was carried out in certain cases where appropriate. The euthanasia procedure was discussed in detail with regard to the well-being of the animal, management of the other animals and also the keepers. The practicality of the post mortem procedure was included in the planning. A thorough post mortem examination including detailed histopathological assessment of tissue samples allowed identification of the cause of clinical findings and the presence of sub-clinical diseases.

Zoo animals can be challenging from a clinical perspective as they often hide symptoms of disease until the disease process has progressed significantly. This model of ageing animal care proved that clinical conditions were found following proactive health assessment which had not been apparent to keepers prior to the examination. Treatment and management of these conditions lead to improvement in individual welfare and cases were followed up appropriately to ensure ongoing monitoring. Examples of cases will be given during this presentation. The use of positive reinforcement training was a key element to the programme. Keeper attitudes changed over time to embrace the programme. Initially sceptical, the team became supportive after seeing the benefits to the animals under their care. Involving the keepers in health assessments, decision making around euthanasia and post mortem examinations also improved teamwork and understanding. Overall, this approach led to the implementation of a positive and proactive programme for assessing the welfare of ageing animals on a regular basis.

Ethical statement

Individual animal assessments were either observation or distance assessment and when intervention was required, this was discussed with the senior animal management team and planned appropriately to reduce risks. The programme was reviewed and supported by the zoo's ethics committees.

References

1. Baine, K. (2012). Management of the geriatric psittacine patient. *Journal of Exotic Pet Medicine*, **21,** 140-148.
2. Longley, L. (2011). A review of ageing studies in captive felids. *International Zoo Yearbook*, **45,** 91-98.
3. Paré, J. A. & Lentini, A. M. 2010. Reptile geriatrics. *Veterinary Clinics of North America: Exotic Animal Practice*, **13,** 15-25.
4. Tidière, M., Gaillard, J.-M., Berger, V., Müller, D. W., Lackey, L. B., Gimenez, O., Clauss, M. & Lemaître, J.-F. 2016. Comparative analyses of longevity and senescence reveal variable survival benefits of living in zoos across mammals. *Scientific reports*, 6.

Competence and Agency as Novel Measures of Zoo Animal Welfare

F.E. Clark

Department of Field Conservation & Science, Bristol Zoological Society, Bristol Zoo Gardens, United Kingdom. fclark@bristolzoo.org.uk

An environmental 'challenge' refers to a problem involving the acquisition of resources, mates, dealing with adverse weather conditions or anything else which may threaten an animal's health and survival. We tend to think that wild animals face the most salient and potentially life-threatening challenges; conversely captive animals face few intrinsic challenges and these are "…problems to avoid rather than as opportunities to enjoy" from our own human experience [1]. However, there is a growing body of evidence that captive animals are motivated to seek and overcome environmental challenges, even when resources are available with little physical or cognitive effort, or there is no real threat to survival [2, 3]. Indeed, successfully dealing with challenges can enhance welfare. In response to this shift in view, the concept of 'cognitive enrichment' is gaining popularity in zoos; this subset of environmental enrichment aims to focus on the stimulation of evolved cognitive skills to enhance welfare [3]. However, cognitive enrichment has not been easy to accomplish in zoos thus far. The key reasons are as follows:

- Fundamental limitations of zoo research. Compared to research in labs and farms, there are considerably smaller sample sizes and relatively uncontrolled experimental conditions.
- Design and implementation of relevant and appropriate cognitive challenges for an animal. 'Relevant' in terms of their evolved cognitive skills (as well as anatomy, morphology), and 'appropriate' in terms of their individual abilities and motivations. There is a lack of fundamental knowledge of the cognitive skills of many exotic taxa.
- It is difficult to assess whether the cognitive skills of an animal are indeed being challenged to a high level (i.e. is the enrichment 'cognitive' or perhaps more physical, social or sensory?), with the caveat that all enrichment overlaps categories to a certain extent.
- Humans can self-report feelings such as satisfaction or frustration in response to challenges; conversely animals cannot tell us how they feel in comparable situations. We therefore rely on making inferences from their behavioural responses.

In response to these difficulties, I present a new framework for assessing zoo animal welfare which hinges on the central assumption that cognitive challenges can enhance welfare. I propose that 'behavioural competence' and 'agency' can be used as measures of welfare in response to cognitive challenge. In other words, these behavioural responses can be used to assess the effectiveness of cognitive enrichment in a zoo setting. The measurement of competence and agency in zoo animals is a new prospect, but could traverse may of the key issues with zoo-based cognitive enrichment outlined above.

I will discuss research undertaken on socially-housed chimpanzees (*Pan troglodyes*), bottlenose dolphins (*Tursiops truncatus*) and ring tailed lemurs (*Lemur catta*) in response to cognitively challenging stimuli. In experiments 1 and 2, chimpanzees and dolphins were provided with novel, 3D maze navigation tasks containing both food and non-food rewards for direct comparison. In experiment 3, lemurs were provided with novel stimuli on A-frames that were not intended to be a task *per se* but rather exposure to construction materials for a future cognitive task.

Competence was measured in relation to the subject's direct engagement with the cognitive stimuli including latencies and durations of usage, persistence and re-visits, and the nature and diversity of inspective problem-solving strategies used. Agency was measured as the subject's engagement with the wider environment, including inquisitive exploration of the wider environment, and lone and social play. The logic was to capture two important responses to challenge: response to the challenge itself as well as subsequent responses to the wider environment. Traditional zoo enrichment studies have failed on both counts; they tend to overlook the intricacies of enrichment use and the wider environmental effects of enrichment exposure in favour of looking at gross changes to the activity budget.

Assessments of competence and agency may positively impact zoo cognitive enrichment studies in the following ways:

- A lack of experimental control is embraced rather than criticized; the natural, dynamic home environment forms a relevant backdrop to cognitive challenge and highly controlled conditions have little relevance to zoo animals.
- The relevance and appropriateness of cognitive challenges are appraised by comparing behavioural responses within and between subjects in a social group and asking, on balance, is the task effective for this group?

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

- Competence is monitored over time to assess how cognitive skills are challenged and develop over time.
- Competence and agency may be particularly useful in large, naturalistic exhibits and/or when subjects do not perform 'classic' markers of poor welfare such as abnormal or stereotypical behaviours.

Ethical statement

All experiments were approved by the ethical review boards of the Royal Veterinary College, Zoological Society of London and Bristol Zoological Society. Research was observational and participation by the study subjects was entirely voluntary.

References

1. Špinka, M., Wemelsfelder, F. (2011). Environmental challenge and animal agency. In Animal welfare CABI International, Wallingford, 27-43.

2. Meehan, C.L., Mench, J.A. (2007). The challenge of challenge: can problem solving opportunities enhance animal welfare? Applied Animal Behaviour Science **102**(3), 246-261.

3. Clark, F.E. (2017). Cognitive enrichment and welfare: Current approaches and future directions. Animal Behavior and Cognition **4**(1), 52-71.

Abnormal Repetitive Behaviours as Indicators of Zoo Animal Welfare: Why, What, When and How to Measure

M. Díez-León[1]

1Department of Pathobiology and Population Sciences, Royal Veterinary College, University of London, London, United Kingdom. mdiezleon@rvc.ac.uk

Many zoo animals of different taxa perform abnormal repetitive behaviours (ARBs) - repetitive, invariant behaviours in either pattern or goal that either appear to serve no function, or are maladaptive, self-injurious or inappropriate [1] - are prevalent among zoo animals. ARBs are widely used as a welfare indicator across species. This is because, when comparing different settings, they typically occur in settings that are conducive to decreased welfare as assessed by other welfare indicators [2]. However, their validity as welfare indicators *within* the same setting is far less clear [2], with recent evidence suggesting non-stereotypic animals' welfare might be equal or worse than that of stereotypic individuals [e.g. 3, 4]. Because conspicuous, these behaviours appear deceptively 'simple' to record. However, in practice, defining what constitutes a bout of ARB can be problematic: what makes it abnormal? how repetitive or invariant should it be?. Furthermore, different measurable aspects of the behaviour (e.g. frequency, prevalence, severity, timing, etc.), and even different forms of it [4] might relate to current welfare state differently, and should therefore be taken into account when recording zoo ARBs as a means to assess welfare. Two additional practical aspects need to be considered. First, as with all behavioural recording, is whether to use live observations or video cameras to record ARBs. While live observations provide context and data is (typically) quickly available for analysis, observer influence on behaviour can be an issue. Live observations are also typically not validated against video recordings. In contrast, while video recording avoids observer effects on behaviour and can yield important measurable aspects of ARBs, it tends to be more time consuming to analyse and prone to technical glitches that can lead to subsequent loss of data, not to mention the great variety of video recording equipment that exists in the market (e.g. cctv systems, go-pro cameras, camera traps). Finally, a common issue when recording ARBs (and other behaviours) of zoo animals is the reliance on several scorers (e.g. keepers, students, volunteers) to obtain data, which makes testing for inter-observer reliability and avoiding unconscious biases a fundamental aspect to ensure data quality. This contribution will therefore review fundamental and practical aspects of recording ARB for use as zoo animal welfare indicators, as well as provide best practice guidelines for recording ARBs and discuss some ways to tackle them in practice.

Ethics statement

Any data that referred to during this presentation would have been approved by the relevant institutional ethical boards.

References

1. Garner, J. (2006). Perseveration and stereotypy - systems-level insights from clinical psychology. In: Stereotypic Animal Behaviour - Fundamentals and Applications to Welfare (Mason, G. and Rushen, J. eds. 2nd ed.), 121-152.

2. Mason, G., Latham, N. (2004). Can't stop, won't stop: is stereotypy a reliable welfare indicator?. Animal Welfare **13**, S57-69.

3. Ijichi, C., Collins, L., Elwood, R. (2013). Evidence for the role of personality in stereotypy predisposition. Animal Behaviour **85**, 1145-1151.

4. Díez-León, M., Bursian, S., Galicia, D., Napolitano, A., Palme, R., Mason, G. (2016). Environmentally enriching American mink (Neovison vison) increases lymphoid organ weight and skeletal symmetry, and reveals differences between two sub-types of stereotypic behaviour. Applied Animal Behaviour Science **177**, 59-69.

Application of Technologies to Animal Welfare Assessments in Zoos

Holly Farmer

Whitley Wildlife Conservation Trust, Paignton Zoo, UK. holly.farmer@paigntonzoo.org.uk

Modern zoological institutions are committed to monitoring and improving the welfare of the animals in their care. Many of the regional associations of zoos and aquariums worldwide are currently developing welfare assessment schemes that can be adopted by their members. However, due to the diverse range of species housed in our zoos, there are practical constraints in adopting a 'one size fits all' approach to assessing animal wellbeing. Limitations in the knowledge of both positive and negative welfare indices for many species, and time restrictions in conducting regular assessments, suggest that new approaches to the field are required. Currently, the majority of welfare assessments in zoos focus on behavioural and health measures, with considerations around environmental factors, and often use a simple scoring system. There are difficulties in adopting many welfare measures applied by the farm and laboratory industries, mainly due to the limitations in transfering knowledge to the wide range of exotic species housed in zoos. In additon, the often invasive nature of laboratory and agricultural wellbeing assessments involving blood draws and anatomical measures of brain morphology, which zoos are not licensed to conduct, and the high cost of equipment and testing also limit application. In response, there has been a move towards incorporating technology to complement behavioural studies, to further consider the behavioural needs and environmental requirements of the animals in our care. This talk will introduce a range of technological approaches currently being adopted to examine animal welfare in zoos and advances in software which can be used to record and share data by animal caregivers. Current technology includes activity trackers to monitor movement and enclosure use, and non-invasive physiological measures such as thermal imaging and recording respiratory and heart rates in response to stressful situations. I will discuss the practical application of technologies from case studies at our three zoo sites in the UK and how they are being incorporated into our welfare management plans. The talk will also consider how an integrated approach to welfare assessments may help to advance the wellbeing of animals housed in zoos and ultimately address the difficulties of conducting welfare assessments in zoos.

Measuring Human-Animal Interactions and Relationships: using psychological and biological approaches

Samantha Ward[1] and Geoff Hosey[2]

1. samantha.ward@ntu.ac.uk United Kingdom Nottingham Trent University; 2. G.Hosey@bolton.ac.uk United Kingdom Bolton University

Human-animal interactions (HAI) and the consequent development of Human-animal relationships (HAR) are increasingly being recognised as of fundamental importance to the welfare of animals in human care. Past research on these has concentrated on the benefits of HARs to people (through interaction with companion animals) or the effects of HARs on animal production (in agricultural animals), but it is important to investigate HARs from both human and animal points of view. Applying these ideas to zoo-housed animals is a recent development, and it carries a number of particular challenges, including dealing with a huge species diversity, issues about the extent of acceptable contact, and the dichotomy between animal contact with familiar (keepers) and unfamiliar (zoo visitors) people. In this talk, we will firstly describe some psychometric approaches we have used to try to define the human side of attachment to animals in zoos. Following on from this we will discuss and evaluate methods used to measure the HAR according to the animals' behavior, which have pinpointed for suitability in accordance with the particular challenges associated with zoo animal research. These include 'latency to perform a behaviour', a voluntary approach test and the adaptation of a current welfare assessment method for this novel application, Qualitative Behaviour Analysis.

Ethical statement

Animal behavioural experiments were approved by the ethical review boards at Nottingham Trent University and granted approval number ARE505.

Using sound observations to measure behaviour

Robert Young and Vinicius Goulart

University of Salford, UK; r.j.young@salford.ac.uk; viniciusdonisete@gmail.com

The quantification of (animal) behaviour is typically done through visual observations of animals or humans either through direct live observations or from video recordings. This technique has generated the data in the majority of studies to date, however, making behavioural observations is a time consuming and expensive process. Some video systems that automatically record and analyse the behaviour of rodents and zebrafish have become available, but systems for automatically recording the behaviour of non-laboratory animals are scarce. One potential alternative to observing animals visually is to make sound observations of their behaviour, emotions, location and physiological responses:

- The recording of animal vocalisations is a well-established technique and can provide much important information. For example, most species of animals have behaviour specific vocalisations such as mating calls, threat vocalisations, distress calls, excitement calls, etc.
- Beyond recording vocalisations actively emitted by animals, the passive sounds they make as a consequence of a behaviour can also be recorded. For example, the sound that an animal makes when it is eating or drinking are often highly distinctive, it is should be possible to indicate the type of food being consumed (e.g., leaf, fruit or nut in the case of a primate) and how much.
- The noise that animals make can also be used to locate them with accuracy in a captive environment and hence to trace their movements and speed (i.e., indications of type of locomotion).
- Many physiological processes are accompanied by sound. For example, the beating of a heart or respiration by the lungs (these parameters can also indicate whether an animal is active, resting or sleeping). Thus, sound can provide insights into animal wellbeing.

Until recently the problem in implementing such as system would have been the need to have it located in the animal's environment (due to weight issues associated with the technology; that is, too heavy to place on an animal). A sound recording station set in an animal's environment would only be suitable for recording loud vocalisations and localising these sounds. The advent of miniature microphones, small microcontrollers, micro-SD memory cards and powerful Wi-Fi or Bluetooth Low Energy (BLE) means that other approaches are now possible. It is now possible to build a device, which measures all of the aforementioned variables at a low price as a piece of wearable technology for animals or humans. The data either being stored on an SD memory card or transmitted wirelessly to a computer for automated analysis. Plus, allowing the Internet of Things in animal behaviour measurements through the live acquisition of welfare states.

The proposed sound observation solution only becomes a practical reality once the issue of automatic data processing has been addressed. The system described could be programmed to sample the aforementioned variables once every second, whilst providing high-quality detailed data this would quickly create huge datasets. This problem can be solved by training computers using neural networks (automated classification) to identify and record the target variables from the device. One issue with this approach is that we do not know how many different subjects would be required to train the computer for the different variables. For example, if the system was bespoke and only to be used for one individual our experience shows us in just a few hours a computer can learn to identify target variables with high accuracy. Commercial systems that respond to human voice commands, often require thousands of individuals in their training set. However, this is due to the great variety of human accents.

At the moment we have produced a proof of concept device and are working on solving the aforementioned problems before going on to test such a device on an animal model (e.g., dogs). However, pilot studies with humans have proven promising. The availability of low-cost components and the possibility of sharing behaviour classification models makes this approach reproducible by other research groups interested in measuring behaviour. Of course, a wearable technology solution will not be appropriate for all animal species. This technology has the potential to fundamentally change how we collect behavioural data on animals and humans (e.g., monitoring people with dementia without the privacy issues associated with video monitoring).

Application of sensor technologies in animal breeding

Farm animals are often kept in large groups. However, in animal breeding, individually collected data is used to improve health, welfare and performance of animals. Keeping animals in large groups makes it difficult to collect individual data. Furthermore, collecting data on changes in health and behavior is a challenge when animals are kept in groups. With the upcoming use of sensor technologies and video imaging, automatic identification and monitoring of animals is possible, which provides opportunities for animal breeding. Breed4Food (B4F) is a consortium established by Wageningen University & Research and four international animal breeding companies. One of the projects of B4F is to develop methods to track and monitor individual animals kept in groups using sensor technologies. A second aim is to use this information to measure traits related to animal behavior, health and nutrient use efficiency. In this symposium, we will bring together animal breeders and researchers using state-of-the art technologies to track and monitor animals in groups. This symposium highlights research that can be implemented in the lab and commercial situations and to different taxa. The presentations will be followed by a plenary discussion to place the presented research in a broader perspective and identify future research directions.

Esther Ellen & Bas Rodenburg, Wageningen University & Research

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

An approach for tracking directional activity of individual laying hens within a multi-tier cage-free housing system (aviary) using accelerometers

Ahmed Ali, Janice Siegford

Animal Behavior and Welfare Group, Department of Animal Science, MSU, USA

Our aim was to use accelerometers to monitor hens' directional activity during light hours, and to track falling and collision during dark hours. 5 hens per each unit were randomly selected, each hen was harnessed with a vest withholding a HOBO® accelerometer with scanning frequency of 20 Hz across 3 days (±3 g) in 3 axes. Loggers were firmly fixed onto vests over hens' bodies, hens' overall activity and for both vertical and horizontal planes during light hours were obtained directly from loggers. For night hours, data were smoothed and filtered from noisy components, acceleration values were post-processed by passing through a moving average low-pass filter and a step function. Falls were recognized by detecting massive shifts in acceleration, the force of collision was calculated using the hen's body weight and the recorded acceleration. Displacement was calculated using initial velocity (assumed zero for static condition), time elapsed, and the sum acceleration during falling down

Introduction

The laying hen industry is phasing out cage-systems in response to consumer demands for alternative housing systems such as aviaries in an attempt to improve hen welfare. The multi-tier complex configurations of aviaries provide additional vertical and horizontal space and resources to groups of hens, including elevated tiers, perches, nests and a litter area on the floor.

However, hen's ability to effectively navigate within the complex configurations of aviaries to utilize resources and distribute throughout the given space has not been fully investigated yet. However, a thorough tracking of these nearly visually identical, and fast-moving hens, through the complex multi-tier aviary design, across light and dark hours using live observation to assess their ability to effectively navigate through multi-tier aviary resources is very challenging, time-consuming and would yield inaccurate results.

Our aim was to innovatively use affordable, and existing device such as accelerometers to monitor hens' directional activity during light hours, and to track falling from higher locations during dark hours throughout the aviaries. Using technology in this way will enable us to understand hens' ability to navigate aviaries vertically and horizontally during light hours, determining hen's ability to effectively use nighttime roosting sites without falling, and to assess distance and the corresponding force of collision of falls that do occur.

Methods

Three treatments were applied to 4 aviary units/treatment (100 Hy-Line W36 laying hens/unit). Control (CON) pullets were raised in floor pens and placed into aviaries at 17 wk old. Floor (FLR) pullets were raised in floor pens and placed into aviaries at 25 wk old. Enriched (ENR) pullets were raised in floor pens and given perches and nests at 17 wk, then placed into aviaries at 25 wk old. To develop a system which was suitable for our research, 5% of hens per each of 4 units/treatments were randomly selected, each hen was harnessed with a different colored trendily designed vest withholding an Onset HOBO Pendant®G acceleration data loggers (18 g) with scanning frequency of 20 Hz across 3 days (−3 g to +3 g) in 3 axes to measure hen's activity.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

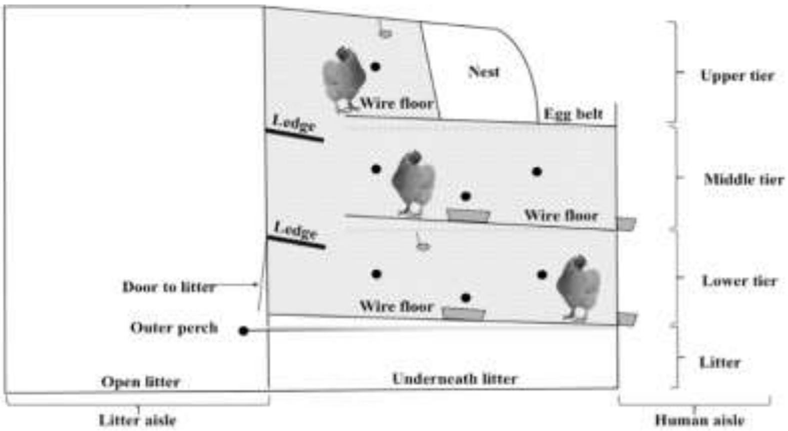


Figure 1: Cross section of aviary unit

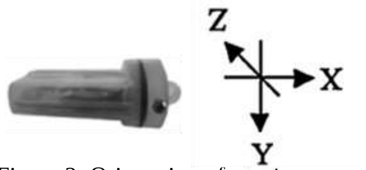


Figure 2: Orientation of accelerometers

Each aviary unit was composed of a litter covered floor area and an enclosure, the latter consisting, of 3 vertical tiers, and 2 solid metal ledges in between for transition. Resources such as feeders, drinkers, perches, and nest are distributed throughout the 3 tiers as shown **Figure 1**.

'Loggers were oriented in the vests on the hens with the X-axis recording forward and backward movement of the bird (craniocaudal movement of the bird), Y-axis recording sideways movement (lateral movement of the bird), and Z-axis recording vertical movement (dorsoventral movement) of hens as shown in **Figure 2**.

Loggers were firmly attached to vests to avoid unnecessary noisy data components due to movement of the loggers themselves or disruption of tilt angles due to misorienting loggers. Loggers recorded hens' activity for 3 consecutive days at Peaklay (36 wk of age) and Midlay (54 wk of age).

Hens' overall activity (A_s) and for both vertical (a_z: represents the dorsoventral activity of the birds), and horizontal (a_x: represents the craniocaudal activity of the birds) planes, during light hours were obtained directly from loggers by summing and averaging activity data (a_y: represents activity due to sideways or lateral movement of the birds).

$$A_s = \sqrt{a_x^2 + a_y^2 + a_z^2}$$

In order to accurately calculate falls and force of collision during night hours, data were first smoothed and filtered from noisy components by removing all acceleration fluctuations that were not due to falling down.

Raw acceleration values (A_j) were postprocessed by passing through an asymmetrical 3 point-moving average low-pass filter (i = the middle point in the 3 point-moving average low pass filter) and through a step function, to define thresholds used to remove such fluctuations (t = threshold values of minor fluctuations). After smoothing data, falls were recognized by detecting massive shifts in acceleration from the static condition of hens during dark hours.

$$A_i = \frac{1}{3}\sum_{j=i-1}^{i+1} A_j \qquad A_i' = \begin{cases} \mu, if\ |A_i - \mu| < t \\ \mu, if\ |A_i - \mu| \geq t' \end{cases}$$

Hens were considered in a stationary state when a constant acceleration (1g), was acting on the hen. Incidence of falling down during night time was detected when the recorded acceleration was lower than the constant/static one. Force of collision (F = N) due to falling down was calculated using the hen's exact body weight (M = kg) and the sum acceleration recorded during falling down ($A_i' = m/s^2$).

$$F = M \times A_i'$$

Finally, the displacement or distance of the fall (D = cm) was calculated using initial velocity ($Vi = m/s$), assumed to equal zero during stationary condition before falling), time (t = s) elapsed during falling down, and the sum acceleration during falling down (A_i' = m/s^2).

$$D = \left\{V_i \times t + \frac{1}{2} A_i' t^2\right\} \times 100$$

All research protocols were approved by the Michigan State University Institutional Animal Care and Use Committee prior to the start of data collection. Data were analyzed using GLMM with treatment, time of day and stage of production as fixed effects and unit and day of observation as random effects (α set at 0.05) with R software.

Results

In the current study, using accelerometers to capture hen's triaxial activity and to detect falls during the night was found to be a feasible method that can be used to investigate hens' ability to effectively navigate within the complex configurations of aviaries. In the current study, triaxial activity was not different among treatments; however, vertical activity (Z-axis) among tiers was higher in ENR (0.8 g) than CON (0.6 g) and FLR hens (0.3 g; P=0.003). CON hens showed the highest fall incidence (16 vs 9 FLR and 5 ENR), while FLR hens showed the highest collision force and acceleration (15 N and 0.8 g vs 10 N and 0.5 g in CON and 5 N and 0.3 g in ENR hens; all P≤0.05). CON and ENR hens fell from higher locations than FLR hens, with average distances of 151, 148 vs 102 cm, respectively (all P≤0.05).

However, various limitations were encountered when using triaxial accelerometers in this way, for instance, analysis showed massive disruption of acceleration and displacement data from 2 accelerometers, which may have been due to misorientation and frequent tilting from vests loosely fitted to hens. Therefore, it is highly recommended to assure complete immobility of accelerometers.

Acceleration data in a single falling incident from dark hours showed an interrupted pattern, and subsequently, the force of collision and displacement were wrongly calculated. The most rational justification might be that the hen had an interrupted fall, with the hen was falling from a certain location in the upper tier and onto a structure in the middle tier when a nearly static acceleration was recorded before the hen continued falling down to the lower tier.

After several trials a scanning frequency of 20 Hz was determined to be optimum to accurately calculate acceleration, displacement, and force of collison during nighttime falling. Scanning at lower frequencies gave inaccurate results while higher frequencies gave similar results to the 20 Hz, but for shorter periods of observation due to the limited storage capacity of the device.

Also, we conducted several trials to calculate the distance covered during light hours using acceleration data. However, since the initial velocity for each recorded acceleration could not be obtained, so far it seems to be impossible to calculate such information.

Finally, physical indicators of welfare (body weight, plumage quality, keel bone damage, foot lesion) for sampled hens were assessed before and after data collection to ensure hens were not negatively affected by wearing the sensors.

Conclusion

Based on our findings we conclude that accelerometers can be used efficiently during nighttime to record falls, the force of collision and displacement of laying hens in multi-tier aviaries. The method can be used to estimate directional activity (i.e., vertical vs horizontal) during light hours which would indicate hens' ability to effectively navigate through the 3-dimensional housing environment that has several vertical tiers. However, a different system is required to perform accurate location tracking during light hours of hens within these complex housing systems to better indicate hens' resource use rather than directional activity during light observations.

Using sensor technologies in animal breeding: reducing damaging behaviour of animals kept in groups

E.D. Ellen[1], M. van der Sluis[1,2], E.N. de Haas[3,4] and T.B. Rodenburg[2,4]

1 Animal Breeding and Genomics, Wageningen University & Research, Wageningen, The Netherlands. Esther.Ellen@wur.nl 2 Department of Animals in Science and Society, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands. 3 Behavioural Ecology Group, Wageningen University & Research, Wageningen, The Netherlands, 4 Adaptation Physiology Group, Wageningen University & Research, Wageningen, The Netherlands.

In animal breeding, individually collected data is used to improve health, welfare and performance of livestock. Furthermore, genetic variation between individuals is utilized to select the best parents for the next generation. There is a tendency in commercial livestock production that livestock are increasingly kept in large group housing systems. To improve health, welfare and performance in such a system is a challenge, because it is difficult to identify and monitor individual animals. Furthermore, one of the main problems in large groups is that social interactions between group members can result in the development of damaging behaviour, such as feather pecking behaviour in laying hens and tail biting in pigs [1]. To improve socially-affected traits, it is important to take both the genotype of the individual itself (direct genetic effect or victim effect; DGE) and the genotype of the group members (indirect genetic effect or actor effect; IGE) into account [2]. Previous studies showed that IGE contribute 70-94% of total heritable variation in survival time and plumage condition of birds [e.g. 3]. However, when animals are kept in large groups it is difficult to identify the actors performing damaging behaviour. A wide variety of technologies has been developed to monitor the activity and behavioural patterns of group housed animals in an automated way. For instance sound analysis, image analysis and IR thermography are used to continuously monitor the activity and presence of animals [4-6]. However, these technologies do not allow monitoring of the activity of individual animals and are only used in research settings. In recent years, accelerometers as well as RFID, ultra-wideband and GIS technology have been developed where animals are equipped with a body worn sensor to monitor their individual behaviour patterns, which allows us to identify and monitor individual animals in groups. However, it is difficult and expensive to upscale these sensor systems to breeding farm level. The sensors often require considerable battery power for the transmission of data [e.g. 7], and it is unknown whether these sensors can survive the harsh environment of the farm house during the entire life span of the animals. Therefore, it is important to investigate sensor technologies that can be used to identify and monitor individual animals kept in commercial housing systems, and to identify the actors performing damaging behaviour. The aim of the Breed4Food project "Individual tracking" is to find the optimal method to track and monitor individual animals in commercial settings. Breed4Food is a consortium of four international breeding companies and Wageningen University & Research, focussing on the genetic improvement of laying hens, broilers, pigs, dairy cattle and turkeys. Within the project "Individual tracking" using different sensor technologies (like ultra-wideband, RFID, imaging, accelerometers) will be compared in a small research setting in PhenoLab [8]. The PhenoLab project, focusing on sensor-based recording of location, activity and proximity in laying hens, is a collaboration between Wageningen University & Research, Breed4Food, Noldus Information Technology BV, and i3B (ICT for Brain Body and Behavior). In the PhenoLab project, we investigate possibilities for tracking of location, activity and proximity of individual laying hens. To meet that aim, we tracked individual hens during a five-minute Open Field test using two different tracking systems: Ultra-wideband tracking using TrackLab and automatic video tracking using EthoVision. Ultra-wideband tracking consists of an active RFID tag that is placed on the bird in a backpack. This tag is then located by triangulation by four beacons. Comparing distance moved between TrackLab and Ethovision yielded 96% similar results (sample of 24 hens). Proof of concept in PhenoLab will be followed by testing on commercial breeding farms. The optimal method will be implemented in the breeding program to automatically collect large-scale data on behaviour, health and performance of animals kept in groups, to reduce damaging behaviour in livestock.

References:

1. Rodenburg, T.B. and Koene, P. (2007). The impact of group size on damaging behaviours, aggression, fear and stress in farm animals. *Applied animal behaviour Science* **103**, 205-214.
2. Bijma, P., Muir, W.M., Ellen, E.D., Wolf, J.B. and van Arendonk, J.A.M. (2007). Multilevel selection 2: estimating the genetic parameters determining inheritance and response to selection. *Genetics* **175**, 289-299.
3. Brinker, T., Bijma, P., Visscher, J., Rodenburg, T.B. and Ellen, E.D. (2014). Plumage condition in laying hens: Genetic parameters for direct and indirect effects in two purebred layer lines. *Genetics Selection Evolution* **46**, 33.

4. Quwaider, M.Q., Daigle, C.L., Biswas, S.K., Siegford, J.M. and Swanson, J.C. (2010). Development of a wireless body-mounted sensor to monitor location and activity of laying hens in a non-cage housing system. *Transactions of the ASABE* **53**, 1705-1713.
5. Mench, J.A. and Blatchford, R.A. (2014). Determination of space use by laying hens using kinematic analysis. *Poultry science* **93**, 794-798.
6. Lee, H.-j., Roberts, S.J., Drake, K.A. and Dawkins, M.S. (2010). Prediction of feather damage in laying hens using optical flows and Markov models. *Journal of The Royal Society Interface* doi 10.1098/rsif.2010.0268.
7. Banerjee, D., Daigle, C.L., Dong, B., Wurtz, K., Newberry, R.C., Siegford, J.M. and Biswas, S. (2014). Detection of jumping and landing force in laying hens using wireless wearable sensors. *Poultry Science* **93**, 2724-2733.
8. Rodenburg, T., van der Eijk, J., Pichova, K., van Mil, B. and de Haas, E. (2017). PhenoLab: automatic recording of location, activity and proximity in group-housed laying hens. *Proceedings of ISAE 2017* (Benelux conference, 12 October 2017).

Application of sensor technologies in animal breeding with a focus on pig production

Tomas Norton, Daniel Berckmans

M3-BIORES, KU Leuven, Kasteelpark Arenberg 30, 3001 Heverlee, Belgium; tomas.norton@kuleuven.be

Introduction

Recent years have seen remarkable developments in technologies to monitor various traits of animals in large groups. This evolution strengthens our capacity to uncover interesting information of behavioural and production dynamics of animals in both natural and managed ecosystems. With specific reference to the latter, the field Precision Livestock Farming (PLF) emerged with the focus on linking these technologies with associated algorithms to empower livestock farmers with capacity to better manage their farming practices. However, there is no doubt that other stakeholders in the value chain can also be implicated in a positive way. Animal breeders, feed developers for example can use the information derived from these technologies to improve the inputs to the farming system so to help farmers balance target for efficiency, welfare and environmental impact.

The animal breeding sector, in particular, has a lot to gain from these animal monitoring technologies. For so long the genetic improvement of production animals was focussed solely on achieving high efficiency (ADG, FCR). However, it is widely known that this has negatively implicated the welfare of group-housed animals, as harmful social interactions between conspecifics is a widespread problem in the industry [2]. Genetic selection to achieve productivity targets while reducing undesirable behaviours is promising. However, it is also challenging because phenotyping of such behavioural and production traits is time consuming and labour intensive, as audio-visual decoding of observational videos are painstakingly time and labour intensive [6]. There is also the challenge of scale, as the breeding sector need to able to obtain direct observations on thousands of animals to make significant impact on the selection of animals. Therefore, there is a growing need for technologies to monitor in an integrated way the different traits of interest to breeding companies.

Here, we present the core principles underpinning the development of sensing systems for animal monitoring followed by the possibilities of animal monitoring technologies that can address the needs of the breeding sector. The focus of this abstract is the pig sector, although a similar analysis can be made for poultry and dairy cow sectors also.

Core principles of developing animal sensing technologies

When developing sensing system from animal monitoring, the behavioural and physiological traits of the animal must be key considerations. It is, however, difficult to frame these conditions within engineering terminology so that solutions that require the integration of engineering, computer science and biological and behavioural disciplines can be developed.

The basic idea in Precision Livestock Farming (PLF) is that any problem giving discomfort to an animal will be first shown by the animal by changing behaviour. Therefore, the conceptualisation of living organisms as Complex, Individually Different, Time-Varying and Dynamic (CITD) has been key in linking engineering terminology with the nature of biological systems [7].

Complex, Individually Different, Time-Varying and Dynamic is what all us living organisms are It is obvious that a living organism is much more complex than any mechanical, electronic or ICT system. The complexity of information transmission in a single cell of a living organism is for example much higher than in most man made systems (e.g. today's most powerful microchip). The concept of the "CITD system" first emerged at the first European Conference on Precision Livestock Farming EC PLF conference (Berckmans, 2003, Berlin), and since has enabled significant progress to be achieved in PLF tool development. The CITD nature of living organisms has an important impact on the type of algorithms we need to develop. It implies that algorithms to monitor these time-varying individuals must continuously adapt to the individual and/or use principles that can be used in real time in the field application to adapt to the time variation of that individual (Figure 1).

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

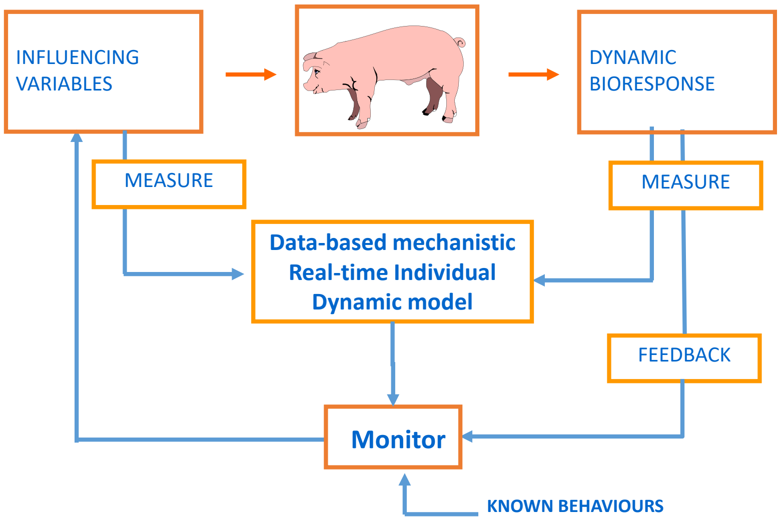


Figure 1. Scheme showing how behaviours of animals can be continuously monitored (Aerts et al, 2003).

Possibility 1: Pig behaviour monitoring at group level

Real-time monitoring of the activity level of animals continuously at group level using video cameras has great potential. One technology called eYenamic (Fancom, BV) was developed and improved during the EU-PLF project (ww.euplf.eu; Project No 311825; [3]) to measure the activity level of pigs in real time and in practical farming conditions. The activity level could be localised by measuring it for different pre-defined zones separately. An arbitrary number of zones could be defined interactively by the user as rectangular boxes in the camera image, defined by their upper-left and lower-right corners. For the experiments, the two pens visible in the camera images were each divided in two zones, covering the left and right hand side, respectively (Figure 2).

Every second, the algorithm logged the camera image and the activity index for each zone, defined as the fraction of the floor space in the pen that was covered or uncovered by pigs in the camera image during a one second period, i.e. the activity index is the fraction of floor space that has 'moving' animal pixels. The activity index for the four defined zones that was measured on-line during a 32 hour testing period. The difference between day- and night activity level can clearly be noticed (Figure 3).

Figure 2. the eYenamic setup interface, showing the 2 x 2 zones that cover the left and right hand side of each of the two pens visible in the camera image.

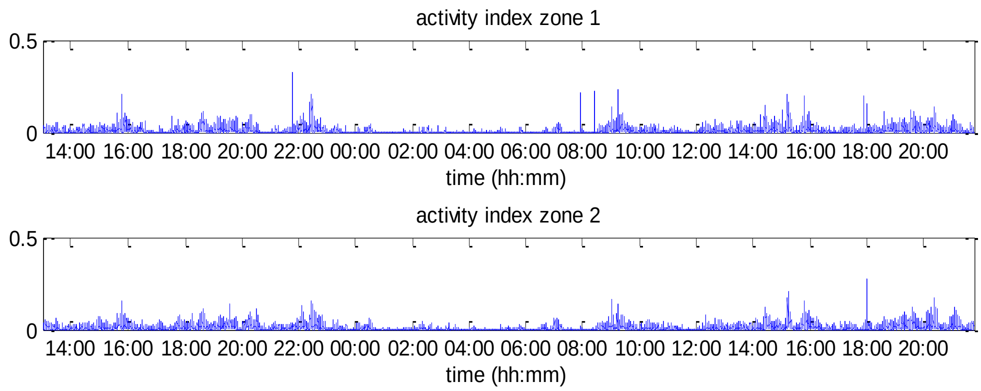


Figure 3. the activity index for 2 of the 4 defined zones during a 24 hour testing period

Possibility 2: Pig growth monitoring and water usage at group or individual level

Individual weight measurement is an important process output and breeding variable that nonetheless suffers from a number of drawbacks when performed manually. Gathering weight data using a manual scale is therefore done sparingly, generally only at the beginning and end of a production period and most often only for a representative subset of animals [8]. Machine vision-based weighing of pigs is a non-intrusive, fast and accurate approach, which could reduce stress for both the animal and the farmer during the weighing process. Since slow weight gain can happen for some of the pigs in a pen, it is important to monitor weight for each pig individually. This helps the farmer or breeding company to identify slow-growing and fat-growing pigs.

Building further upon the idea of Schofield, an algorithm was developed to monitor weight using machine vision techniques [4]. First, the pig image was segmented and the corpus image was separated from the head by using the same ellipse fitting algorithm (Figure 4).

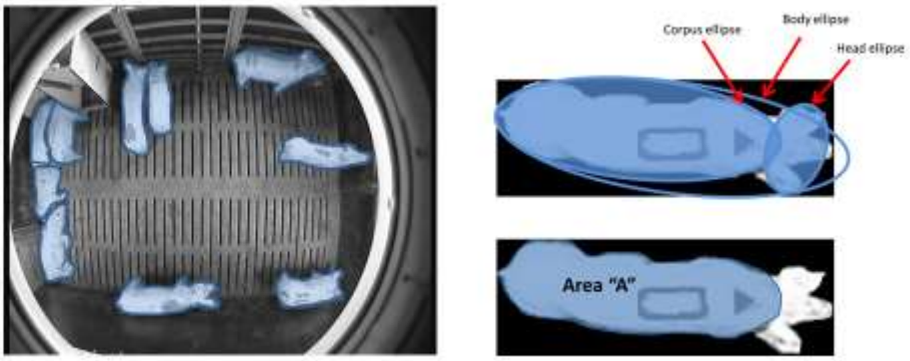


Figure 4. Individual weight monitoring of pigs in farm conditions

The objective of the next step was to quantify the dynamics of body area and to relate it to the gold standard body weight (BW). The relationship between known image features and BW is fixed and known. Figure 5 shows the measured actual weights versus the predicted weights over six days of measurements for all four pens and ten pigs per pen (240 data points). The ideal case was that all of the data points align with the identity line (R^2 of 100% which means for every data point, predicted weight would equal the measured weight). This means the more erratic the points are, the lower R^2 and accuracy of weight prediction will be.

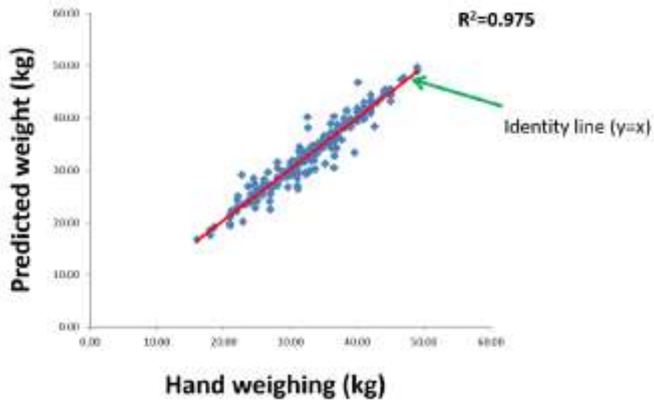


Figure 5. Measured weights versus predicted weights over six measurement days of all four pens with ten pigs per pen (Kashiha et all, 2014a).

Drinking behaviour and water volume usage are also important indicators of satisfactory performance and health. For example, feed and water volume usage are closely related and solid feed intake must be accompanied by water intake. Kashiha and colleagues [5] developed a drinking behaviour monitor for pigs could be useful in detection of diseases and other production related problems too. The goal was to estimate half-hourly water volume usage (in litre) in a pig barn by analysing half-hourly duration of drink nipple visits. To achieve that purpose, a data-based dynamic model was developed to quantify the dynamics of water usage and to relate it with half-hourly duration of visits, as detected from a video stream (Figure 6).

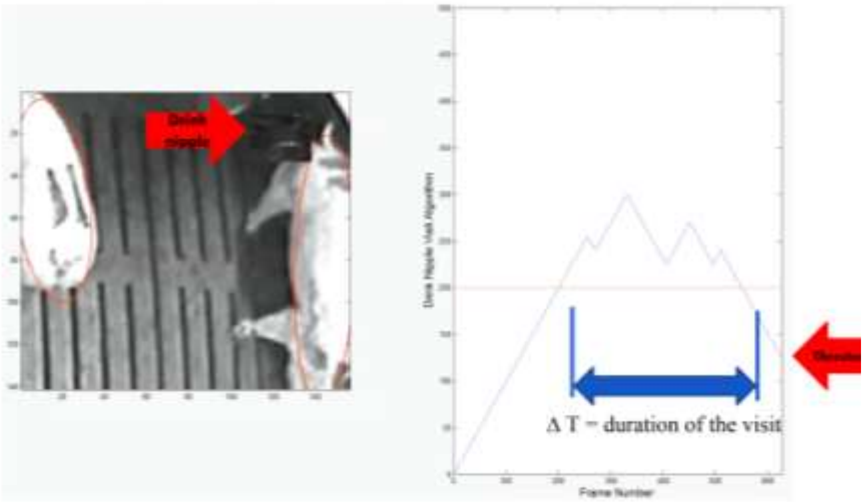


Figure 6. Water monitoring using video analysis (Kashisha et al., 2014b).

Possibility 3: Pig aggression monitoring

Aggression in group housed pigs is significant challenge for the breeding and farming sector alike. Pigs engaged in non-productive activities such as fighting, increase energy production without production gains. Video tracking of aggression is possible but requires a sensor system that can correctly identify many specific features and implement algorithms to translate this into a meaningful a behavioural phenotype that could be used in breeding programmes. Automatic detection of aggressive behaviours in pigs has already been proven to be possible by characterizing pigs' movement with respect to intensity and distribution of motion [9]. The first step in this approach is to label key behaviours so that the meaningful features can be extracted from the image (Figure 7a). This step requires significant effort. For example, decoding of 8 h of video to train a system to detect aggressive behaviour in pigs required 90 person hours [6].

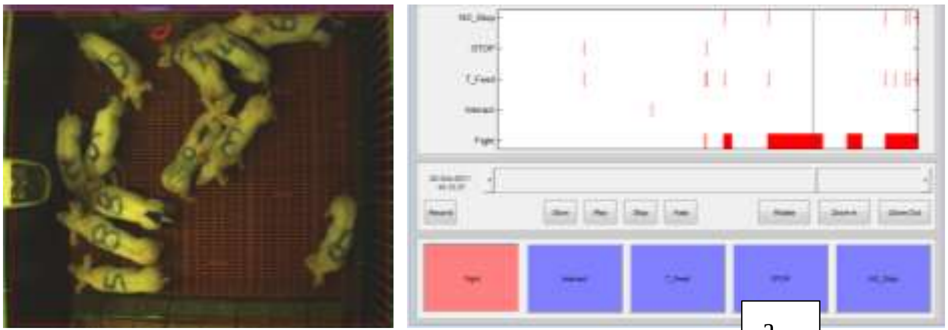
Once the features are extracted they are then grouped and classified according to the severity to the interaction. The model of Viazzi et al [9] successfully classified aggressive and non-aggressive interactions with an accuracy

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

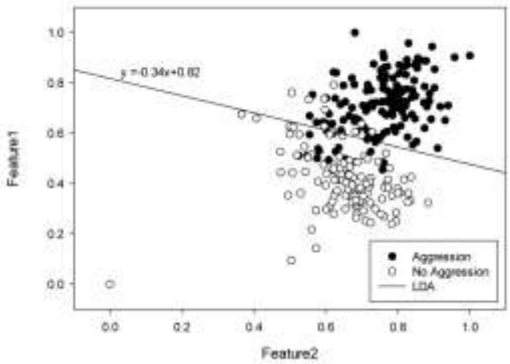
of 89% on an experiment farm (Figure 7b). Further effort is needed to determine their robustness under more realistic production conditions. For example, if commercial breeders wish to incorporate automated behavioural phenotyping on farm, systems must be able to work with large areas and groups of pigs, low or uneven lighting, varying backgrounds, and obstacles (e.g., feeders and drinkers).

Labelled variables:

- Duration of aggression (start-finish time)
- Time of feeder activation
- Aggression interrupted/not interrupted
- Type of aggressive interaction
- Aggression initiator and receiver
- Individual reaction to feeder activation



a.



Sensitivity: 88 %
Specificity: 89 %
Accuracy: 89 %

		Class	Predicted Group No Aggression	Predicted Group Aggression	Total
Cross-validated	Count	No Aggression	104	16	120
		Aggression	15	133	148
	%	No Aggression	89.3	10.7	100.0
		Aggression	11.3	88.7	

b.

Figure 7. (a) the labelling requirements and tool used, (b) the results of the classification of non- and aggressive interactions (Viazzi et al, 2014).

Conclusions

Sensor-based animal monitoring technologies are finding more widespread use across commercial farms, and can create value for the breeder companies to improve their genetic lines. From there we have to discover how they can create value for the animals and the farmers in the first place. The fundamental advantage is that these Precision Livestock Farming systems can monitor continuously, for example the frame-rate of a typical video is 25 images per second. While a human eye can observe up to 3 times this rate it is not possible to expect an observer to do this 24 hours a day, 7 days a week, 365 days a year. From the examples available today, it is clear that monitoring technologies can create value for the farmer, breeder as well as other stakeholders. Opportunities abound to share data-streams between these stakeholders to ensure a future for sustainable and resilient livestock production. Nevertheless, it is fair to say that more work is required to develop these technologies so they can be implemented in the commercial livestock production facilities of today.

References

1. Aerts, J-M, Berckmans, D. (2002). Integration of Biological Responses into Process Management. KU Leuven Masters Course.
2. Camerlink, I., Turner, S.P., Ursinus, W.W., Reimert, I. and Bolhuis, J.E., 2014. Aggression and affiliation during social conflict in pigs. PLoS One, 9(11), p.e113502.
3. EU-PLF Project, Bright Farm by Precision Livestock Farming, Grant agreement no: 311825 (M3-BIORES, KU Leuven, Belgium; The Swedish University of Agricultural Sciences, Sweden; The University of Bristol, UK; Universtità degli Studi di Milano, Italy; The Royal Veterinary College, UK; Stichting Dienst Landbouwkundig Onderzoek, the Netherlands; Wageningen Universiteit, The Netherlands; Institut National de la Recherche Agronomique, France; The Agricultural Research Organisation of Israel – The Volcani Centre, Israel; Teagasc, Ireland; European Federation of Animal Science, Italy; Fancom BV, The Netherlands; Nuscience, Belgium; GEA Farm Technologies, Germany; SoundTalks, Belgium; PLF Agritech, UK; Xenon New Technologies, Belgium; Abrox, Spain; Syntesa, Faroe Islands; M&M Corporation, Belgium).
4. Kashiha, M.; Pluk, A.; Bahr, C.; Vranken, E.; Berckmans, D. (2014). Development of an early warning system for a broiler house using computer vision. BIOSYSTEMS ENGINEERING Volume: 116 Issue: 1, Pages: 36-45 Published: SEP 2014
5. Kashiha, M.; Bahr, C.; Haredasht Amirpour, S.; Ott, S; Moons, C.; Niewold, TA; Odberg, FO; Berckmans, D. (2014) The automatic monitoring of pigs water use by cameras COMPUTERS AND ELECTRONICS IN AGRICULTURE Volume: 90 Pages: 164-169
6. Oczak, M., Viazzi, S., Ismayilova, G., Sonoda, L.T., Roulston, N., Fels, M., Bahr, C., Hartung, J., Guarino, M., Berckmans, D. and Vranken, E., 2014. Classification of aggressive behaviour in pigs by activity index and multilayer feed forward neural network. Biosystems Engineering, 119, pp.89-97.
7. Quanten, S., De Valck, E., Cluydts, R., AERTS, J. and Berckmans, D., 2006. Individualized and time-variant model for the functional link between thermoregulation and sleep onset. Journal of sleep research, 15(2), pp.183-198.
8. Schofield, C.P., 1990. Evaluation of image analysis as a means of estimating the weight of pigs. Journal of Agricultural Engineering Research, 47, pp.287-296.
9. Viazzi, S.; Bahr, C.; Van Hertem, T.; Schlageter-Tello, A.; Romanini, CEB.; Halachmi, I.; Lokhorst, C. and Berckmans, D. Comparison of a three-dimensional and two-dimensional camera system for automated measurement of back posture in dairy cows. 2014. COMPUTERS AND ELECTRONICS IN AGRICULTURE. Volume: 100, Pages: 139-147.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Perspectives of using deep learning and computer vision algorithms for continuous behavioural monitoring of dairy cows.

Oleksiy Guzhva[1]

1Swedish University of Agricultural Sciences, Department of Biosystems and Technology, Alnarp, Sweden.
oleksiy.guzhva@slu.se

In modern dairy production, loose housing or free stalls are the most common practices, which, while widely implemented along with different management routines, do not always include the adjustments necessary for assuring natural behaviour of animals. Visual monitoring of interactions occurring between dairy cows is, however, very time consuming and has the possibility of observer bias. A possible solution may thus be to use the computer vision-based techniques that could create the potential to carry out behavioural research more efficiently and reliably.

The behavioural traits of dairy cows related to “ease of handling”, as well as “reactiveness” have traditionally been the main factors allowing domestication and use of dairy cows by humans. Dairy cows are social animals, and the “group” is an essential resource influencing health state, welfare and production performance. Studies [5] show that a safe and positive social environment in the cow barn has both short- and long-term effects linked to various aspects of cow health, as well as production performance, longevity and stress levels on a group level. Analysis of interactions occurring between cows in dairy barns and their effect on health and performance is of great importance for sustainable, “animal-friendly” production. There is data [6] showing that consequences of regrouping and changes in social groups of dairy cows and the relation to possible health and stress-related deviations are not always considered in everyday farm practice. Therefore, these problems require more knowledge to create management strategies to ensure “animal-friendly” environments.

To improve the existing breeding strategies, one should use the precise and uniform definition of behavioural traits defining “a good cow”, as well as find reliable tools for monitoring and analysing those traits. However, no practically reliable technology to monitor individual animals in new, sometimes “de-syncronized” environments, is available. The vast majority of current computer vision solutions are still in the developing phase and do not provide the flexibility/functionality required for continuous monitoring of animals. The key-concepts forming the framework needed for robust solutions for automated and accurate tracking/identification of animals, as well as extended behavioural analysis features are also not fully established yet. Therefore, investigating the opportunities and limitations of recent advances in computer vision and deep learning will facilitate the development of modules capable of monitoring animal welfare, health and behaviour related parameters at low computational cost.

The collective aim of our work [1, 2, 3, 4] was to investigate the possibilities and limitations of recent advances in computer vision and deep learning for studying dairy cattle behaviour and cow traffic. Subsequently, this knowledge was used when taking the first step towards a fully automated system for continuous surveillance in modern dairy cows. The outcome is the so-called, “WatchDog” system, consisting of several state-of-the-art modules. These modules were developed and tested in the following order: an algorithm for the behavioural analysis ("Behavioural Detector" module), an algorithm for virtual/real floor assessment and claw placement ("Floor/Claw" module), an algorithm for filtering video material (actual "WatchDog" module) and at last an algorithm for tracking and identification ("Tracker" module).

The “WatchDog” system was evaluated on several levels: cows, frames and interactions. Single cows were detected with a hit rate of 97% and a false alarm rate of 2.9%. Single frames were perfectly interpreted 92.8% of the time. Frames with only zero or one cow present were correctly identified 99.5% of the time (at any number of given examples). All interactions were detected while discarding 60% of the video as uninteresting. The overall accuracy of the “Behavioural Detector” module varied depending on different test-conditions:

- no “line border features”, single frame: accuracy 79.2%,
- with “line border features”, single frame: accuracy 83.1%,
- with “line border features”, three consecutive frames: accuracy 85.1%.

These different test-conditions were mainly related to the complexity of the geometrical analysis and number of frames from the video sequence used for the evaluation. Since the “Behavioral Detector” module operates with earlier developed seven-point shape-model [3], there is some freedom with relation to how many of those seven points (and lines that they potentially form) could be included into the analysis.

For the “Tracker” module performance, in total 26 tracks were considered, and 23 were correctly tracked, while three were lost at some point (no longer possible to confirm real-ID). Given those 26 starting points, the tracker was able to maintain the correct position in a total of 101.29 minutes or 225 seconds on average per starting point.

Future work should aim to improve the performance and flexibility of the modules from the "WatchDog" system as well the development of new "building blocks" or modules, which will improve the functionality of the proposed solution by:

- source code optimisation to achieve the opportunity to use "WatchDog" on different platforms and embedded hardware at lower computational cost without losing the functionality.
- adding the additional features to "Behavioural Detector" module that will allow spatial analysis as well as social network analysis based on proximity between the individuals. This will allow the studies investigating hierarchical order within the herd, maternal relationship and preference tests (both towards other individuals and specific areas of interest).
- extending the range of functions for the "Behavioural Detector" module by adding features that will allow calving detection, simple gait recognition and benchmarking of "scene complexity".
- investigating the opportunities to create a smartphone-based application with Augmented Reality (AR) functionality for behavioural evaluation on-site (as well as some lameness detection based on back ridge angle).
- making a "For Research Purposes" only graphical user interface (GUI) for "WatchDog" module responsible for video filtering, allowing more efficient work with the recordings.
- investigating the opportunity for adjusting the developed shape-model to include other animal species into the algorithm.

References:

1. Ardo, H., Guzhva, O., Nilsson, M. and Herlin, A. (2016). Cows on concrete slats of the waiting area in a dairy barn estimated by use of image analysis (manuscript).
2. Ardo, H., Guzhva, O., Nilsson, M. and Herlin, A. (2017). A CNN-based Cow Interaction Watchdog. *IET Computer Vision* **2017**, *pp. 1-7.*
3. Guzhva, O., Ardo, H., Herlin, A., Nilsson, M., Astrom, K. and Bergsten, C. (2016). Feasibility study for the implementation of an automatic system for the detection of social interactions in the waiting area of automatic milking stations by using a video surveillance system. *Computers and Electronics in Agriculture* **127**, pp. 506-509.
4. Guzhva, O., Ardo, H., Nilsson, M. and Herlin, A. and Tufvesson, L. (2018). CNN-based animal detector with real-time tracking and identification features. *Biosystems Engineering*, (in press).
5. Kikusui, T., Winslow, J. T., & Mori, Y. (2006). Social buffering: Relief from stress and anxiety. *Philosophical Transactions of the Royal Society B*, **361**, pp. 2215 – 2228.
6. Seyfarth, R., & Cheney, D. (2015). How sociality shapes the brain, behaviour and cognition. *Animal Behaviour* **103**, pp. 187 –190.

Finding hens in a haystack: Consistency of movement patterns within and across individual laying hens maintained in large groups

Christina Rufener[1], Berezowski John[1], Filipe Maximiano Sousa[1], Yandy Abreu[2], Lucy Asher[3] and Michael Toscano[1]

1. University of Bern, Switzerland; 2. National Centre for Animal and Plant Health, Cuba; 3. Newcastle University, UK; christina.rufener@vetsuisse.unibe.ch

Animals of various species, including poultry, have structured behavioural routines with variations relating to individuality. Efforts focusing on poultry have been largely limited to small (< 25 hens), non-commercial groups or the presence on exterior ranges rather than inside the house. Given existing knowledge of hen behavioural patterns and a growing need for animal monitoring within ever larger commercial flocks, we sought to investigate methodological tools that could objectively quantify the recorded movement and location patterns and allow comparison across and within hens. All experimental animal work was approved by the Bern Kantonal authority (BE-31/15) and all methodologies were performed in accordance with the relevant national and kantonal guidelines and regulations. For the effort, we used a custom designed tracking system to monitor the location within five pen zones (Figure 1) of a commercial aviary for 13 hens (Lohmann Selected Leghorn or Lohmann Brown) within a flock of 225 animals for a contiguous period of 11 days. Most hens manifested a hen-specific pattern that was (visually) highly consistent across days, though, within that consistency, manifested stark differences between hens (Figure 2). Some visual features that stood out between hens were: patterns of specific, repeating transitions at a relatively similar rate/time; extended duration of time within a single specific zone; lack of entry into an area; and isolated periods of activity within extended periods of inactivity. To test the hypothesis that daily movement patterns were more similar within hens than between hens and evaluate methodological tools capable of such comparisons, three different methods were used to classify individual daily datasets into groups based on their similarity. The three methods were: i) Linear Discriminant Analysis based on six summary variables (transitions into each zone) and total transitions; ii) naïve clustering analysis, Hierarchical Clustering, applied to summary variables and iii) Hierarchical Clustering applied to dissimilarity matrices produced by Dynamic Time Warping. The three methods were highly successful in correctly classifying hens with each method correctly classifying more than 85% of the hen days. In summary, our methods provide a unique means to assess behavior of a system that was previously not accessible but indicates the presence of an enormous degree of complexity and structure. We believe the current effort is the first to document these behaviors within a large, complex commercial system and thus has enormous potential to influence the assessment of animal welfare, health, and productivity. Future work will need to re-evaluate these methods over extended periods of time and in relation to variables of interest, e.g., egg production and hen mortality.

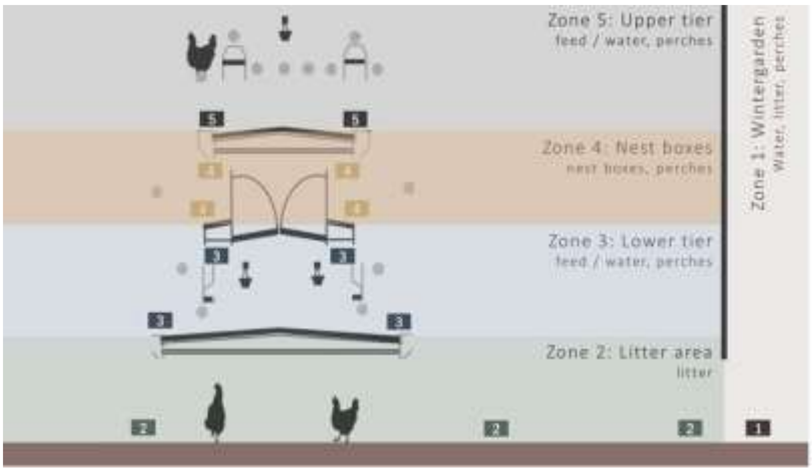


Figure 1. Representation of the aviary system and the five coverage zones showing available resources in each zone as well as emitter positioning (boxes).

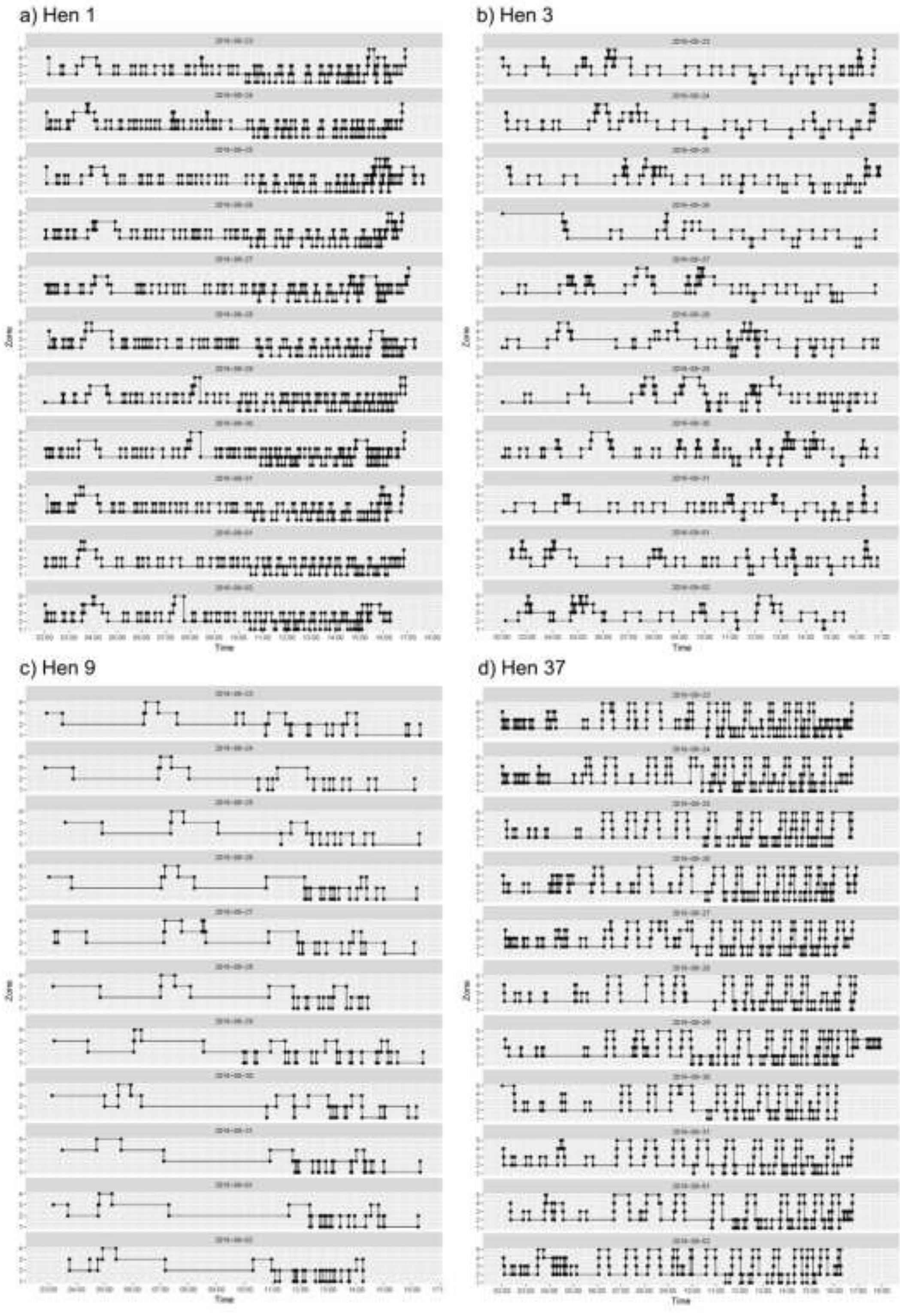


Figure 2. Location graphs of four focal hens over 11 consecutive days. Each panel represents a single hen (1, 3, 9, 37). Within each graph, each line (top to bottom) represents location data for a single day where lights were on between approximately 0200 and 1700. The Y-axis represents the zone in which the animal was present and the X-axis represents the time of day.

Radiofrequency identification systems: Advantages and constraints for tracking and monitoring of individual animals

M. van der Sluis[1,2], E.D. Ellen[1], Y. de Haas[1] and T.B. Rodenburg[2,3]

1 Animal Breeding and Genomics, Wageningen University & Research, Wageningen, The Netherlands. Malou.vandersluis@wur.nl; 2 Department of Animals in Science and Society, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands; 3 Adaptation Physiology Group, Wageningen University & Research, Wageningen, The Netherlands.

To track and monitor individual animals in groups, it is possible to use radiofrequency identification (RFID) systems. RFID encompasses all wireless communication systems that use radiofrequency fields [1]. RFID systems consist of tags and readers [2]. The tags contain unique identification data and can be attached to the item or, in the proposed application, to the animal that needs to be tracked [2]. The readers are used to read and identify the tags via radiofrequency fields [2]. The development of RFID technology started in the late 1960s and RFID systems are currently used for a large range of applications, including electronic door locking systems, contactless smart cards, and animal identification [2,3,4]. Further progress in the technology of chip manufacturing now makes RFID suitable for novel applications and available at a lower cost [4,5]. Therefore, the existing RFID technology may be applicable for animal tracking and monitoring as well. Here, different types of RFID systems will first be discussed for their applicability in individual animal tracking and monitoring, after which future work using RFID to track individual animals will be presented.

Many different types of RFID systems exist. The main characteristics by which RFID systems can be categorized are 1) power supply of the tags and; 2) operating frequency and associated reading ranges. Tags can be classified as active or passive, depending on their power supply. Passive tags obtain their power from the field of the reader and therefore do not have their own power supply in the form of a battery [6]. Active tags contain a battery that powers their signalling and allows them to initiate communication, and often have larger potential reading ranges than passive tags [4,6]. As a consequence, active tags can signal their location in real-time, while passive tags are only registered when they are near a reader. This results in active tags often providing more precise locational information than passive tags. Generally, active tags are heavier and more expensive than passive tags. Different operating frequency classes of RFID systems are distinguished: low frequency (± 134.2 kHz), high frequency (± 13.56 MHz), ultra-high frequency (866-868 MHz (EU) and ± 915 MHz (US)), microwave (> 3GHz) and ultra-wideband (wide range of frequencies, using low-power signals) [2,7]. Microwave and ultra-wideband systems are mainly limited to active tags, and low frequency systems to passive tags [2,4,8]. The operating frequencies differ in their sensitivity to metals and water in the environment, and have different data reading rates and ranges (Figure 1). When assessing the value of different RFID systems for tracking and monitoring of individual animals, several requirements need to be considered. For example, multiple animals have to be tracked and monitored at the same time, while housed in groups on site, i.e. on farms. Also, the animals should not be affected by the RFID tag, either physically or in terms of behaviour, and the systems should be able to function with little or no interventions or adjustments after implementation, to measure undisturbed behaviour and to be of value for larger scale applications. Therefore, we need a system that 1) can read a large number of tags in a short time frame; 2) is not strongly affected by interference from water or metal; 3) has small, lightweight tags (especially of importance for smaller animals), and; 4) does not require frequent replacement of batteries. It is concluded that passive high-frequency (HF) RFID systems appear to be best applicable for monitoring individual animals housed in large groups, based on the characteristics described in Figure 1. However, what type of RFID system is preferable strongly depends on the desired traits to be measured and the characteristics of the animals under study.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

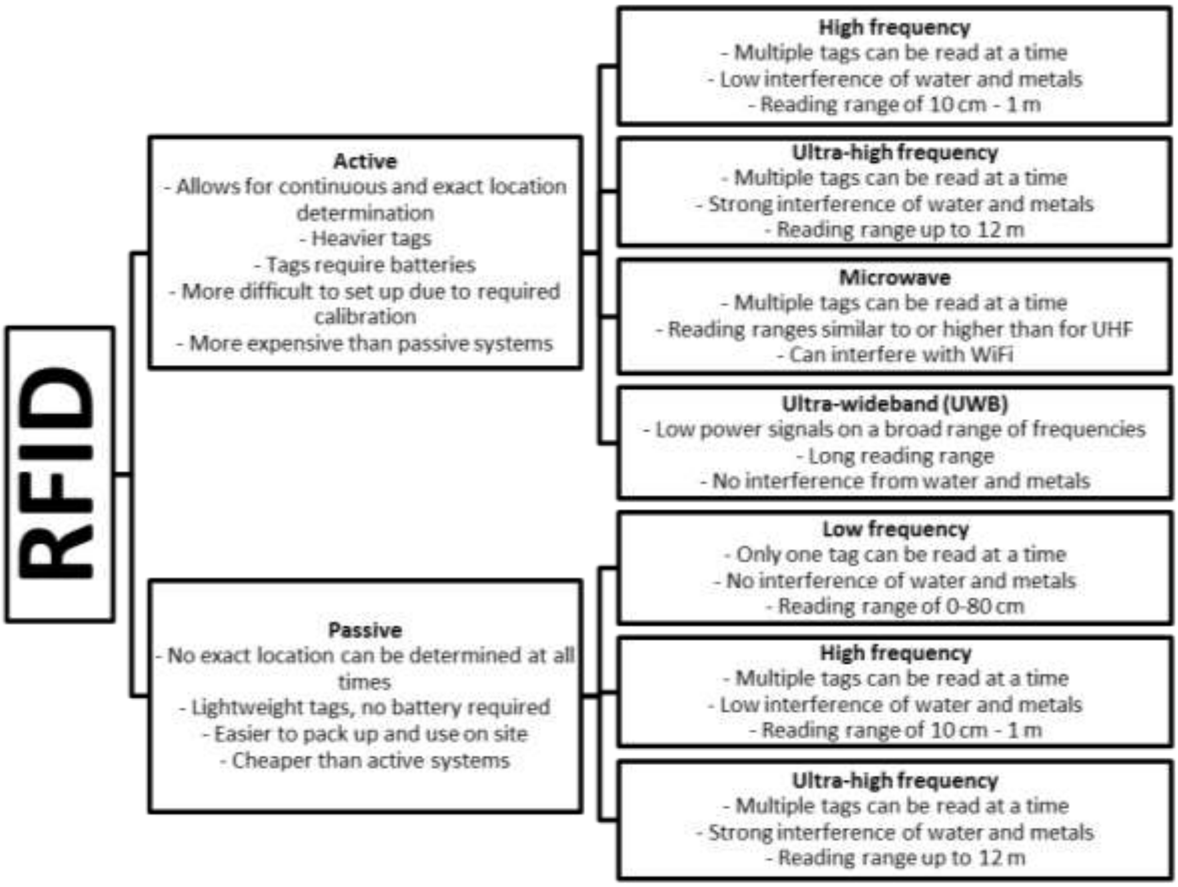


Figure 1: Overview of the different possible RFID systems, based on [2,4,7,8].

Animals in livestock production systems are kept in increasingly large groups, which makes identifying individuals difficult and monitoring their health and behaviour time-consuming [9]. However, information on individual traits is crucial for adequate monitoring of health and welfare of the animals, and is essential for breeders as input for breeding programmes and for evaluation of animals in group-housing. Therefore, in our proposed study, we aim to monitor health, production, and behavioural traits of individual animals of different livestock species that are housed in large groups, using radiofrequency identification systems. We will focus on two main areas: 1) activity in poultry, and; 2) proximity in pigs. For the activity studies in poultry, we will apply passive HF RFID to track and monitor individual poultry in small groups, possibly in combination with other systems (e.g. accelerometers). To this end, we will construct a grid of HF RFID antennas under the flooring of the pen in which the animals are housed and all animals will be fitted with a passive HF RFID tag on their leg rings. By looking at the number of antennas visited or the time in between registrations at different antennas, the level of activity of individual animals can be determined. After applying the system in small groups in a controlled setting, we will apply the system in a commercial breeding setting, to investigate the on-farm application of RFID. Using the proposed passive HF RFID system, differences in activity between genetic lines will be studied, as well as links between activity and gait score of the animals. To study proximity (i.e. interactions between individuals) in pigs, ultra-wideband (UWB) tracking methods will be applied. The reasoning behind this is that pigs are larger animals and are thus able to carry the weight of active tags. Also, using UWB tracking, a more exact location of individual animals can be determined, which is of importance in assessing the distance between individuals. By looking at the distances between different dyads of animals over time, a social network can be constructed. UWB tracking can also be combined with video tracking to study problem behaviours, by looking at the type or direction of interaction (e.g. head-to-tail interaction may indicate tail biting in pigs). Overall, the data obtained with these RFID systems can help to detect problem behaviour and/or gait problems early on and can aid in precision phenotyping. In this way, we hope to contribute to breeding healthy animals with good welfare that perform well in group-housing systems.

References:

1. Debouzy, J.-C., Perrin, A. (2012). RFID. In Perrin, A. & Souques, M. (Eds.), *Electromagnetic Fields, Environment and Health* (pp. 81-87). Paris, France: Springer-Verlag France.
2. Finkenzeller, K. (2010). *RFID handbook: fundamentals and applications in contactless smart cards, radio frequency identification and near-field communication*. Chichester, United Kingdom: John Wiley & Sons.
3. Rossing, W. (1999). Animal identification: introduction and history. *Computers and Electronics in Agriculture* **24**, 1-4.
4. Weis, S. A. (2007). RFID (radio frequency identification): Principles and applications. *System* **2**, 1-23.
5. Sower, V., Green, K., Zelbst, P., Thomas, M. (2012, 11th of June). U.S. Manufacturers Report Greater RFID Usage. *The RFID Journal*. Accessed on 30th of January 2018. Retrieved from http://www.rfidjournal.com/articles/view?9589/3
6. Ilie-Zudor, E., Kemeny, Z., Egri, P., Monostori, L. (2006). The RFID technology and its current applications. *Proceedings of the Modern Information Technology in the Innovation Processes of the Industrial Enterprises,* ISBN 9638658657, 29-36.
7. Brown-Brandl, T.M., Maselyne, J., Adrion, F., Kapun, A., Hessel, E.F., Saeys, W., Van Nuffel, A., Gallmann, E. (2017). Comparing three different passive RFID systems for behaviour monitoring in grow-finish pigs. In Berckmans, D. & Keita, A. (Eds), *Precision Livestock Farming '17: Papers presented at the 8th European Conference on Precision Livestock Farming* (pp. 622-631). Nantes, France: Self-published.
8. Ruiz-Garcia, L., Lunadei, L. (2011). The role of RFID in agriculture: Applications, limitations and challenges. *Computers and Electronics in Agriculture* **79,** 42-50.
9. Siegford, J.M., Berezowski, J., Biswas, S.K., Daigle, C.L., Gebhardt-Henrich, S.G., Hernandez, C.E., Thurner, S., Toscano, M.J. (2016). Assessing activity and location of individual laying hens in large groups using modern technology. *Animals* **6**, 10.

Measuring behaviour in elite sport

Measuring behavior is essential in sport for improving performance, reducing injury risks and establishing appropriate regulations. Behavioural measurements are used to guide athlete preparation, training interventions and inform strategy going into a competition; as well as to assess the impact of changes to equipment. Other aspects of athlete behavior, inside and outside the performance arena, are important for athlete health and injury risk reduction. The design of effective safety equipment and procedures is also reliant on accurate measurement of athlete behavior. Traditionally, measurements were constrained to laboratories and expensive motion capture systems, but the emergence of low-cost wearable sensors and camera systems has helped to facilitate in-field testing.

This symposium presents research from a range of sports and considers the impact that measuring behavior has had on elite sport.

Tom Allen, Manchester Metropolitan University

Measuring Performance at the Great Britain Cycling Team: A Case Study of Optimizing Bicycle Setup Parameters

P.R.Barratt

Performance Support Team, Great Britain Cycling Team, UK

Introduction

Cycling speed is governed by power output, and so interventions that act to increase cycling power are highly relevant in the support of elite cyclists. Bicycle setup parameters, such as saddle height, crank length and cleat position, have the potential to influence overall cycling power output by altering the region of the length-tension and force-velocity relationships used by the major power-producing muscle groups. The optimization of bicycle setup parameters therefore represents a valuable area of research with the applied goal of increasing overall cycling power. At the Great Britain Cycling Team, laboratory-based experimental methods and field-based observational methods are combined to investigate how changes in bicycle setup parameters influence cycling power output. The purpose of this abstract and presentation is to describe these two measurement paradigms, together with a discussion on the benefits of addressing research questions by combining two complimentary measurement paradigms.

Methods

Experimental data are collected on a static laboratory ergometer that provides an isokinetic resistance. Two-dimensional kinematic and kinetic analyses are used to describe joint-level mechanics. Using this approach, adjustments in bicycle setup parameters can be assessed in terms of the performance outcome of overall power output, together with a detailed description of the joint-level contributions. Observational field data are collected by instrumenting a bicycle to record one-dimensional crank-level kinetics data. These instrumented cranks are permanently installed on the bicycle to allow data collection at every training session.

Discussion

The experimental laboratory environment offers a highly repeatable and controlled measurement paradigm, which is clearly desirable for measurement precision as well as allowing for a more detailed analysis of the pedalling movement. The extent to which we can assume that data from this environment are representative of the actual performance environment however, is questionable. Specifically, there is a requirement for a cyclist to balance a moving bicycle when cycling on a velodrome, together with a wealth of additional visual, auditory and other contextual infomation. These altered constraints may result in the emergence of different movement co-ordination patterns which may in turn result in different optimal values of bicycle setup parameters. In comparison to the experimental laboratory environment, the observational field environment has a high level of representativeness. Indeed the minor level of athlete interference resulting from the observational field instrumentation means that information can be collected in competition and well as in training. However, the considerable additional variabililty in this measurement environment results in a relatively poor level of repeatability and precision. The experimental laboratory environment and observational field environment are clearly complimentary measurement paradigms. The strength of one is the weakness of the other, and vice versa. Rather than choosing one method to address our research question of optimizing bicycle setup parameters therefore, we adopt a mixed approach by using both paradigms to address our problem of optimizing bicycle setup parameters. In building a weight of evidence in this way we believe we are able to have a greater confidence in our results, and ultimately deliver a better support service to our elite athletes and coaches.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Estimation of performance and infringements during race-walking field tests using a wearable inertial sensor system

T. Caporaso[1], S.Grazioso[1], G. Di Gironimo[1] and A. Lanzotti[1]

[1]Department of Industrial Engineering, Fraunhofer Joint Lab IDEAS, University of Naples "Federico II", Italy.
teodorico.caporaso@unina.it

This study presents a kinematics based system for the assessment of performance and infringements in race-walking. To this purpose, we develop a strategy customized for élite race-walkers based on wearable inertial sensor data. For the evaluation of performance and infringements, five parameters (i.e., loss of ground contact time, loss of ground contact step classification, step length ratio, step cadence and smoothness) are selected. Their assessment is carried out according to the literature on outdoor tests. From these parameters, we derive custom indices, whose representation on a polar diagram allows an intuitive analysis of performance and infringements. One athlete, member of the Italian national team, participates to the experimental validation. He performs four outdoor-field-tests at different velocities. Data are collected using an inertial sensor placed at L5/S1 vertebrae. The results of this work are promising toward using the derived indices and the wearable system for assisting training and judgment in race-walking.

Keywords: race-walking, performance, infringements, field tests, wearable sensors

Introduction

Race-walking is a historical discipline which is part of the athletics sporting events. The rule 230 of IAAF Competition rules [1] defined through two commas the possible infringements: "bent knee" and "Loss Of Ground Contact" (LOGC). Thus, these infringements must be avoided but at the same time the best chronometric performance are required.

The current practice during competitions is to entrust the infringements to the subjective human observations made by judges, only with their eyesight. The very short duration of the LOGC events (in the order of few hundredth of second [2]), generates difficulties in a proper identification due to human psychophysiological limitations of vision [3]. Despite the literature underlines how LOGC duration is directly proportional to the velocity [2], the analysis of the performance from the beginning to nowadays shows a continuous increasing of the race speed [4]. In addition, other practical difficulties for the judgments during competition are due to: (i) the restricted period of assessment available for the judges; (ii) the situation in which multiple athletes are close to each other's (this often happens, especially during 20 km élite race). Thus, the judgments are often based on biomechanics patterns (i.e. too much knee lift, interruption of the line "trunk-pelvis-pushing leg) [5]. The consequences of a missed or incorrect disqualification could generate controversies which can be devastating to the individual athlete. On the other side, reaching the best performance while avoiding disqualification is the main goal of every professional racewalker. The literature underlines the correlation between performance and kinematic parameters. In [6] the authors pool together the data from eleven different studies, showing a linear descriptive equation between the Step Cadence (SC), Step Length (SL) and the race-walking speed. Another important parameter is the smoothness. Indeed, studies show that it significantly improves when the athlete's performance level increases [7].

In the context of race-walking, in particular for élite athletes, tools for measuring and monitoring both the performance and infringements are needed. The assessment of the performance is possible in different scenarios, as laboratory or field testing. In laboratory condition (with [8] and without [9] the use of treadmill), many authors have studied performance and infringements (using kinematic, kinetic and energic data) through powerful tools (i.e. motion capture and force platform). However, field data represents the benchmark for the analysis of the gesture since they allow to study the phenomena with the real ground interaction. Field tests allow collecting a larger number of steps, with varying velocities. In this scenario, the critical points are due to the limitations of the available instrumentations as well as to the quality of data and the more variable conditions. For the evaluation of the infringements in field condition the related work present different solution. Authors in [10], proposed a system consisting of a pair of insoles made up of piezoelectric sensor to identify gait temporal event and consequently the LOGC. For the same aim, another device is the inertial sensor, which allows to assess gait temporal events using a kinematics approach [11, 12]. With this approach, the LOGC timing was carried out and two different step classifications (to define legal or illegal steps) are proposed.

For the evaluation of the performance the researchers usually collected data through high speed camera [2]. The video analysis allows to evaluate kinematic parameters (also related with human joint) but often this tool offers a restricted period of assessment. Therefore, in the context of road test, this study presents: (i) the use of a wearable inertial sensor system for the monitoring of both infringements and performance; (ii) the development of an evaluation methodology customized for elite race-walkers; (iii) the derivation of a simple output useful for a quick evaluation in real typical training or competition scenario for the race-walking. The paper presents the selection of five parameters related with infringements and performance. For the assessment of these parameters in field scenario, methods are chosen from the literature. For each one a parameterization with a scale between 0 and 1 is proposed. The results are plotted on polar diagram graph. Finally, an experimental validation shows the useful application of the system in a real filed scenario.

Methodology

In this section we describe the assessment of parameter for the evaluation of infringements and performance and the following representation on a polar diagram.

2.1 Assessment of parameters for infringements and performance

The methodology is based on quantitative parameters for the evaluation of the infringements and performance. The following parameter are chosen for this aim: the LOGC with the assessment of the timing ($LOGC_T$) and the step classification ($LOGC_C$), the smoothness for anterior/posterior linear movement (S), the step cadence (SC) and step length ratio over athlete's height ratio (SLR). The first two parameters are strictly connected with the infringements, while the last three with the performance. According to the IAAF recommendations, the judgments must consider a sequence of steps instead of a single, such that all parameters are related to the mean values of a sequence of 30 steps. The assessment of $LOGC_T$, SC and SLR was carried through the definition of the gait temporal events (toe-off and heel strike) according to [12]. Hence, the previous parameters are evaluated with following equations:

$$LOGC_T = \text{mean}\left(\sum_1^{30} t_{max,i} - \left(t_{min,i} + E\right)\right) \quad (1)$$

$$SC = \text{mean}\left(\sum_1^{30} \frac{1}{(t_{max,i+1} - t_{max,1})}\right) \quad (2)$$

$$SLR = (SC * v_{mean})/h \quad (3)$$

where t_{min} is time instant of the minimum value of the vertical component of the center of mass' acceleration; t_{max} is time instant of the consecutive peak of anterior-posterior acceleration, corresponding to the heel-strike event (HSE), E is a threshold value, which was fixed to 30 ms, v_{mean} is the mean test mean velocity and h is the athlete's height. For the assessment of $LOGC_C$, the steps were classified as 'legal' or 'illegal' according to the classification proposed in [9] as

$$\begin{cases} LOGC_T > 40\ ms & Illegal\ step \\ LOGC_T \leq 40\ ms & Legal\ step \end{cases} \quad (4)$$

So, $LOGC_C$ for each sequence of steps was fixed equal to:

$$LOGC_c = \frac{Illegal\ steps}{30} \quad (5)$$

Finally, the smoothness parameter S is evaluated using Normal Jerk according to [13] through the following equation:

$$S = mean\left(\sum_{i=1}^{30} \sqrt{\frac{(t_{max,i+1} - t_{max,i})^5}{(CS * v_{mean})^2} \int_{t_{max,i}}^{t_{max,i+1}} j^2(t)dt}\right) \quad (6)$$

where j(t) is the jerk related to the anterior/posterior acceleration.

Polar diagram representation

Starting from the previous defined five parameters, each one is associated to a scale between 0 (best score) and 1 (worst score). We considered the feature of inertial device and fixed this criteria correlation for $LOGC_T$:

$$\delta(LOGC_T) = \begin{cases} \delta = 1, & LOGC_T \geq (40\,ms + \frac{2}{SF}) \\ \delta = \frac{1}{40-\frac{4}{SF}} * \left(LOGC_T - \left(40\,ms - \frac{2}{SF}\right)\right), & (40\,ms - \frac{2}{SF}) > LOGC_T > (40\,ms + \frac{2}{SF}) \\ \delta = 0, & LOGC_T \leq (40\,ms - \frac{2}{SF}) \end{cases} \quad (7)$$

where SF is the sample frequency expressed as $1/T_s$ (T_s is the sample timing) of the wearable inertial system.

For the SLR and SC we used the linear regression presented in [2, 6] which, starting from elite competition data, derived the following equations:

$$\{SLR = 2.47 * v + 32.73 \qquad \{SC = 0.259 * v + 2.253 \quad (8,9)$$

From these equations we obtain for all type of race competition the optimal value (SLR_0) and the passing one ($SLR_{0,4}$):

$$\begin{cases} SLR_{0,4}\ with\ v = v_E \\ SLR_0\ with\ v = v_R \end{cases} \qquad \begin{cases} SC_{0,4}\ with\ v = v_E \\ SC_0\ with\ v = v_R \end{cases} \quad (10,11)$$

where v_E is the velocity of the entry standard time for the last World Championship and v_R is the velocity of the world record. All velocities in the equations are expressed in km/h for the equations (8,10) and in m/s in (9,11). Finally, the definition of the ρ(SLR) and γ(SC) are carried through the following system equations:

$$\begin{cases} \rho(SLR) = -0.4\frac{(SLR-SLR_{0,4})}{(SLR_0-SLR_{0,4})} + 0.4, \\ for\ SLR > SLR_0, & \rho(SLR) = 1 \\ for\ SLR < -1.5 * SLR_0 + 2.5 * SLR_{0,4}\,, & \rho(SLR) = 0 \end{cases} \quad (12)$$

$$\begin{cases} \gamma(SC) = -0.4\frac{(SC-SC_{0,4})}{(SC_0-SC_{0,4})} + 0.4 \\ for\ SC > SC_0, & \gamma(SC) = 1 \\ for\ SC < -1.5 * CS_0 + 2.5 * CS_{0,4}\,, & \gamma(SC) = 0 \end{cases} \quad (13)$$

Also, for the assessment of $\mu(S_{AP})$ we used a similar equation where S_{Min} is equal to 1 (ideal best value of smoothness) and S_{Max} was fixed equal to 10 (since no references are provided).

$$\mu(S) = \frac{(S-S_{Min})}{(S_{Max}-S_{Min})} \quad (14)$$

Instead, the parameter $LOGC_C$ is just defined between 0 and 1.

$$\alpha(LOGC_C) = LOGC_C \quad (15)$$

All parameters are shown in a synthetic polar diagram graph shown in the example figure below. The evaluation of the polygon area (Area) allows to obtain a synthetic index (ε) to give in consideration both infringements and performance. This index is expressed as:

$$\varepsilon = \frac{Area}{Area_{max}} \quad (16)$$

where $Area_{max}$ is the maximum possible area (area of a regular pentagon with unitary radius). Furthermore, we fixed the minimum condition of the correct gesture (assuming the threshold values of 0.4 for the infringements parameters δ and α) and we carried the best acceptable ε value (ε_{opt}):

$$\begin{cases} \varepsilon_{opt} = \dfrac{Area}{Area_{max}} \\ \alpha \leq 0.4 \quad \delta \leq 0.4 \end{cases} \tag{17}$$

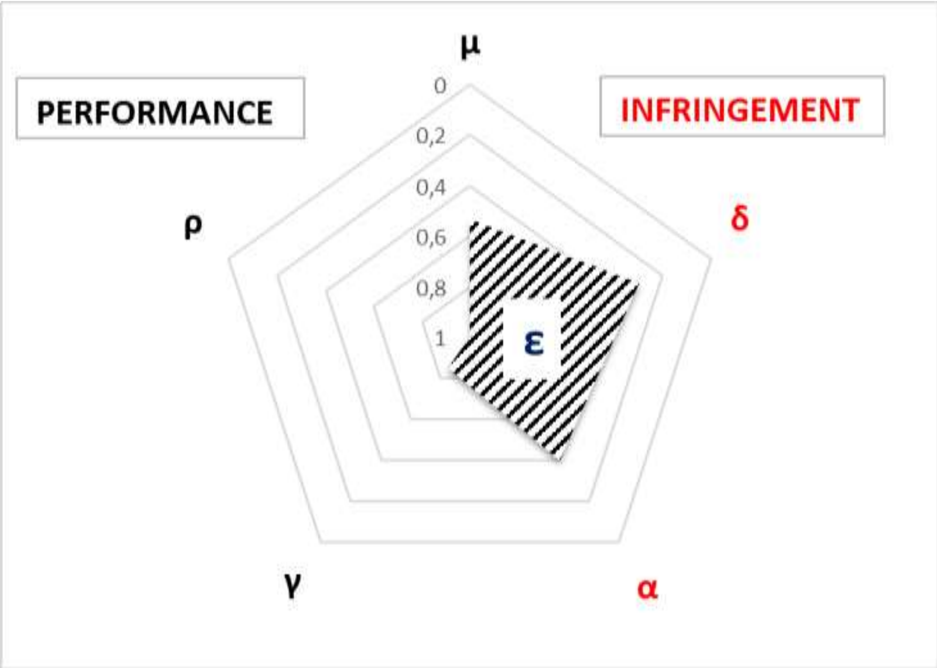


Figure 1 In the figure the red indices are related to the infringements, the black ones to the performance. The underlined area (ε) represents the synthetic index.

Experimental results

In this section the experimental validation phase is described, and the previous indices are evaluated, in this scenario. Furthermore, we highlight the main outcomes of the paper.

Experimental setup

An athlete, specialized on the distance of 50 km, member of the Italian national team, participated to the experimental session. He signed a written informed consent, in accordance with Etic Committee of the University of Naples Federico II. After an initial briefing, the test leader collected racewalker anthropometric data and personal details. Data were collected using an inertial system (i.e. the model type G-Sensor2, BTS) set at SF of 200Hz (1/5ms), ±8g for the tri-axis accelerometer, ±300gps for the tri-axis gyroscope sensor, and ±6 Gauss for the tri-axis magnetic sensor. The sensor was located at the bottom of athlete vertebral column in correspondence of the L5–S1 inter-vertebral space. Trials were performed on a long-paved road, which is straight and flat in according to the IAAF recommendations about race walking courses [10]. The test protocol consists of four trials of three hundred meters race-walking each, at 4 different velocities (mean values of 12.0, 12.9, 13.7, 14.6 km/h that represent, respectively, the 93%, 100%, 106% and 113% of the athlete's racing pace (RP)) evaluated respect the best results of the athlete in the last two seasons. For each race-walking test, excluding the initial acceleration phase of the athlete, 180 consecutive steps (six sequences of step for each trial) were considered.

Results and discussion

Table 1 shows the mean value of each parameter for the four trials (cf. 2.1). In relationship with the race distance of the athlete v_E and v_R are fixed respectively 12.20 km/h and 14.11 km/h. Consequently $SLR_{0,4}$ and SLR_0 are setted equal to 62.8 and 67.6; $SC_{0,4}$ and SC_0 equal to 3.130 step/s and 3.268 step/s. Finally, in Table 2, we report the indices evaluated according to the proposed methodology (cf. 2.2). Figure 2 shows the polar diagram representation of the evolution of all parameter in the different trials. Figure 2 and Table 2, according to the literature [2, 6], show a decreasing trend for infringements indices (δ and α) and, on the other side, an increasing of the performance indices (ρ, γ and μ) when RP increased. Specifically, at greater speed (RP≥ 106%) the infringement scores appear to be critical (with value over 0.4). The analysis of the performance indices underlines how in this athlete, the step length values are better than step cadence ones. Furthermore, we can underline how the smoothness index improve less than the step cadence with the race speed. Finally, ε (as shows in Table 2) underlines how at greater RP the graph area increases. However, this is not the best compromise, since it allows improvements in performance but with critical (incorrect) infringement indices. Thus, probably the optimal RP for this athlete is between 100% and 106%, where is placed the ε_{opt} (that allow to both maximize ε and guarantee an acceptable level of correct technique (with $LOGC_T$ and $LOGC_C$ under 0.4).

Table 1 Performance and infringements parameters data collected during trials

RP [%]	SC [steps/s]	SLR [%]	$LOGC_T$ [ms]	$LOGC_C$ [-]	S [-]
93	3.027	66.5	35	0.14	6.935
100	3.063	70.6	37	0.26	6.815
106	3.165	72.3	41	0.50	5.833
113	3.215	76.1	44	0.50	5.197

Table 2 Performance and infringements indices evaluated starting from collected data

RP [%]	γ	ρ	δ	α	μ	ε [%]
93	0.70	0.09	0.25	0.14	0.66	34.8
100	0.59	0	0.35	0.26	0.64	35.4
106	0.30	0	0.55	0.50	0.54	38.9
113	0.15	0	0.75	0.50	0.47	42.3

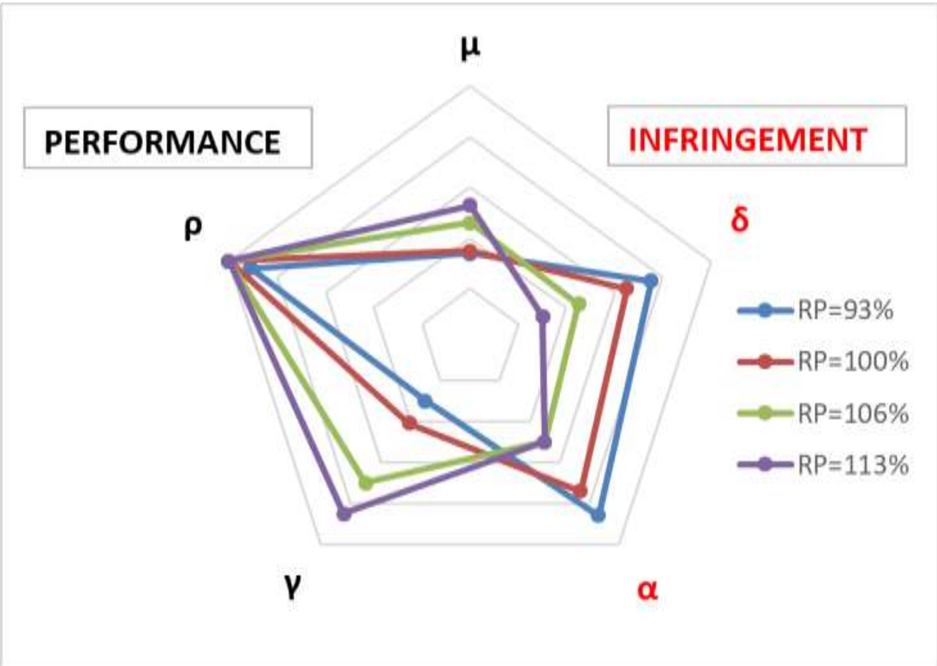


Figure 2 Polar diagram for the experimental tests. The red indices are related to the infringements, the black ones to the performance. The coloured lines graphically show the trend of the indeces at different RP percentages.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Conclusions

The results of this paper agree with the literature and confirm the possibility to use wearable inertial system both for the assessment of infringements and performance in field condition. The proposed method allows to individuate when the race pace becomes critical for correct execution of the gesture. For the analysis of the performance, the polar diagram allows to understand the strong and critical point that characterized the technique of the single athlete. The synthetic index (in combination with an acceptable level of infringement parameters) could allow to individuate the best quality of gesture with the possible optimal race pace speed. Further developments will be centred on: (i) testing with a greater number of athletes with different anthropometric characteristics and also in race conditions; (ii) including a smoothness rotation index in the evaluation parameters; (iii) deriving customized strategies diversified for the main type of race competition (men's and women's 20 and 50 km).

References

1. IAAF (2006). The judging of Race Walking and the organisation of a Race Walking event. A guide for judges, officials, coaches and athletes. 5th Edition.
2. Hanley, B., Bissas, A., Drake (2008). A. The biomechanics of elite race walking: technique analysis and the effects of fatigue. *New Studies in Athletics*; **24(3)**: 17-25.
3. Loschky L.C., McConkie G.W., Yang J., Miller M.E. (2005). The limits of visual resolution in natural scene viewing. *Visual Cognition*;**12(6)**: 1057-1092.
4. Butler, M. (2011). 13th IAAF World Championships In Athletics: IAAF Statistics Handbook.
5. Di Gironimo, G., Caporaso, T., Del Giudice, D.M., Tarallo, A., Lanzotti, A. (2016). Development of a new experimental protocol for analysing the race-walking technique based on kinematic and dynamic parameters. *Procedia Eng*, **147**: 741-746.
6. Pavei, G., Cazzola, D., La Torre, A., Minetti, A. E. (2014). The biomechanics of race walking: Literature overview and new insights. *European journal of sport science*, **14(7)**: 661-670.
7. Choi, A., Joo, S.B., Oh, E., Hwan Mun, J. (2014). Kinematic evaluation of movement smoothness in golf: relationship between the normalized jerk cost of body joints and the clubhead. *BioMedical Engineering OnLine*, **13(1)**:20.
8. Hanley, B. (2015). Gait alterations during constant pace treadmill racewalking. *The Journal of Strength & Conditioning Research*, **29.8**: 2142-2147.
9. Di Gironimo, G., Caporaso, T., Del Giudice, D.M., Lanzotti, A. (2017). Towards a new monitoring system to detect illegal steps in race-walking. *IJIDeM*, **11(2)**: 317-329.
10. Santoso, D. R., Setyanto, T. A. (2013). Development of Precession Instrumentation System for Differentiate Walking from Running in Race Walking by Using Piezoelectric Sensor. *Sensors & Transducers*, **155(8)**: 1726-5479.
11. Lee, J.B., Mellifont, R.B., Burkett, B.J., James, D.A. (2013). Detection of illegal race walking: a tool to assist coaching and judging. *Sensors*, **13(12)**: 16065-16074.
12. Di Gironimo G., Caporaso, T., Amodeo, G., Del Giudice, D.M., Lanzotti, A., Odenwald, S. (2016). Outdoor tests for the validation of an inertial system able to detect illegal steps in race-walking. *Procedia Eng*, **147**: 544-549.
13. Balasubramanian, S., Melendez-Calderon, A., Burdet, E. (2012). A robust and sensitive metric for quantifying movement smoothness. *IEEE transactions on biomedical engineering*, **59(8)**: 2126-2136.

Validity and Repeatability of wrist joint angles in Boxing using the Polhemus markerless motion capture system

Ian Gatt[1], Jon Wheat[2], Tom Allen[3]

1. ian.gatt@eis2win.co.uk United Kingdom English Institute of Sport; 2. J.Wheat@shu.ac.uk United Kingdom Sheffield Hallam University; 3. T.Allen@mmu.ac.uk United Kingdom Manchester Metropolitan University

Introduction

Knowledge of wrist kinematics during the impact phase of punching in boxing is required considering injuries at hand-wrist account for the highest incidence in both training and competition [1]. To date, studies in boxing have investigated shoulder and elbow joint kinematics, but not the wrist [2]. These studies used reflective surface markers, which we considered unfeasible for wrist investigation due to protective equipment covering these areas. We therefore assessed the validity and reliability of a new method for quantifying wrist motion during boxing activities using an electromagnetic tracking system (Polhemus).

Methods

This study included three components; a) a mechanical jig based investigation, using a polyamide hand and forearm shape surrogate, b) an *in-vivo* quasi-static measurement of the wrist, and c) an *in-vivo* measurement of the wrist during boxing punching activities. Three receivers were fixed to either the jig or participants of this study. Participants were 29 Great Britain Boxers (23 men and 6 women). All participants were orthodox stance boxers (left hand leading). The relative orientation of the left upper limb segments (upper arm, forearm, hand) were defined using Cardan angles [3]. For the jig and quasi-static testing, multiple positions of wrist angle were determined during repeat sessions, analysed using the Polhemus. and video footage processed through a computer-based software (Kinoveo). For the punch testing, two types of boxing shots were thrown repetitively against a training bag with their left leading hand; Jab (straight shot) and Hook (bent arm shot). Wrist motion for the punch testing was analysed with the Polhemus.

Figure 1. Field testing showing the Polhemus system set-up with one of the participants during the punch testing component of this study

Results and Discussion

The Polhemus provided valid estimates of wrist kinematics, demonstrating excellent correlation with the marker based video analysis in both the jig and quasi static testing by <0.2° (p = <0.001) and 2-6° (p = <0.001) respectively (Paired *t*-tests). Both systems showed high intraclass coefficient of reliability (ICC), with the quasi-static data also meeting the assumptions of Bland-Altman graphs. For the punch testing, the Polhemus testing showed an almost

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

perfect reliability when assessing wrist motions during both repeated jab and hook shots. This study therefore supports future in-field work in this sport aiming at reducing injury risks at the hand-wrist region.

References

1. Loosemore M., Lightfoot J., Gatt I., et al (2016). Hand and wrist injuries in elite boxing: A longitudinal prospective study (2005-2012) of the Great Britain Olympic boxing squad. *American Association of Hand Surgery* 1-7.
2. Piorkowski B.A., Lees A., and Barton G. J. (2011) Single maximal versus combination punch kinematics. *Sports Biomechanics 10 (1): 1–11*
3. Metcalf C. D., Notley S.V., Chappell P.H., et al., (2008) Validation and application of computational model for wrist and hand movements using surface markers (2008) *IEEE Transactions on Biomedical Engineering 55 (3): 1199-1210*

Monitoring and encouraging participation in sport and physical activity

Technology is playing an increasingly prominent role in sport and physical activity. Many of us wear or carry sensors which track our movements and monitor our heart rate — whether we're running in the park or competing at the elite level. At the same time, the Western World is facing a public health crisis. Levels of obesity (in adult and childhood populations) are predicted to increase, alongside associated physical and mental health conditions. Exercise and physical activity is one way to increase public health and has been shown to have positive effects on a number of physical and mental conditions. Technology can be our ally when encouraging others to exercise. It can be used to create new environments for exercise or to reveal effective ways to encourage participation.

This symposium explores the ways we can meet the challenges of increasing physical activity and encouraging participation.

Simon Choppin & John Kelley, Sheffield Hallam University

Representative Movement Behaviour Measurement in Sport: Gold Standard Measurement versus Gold Standard Data

Emma Anderson [1,3], Joseph Stone [2], Mohsen Shafizadehkenari [2], Marcus Dunn [1], Ben Heller [1]

[1] Centre for Sport Engineering Research (Sheffield Hallam University); [2] Academy of Sport (Sheffield Hallam University); [3] Lawn Tennis Association

Newell's constraints-led model [1] suggests that boundaries or instructions (constraints) placed on the task, performer and environment affect the affordances (opportunities for action) of movement production. Therefore, when measuring movement behaviour, the data recorded represents the interaction between the imposed constraints in those experimental conditions. Here, using the theoretical framework of ecological dynamics (i.e., constraints-led model and representative design) we propose the idea of 'gold standard data', meaning data collected in an everyday sport environment with no interference to the typical constraints of the performer, environment or competitive/training task. We propose that a continuum exists between 'gold standard data' and 'gold standard measurements' (i.e. highly accurate systems that can measure kinematics with precision which are often used to validate other technologies [2, 3]); suggesting this continuum as a method to help movement scientists reflect on the measurement techniques used in research and applied practice (see Figure 1). Finally, we suggest the term 'co-operative data' to describe data collected during experiments that negotiate a balance between measurement precision and representative design (see Figure 1).

This continuum has been informed by a systematic survey of human movement literature, where the most common method identified for measuring movement behaviour during sporting tasks was marker-based motion-tracking (19% of sample). Researchers may consider this type of system as the 'gold standard', based on its ability to measure kinematics with precision and reliability [2, 3]. However, to enable marker-based motion-tracking systems to achieve high accuracy and resolution, invasive measurement techniques that are impracticable within everyday sports performance environments are likely to be implemented. To measure athlete movement in everyday performance conditions, investigators may use methods that are not 'gold standard' but impose little or no impedance to the athlete, task or environment. This was evidenced in our survey by global positioning systems (GPS; 16% of sample) and notational analyses (8% of sample) being the subsequent most popular techniques used to measure human movement behaviour during sporting tasks.

Depending on the resources available and the performance constraints of the sport, some experiments may be able to collect 'gold standard data' using 'gold standard measurement' systems (see [4]). However, the current literature demonstrates a prevalence of experiments that use methods (for example, marker-based motion-tracking) likely to introduce less representative constraints; and compromise action fidelity (that experimental situations adequately simulate a performance system) and functionality of task design (maintenance of the coupling of cognitive, perceptual and action processes) [5], making experimental data inappropriate for practitioners or researchers to generalize to everyday sport conditions. Hence, representative but less precise data may be more informative than data collected using gold standard measurement techniques; because it represents the constraints experienced by athletes as they perform training or competitive tasks in typical conditions. Careful consideration is therefore required concerning the prescribed task, experimental environment and measurement techniques employed during data collection [5, 6]. We suggest that the use of the terms 'gold standard measurement', 'co-operative data' and 'gold standard data' might help researchers to contextualize experiments and help practitioners to appropriately generalise data in relation to typical sports performance constraints.

R.A Grant *et al.* (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

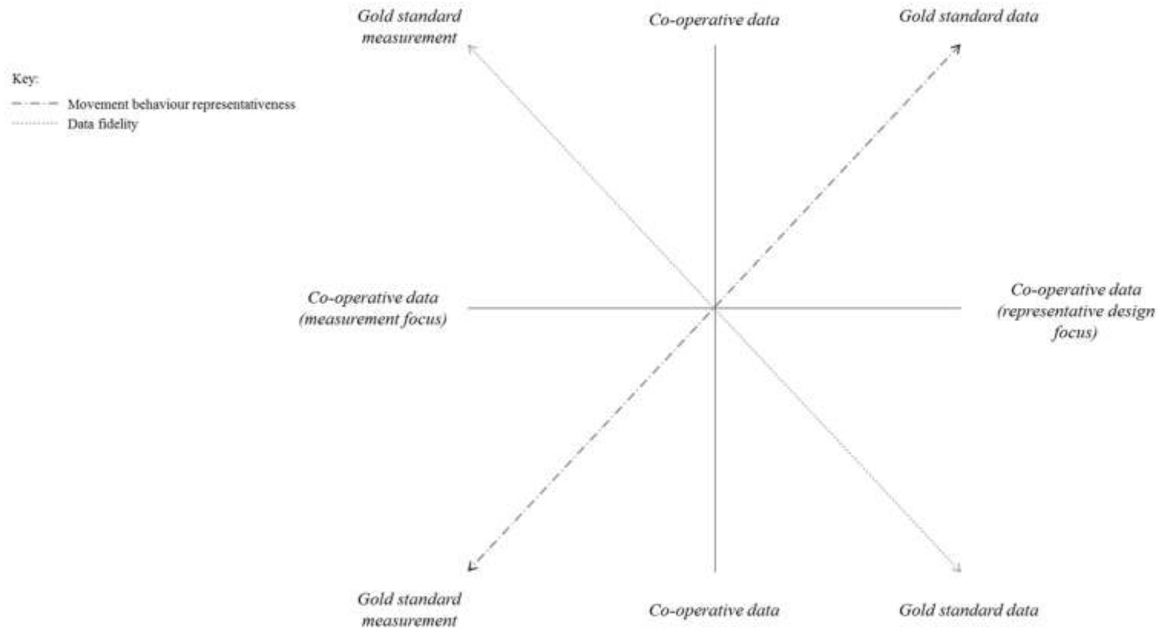


Figure 1. Proposed continuum of 'gold standard measurement' and 'gold standard data'

References

1. Newell, K.M. (1986) Constraints on the development of coordination, in Motor Development in Children: Aspects of Coordination and Control, M.G. Wade and H.T.A. Whiting, Editors. Springer Netherlands. p. 341-360.

2. Ogris, G., et al. (2012) Accuracy of the LPM tracking system considering dynamic position changes. J Sports Sci **30**(14): p. 1503-11.

3. Kelley, J., D. Higham, and J.S. Wheat (2016) Technology for studying interpersonal coordination in social collectives, in Interpersonal coordination and performance in social systems, P. Passos, K. Davids, and J.Y. Chow, Editors. Routledge: Abingdon, Oxon; New York, NY. p. 305-321.

4. Rabita, G., Dorel, S., Slawinski, J., Sàez-de-Villarreal, E., Couturier, A., Samozino, P., & Morin, J. (2015). Sprint mechanics in world-class athletes: A new insight into the limits of human locomotion. Scandinavian Journal of Medicine & Science in Sports,25(5), 583-594.

5. Pinder, R.A., et al. (2011) Representative Learning Design and Functionality of Research and Practice in Sport. Journal of Sport and Exercise Psychology **33**(1): p. 146-155.

6. Hristovski, R., K. Davids, and D. Araújo (2006) Affordance-controlled bifurcations of action patterns in martial arts. Nonlinear Dynamics, Psychology, and Life Sciences **10**(4): p. 409-444.

Non-invasive Motion Capture of Swimmers

G. Ascenso[1], C. Payton[2], T. Allen[1], and M. H. Yap[3]

1. School of Engineering; 2. Department of Sports Science; 3. School of Computing, Maths and Digital Technology, Manchester Metropolitan University, Manchester, UK. guido.ascenso@stu.mmu.ac.uk; t.allen@mmu.ac.uk

Motion analysis studies of swimming have focused predominantly on 2-D kinematics (i.e. joint angles, velocities, and accelerations in the sagittal plane) [1]. However, swimming actions involve movements in all three planes, making a 2-D approach prone to considerable perspective errors. A more complete analysis of swimming technique would require 3-D kinematical data, which would eliminate this source of error and provide more detailed information [2]. Such data, however, are challenging to obtain in a swimming pool and have rarely been captured satisfactorily. Therefore, the aim of this project is to develop a tool that will allow the measurement of whole-body 3-D kinematical parameters without the aid of markers.

The visual hull algorithm will constitute the core of this tool. First proposed by Laurentini in 1994, the visual hull approximates the volume of a 3-dimensional object by using its silhouettes (recorded by cameras positioned around the capture volume and extracted from the images using specialized algorithms) as constraints for where the object can lie in 3-D space [3]. Because the accuracy of this reconstruction is strongly dependant on the quality of the silhouettes seen by the cameras, silhouette extraction is a fundamental step for this algorithm. Therefore, in the first phase of the project the current state of the art algorithms for contour detection have been compared quantitatively and qualitatively, using data obtained from participants fully submerged and recorded from different camera angles. Because contour detection algorithms do not output semantically meaningful closed lines [4], but rather give a probability that a pixel belongs to a contour or not, additional algorithms that are able to connect these contours into closed lines (i.e. silhouettes) are currently being explored.

In a second phase, having selected an optimal contour detection algorithm, the visual hull of complex objects (i.e. humans) will be reconstructed in a dry-land setting to gain familiarity with the algorithm without the added complexity of underwater capture. The visual hull, which by definition is an over-approximation of the actual volume of the object, will then be refined using available algorithms, which use photometric constraints to further "carve" the visual hull into a more accurate model. Specialised software (MAYA) will then be used to articulate the visual hull, inserting virtual joints into the 3-D model. We will then be able to compute kinematical parameters such as joint angles, velocities, and accelerations.

Having tested the entire pipeline in a dry-land lab, it will be adapted for underwater capture. In particular, camera calibration will have to take into account the different refractive index of water and the possible presence of a glass housing for the cameras, which introduces a second degree of refraction.

Ultimately, this tool would represent the first method to record high-quality 3D kinematics of swimmers in a completely non-invasive fashion, thus possibly enabling a deeper understanding of the technique and performance of elite and recreational swimmers. It will also be the first study to compare how the choice of different silhouette extraction techniques affects the accuracy of the visual hull.

References

1. Toussaint, H. and Truijens, M., 2005. Biomechanical aspects of peak performance in human swimming. Animal Biology, 55(1), pp.17-40.
2. Bartlett, R., Wheat, J. and Robins, M., 2007. Is movement variability important for sports biomechanists? Sports biomechanics, 6(2), pp.224-243.
3. Laurentini, A., 1994. The visual hull concept for silhouette-based image understanding. IEEE Transactions on pattern analysis and machine intelligence, 16(2), pp.150-162.
4. Arbelaez, P., Maire, M., Fowlkes, C. and Malik, J., 2011. Contour detection and hierarchical image segmentation. IEEE transactions on pattern analysis and machine intelligence, 33(5), pp.898-916.

Examining continued participation in Parkrun

Simon Choppin[1], John Kelley[1], Jeff Breckon[1], Mike Graney[2], Steve Haake[1]

1. Sheffield Hallam University, Sheffield, UK; 2. Parkrun; s.choppin@shu.ac.uk

The proportion of the UK, adult population that meets official activity guidelines is low. Health interventions have been implemented to try to increase activity in the general public but evidence of success is limited -- high proportions of participants return to relative inactivity soon after starting.

As a weekly, free, 5 kilometre run, Parkrun may contribute to increasing physical activity among its community members. Parkrun provides a relaxed setting in which to exercise which mixes elements of social-support and perceived accessibility with the opportunity to set goals and self-monitor progress (the runs are timed and recorded for participants) -- both have shown to be effective in maintaining behaviour change with regards to physical activity.

This study examines the participatory behaviour of 1.33 million registrants to the Parkrun event. This includes the effects of demographics, what level of engagement is likely to result in continued participation and the physical effect (through run time) on continued participation.

While a large proportion of registrants to Parkrun will only run once, those that do continue to participate for three years or more demonstrate an increase in physical fitness through improved run-times. The age, level of deprivation and activity levels of registrants all affect the likelihood of continuing to participate in the event. Given the positive benefits of participation in Parkrun, ways should be explored to increase the appeal to less active and more deprived registrants.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Multivariate Body Area Network of Physiological Measures “In the Wild”: A case study with zipline activity

D. Dupré[1], N. Andelic[2], D. Moore[1], G. Morrison[2] and G. McKeown[1]

1 School of Psychology, Queen’s University Belfast, Belfast, United Kingdom. {damien.dupre,dmoore33,g.mckeown}@qub.ac.uk 2 Sensum, Belfast, United Kingdom. {n.andelic,gawain}@sensum.co

With the development of wearable sensors and algorithmic improvements, it is now possible to assess the dynamic progression of physiological rhythms such as heart rate, breathing rate or galvanic skin response in ways and places and during certain activities that were previously not possible due to cost and reliability. This paper investigates the physiological changes when participating in a zipline activity. Despite the advances in sensor technology, the statistical analysis of such physiological signals remains a challenge for data analysts. This paper presents a workflow encompassing the whole process with the goal of obtaining a range of best fitting models to analyse the patterns given by these measurements ‘in the wild’. Here we observe significant changes in heart rate, breathing rate and skin conductance level in anticipation, during and after the zipline activity.

Introduction

Despite technical advancements of wearable biometric sensors, little is known about the psychophysiological modifications triggered by physical activities [1]. This is particularly true in the case of physical activities ‘in the wild’ which also involve emotional experiences such as skydiving [2] or kitesurfing [3]. Among the different outdoor sports to analyse, zipline activity has particularities which make it suitable to investigate physiological changes. Zipline activity can trigger intense emotions due to the speed and height, whilst also being a controlled activity: The participants are performing the same quasi-linear pattern and are unable to deviate from it. Sensorial inputs (mainly visual but sound and vibrations can also be used) are appraised differently by individuals and are triggering specific emotional reactions [4]. People use a combination of experience and individual skills in emotion regulation to manage the potentially elicited danger as quickly as possible.

During this process, changes in the physiological rhythms can be observed. These changes can be antecedent or consequent to an event [5]. Antecedent physiological changes relate to the behavioural modifications before the event. In fact, physiological activation facilitates the accurate response to the event [6]. Increase in heart rate and breathing rate leads to better blood irrigation of the muscles to provide the best behavioural response to the triggering event. Skin temperature changes and galvanic skin response are typically a side effect of the increase in heart rate and breathing rate. They aim to regulate and limit the physiological changes by cooling the body. Thus antecedent physiological changes are triggered not only by the comparison between sensorial inputs and behavioural response expectation but also by the uncertainty of the results of the future behavioural response. Consequent physiological changes happen because of changes in the physiological rhythms after the event. These changes are the results of the behavioural response to the event in order to adjust a new accurate behaviour depending on the consequences of the previous behaviour [7]. Indeed, these physiological changes infer homeostatic deviations which are associated with different psychological states [1,8].

By analysing the dynamic changes of physiological measures simultaneously, it is possible to identify these individual’s triggered reactions. Consequently, Zipline activity provides us with an opportunity to measure a high-intensity controlled activity “in the wild”.

Multivariate Body Area Network

As they allow continuous and real-time monitoring of physiological changes, wearable sensors are efficient tools for monitoring patients’ physical [9,10] and psychological health [11,12]. The use of wearable sensors to exercise and sport monitoring have become more and more common as the technology of the devices have improved [13]. In fact, there are currently over 300 devices that are commercially available for the purpose of tracking exercise or sport [14]. The sensors can be classed into three categories: contextual measures, direct measures of physiological rhythms, and indirect measures of physiological rhythms. Contextual sensors deliver information about the environment such as the geographic position with geolocalisation data or sensing movement with accelerometer data. Direct measures of physiological rhythms provide the information of the frequency and

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

intensity of these rhythms such as with electrocardiogram (ECG), heart rate (HR) or breathing rate (BR). Indirect measures of physiological rhythms provide information of the general bodily activity such as galvanic skin response (GSR) or skin temperature (ST).

In order to better understand physiological changes, "Multivariate Body Area Network" was used since a long time. Multivariate Body Area Network describes a setup in which multiple data stream sensors are simultaneously recording physiological changes. Most common academic setup involves wired sensors and is used in laboratory conditions. However the rise of wearable devices made the use of Multivariate Body Area Network easier to evaluate physiological changes in the wild.

A major issue when analysing physiological changes in the wild is the measure and the synchronization of multivariate time-series (i.e. HR, BR, ST and GSR). The time it takes to put sensors on and off can lead to a desynchronisation of the different measurements. In addition to the time desynchronisation, each sensor has a different recording frequency which makes data analysis more difficult. In order to solve these problems, a synchronization application hosted on Android OS was developed by the Sensum Co[1]. With this application, the signals provided by the sensors used in the Multivariate Body Area Network are synced where all data streams start and terminate simultaneously.

Dynamic analysis of multivariate measurements

Physiological time-series are particularly challenging to analyse. The main issue is related to the residual distribution. For example, the existence of a potential pattern in the residuals can indicate that linear models are not suitable for fitting physiological measurements. Therefore, a new model with covariates and random effects should be implemented using Generalized Additive Mixed Models (GAMMs) to fit with non-linear patterns. By estimating the degree of smoothness of a Bayesian spline smoothing using restricted maximum likelihood estimation (REML) [15,16], GAMMs allow the identification of dynamic patterns underlying time-series while taking into account participants' idiosyncratic response.

By using GAMMs to analyse Multivariate Body Area Network measures in zipline activity, it is possible to identify subtle changes nested in individual's idiosyncratic physiological changes and thus, to infer their emotions.

Methods

Participants

After providing their written informed consent, 30 participants (12 females and 18 males, age *M* = 28.3, *SD* = 6.33) volunteered for the study. The recruitment process included a medical check to insure that none of the participant had a history of previous cardiac abnormalities and none of them were on any cardio-active medication. Participants were informed that they can change their mind and withdraw their consent to participate in this study at any time. Two participants decided to not take part in the experiment before the zipline.

Measurements

For this experiment the participants wore an Equivital EQ02 sensor belt, a Shimmer GSR device, a GoPro Hero 4 camera and a Sennheiser lavalier microphone. They were also wearing an iPhone on their right arm to record audio from Apogee app and a One Plus X Android smartphone on their left arm running the Sensum application to control, record and sync the output from the wearable devices (Fig 1). The GPS coordinates of the participants were also recorded with the iPhone.

Out of the available wearable devices, the Equivital EQ02 sensor belt was chosen based on the accuracy and robustness of its measurement. We recorded two categories of measures: the physiological measures (GSR, ECG, HR, BR and ST) and the context measures (Accelerometer and GPS). The context measures were used as predictors of the physiological changes.

[1] https://sensum.co/

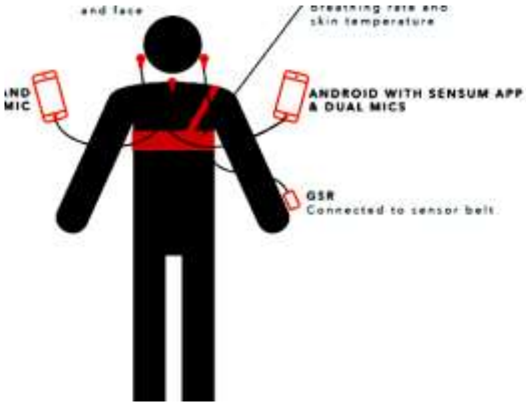


Fig. 1. Description of the Multivariate Body Area Network (MBAN) set up.

Procedure

The experiment took place in Todd's Leap Activity Centre based in Ballygawley, Co. Tyrone, Northern Ireland[2]. The Todd's Leap zipline hangs 50 meters above the ground and is 500 meters long. Prior to participants' activity, they were harnessed up and given full safety advice by qualified instructors. A period of 40 seconds was selected before and after the zipline experience to evaluate the difference between them (Fig. 2).

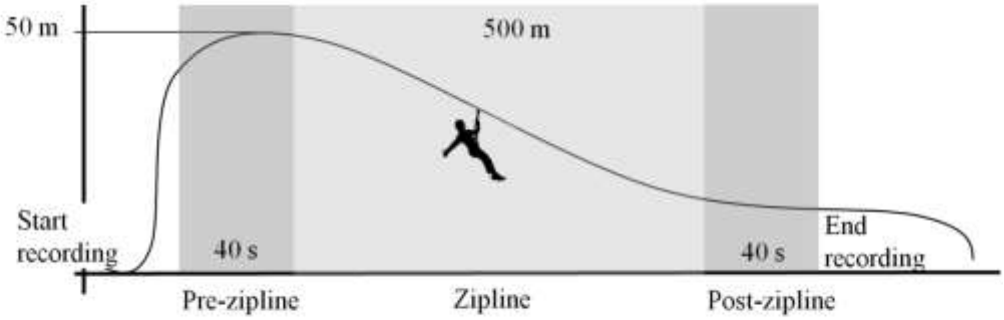


Fig. 2. Timeline of measurements.

Signal Pre-processing

In order to have the cleanest data streams for the analysis, participants' were omitted from further analysis if one or more physiological measurement reached exclusion criteria [17]. See Table 1 for a detailed description of exclusion criteria.

Table 1. Exclusion criteria given by unreasonably low or high physiological measures. bpm = beat per minute, rpm = respiration per minute, °C = degree Celsius, mS = mirco Siemens

Measurement	Exclusion criteria
Heart Rate (HR)	HR < 50 bpm or HR > 220 bpm
Breathing Rate (BR)	BR < 10 rpm or BR > 60 rpm
Skin Temperature (ST)	ST < 30°C or ST > 40°C
Galvanic Skin Response (GSR)	GSR < 2mS or GSR < 20mS

Following these criteria, participants r05, r12 and r23 were removed because of too extreme HR, participants r11 and r16 were removed because of too extreme BR and participants r09, r10, r11 and r20 were removed because of too extreme GSR.

Results

In order to evaluate how the zipline context influences participants' physiological responses, we first used descriptive statistical analysis. After removing outliers and artefacts, we used a GAMM analysis to evaluate significant physiological changes across the participants.

To analyse the ECG signal, an R-peak detection algorithm was applied. Heart Rate Variability (HRV) is extracted from the R-peak detection using a frequency-domain analysis technique with least asymmetric Daubechies

[2] http://www.toddsleap.com/

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

wavelets of width 8 samples for a higher temporal resolution. The High Frequency HRV has been identified as a relevant feature to extract [18]. Indeed, the power temporal evolution in the High Frequency band is correlated with participant's emotional state [19,20].

The GSR also provides relevant features to analyse participants' psychological state with the extraction of Skin Conductance Level (SCL) and Skin Conductance Response (SCR) [21–23]. The SCL represents the tonic level of GSR which slowly varies according the time, and is a representation of long term responses to an event. The SCR represents the phasic response of GSR. It is compound by GSR peaks corresponding to immediate events.

From descriptive statistics, we observed homogeneous variations in the participants' physiology (Fig. 3). HR (M = 116, SD = 28.5), BR (M = 23, SD = 4.7) and SCL (M = 7.80, SD = 3.22) show a consistent pattern in value distribution density across the participants. ST (M = 36, SD = 1.1) and HRV (M = 13480, SD = 22488) appears to have a participant with extreme values but still within a reasonable physiological range. SCR values (M = 0.002, SD = 0.43) are centred around 0, where deviations from 0 indicate a response.

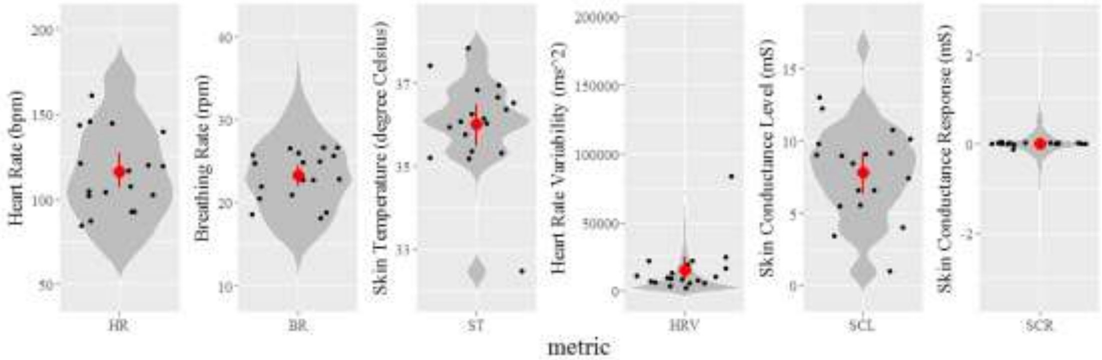


Fig. 3. Dispersion of physiological time-series for every participant (grey area). Small dots indicate participants' average and red dots indicates the overall average.

Multivariate Analysis of Physiological Measures

In order to obtain the optimal structure, it appears that a local regression smoother (LOESS) should be added to the current model. One way to use this LOESS parameter is to implement the GAMMs. This LOESS is a weighted linear regression that enables a fixed factor to have a gradient divided in segments that fits the data for each segment. The new model can be described as follows:

$$Y_i = \alpha + f(X_{1i}) + a_i + \varepsilon_i$$
$$\text{where } a_i = \beta_2 \times X_{2i}$$
$$\text{and } \varepsilon_i \sim N(0,\sigma^2)$$

The new pattern of the GAMM is the LOESS parameter $f(X_{1i})$. Degrees of freedom for the smoother pattern above 1 indicate the importance of the "smooth" term to estimate the variability of the data.

The results provided by the GAMMs for HR ($F(7.36) = 7.80$, $p < .001$), BR ($F(6.82) = 2.37$, $p < .05$), HRV ($F(2.66) = 6.96$, $p < .001$) and SCL ($F(8.63) = 49.40$, $p < .001$) show significant changes according the time whereas ST ($F(1.00) = 0.31$, $p = .58$) and SCR ($F(1.87) = 1.15$, $p = .39$) are stable according the time.

Even though GAMMs allow the assessment of time-series changes, it does not provide a statistical analysis of where these changes happen. Therefore, a Significant Zero Crossing of the Derivatives (SiZer) approach offers a method to identify significant changes in the GAMM predicted values. SiZer methods enable meaningful statistical inference, while doing exploratory data analysis using statistical smoothing methods [24]. A SiZer analysis was performed on the first derivatives of GAMMs with a 99% point-wise confidence interval (Fig. 4).

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

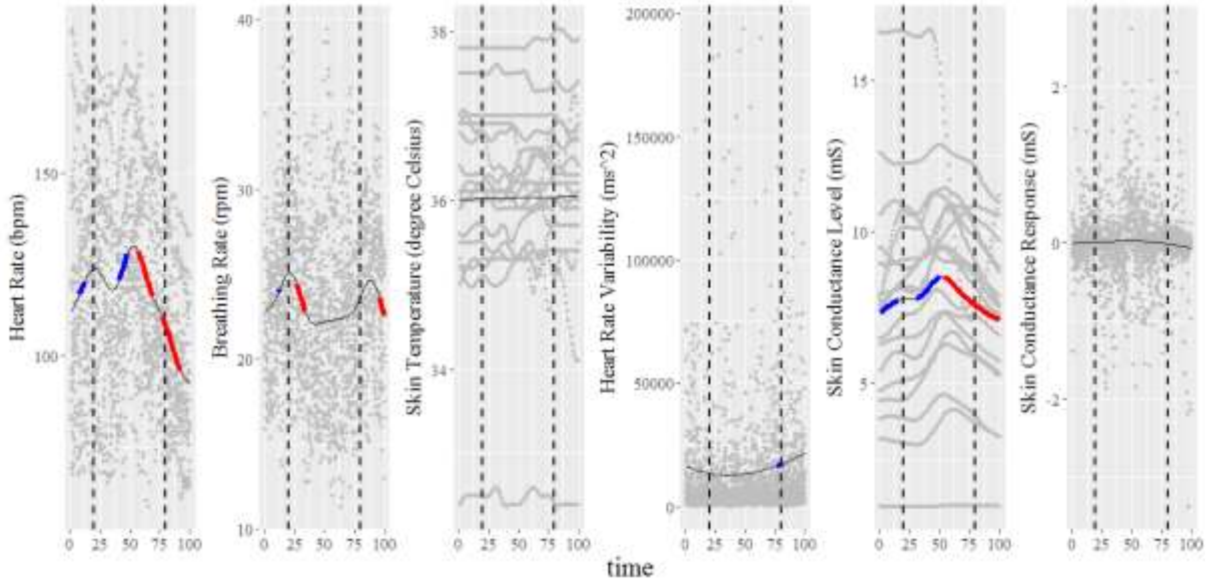


Fig. 4. SiZer analysis of the GAMM predicted values after signal pre-processing treatment and feature extraction. A 99% point-wise confidence interval is shown. Significant periods are extracted from the first derivatives and reported on the actual GAMM predicted values where red periods indicate a significant decrease and blue periods a significant increase. the period between the dashed black lines denotes the zip-line activity.

Using GAMMs and Significant Zero Crossing of the Derivatives methods, results indicate significant local changes in physiological measures. Before the beginning of the zipline, results show a significant increase in HR, BR and SCL indicating an increase in participants' arousal due to the appraisal of the potential 'danger'. During the first part of the zipline, results show a significant increase in HR and SCL as well as a significant decrease of BR. These changes can be explained by the thrill of jumping out and by the zipline acceleration. In the second part of the zipline, with speed decreasing participants' results show a significant decrease in HR and SCL. It should be noticed that HRV increases only when the participants are reaching the platform. This increase can be interpreted as a negative emotional state due to the potential difficulty of reaching the platform. Finally, a significant decrease in HR, BR and SCL can indicate participant's relief after the zipline.

Discussion

In the current study wearable and a prototype controller app (Sensum app) were tested to evaluate physiological changes in Zipline activity. By using this we were able to establish the feasibility of identifying a meaningful data signal in sports environments, and make recommendations on the best technical setup.

Physiological data recorded in the wild

By monitoring physiological changes, it is possible to analyse the optimal patterns in order to infer individuals' physical and psychological response. The data analysed in the current study is not restricted by a lab environment but close to the 'ground truth' of the physiological changes from 'in the wild' activities. The aim of Multivariate Body Area Network is to provide accurate feedback to individuals about their performance, such as an unexpected increase or an expected decrease of the physiological activity. Moreover, wearable technologies have the advantage of being small and unobtrusive which is particularly important so that individuals are not uncomfortable or limited in their performance.

Limitations

Despite technical improvements in multimodal sensor recording and their decreased size, for the belts or the wristband some upgrades can still be made towards their shape and size, particularly in the field of sport monitoring. Indeed, sensors can affect individuals' performance due to their inconvenience (e.g. the sensors can cause discomfort and being aware of the data being recorded can make the individuals uncomfortable). For example, the GSR sensor on the finger may bother or limit the range of the movement for the user. There is room for improvement to create less obtrusive sensors, especially in the field of sport monitoring [25]. For future research, one possibility is to change the sensor placement. For example, the thenar and hypothenar eminence sites could be used instead of proximal phalanx to measure GSR activity. This may reduce distortion in the data due to the sensors rubbing against the zipline equipment.

Recording physiological measurement in the wild is particularly difficult with sport activities. The context itself is a challenge due to the vibration which can interfere with the measurements taken. Contrary to the lab experiments, field experiments bring a very high percentage of artefact and corrupted data. For example, some incoherent data from the GPS, HR, ECG and GSR had to be removed. Other technical limitations such as battery life of the sensors and potential network disconnection provide additional problems that need to be solved.

Finally, even if contextual and technical variables were controlled between the participants, inter-individual variables continue to be a potential limitation. For example, even if the size of the Equivital Belt was adjusted according to participants' anatomy, it is difficult to control for their overall health, alimentation, resistance to cold weather and even their general mood which can influence the recording of physiological measurement.

Conclusions

Wearable technologies are able to provide substantial physiological data, allowing individuals to monitor their performance. In this perspective, zipline appears to be a relevant activity to evaluate the variability of the physiological rhythms during experiences which are highly similar across several trials. As antecedent and consequent physiological changes are related to contextual events, their variability is a relevant indicator of whether athletes' decision making process is correct or incorrect. By using this perspective, physiological measures can provide a baseline for athletes and be used to evaluate their progress.

References

1. Kreibig, S.D., Samson, A.C., and Gross, J.J. (2013) The psychophysiology of mixed emotional states. *Psychophysiology*, **50** (8), 799–811.

2. Cavalade, M., Papadopoulou, V., Theunissen, S., and Balestra, C. (2015) Heart rate variability and critical flicker fusion frequency changes during and after parachute jumping in experienced skydivers. *European Journal of Applied Physiology*, **115** (7), 1533–1545.

3. Bourgois, J.G., Boone, J., Callewaert, M., Tipton, M.J., and Tallir, I.B. (2014) Biomechanical and physiological demands of kitesurfing and epidemiology of injury among kitesurfers. *Sports Medicine*, **44** (1), 55–66.

4. Ellsworth, P.C., and Scherer, K.R. (2003) Appraisal processes in emotion, in *Handbook of Affective Sciences* (eds. Davidson, R.J., Scherer, K., and Goldsmith, H.H.), Oxford University Press, New York, NY, pp. 572–595.

5. Lazarus, R.S., and Folkman, S. (1986) Cognitive theories of stress and the issue of circularity, in *Dynamics of Stress. The Plenum Series on Stress and Coping* (eds.Appley, M.H., and Trumbull, R.), Springer, Boston, MA, pp. 63–80.

6. Zilio, D. (2016) On the autonomy of psychology from neuroscience: A case study of Skinner's radical behaviorism and behavior analysis. *Review of General Psychology*, **20** (2), 155–170.

7. Fanselow, M.S. (2018) Emotion, motivation and function. *Current Opinion in Behavioral Sciences*, **19**, 105–109.

8. Panksepp, J. (2004) *Affective neuroscience: The foundations of human and animal emotions*, Oxford university press, New York, NY.

9. Poh, M.-Z., Loddenkemper, T., Swenson, N.C., Goyal, S., Madsen, J.R., and Picard, R.W. (2010) Continuous monitoring of electrodermal activity during epileptic seizures using a wearable sensor. *International Conference of Engineering in Medicine and Biology*, 4415–4418.

10. Jain, P.C. (2011) Wireless body area network for medical healthcare. *IETE Technical Review*, **28** (4), 362–371.

11. Picard, R.W., Fedor, S., and Ayzenberg, Y. (2016) Multiple arousal theory and daily-life electrodermal activity asymmetry. *Emotion Review*, **8** (1), 62–75.

12. Picard, R.W. (2009) Future affective technology for autism and emotion communication. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **364** (1535), 3575–3584.

13. Etemadi, M., Inan, O.T., Heller, J.A., Hersek, S., Klein, L., and Roy, S. (2016) A wearable patch to enable long-term monitoring of environmental, activity and hemodynamics variables. *IEEE transactions on biomedical circuits and systems*, **10** (2), 280–288.

14. Szczsna, A., Nowak, A., Grabiec, P., Paszkuta, M., Tajstra, M., and Wojciechowska, M. (2017) Survey of wearable multi-modal vital parameters measurement systems, in *Innovations in Biomedical Engineering* (eds. Gzik, M., Tkacz, E., Paszenda, Z., and Piętka, E.), Springer, Cham, Switzerland, pp. 323–329.

15. Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73** (1), 3–36.

16. Lin, X., and Zhang, D. (1999) Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **61** (2), 381–400.

17. Cacioppo, J.T., Tassinary, L.G., and Berntson, G. (2007) *Handbook of Psychophysiology*, Cambridge University Press, New York, NY.

18. Task Force of the European Society of Cardiology (1996) Heart rate variability standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, **17**, 354–381.

19. Godin, C., Prost-Boucle, F., Campagne, A., Charbonnier, S., Bonnet, S., and Vidal, A. (2015) Selection of the most relevant physiological features for classifying emotion. *International Conference on Physiological Computing Systems*.

20. Lane, R.D., McRae, K., Reiman, E.M., Chen, K., Ahern, G.L., and Thayer, J.F. (2009) Neural correlates of heart rate variability during emotion. *Neuroimage*, **44** (1), 213–222.

21. Nakasone, A., Prendinger, H., and Ishizuka, M. (2005) Emotion recognition from electromyography and skin conductance. *International Workshop on Biosignal Interpretation*, 219–222.

22. Boucsein, W. (2012) *Electrodermal activity*, Springer, New York, NY.

23. Öhman, A., Esteves, F., Flykt, A., and Soares, J.J. (1993) Gateways to consciousness: Emotion, attention, and electrodermal activity, in *Progress in Electrodermal Research* (eds. Roy, J.C., Boucsein, W., Fowles, D.C., and Gruzelier, J.H.), Springer, Boston, MA, pp. 137–157.

24. Chaudhuri, P., and Marron, J.S. (1999) SiZer for exploration of structures in curves. *Journal of the American Statistical Association*, **94** (447), 807–823.

25. Peter, C., Ebert, E., and Beikirch, H. (2005) A wearable multi-sensor system for mobile acquisition of emotion-related physiological data. *International Conference on Affective Computing and Intelligent Interaction*, 691–698.

General Sessions

All abstracts are arranged for the general session below alphabetically, by the first author's surname

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

A measure for structural similarities in dyadic conversations

Nikolai W.F. Bode[1]

[1]Department of Engineering Mathematics, University of Bristol, UK. nikolai.bode@bristol.ac.uk

Verbal communication is crucial to human social behavior and it is therefore important to understand the patterns of how conversations are conducted [1]. Much previous work on patterns in conversations relies on manual coding [1], is qualitative [2] or focuses on individuals rather than turn-taking, turn duration or other hallmarks of interaction (e.g. speech recognition [3]). Despite considerable advances, a general understanding of whether conversations patterns depend on factors such as age, gender, familiarity of conversation partners, health status, education status, etc. is still missing. Such population-level questions call for a largely automated exploratory analysis and, most importantly, a way of quantitatively measuring the similarity between different observed conversations. Specifically, we would like to measure whether a given conversation between persons A and B is more similar to a conversation between persons C and D or to a conversation between persons E and F. If a reliable measure for this scenario was available, it could be used in many ways. For example, it could be used for a general investigation of how verbal communication differs across conversation partners. It may also be useful to assess how verbal communication changes with age or even mental health.

Here, we present a measure for the structural similarity of different conversations. Our approach consists of two steps. We describe each of these steps below and illustrate our approach using publicly available audio data on free conversations between pairs of individuals [4]. The main novelty of our approach is that we represent interactions as sequences of labels and measure differences across patterns contained within these sequences. In this way, we measure the structural similarities in verbal interactions, rather than focusing on specific aspects, such as content or turn taking [1].

In the first step of our analysis, we infer who speaks when in a given audio recording, a process called speaker diarization [5]. We split a recorded conversation into 1-second segments (Figure 1A) and label each segment using four labels 1-4 that indicate whether both speakers are silent – 1, speaker A talks – 2, speaker B talks – 3 or both speakers talk – 4 (Figure 1A). Testing the performance of our algorithm using eight manually coded conversations, we find a 20% error rate for labels. For comparison, a review of diarization algorithms found error rates of 5-32% for different data sets [5]. This analysis yields a sequence of labels for each recorded conversation.

In the second step of our analysis, we find the closest localized match between two conversation sequences (Figure 1B, we use the Smith-Waterman algorithm [6]). This match is associated with a score that depends on how long the matched pattern is and how many elements of the pattern are identical across the two sequences. The score for the pattern provides a measure for the structural similarity between two conversation sequences (labels 2 and 3 may be swapped in one of the sequences). All sequences we analyze are the same length and we can therefore normalize the alignment score into a distance measure taking values between 0 (perfect match) and 1 (no match).

We demonstrate our approach on 46 five-minute recordings of conversations [4]. Each conversation is between two people on a topic of their choice. They were recorded by the BBC to archive a snapshot of our relationships and lives today. Using our distance measure, we perform hierarchical clustering on this set of sequences, which leads to two well-separated clusters, as well as a cluster of three outliers (Figure 1C). Focusing on the two large clusters, we find that conversations in one of them contain more silent intervals, fewer intervals in which both conversation partners speak and more conversations conducted by related individuals. There was no difference in the age, gender or speaking time of conversation partners across clusters (all differences significant; assessed using permutation tests).

These preliminary findings suggest that our approach can detect general structural characteristics in large data sets of dyadic conversations. Our bioinformatics-inspired approach can be extended, for example using multiple sequence alignment, to identify specific patterns that are preserved within a cluster of similar conversations. Our analysis is largely automated and thus facilitates a high-throughput analysis of data on verbal communication.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

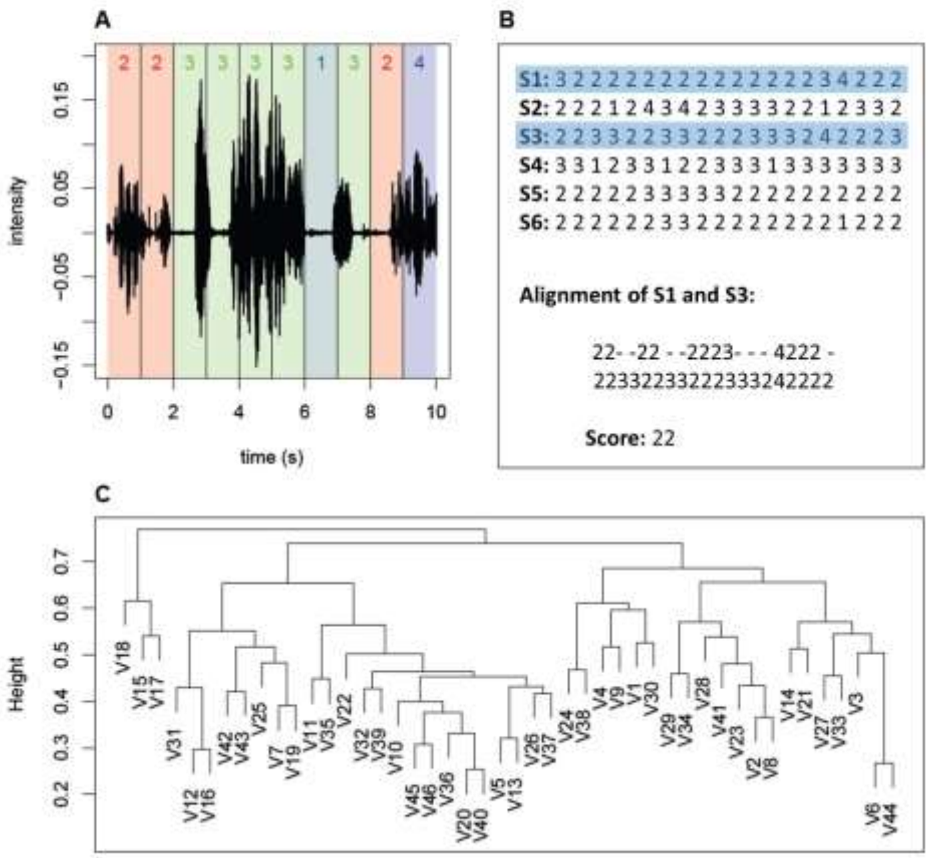


Figure 1: Analysis steps and illustrative results. (A) Audio signal. The colors and numbers indicate labels for one-second intervals obtained using speaker diarization (1 – no-one speaks, 2 – speaker A talks, 3 – speaker B talks, 4 – both speakers talk). (B) Applying the analysis from panel A to audio recordings of many conversations yields a collection of sequences, one for each conversation (illustrated by sequences S1-S6). We perform local sequence alignment for pairs of sequences using the Smith-Waterman algorithm. This produces an alignment indicating the closest local match between the two sequences, allowing for some gaps, '-', in the alignment. We convert the score of the alignment into a distance measure between sequences. (C) Using pairwise local alignment scores, we perform hierarchical clustering on a set of 46 five-minute intervals from free conversations. This can be used to study structural similarities between conversations.

References:

1. Stivers, T., Enfield, N.J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., De Ruiter, J.P., Yoon, K.E., Levinson, S.C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences* **106**, 10587-10592.

2. Hutchby, I. (2008). *Conversation analysis*. John Wiley & Sons, Ltd.

3. Rabiner, L.R., Juang, B.H. (1993). *Fundamentals of speech recognition (Vol. 14)*. Englewood Cliffs: PTR Prentice Hall.

4. Listening Project. BBC, UK (2012) [digital audio files] British Library, C1500. Available online at https://sounds.bl.uk/Accents-and-dialects/The-Listening-Project>. Accessed 10 April 2018.

5. Anguera, X., Bozonnet, S., Evans, N., Fredouille, C., Friedland, G., Vinyals, O. (2012). Speaker diarization: A review of recent research. *IEEE Transactions on Audio, Speech, and Language Processing*, **20**, 356-370.

6. Smith, T.F., Waterman, M.S. (1981). Identification of common molecular subsequences. *Journal of Molecular Biology* **147**, 195-197.

Using Human Operators to Understand Human Operator Behavior: Adding Meaning to Behavior in Complex Environments using Ethnographic Auto-Confrontation

Erwin R. Boer[1], Deborah Forster[2] and Natasha Merat[1]

1. University of Leeds, UK; 2. UCSD USA; e.r.boer@tudelft.nl

Understanding human behavior in rich natural contexts is best explained by the humans doing the actual behaving. Ethnographic techniques such as auto-confrontation are employed to establish what elements and factors influence driver behavior in natural complex driving situations. This technique explicitly exposes how drivers make meaning of their behavior and especially of outlier behavior. We used these data on natural driving behavior to design an autonomous controller that has the same sensitivities to road environment factors as drivers do. These autonomous controllers were then evaluated using the same auto-confrontation technique. The approach offers a richness of understanding that goes well beyond that of typical automotive designers and engineers and finds parallel in many areas of where human behavior can/should be augmented, complemented or automated to enhance safety and enjoyment.

Experiments with human being are generally constrained in three ways: i) a limited number of stimuli or constraints are presented in a sparse world, ii) the possible behaviors in response to the stimuli and constraints are limited, and iii) the instructions to the participants are highly detailed so as to elicit specific behavior. While such experiments are highly suitable to quantify human ability they are less suitable for understanding human behavior in the wild. When observing humans in a natural environment we project our own experience in trying to understand their behavior. However, because everybody has a unique natural make up as well as a unique life experience, their understanding of their own behavior may not match our interpretation of their behavior. These discrepancies are especially relevant when technologies are involved that may require some understanding of the technological capabilities and limitations such as in driver support systems. The question is: how do we capture how different people make meaning of their own behavior.

Ethnography offers a suite of behavior assessment techniques that are highly suitable to let human operators make meaning of their own behavior. One such approach is auto-confrontation where human operators are confronted with re-experiencing a previously exhibited behavior generally grounded in video replay but often enriched with other modalities such as sound or even motion. By exposing human operators to multi-modal traces of their own behavior one can through non-biasing interview techniques gain deep insights into how they organized their behavior and why they behaved the way they did.

In a UK project with the aim to create an autonomous vehicle that drives like a human, we were faced with the question what it means to drive like a human. The operational hypothesis is that when a human is driven autonomously with a car that adopts their own driving style that acceptance will be higher than when they are driven in a fashion that is purely designed from an engineering perspective. We collect data of naturalistic driving in complex driving environments in driving simulators as well as in instrumented vehicles. The question was how we can establish what it is in the driving environment that causes drivers to adapt their speed, lateral position as well as level of attention.

To capture how drivers behave in natural environments we employed three different triggers for what portions we used for the auto-confrontations: i) around their own button presses during their driving which they press when they feel that something out of the ordinary occurred, ii) around stronger and/or faster than normal control responses which we automatically identify and iii) where observed behavior deviates significantly from predictions created by a computational model that is based on pilot data. The data obtained from these auto-confrontations provided not only a set of common factors that influence behavior and that specifically cause difficulty but more interestingly also factors that are unique to specific sub-sets of participants. These data provided the necessary detail to know what elements in the environment autonomous vehicle need to be sensitive to and what additional elements may be needed to achieve a highly personalized autonomous driving. We did identify groups with

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

different driving styles and continue to explore the relationship between how drivers make meaning of their own behavior and how they actually drive.

For each class of driving styles, a personalized autonomous vehicle controller was created. Subsequently drivers were exposed to autonomous driving in their own group's style or in the style of a different group. Drivers again were asked to press a button when they experienced that the autonomous car behaved unnatural or unsafe. The windows of time around these button presses were then analyzed again using auto confrontation to elicit how drivers make meaning of the discrepancy between how they would have driven and how the autonomous car drove. This study was conducted in a highly advanced driving simulator but the same approach will also be employed in the real world.

In the paper we will explicitly explain how auto-confrontation aids in guiding design of human-operator support systems in 4 ways: i) understanding of outlier behavior as recognized by human operator, ii) understanding of outlier behavior compared to observed distributions of situated behavior, iii) understanding of mismatch between model of human operator behavior and actual human operator behavior and iv) understanding of the degree to which a support function (including autonomy) is acceptable. They key is that the driver explains why rather than the arguably biased designer or experimenter. These 4 ways were all employed in our current project on the design of an autonomous vehicle that not only drives smoothly but also has the same risk sensitivities and sensibilities to other-road-user as humans have in terms of adapting their behavior to assure sufficient safety margins and accommodation.

The auto-confrontation technique has proven invaluable to understand the environmental elements that impact behavior in a vast set of different contexts. It also provided clear insight into the reasons why a particular response either from the driver or from the autonomous vehicle was deemed undesirable and therefore providing clear guidance for the design of future autonomous vehicles.

The final paper will provide a full detailed exposition of the stages of auto-confrontation as well as details of what types of insights were obtained as well as how they were used in shaping the design or motivating changes to future design of driver support systems and autonomous vehicles. The paper will finish with an exposition of other domains where these techniques have been or are expected to be useful. We will also explore ways in which these human-grounded approaches may also be suitable in different incarnations to study different species.

Observing the Fine-Scale Behaviour of Cleaner Fish in Commercial Salmon Net Pens Using Passive-Acoustic Telemetry

A. J. Brooker, A. Davie, H. Migaud

Institute of Aquaculture, University of Stirling, Stirling, UK. a.j.brooker@stir.ac.uk

Introduction

In recent years, cleaner fish have been implemented at a large scale in the North Atlantic salmon industry as a strategic component to the sustainable control of the potentially devastating salmon louse *Lepeophtheirus salmonis*. The key to the continual improvement of husbandry practices in any farming system is the observation of the livestock and the ability to rapidly identify deviations from normal behaviour and natural rhythms. However, the monitoring of cleaner fish deployed at low densities within commercial Atlantic salmon net pens is difficult and remains largely anecdotal. We used passive-acoustic telemetry to observe the fine-scale swimming behaviour of cleaner fish in commercial salmon net pens in response to changes in husbandry and environmental conditions.

Methods

We used a passive-acoustic telemetry system (HTI-Vemco Inc., Seattle, WA, USA) to record the positions of acoustic-tagged ballan wrasse (*Labrus bergylta*) and lumpfish (*Cyclopterus lumpus*) over periods of several months. An array of eight underwater hydrophones was deployed by divers around the perimeter of two 24 × 24 × 15 m net pens, comprising four surface hydrophones suspended at 1 m depth and four deep hydrophones fixed at 20m depth. The hydrophones were hardwired to an acoustic tag receiver housed on the cage structure, time synced via GPS satellite and with wireless connectivity to a shore-based PC for system control and raw data storage. The acoustic tags (6.8 × 20.0 mm; 0.55 g in water) emit at a single frequency (307.2 kHz), and each one is programmed with a unique pulse rate interval (PRI) to allow tag identification. The tags were surgically implanted into the abdominal cavity of experimental fish under anaesthetic, and all experiments were approved by the University of Stirling Animal Welfare Ethical Review Board. Control tags were fitted to cleaner fish hides located in the pen corners to measure system performance. Using proprietary software (HTI-Vemco Inc.), raw acoustic data were filtered and processed to identify tags, and the 3D position of each tag pulse to sub-metre resolution was calculated by measuring its time delay to at least four hydrophones and triangulating its position. Fish position data were then analysed using R to derive swimming parameters (speed, heading, turn angle, turn rate), calculate preferred locations, and identify changes in behaviour in response to time of day, tide, season, salmon feeding times and farm activities.

Results and Discussion

Despite regular underwater farm noise (e.g. the predator acoustic deterrent device and workboats) and a multitude of physical barriers (e.g. farm structure, nets, salmon, etc.) that may block or attenuate acoustic tag signals, the system performance was not significantly compromised, as confirmed by validating control tag data. Using this setup, we have characterised the different behaviours of ballan wrasse and lumpfish in terms of (a) spatial patterns, e.g. pen location preferences and home ranges based on kernel utilisation distributions, and (b) temporal patterns, e.g. behaviour periodicity based on continuous wavelet transforms. Furthermore, we have defined behaviour states in individual fish based on changes in behaviour patterns over time and identified individual cleaner fish personality types. Using this information to predict the incidence of different personalities within a given cleaner fish population, we are exploring how to adapt husbandry practices (e.g. by acclimating fish before deployment) to shift this range towards more desirable personalities where delousing is a prominent behaviour. The further development and application of behavioural monitoring through this research programme will inform farm management practices to improve the welfare of cleaner fish populations and the predictability of biological delousing in salmon aquaculture.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Combining two observational methods to describe mealtime episodes in mother-infant dyads during complementary feeding

P. Brugaillères, C. Chabanet, S. Issanchou and C. Schwartz

Centre des Sciences du Goût et de l'Alimentation, AgroSup Dijon, CNRS, INRA, Université Bourgogne Franche-Comté, F-21000 Dijon, France. pauline.brugailleres@inra.fr, camille.schwartz@inra.fr.

Introduction

Studies conducted in adults have shown that individual cumulative food intake curves (based on several consumption parameters like meal duration, bite size, bite frequency and change in eating rate) reflect the eating rate and are stable and consistent within subjects [1]. Yet, a faster eating rate may be a determinant of obesity risk in adults [2]. Considering that eating habits track until childhood and even later until adulthood [3], the description of eating behavior in early life is of importance, especially considering that learning processes related to eating behavior are numerous during this period [4]. Depending on studies, different methods have been used to describe the sequence of eating behaviors during mealtime episodes in children. Thus, videotape recordings or a connected weighing scale which continuously records the weight of plate have been used to analyze eating behavior. For example, in children aged from 10 to 13 years old, some authors assessed the children's eating rate using a connected weighing scale, both in subjects with Prader-Willi syndrome in hospital [5] and in healthy children at schools [6]. Other researchers used the analysis of videotape recordings of meals taken at home to assess the number of bites [7]. Recently, an experimental study conducted in adolescents associated both a videotape recording strategy to identify the type of selected foods in an *ad libitum* buffet and a connected weighing scale to quantify the amount of food eaten [8]. In younger children (4 to 5 years old), the video-recorded data of *ad libitum* lunches taken in laboratory revealed the existence of slow and fast eaters [9]. In this study, the authors went further in assessing the children's eating behavior by describing the microstructure of their eating behavior (i.e., bite size, number of chewing and swallows). In infants younger than one year old, whereas some studies have investigated the eating behavior in formula fed infants [10], no study was performed in complementary fed infants. The main feature of mealtime episodes during complementary feeding is the fact that infants remain dependent on a caregiver to be fed. This specific aspect has to be considered to describe mealtime episodes in mother-infant dyads during complementary feeding. Indeed, the meal initiation and cessation are highly linked to the parent-infant dyad functioning [11]. To start the meal when the infant is hungry and to stop the meal when the infant is full, the parent has to interpret correctly hunger and satiation signals that should be clear enough from the infant [12]. Hence, the infant's satiation signals play a major role in the cessation of the meal but their nature, and thus their clarity, could be highly variable between infants [13, 14]. From the parent's side, the perception of the satiation cues could also be modulated depending on several factors like maternal and infants' characteristics (e.g. sex of the infant, maternal weight status) [11]. In this regard, we assumed that combining two observational methods, the videotape recording and the use of a connected weighing scale, could be a relevant way to describe the dyadic processes occurring in mealtime episodes during complementary feeding. Thereby, the aim of this study was to develop and validate such mixed observational method as a tool to describe dyad functioning. Through this combined approach, changes in different parameters like the eating rate before and after the first refusal to consume, as a first cue of satiation, could be explored to describe whether the parents and the infants are synchronized and responsive to each other.

Materials and methods

Data collection

This study was part of a larger research project conducted from May 2015 to December 2016 and aimed at understanding the etiology of infants' eating behavior. This study was conducted according to the guidelines established in the Declaration of Helsinki and was approved by the local ethical committee (Comité de Protection des Personnes Est I Bourgogne, 2015-A000014-45). This trial was registered at clinicaltrials.gov as NCT03409042. Written informed consent was obtained from both parents. When the infants were nearly 1 year old, the mother-infant dyads were invited to come twice to the laboratory on two non-consecutive days to participate in videotaped meals. For the purpose of the study, the infants were offered a food preload followed by an *ad libitum* meal. Here, we focus on the consumption of the *ad libitum* meal; thus, the preload consumption will

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

no longer be mentioned. Two foods (i.e. 300g of a vegetable and meat or fish puree – here called as 'vegetable puree'– followed by 195g of fruit puree) were offered as the *ad libitum* meal. Each food was served in large opaque bowl, so that mothers could not get an accurate idea of the offered quantities. The infants were seated in a high chair in front of their mother in a dedicated study room. The mothers fed their infant as they usually did (i.e., free to interact), using their own spoon. They were instructed to remain attentive to the refusals to consume emitted by their infant. Indeed, they had for instruction not to encourage nor restrict consumption and to stop offering each food of the meal (vegetable and fruit puree) when the infant emitted 2 consecutive refusals (e.g., the infant refused to open the mouth, pushed away the spoon or shook his/her head). Via a door equipped with a two-way mirror, the experimenters could observe the mother-infant dyads without being seen and thus ensuring that instructions were followed by the mothers.

Connected weighing scale

During consumption, the opaque bowls were placed on a digital weighing scale of 0.1g sensitivity (Adam® PGL-12001 or NBL-4602e depending on the study room) connected to a computer which recorded the weight every second (Figure 1). Hence, mothers were asked to manipulate the spoon without handling the bowl. The experimenter started off and stopped the recording manually for each food. For this, the experimenter made sure that both start and end time corresponded to the same physical condition (i.e., spoon placed into the bowl + food content).

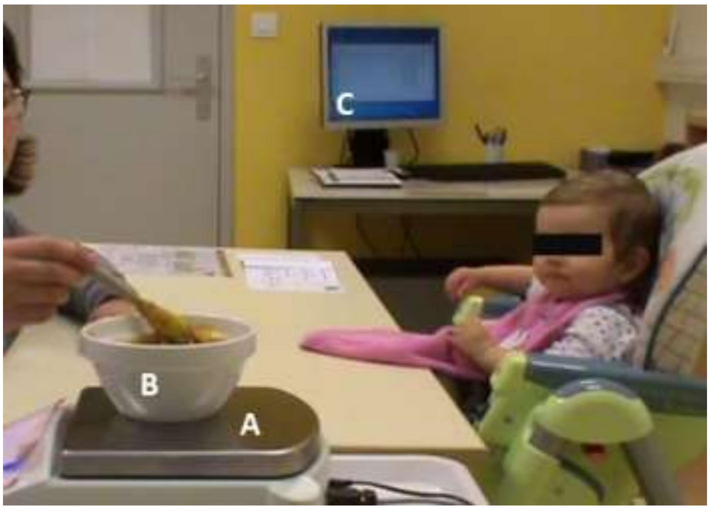


Figure 1. Recording device with the digital weighing scale (A), the bowl (B) and the computer (C).

The main output variables directly obtained were the total weight intake (g) and the total duration of the food consumption (s). Both were respectively obtained by the difference between the final and the initial weight and time recorded. Moreover, the device allowed the collection of numerous other variables like the number of spoonfuls (i.e., each time when the mother took off a full spoon from the bowl) or even the weight of spoonful content and the time at which each spoonful was taken off. Thus, the eating rate could be derived and was expressed in gram or in number of spoonfuls per second.

Videotape coding

All meals were video recorded using two cameras (Sony Handicam® DCR-SX31) filming the infant of the front and profile. The videotape coding was performed by four trained experimenters using the Noldus The Observer® software. For each food, four behaviors and one methodological parameter were coded (Table 1).

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Table 1. Video-coded behaviors, methodological parameter and output variables

	Definition	Coding scheme	Output variable
Behaviors			
1. Consumption duration	The start of the consumption corresponded to the first lip-food contact, when the infant closed his mouth on the first accepted spoon The end of the consumption corresponded to the last lip-food contact, when the last accepted spoon got out of the mouth of the infant	State event	Effective consumption duration (s) (Consumption duration – duration of meal's breaks)
2. Meal's break	The start of each break was coded when the mother stopped to offer the food in response to an external or unusual event (e.g., spilled food requiring to clean the table and/or the infant) The end of the break was coded when the mother went back to feeding her infant	State event	
3. Bite	When each accepted spoon got out of the mouth of the infant	Point event	Number of bites
4. First refusal to consume	First refusal to consume the offered spoon by the infant. This behavior being highly variable between individuals, we established a list of behaviors based on a published list [13]. We completed this list by previewing some videos. Examples of first refusals to consume are shown in Figure 2	Point event	Duration from the start of consumption and the appearance of the 1st refusal to consume (s)
Methodological parameter			
1. First removed spoon from the bowl	Time when the first spoon was removed from the bowl (i.e., this corresponds to the first decrease of weight recorded by the connected weighing scale). This point event was necessary to synchronize data from the videotape with data from the connected weighing scale	Point event	Duration from the start of the videotape and the start of the connected weighing scale recording (s)

Figure 2. Examples of refusals to consume: pushing away the spoon (A), turning the head away (B) and refusing to open the mouth (C).

Data analysis

The analyses were performed using the R software for Windows (version 3.4.0). Statistical significance was set at $P < 0.05$. To exploit the connected weighing scale data, starting from the weights recorded each second, the main task was the selection of one measurement per spoonful (Figure 4). Thus, series of consecutive equal weights (runs) were identified and two runs were collapsed if the difference was lower than 0.55g. Runs corresponding to periods where the spoon was left in the bowl were automatically removed. Then, the measurement at the center of each remaining run was selected. Moreover, the recording was not taken into account if less than two measurements were selected or if the weight range was less than 2g over the meal. Null weights obtained when the scale stopped functioning were detected and removed. Afterwards, the weight of the spoon (manually weighed by the experimenters) was subtracted from the first selected measurement since the spoon was left in the bowl at the beginning of the recording. Finally, when the selected measurements increased between two consecutive values, the first of the consecutive measurements was cancelled. This could occur when the spoonful was not consumed and the next spoonful was smaller. R functions were written to perform all those steps automatically (downloadable R functions on: https://prodinra.inra.fr/record/425131) [15]. Reliability of the connected weighing scale was assessed by evaluating the consistency between the videotape coding data and the connected weighing scale data using both graphical comparisons and Kendall correlations.

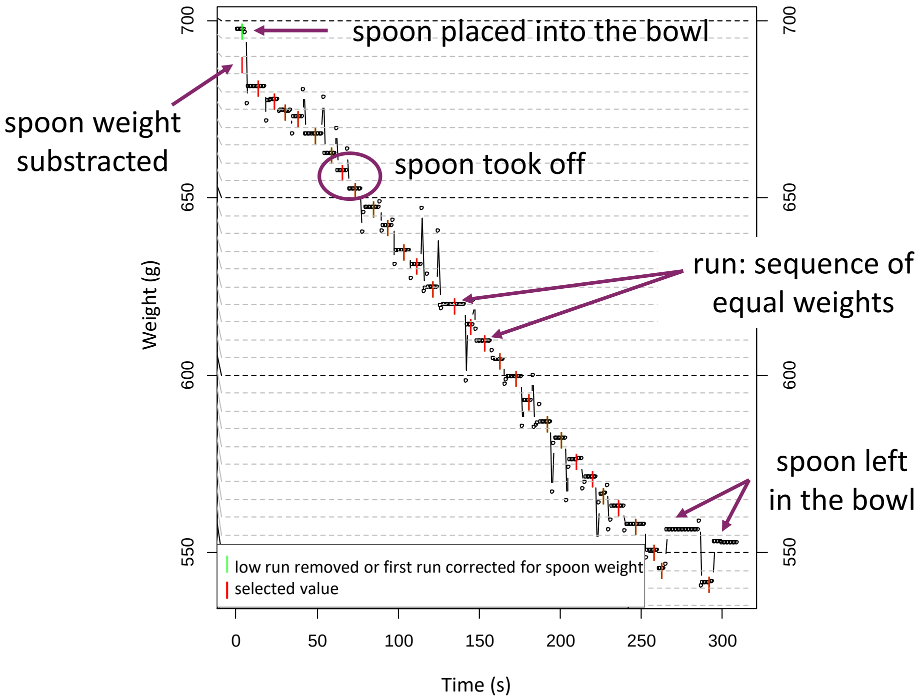


Figure 3. Example of a typical recording from the connected weighing scale illustrating different steps performed by the R functions. Each spoonful is represented by a red vertical trait.

Results

Among the 46 mother-infant dyads who participated in the laboratory meals, videotape and weighing scale data were obtained for n=37. For them, data were obtained from both the videotape coding and from the connected weighing scale for at least one food (vegetable and/or fruit puree) over the two meals. Thus, we analyzed data of 137 food consumption episodes.

Reliability of the data from the connected weighing scale

The results showed good agreement and strong positive correlations between the data obtained from the videotape coding and the connected weighing scale regarding respectively the effective and total duration of food consumption (τ=0.89, df=135, $p<0.001$). We also observed a positive correlation between the number of bites from the videotape coding and the number of spoonfuls from the connected weighing scale (τ=0.89, df=135, $p<0.001$) (Figure 4). These results confirmed the reliability of the developed R functions and the accuracy of the synchronization between the connected weighing scale and the videotape records.

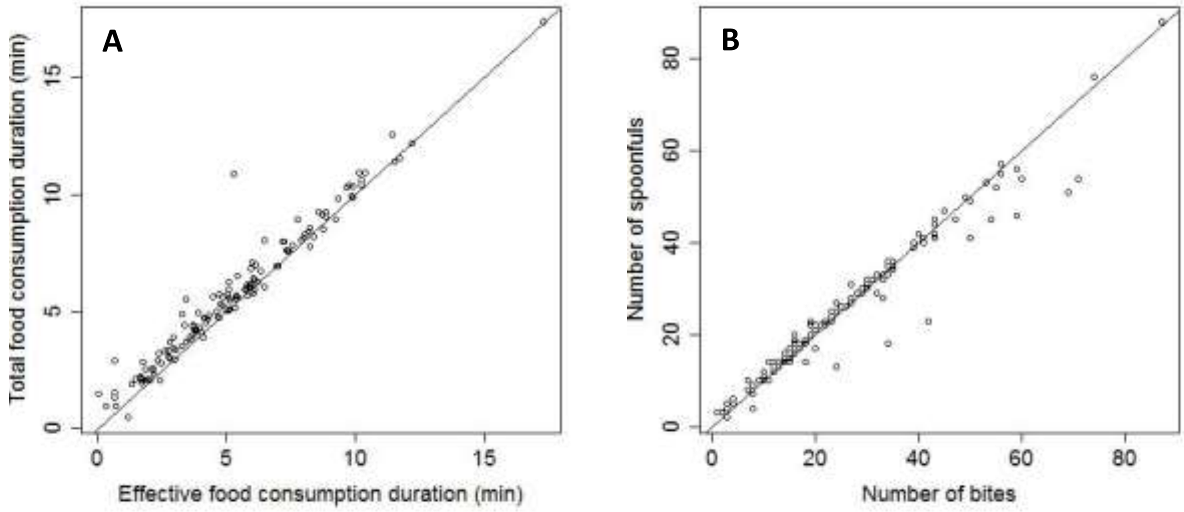
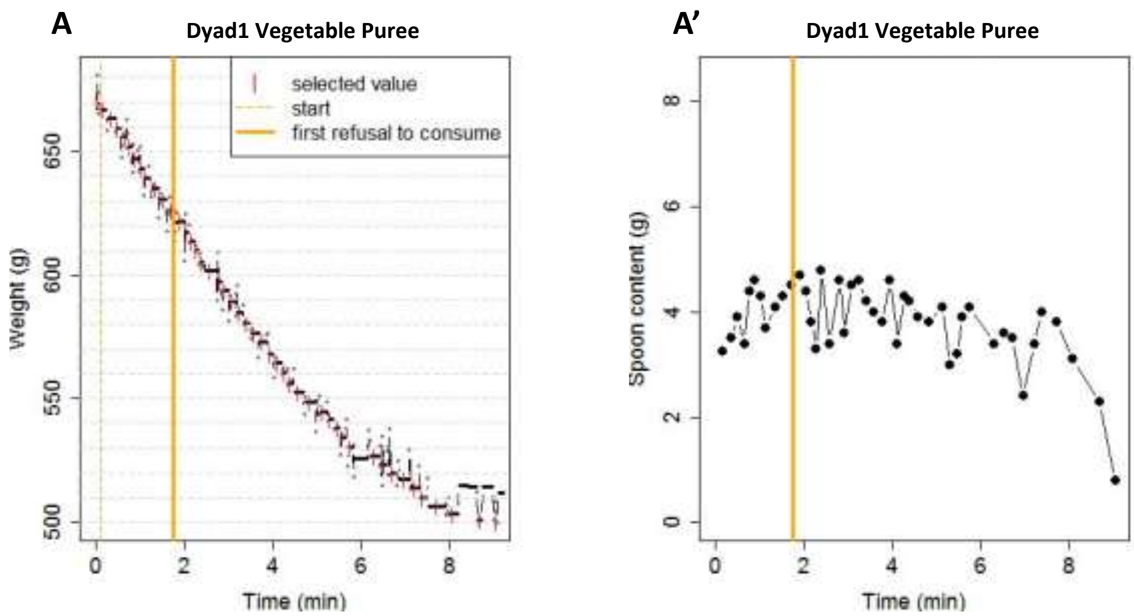


Figure 4. Graphic representations of the effective food consumption duration (from the videotape coding) and the total food consumption duration (from the connected weighing scale) (A) and the number of bites (from the videotape coding) and the number of spoonfuls (from the connected weighing scale) (B).

The overvaluation of the total consumption duration (from the connected weighing scale) compared to the effective consumption duration (from the videotape coding) (Figure 4A) is explained by the fact that the connected weighing scale, contrary to the videotape recording analysis, included the meal's breaks as part of the course meal. Similarly, as shown in Figure 4B, the number of spoonfuls tended to be lower than the number of bites; this reflects that one spoonful may be consumed in several bites before the spoon returned to the bowl.

Examples of different mother-infant dyadic functionings revealed by the mixed observational method

Below, graphic representations illustrate some mealtime episode parameters from 3 different mother-infant dyads. These are presented as typical cases of different functionings (Figure 5).



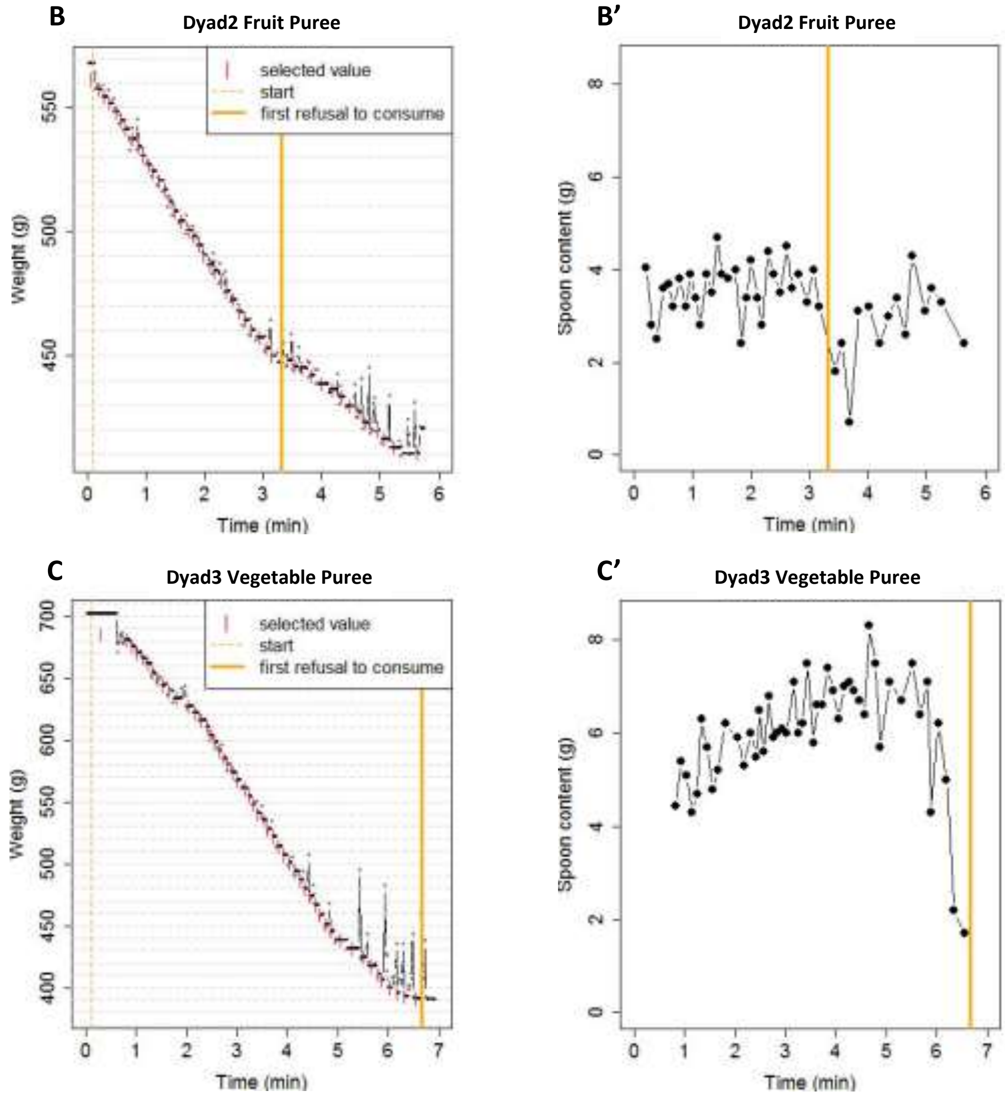


Figure 5. Graphics (A), (B) and (C) illustrate the eating rate (g/min) to consume a food (vegetable or fruit puree) obtained with the connected weighing scale. Each spoonful is represented by a red vertical trait. The time of appearance of the first refusal obtained with the videotape record is represented by a vertical yellow line. The axis scales were adapted according to the weight consumed and the duration of consumption which were variable between dyads. The initial weight for the vegetable puree could vary slightly depending on the used bowl (A), (C). For each dyad, the corresponded spoonful content over time is presented in (A'), (B') and (C') graphics.

The comparative observation of the (A), (B) and (C) graphics clearly revealed that the duration from the start of the consumption and the appearance of the 1st refusal to consume was highly variable between dyads. The first refusal appeared in the first third of the course of the consumption episode for the dyad 1 (A), whereas it appeared in the middle for the dyad 2 (B) and by the end for the dyad 3 (C). Moreover, this case analysis also showed a change in the dynamic occurring around the appearance of this first refusal. Contrary to the dyad 1 (A), the first refusal was followed by a decrease of the eating rate in the dyad 2 (B).

Inter-dyad variability regarding the mean spoon content could also be revealed: overall, the spoon content seemed to be greater in dyad 3 (C') than in the other dyads (A'), (B'). The dyad 3 (C') also showed an increase of the spoonful content over time. Such dynamic was not observed so clearly in the other dyads (A') and (B').

Finally, the intra-dyad parameters comparison highlighted another interesting fact: in dyad 2, the time of the first refusal to consume seemed to be associated with a punctual decrease of the spoonful content and a greater time-lapse between successive spoons (B/B'). This suggests that the parent and the infant are responsive to each other.

Applications & perspectives

The data from the connected weighing scale and the videotape recordings did not evaluate strictly the same meal's episode parameters (i.e., total *versus* effective food consumption duration and number of spoonfuls *versus* number of bites). However, the strong positive correlations obtained between data from these two methods provided support that both methods allowed to obtain similar meal episode patterns. Considering that the coding of videotaped meals require around 30 minutes per food consumption episode (i.e., almost 70 hours of coding were necessary in the present study), we assume that the use of the connected weighing scale could be a reliable substitute for coding automatically some meal episode parameters like the total consumption duration and the number of spoonfuls. This also enables the collection of other variables like the spoonful weight of each offered spoonful which provides a unique insight to describe the food consumption episode; this would not be possible through the videotape recording method only. However, the videotape recording method is still an essential tool for qualitative aspects like the appearance of the first refusal to consume. Yet, this parameter is especially important when exploring whether the parent has adapted and changed his/her feeding style. As a complementary tool, it also allows to understand artefacts (due to the pressure applied with the spoon to the bowl for example) obtained with the connected weighing scale. Thus, this study highlighted the importance of using a mixed observational method to describe mealtime episodes in mother-infant dyads during complementary feeding. Studying eating behavior by an observational approach has been proved to be relevant by allowing a unique insight in all contextual and social aspects of feeding [16]. To go further and explore more aspects of the sequence of events, a particular focus could be made on the different changes surrounding the first refusal to consume. Also, depending on the question to explore, coding more qualitative aspects of the eating episode could be considered for future studies (e.g., visual attention – target and duration of gazes, joint attention – or verbal communication) [17]. Very interestingly, the present mixed observational method revealed the existence of different mother-infant dyadic functionings in the context of a mealtime episode. Data from the literature have underlined that a higher eating rate was positively associated with a higher Body Mass Index (BMI), through self-reported data in 30-42 months children (mothers answered questionnaires regarding their children) [18] but also through the analysis of meals' videotapes in 4-5 years old children in an *ad libitum* buffet context [19]. In a context of increasing prevalence of obesity [20], understanding the development of eating patterns (among which mother-infant dyadic functioning) and their links with early weight gain trajectories is needed. In their model, DiSantis and colleagues [12] theorized that a discordant caregiver-infant dynamic – also referred as to a non-responsive feeding style – (e.g., inappropriate parental response to their infant's hunger and satiation cues and/or unclear signal emission by the infants) can result in increased feedings (in terms of amount and/or frequency). This might lead to an accelerated weight gain and in turn to overweight. Considering this, different profiles of mother-infant dyadic functioning and infants' anthropometrics could be seen together as an interesting perspective. Very recently, we have also adapted a paradigm to evaluate infant's caloric-compensation abilities [21]. Another perspective is then to explore whether different mother-infant dyadic functionings in the context of mealtime are linked with different infant's caloric compensation abilities.

Funding

This work was supported by grants from ANR (ANR-15-CE21-0014); the Regional Council of Burgundy Franche-Comté (FABER) and the French Society of Nutrition (SFN 2015).

Acknowledgements

We thank Marion-Eve Guillard, Marceline Anthoine-Milhomme and Justine Penin who made the videotape coding. We thank Jacques Maratray for developing the connected weighing scale software. We thank Fabienne Bouillot (ChemoSens) for her help in recruiting participants, Valérie Feyen for her help in the data collection and the mother-infant dyads who have accepted to participate in this study.

References

1. Westerterp-Plantenga, M.S. (2000). Eating behavior in humans, characterized by cumulative food intake curves - a review. *Neuroscience & Biobehavioral Reviews* **24**(2), 239-248.

2. Ohkuma, T., Hirakawa, Y., Nakamura, U., Kiyohara, Y., Kitazono, T., Ninomiya, T. (2015). Association between eating rate and obesity: a systematic review and meta-analysis. *International Journal of Obesity* **39**(11), 1589-1596.
3. Nicklaus, S., Remy, E. (2013). Early origins of overeating: Tracking between early food habits and later eating patterns. *Current Obesity Reports* **2**(2), 179-184.
4. Schwartz, C., Vandenberghe-Descamps, M., Sulmont-Rossé, C., Tournier, C., Feron, G. (2017). Behavioral and physiological determinants of food choice and consumption at sensitive periods of the life span, a focus on infants and elderly. *Innovative Food Science & Emerging Technologies*.
5. Lindgren, A.C., Barkeling, B., Hägg, A., Ritzén, E.M., Marcus, C., Rössner, S. (2000). Eating behavior in Prader-Willi syndrome, normal weight, and obese control groups. *The Journal of Pediatrics* **137**(1), 50-55.
6. Zandian, M., Ioakimidis, I., Bergström, J., Brodin, U., Bergh, C., Leon, M., Shield, J., Södersten, P. (2012). Children eat their school lunch too quickly: an exploratory study of the effect on food intake. *BMC Public Health* **12**(1), 351.
7. Llewellyn, C.H., van Jaarsveld, C.H.M., Boniface, D., Carnell, S., Wardle, J. (2008). Eating rate is a heritable phenotype related to weight in children. *American Journal of Clinical Nutrition* **88**(6), 1560-1566.
8. Langlet, B., Fagerberg, P., Glossner, A., Ioakimidis, I. (2017). Objective quantification of the food proximity effect on grapes, chocolate and cracker consumption in a Swedish high school. A temporal analysis. *PloS ONE* **12**(8), e0182172.
9. Fogel, A., Goh, A.T., Fries, L.R., Sadananthan, S.A., Velan, S.S., Michael, N., Tint, M.-T., Fortier, M.V., Chan, M.J., Toh, J.Y. (2017). Faster eating rates are associated with higher energy intakes during an ad libitum meal, higher BMI and greater adiposity among 4· 5-year-old children: results from the growing up in Singapore towards healthy outcomes (GUSTO) cohort. *British Journal of Nutrition* **117**(7), 1042-1051.
10. Agras, W.S., Kraemer, H.C., Berkowitz, R.I., Korner, A.F., Hammer, L.D. (1987). Does a vigorous feeding style influence early development of adiposity? *The Journal of Pediatrics* **110**(5), 799-804.
11. McNally, J., Hugh-Jones, S., Caton, S., Vereijken, C., Weenen, H., Hetherington, M. (2016). Communicating hunger and satiation in the first 2 years of life: a systematic review. *Maternal & Child Nutrition* **12**(2), 205-228.
12. DiSantis, K.I., Hodges, E.A., Johnson, S.L., Fisher, J.O. (2011). The role of responsive feeding in overweight during infancy and toddlerhood: a systematic review. *International Journal of Obesity* **35**(4), 480-492.
13. Hodges, E.A., Johnson, S.L., Hughes, S.O., Hopkinson, J.M., Butte, N.F., Fisher, J.O. (2013). Development of the responsiveness to child feeding cues scale. *Appetite* **65**, 210-219.
14. Hodges, E.A., Wasser, H.M., Colgan, B.K., Bentley, M.E. (2016). Development of Feeding Cues During Infancy and Toddlerhood. *MCN The American Journal of Maternal Child Nursing* **41**(4), 244-251.
15. Chabanet, C., Maratray, J., Brugaillères, P., Issanchou, S., Schwartz, C. (2018). R functions to extract data from a connected weighing scale used during a meal. http://prodinra.inra.fr/record/425131.
16. Pesch, M.H., Lumeng, J.C. (2017). Methodological considerations for observational coding of eating and feeding behaviors in children and their families. *International Journal of Behavioral Nutrition and Physical Activity* **14**(1), 170.
17. Hetherington, M.M. (2017). Understanding infant eating behaviour – Lessons learned from observation. *Physiology & Behavior* **176**, 117-124.
18. Okubo, H., Miyake, Y., Sasaki, S., Tanaka, K., Hirota, Y. (2017). Rate of eating in early life is positively associated with current and later body mass index among young Japanese children: the Osaka Maternal and Child Health Study. *Nutrition Research* **37**, 20-28.
19. Fogel, A., Goh, A.T., Fries, L.R., Sadananthan, S.A., Velan, S.S., Michael, N., Tint, M.T., Fortier, M.V., Chan, M.J., Toh, J.Y., Chong, Y.-S., Tan, K.H., Yap, F., Shek, L.P., Meaney, M.J., Broekman, B.F.P., Lee, Y.S., Godfrey, K.M., Chong, M.F.F., Forde, C.G. (2017). A description of an 'obesogenic' eating style that promotes higher energy intake and is associated with greater adiposity in 4.5 year-old children: Results from the GUSTO cohort. *Physiology & Behavior* **176**, 107-116.
20. Verdot, C., Torres, M., Salanave, B., Deschamps, V. (2017). Corpulence des enfants et des adultes en France métropolitaine en 2015. Résultats de l'étude Esteban et évolution depuis 2006 // Children and adults body mass index in France in 2015. Results of the Esteban study and trends since 2006. *Bulletin épidémiologique hebdomadaire* **13**, 234-241.
21. Brugaillères, P., Issanchou, S., Chabanet, C., Nicklaus, S., Schwartz, C. (2017). Capacité d'ajustement énergétique à court terme chez le nourrisson : évolution autour de 1 an. *Proceedings of Journées Francophones de Nutrition* (Nantes, 13-15 December).

Empirical game theory of pedestrian interaction for autonomous vehicles

Fanta Camara[1,2], Richard Romano[1], Gustav Markkula[1], Ruth Madigan[1], Natasha Merat[1] and Charles Fox[1,2,3]

1 Institute for Transport Studies (ITS), University of Leeds; 2 School of Computer Science, University of Lincoln; 3 Ibex Automation Ltd

Autonomous vehicles (AV's) are appearing on roads, based on standard robotic mapping and navigation algorithms. However their ability to interact with other road-users is much less well understood. If AVs are programmed to stop every time another road user obstructs them, then other road users simply learn that they can take priority at every interaction, and the AV will make little or no progress. This issue is especially important in the case of a pedestrian crossing the road in front of the AV. The present methods paper expands the sequential chicken model introduced in (Fox et al., 2018), using empirical data to measure behavior of humans in a controlled plus-maze experiment, and showing how such data can be used to infer parameters of the model via a Gaussian Process. This providing a more realistic, empirical understanding of the human factors intelligence required by future autonomous vehicles.

Introduction

Autonomous vehicle technology is currently claimed by many commercial organizations involved in operating experiments which apply robotic navigation algorithms to on-road vehicles. Robotic mapping and navigation are mature research areas (Thrun, 2005) and open-source software stacks are now available to enable a car to pick and follow a route around static environment (Kato et al. 2015). In contrast, the human factors involved in real-world driving in environments containing other road users are not well understood. Other road users include human drivers and pedestrians, and eventually other autonomous vehicles, which may or may not be able to communicate with one another. A recent study operated an autonomous minibus continuously on a real commuter route around La Rochelle, France and Trikala, Greece, for several weeks (Madigan et al., in preparation) and suggested that once the local population has learned that the vehicle was programmed to be perfectly safe and stop at any obstacle, pedestrians and other vehicles would regularly take advantage of the AV to push in front of it. In some cases pedestrians did this simply for fun – delaying the vehicle and its real-world commuter passengers on their way to work. Human drivers do not simply give way to all such threats. Instead they appear to use their knowledge of and predictions about the psychology of other road users to interactively negotiate for priority. This can include dominant behavior – driving straight toward a pedestrian at full speed, or even accelerating, to encourage them to yield and return to the pavement; or polite behavior – backing off to allow them space to cross. A well-known scenario is for two pedestrians to meet similarly and make several attempts dodging from side to side in order to agree on how to pass. Two polite road users can be almost as dangerous as two dominant ones – by continually trying to yield to one another they get closer to a collision without the decision being made.

A recent study (Fox et al., 2018) proposed an abstract mathematical model of two agents – which may be AVs, human drivers, and pedestrians - - meeting at an unsigned intersection and negotiating for priority. This model is deliberately simple and intended as a foundation whose structure can be extended with many additional details. In addition, it contains free parameters decrypting human preferences. The present study illustrates how such parameters can be fit to human data as a method of measuring behavior which combines empirical Game Theory and a Bayesian Gaussian Process analysis.

Game theory of human driven vehicles is used extensively in macroscopic traffic modeling via Wardrop equilibrium in flow networks (Bolland et al., 1979) with focus on route selection in large, economy-like, markets of many road users rather than microscopic pairwise interactions. Where game theory has been applied to pairwise traffic decisions, it has mostly been at the level of simple single-shot games as reviewed in (Elvik, 2014). In a few cases such as lane-changing (Meng et al., 2016; Kim and Langari, 2014) and merging (Kita, 1999) it has been extended to sequential games as used here, but not for AV-pedestrian interactions as here.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Methods

Sequential Chicken model

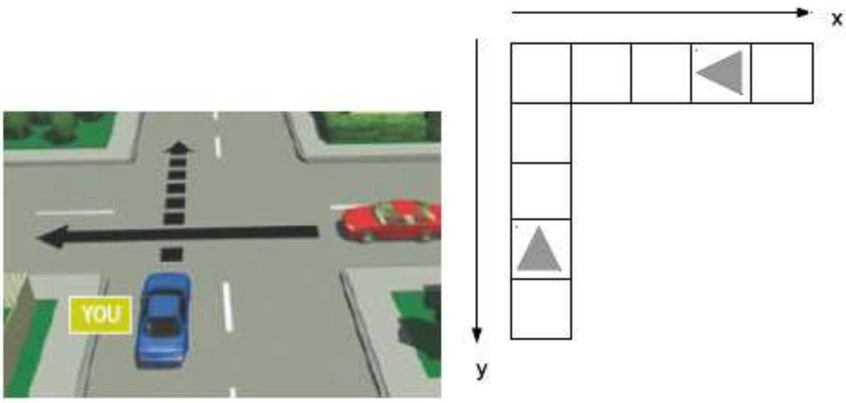


Figure 1 (left): two agents negotiating for priority at an intersection. (right): game theory model of the scenario.

The "Sequential Chicken" model of (Fox et al., 2018) represents a simplified version of the class of scenarios shown in fig. 1 (left) and fig. (2). In these scenarios, two agents (which may each be an AV, human driver of various types, and/or pedestrian) called *Y* and *X* approach the same intersection. If neither agent yields, they will collide and each receive a negative utility U_{crash}. Otherwise, they each reach their destination at some time delay, *T*, from the start of the game. The model assumes a linear value of time, so the total (negative) utility of an agent arriving at time *T* is $-TU_{time}$. The basic model assumes both agents share the same parameters U_{crash} and U_{time} and both know this to be the case. It assumes that no lateral motion is permitted, and that there is no communication between the agents other than seeing each other's positions. It discretizes positions into a plus maze structure of grid squares as in fig. 1 (right), and discretizes time into turns in which both players must select from two speeds, of either 1 square per turn or 2 squares per turn. The model is similar to a board game where each turn, both players make their choices in secret then and reveal them together before acting on them.

The model is called "Sequential Chicken" because its theoretical optimal strategies are solvable using Game Theory, and related to the famous game theory model of Chicken. In (non-sequential) Chicken, two drivers *Y* and *X* face each other and choose (as a single decision) whether to drive straight at each other or swerve away[3]. Swerving makes one the nominal "loser", but in practice, penalties for both are much worse if neither swerves and they collide. The game is considered a draw if both swerve. In game theory, this is represented by a payoff matrix, showing numerical utilities to the two players as in table 1.

$Y \setminus X$	$a_X = swerve$	$a_X = straight$
$a_Y = swerve$	(0,0)	(-1, +1)
$a_Y = straight$	(+1, -1)	(-100,-100)

Table 2: The game of Chicken.

Standard Game Theory provides (Nash) *equilibria* for games in this matrix form, which are configurations of action selections such that neither player would change if they knew the other's choice. (Fox et al., 2018) provide arguments for selecting between these equilibria in real-time traffic situations so as to choose the most "rational" strategy. In general, the best strategies are probabilistic, specifying that each player should choose between the two actions with certain probabilities. (If they were deterministic, then each player could know the other's move in advance and exploit that knowledge, contradicting their supposed optimality.)

Chicken is a one-shot game – both players make just one decision then act on it. Sequential Chicken can be viewed as a sequence of one-shot (sub-)games, which can be solved similarly. Write (*y*,*x*,*t*) to describe the player's current

[3] It is convenient to use *Y*, *X* ordering throughout so that matrix indices $M_{y,x}$ correspond to visualizations.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

locations as in fig. 1 (right) at turn t and let their available actions be the speed choices $a_Y, a_X \in \{1,2\}$. Rather than winning, losing, or crashing, the outcomes of the game at state (y,x,t) are that the players reach a new state, ($y+a_Y$, $x+a_X$, $t+1$). This may be a "game-over" state, where either one player has reached the intersection and/or a crash has occurred, or it may be another intermediate state leading to another game in the sequence. Define the *value* $v_{y,x,t} = (v_{y,x,t}{}^{Y}, v_{y,x,t}{}^{X})$ of state (y,x,t) as the expected utility of the game to each of the two players, assuming that both play optimally from the state. Then the sub-game at time t can be written as a standard game theory matrix,

$$v_{y,x,t} = v\left(\begin{bmatrix} v(y\text{-}1, x\text{-}1, t\text{-}1) & v(y\text{-}1, x\text{-}2, t+1) \\ v(y\text{-}2, x\text{-}1, t+1) & v(y\text{-}2, x\text{-}2, t+1) \end{bmatrix}\right)$$

which can be solved using recursion, Game Theory, and equilibrium selection to give values and optimal strategies at every state (given in Fox et al., 2018). Under a mild approximation (Fox et al., 2018) the solution may further drop all dependency on time t so that state values and optimal strategies becomes functions only of the player positions (y,x).

The model shows that if the two players play optimally, there must exist a non-zero probability for a collision to occur. This corresponds to the intuition that a perfectly safe vehicle which yields in every case will make no progress. Players must include some credible threat of collision to avoid being taken advantage of.

Figure 2: A "Barnes dance" or "scramble" crossing between Hollywood and Highland, Los Angeles. Real-world pedestrian-pedestrian interactions similar to the experiment occur between users of the two diagonal crossings.

Experiments

The model allows for computation of optimal strategies given parameters (U_{crash}, U_{time}) but does not specify numerical values of these human preference parameters. Eventually we would like to measure such parameters from real-world traffic data such as CCTV of vehicles and pedestrians on roads. The present study is intended to illustrate only the method for measuring such behaviors rather than report results, so instead uses a simplified proxy. Rather than interact physically, pairs of human subjects play a board-game version of Sequential Chicken as above. Their moves are recorded and used to show how to fit the parameters.

Sixteen participants aged from 22 to 48 were divided into 8 groups of 2 players. Each group's experiments took place with players and an experimenter seated around a table in university room and last for about 10 minutes. All experiments were conducted in accordance with University of Lincoln research ethics policy.

Natural game. Two players, one designated as player Y and the other as player X were sitting around a table, they were asked to play the chicken game on a plus-maze shaped board with discrete space as shown in Figure 1(right). Player Y was starting from y=10 and player X from x=10 such that they were both starting 10 squares before the intersection. Players were first told that their goals were to avoid a collision and to be the first to cross over this intersection as quickly as possible as if they were trying to commute to their office to work there in a morning. Players were each given two cards containing the numbers 1 and 2, and told that at each turn they should select one in secret then both reveal their cards and make their moves on the board simultaneously.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Chocolate game. After playing 3 natural games, each group played a further 3 games in which specific rewards were specified in advance. Players were told that the winner is the first to pass the intersection, and will receive two chocolates (Celebrations, Mars Inc.), while the second to pass the intersection will receive one chocolate, that both will receive no chocolates if there were any collision. Chocolates were provided after each game as in the instructions.

Measuring behavior

We wish to infer the parameters $\theta = (U_{crash}, U_{time})$ from the observed data, *D*, assuming that players play Bayes-optimally as prescribed by the model, M. By Bayes' theorem, the posterior belief over these is given by,

$$P(\theta|M, D) = \frac{P(D \mid \theta, M)P(\theta| M)}{\sum P(D| \theta', M)P(\theta'| M)}$$

Given a set of possible θ values. assume a flat prior over θ, so that,

$$P(\theta|M, D) \propto P(D \mid \theta, M).$$

H

ence we can work only with the data likelihood term, $P(D \mid \theta, M)$, which is given by,

$$P(D \mid \theta, M) = \prod_{game}\prod_{turn} P\left(d_Y^{game,turn} \middle| y, x, \theta, M\right) P\left(d_X^{game,turn} \middle| y, x, \theta, M\right),$$

where $d_{player}^{game,turn}$ are the observed action choices, *y* and *x* are the observed player locations at each turn of each game. The action probabilities $P\left(d_Y^{game,turn} \mid y, x, \theta, M\right)$ can be read directly from the optimal strategies.

The above assumes that model M is perfect. However, *M* assigns zero or near-zero probabilities to many events which are found to occur in the data. Uncontrolled, this can be catastrophic for any Bayesian model of real data. We know that real humans may behave near-optimally but can make mistakes in perception (of the state of the game, such as the values of *y* and *x*) and in action selection. Thus, we weaken the model used in the analysis to a noisy but more forgiving version *M'*. This is a mixture model which generates each observation according to *M* with some probability *s*, and generates unbiased random actions otherwise,

$$P(D \mid \theta, M') = \prod_{game}\prod_{turn}\left[(1\text{-}s)P\left(d_Y^{game,turn} \middle| y, x, \theta, M\right) P\left(d_X^{game,turn} \middle| y, x, \theta, M\right) + s\left(\frac{1}{2}\right)\right].$$

These equations are unlikely to have analytic solutions for θ, as they include elements of the optimal strategies computed via a long game theoretic process. To compute with them, we may sample various values of θ and observe $P(D \mid \theta, M')$. In practice, we work in log-space to reduce numerical errors, and consider $\log P(D \mid \theta, M')$. Given a sample of these log-likelihoods, we may then fit a Gaussian Process (Rasmussen & Williams, 2006) regression model to them. Assuming that log-likelihood is a smooth function of the parameters, this provides estimates of the log-likelihoods at locations between the sampled values and can also show the uncertainty in these estimates. A Radial Basis Function kernel was used in the Gaussian Process model, with variance set to a minimum after obtaining smooth likelihood surfaces.

Results

Before performing the full analysis, we show in table 1 a rough overview of the game results. More collisions have occurred in the chocolate game than in the natural game. Define "delay" as the distance in boxes, or time in seconds, between the arrival at the intersection of the first player to arrive and the second player to arrive. The average delay in distance (box) is about 1.368 for the natural game but is reduced in the chocolate game (1.125). The time delay is computed following the formula: $d = vt$ where d is the distance, v the speed is assumed to be the maximum (2) and t is the time delay.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Type of Game	Number of Games	Collisions	Average Delay (box)	Average Time Delay (s)
Natural	24	5	$\frac{26}{19} \approx 1.368$	$\frac{26}{19*2} \approx 0.684$
Chocolate	24	8	$\frac{18}{16} = 1.125$	$\frac{18}{16*2} = 0.5625$
Total	48	13	$\frac{44}{35} \approx 1.257$	$\frac{44}{35*2} \approx 0.628$

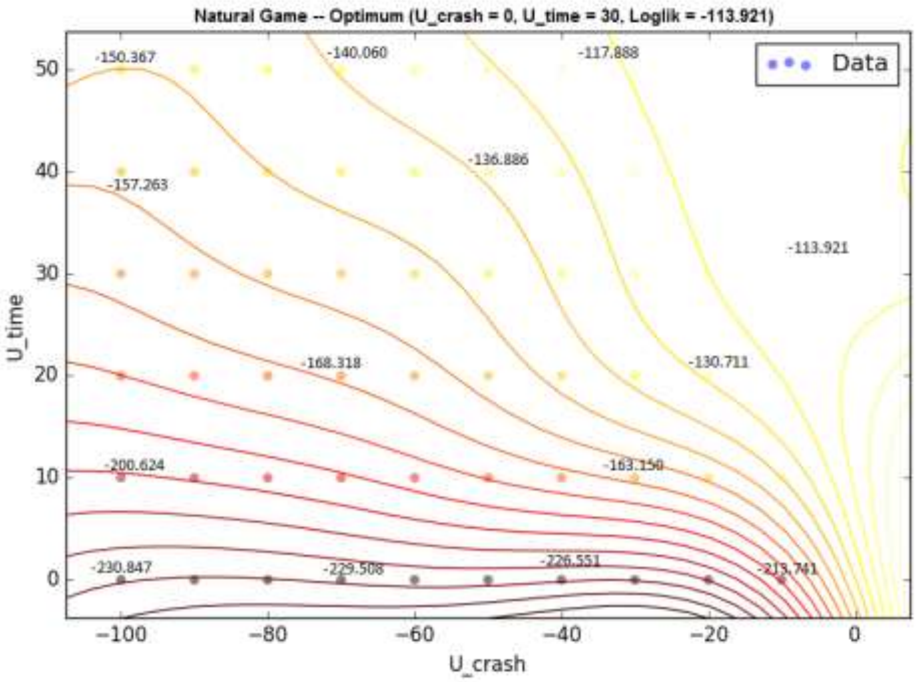
Table 2: Number of collisions and average time delay per game

Fig. 3 shows the results of fitting the Gaussian processes to likelihoods sampled along a regular grid, using s=1/10. The surfaces have a roughly radial structure. By visual inspection, the peak of the smoothed Gaussian Process surface in the natural case appears to lie somewhere in the region of $(U_{crash}, U_{time}) = (0,30)$ and for the chocolate case, around (0, 30). Fig. 4 shows a 1D slice through the 2D surface of Fig. 3(natural) which provides scale and uncertainty information.

Discussion

We have illustrated a new method for measuring and modelling human road user behavior, which is intended as part of a method to imitate or exploit such behavior in autonomous vehicle controllers. The model can be trained on human behavior then transferred into a controller to specify new robotic behaviors.

The particular experiment here has inferred posterior beliefs about the parameters of crash utility and time delay for human participants, in a simplified and artificial version of a real-world intersection negotiation task. We do not consider either the model or the experimental setup to be especially realistic versus real-world interactions, but they are sufficiently complex to illustrate the general method. This method is presented as an invitation for future work to extend both the model (e.g. with continuous positions and speeds, uncertain beliefs in opposite player utility function, lateral motion) and the experimental data collection (e.g. use of "big" real-world CCTV video of road-user interactions) stages to be more realistic.



R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

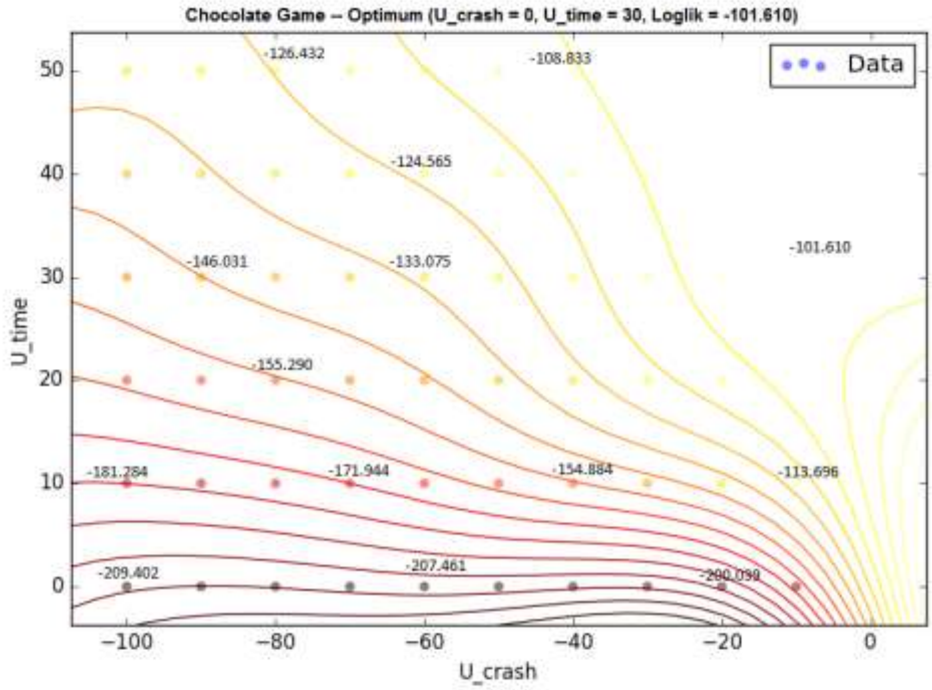


Figure 3: Gaussian process regression shows the likelihood (and hence, posterior belief) surfaces for the parameter values(U_{crash}, U_{time}), for the natural and chocolate experiments. The sampled parameter locations are shown by the dots. (This figure should be viewed in color).

Some conclusions from the results of the particular experiment analyzed which may be of interest, despite its artificial setting: From the viewpoint of real-world interactions, it is surprising to see a preference towards time-saving over crash avoidance in the players. In the real world, the cost of crash is usually millions of times more than the value of a second of commuting time, but here the players behave the opposite way. The effect is more pronounced in the chocolate game than in the natural one. This may be because the board-game-like nature of

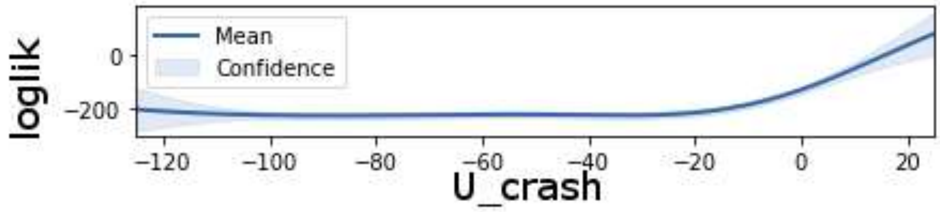


Fig. 4. Horizontal slice though the Gaussian Process of fig. 3(natural), at U_time=10, showing mean and standard deviation of belief about the log likelihood, $\log P(D \mid \theta, M')$ over the parameter space.

the experimental setting encourages the players to think in terms of winning and losing rather than just optimizing their personal utility. In the chocolate game this is made to be explicit by rewards the players specifically as winner and loser in the interaction. This leads to more collisions in this more competitive game.

The shapes of the likelihood surfaces are roughly radial, which suggest that it may only be the ratio between the two parameters which is important rather than their absolute values. Possibly the angles of the contour lines define equivalence classes of parameters. The number of experiments performed was quite small, and future work including further experiments may produce more peaky posterior distributions.

By using this method to fit more realistic models and data to human pedestrian-vehicle and vehicle-vehicle interactions, we can understand and predict behavior of road users to allow autonomous vehicles to more optimally interact with them. The present model requires a non-zero probability of collision to exist as a threat in all interactions. Future work could extend the model with additional, smaller negative utilities such as humiliating annoying pedestrians with horns or spraying them with water jets, in order to reduce or eliminate the necessity of actually hitting them whilst still allowing autonomous vehicles to make progress.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Acknowledgements
This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreements No 723395 (InterACT) and No 645376 (FLOBOT).

References

Bolland, J., Hall, M., Van Vliet, D., and Willumsen, L. (1979). Saturn: Simulation and assignment of traffic in urban road networks. In Proceedings of the International Symposium on Traffic Control Systems, Berkeley, Calif. D, volume 2, pages 99–114.

C.W. Fox, F. Camara, G. Markkula, R.A. Romano, R. Madigan and N. Merat. (2018) When should the chicken cross the road?:Game theory for autonomous vehicle - human interactions. In Proceedings of VEHITS 2018, Funchal, Madeira, Portugal.

Elvik, R. (2014). A review of game-theoretic models of road user behaviour. Accident Analysis & Prevention, 62:388–396.

Kato, S., Takeuchi, E., Ishiguro, Y., Ninomiya, Y., Takeda, K., and Hamada, T. (2015). An open approach to autonomous vehicles. IEEE Micro, 35(6):60–68.

Kim, C. and Langari, R. (2014). Game theory based autonomous vehicles operation. International Journal of Vehicle Design, 65(4):360–383.

Meng, F., Su J., Liu, C., and Chen, W.-H. (2016). Dynamic decision making in lane change: Game theory with receding horizon. In Conference: 11th UKACC International Conference on Control.

Rasmussen, C. E., & Williams, C. K. (2006). Gaussian processes for machine learning. 2006. *The MIT Press, Cambridge, MA, USA*, *38*, 715-719.

Thrun, S., Burgard, W., and Fox, D. (2005). Probabilistic robotics. MIT press.

Madigan, R., Nordhoff, S., Fox, C., Ezzati Amini, R., Louw, T., Wilbrink, M., Schieben, A., and Merat, N. (in preparation). Understanding interactions between automated vehicles and other road users. Manuscript in preparation.

Using Low-Level Interaction Data to Explore User Behaviour in Interactive-Media Experiences

Jonathan Carlton[1,2], Andy Brown[1*], John Keane[2] and Caroline Jay[2]

[1]BBC R&D, Salford Quays, Manchester, UK. [1]jonathan.carlton@bbc.co.uk, [1*]andy.brown01@bbc.co.uk
[2]School of Computer Science, University of Manchester, Manchester, UK. {first.last}@manchester.ac.uk

Introduction

Understanding users of computer-based systems through user testing is an expensive, time consuming, and difficult process. However, low-level interaction data (such as mouse movements and keyboard strokes) present an opportunity to gather (in a non-intrusive way) the same information, which has many potential advantages such as large sample size, remote data collection, and ecological validity. Being able to derive meaning from low-level interaction data has relevance for many real-world applications, from biometric authentication systems [1, 2] to alerting to cognitive impairment [3]. The challenge comes in handling the low-level and voluminous character of such data, which is particularly apparent when attempting to mine human interactions with software and web-based technologies. Considering events in isolation can reveal little about the behaviour of the user, apart from context-specific information stored just after the event occurred (such as time of occurrence and event-type). Rather it is necessary to synthesise the collected data, to abstract useful, semantically-meaningful behavioural characteristics from it.

The work presented in this paper explores how these techniques may be applied to a new domain – interactive media experiences. Being able to detect and evaluate the behaviour of the users of such systems has broad implications for media creators, as creating original content is expensive and time-consuming, and learning as much as possible from pilots is crucial. In this context, it is desirable to understand how users interact with an experience, to assess their levels of engagement, to understand which parts they value the most (or least), and to identify usability barriers. This paper reports a first step in this direction: an exploratory analysis of low-level interaction data collected during an evaluation of user experience of a novel interactive cookery show. The work described here is an initial step into extracting meaningful behavioural characteristics from pilot products that are exposed to public use for feedback. The specific question addressed in this paper is: can it be determined, from low-level interaction data, whether a user was previewing, reviewing or experimenting with the experience, or whether they were using it to support the cooking process as intended. Even this relatively basic information is valuable as it can help producers to better understand and give weight to user feedback. Applying a well-established data mining methodology to the collected interaction data shows that there are detectable differences exhibited by participants during the user experience study. This is a promising step into mining data collected from interactive media experiences and in understanding how people consume media content in this new format.

The paper is structured as follows: *Related Work* overviews work in other domains in which low-level interaction data have been analysed; this is followed by a description of the *Experimental Setup* used from which the data were collected; the *Methodology* section presents data mining methodology; the *Discussion* presents results of the applied methodology; finally, the *Conclusion and Future Work* section considers avenues for future research.

Related Work

Extraction of behaviours from low-level interaction events, such as mouse movements, clicks, and scrolling, to describe users, or sub-sets of users, has a wide range of contexts from security and medical systems to search engines. For example, collecting and synthesising this type of data allows leveraging of the low-level event data to raise alerts for mild cognitive impairment [3] and help identify difficulties in navigating online media [6]. Personality traits of users, during a visual search task (*Where's Waldo*), can also be extracted as demonstrated in [5], whereby applying well-known data mining algorithms (Support Vector Machines (SVM) and Decision Trees) to mouse movement and click data can successfully classify users based on locus-of-control, extraversion, and agreeableness. Low-level event data can also be used to infer problematic situations, such as identifying coping

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

tactics employed by people with a visual disability [11]. Through developing a social network, designed specifically for cognitively disabled people, and collecting interaction metrics (such as the number of clicks and the delays between clicks), clusters of users can be detected based on their actions, which can provide information to management staff on how patients are coping [12].

Further there are examples in the literature that demonstrate use of low-level interaction data to determine the original intent of a user. Mouse movements are the sole focus in [8], where trajectories are extracted and used to train a model to infer user intent while performing queries in a search engine. It is found that their model outperforms, in terms of accuracy, a model trained solely on click data. In a slightly different approach, [7] predicts user intention from interaction using a more comprehensive range of mouse events, including movement, clicks, and positioning. They produce a set of descriptive statistics to train a multimodal model (consisting of action sequences and interaction features) which can perform with a high level of accuracy (70%). The goals of search engine users have been an active research area for some years, due, in part, to the potential advertising revenue that this information may generate. Through mining interactions, such as mouse movements and scrolling, work in [9] detects user goals while using a search engine and finds that user interactions are the most important feature as the classifier accuracy drops significantly without them (from 96.7% to 86.7%).

Finding unique user signatures, based on their low-level interactions, has been utilised in authentication-based systems. By creating abstractions of mouse behaviour, such as click elapsed time, movement speed, and acceleration, work in [2] continuously re-authenticates users during a logged-in session on a website. The approach trains a model using characteristics of the mouse behaviour, which, when evaluated, can outperform sequential behavioural patterns of events and actions. Unique behavioural signatures, exhibited by users while engaged in verification for access to a secure system, are also leveraged in [1]. Here a typing style test consisting of metrics such as how long a key on a keyboard is pressed, the latency between button presses, along with a mouse-based signature matching test is combined to produce an authentication system. The system achieves a low fraudulent access rate of 4.4% and a high authentic user access rate of 99%. Finally, the ability to uniquely identify players of video games through their mouse and keyboard interactions, and the unique signature they exhibit, is presented in [13]. The work here demonstrates that it is possible to detect unique and individual behaviour signatures produced by users, which indicates that models can be built to identify particular types of behaviour or categories of users and to ultimately find related users.

While low-level interactions provide the source to extract behaviours, rediscover user intent, and identify unique signatures exhibited by users, a common step underpinning the resulting models (in each of the papers mentioned) is the extraction of appropriate descriptive statistics. The purpose of this is to select appropriate descriptive features with which to train models. The related work also demonstrates that common patterns of behaviour can be extracted from low-level data, which can reveal underlying information about users and subsequently be mapped onto new data to find similar patterns and identify similar users.

The work here extends that in [4] and [10] in that it uses the same approach to extract the same behaviours. These studies aggregate low-level interaction data, such as mouse movements and keystrokes, among others, into the micro-behaviours users perform while interacting with a website. However, whilst this earlier work focuses on the evolution of behaviour for returning users over a period of 12-months, the current research focuses on the behavioural statistics exhibited by users within a short, controlled study with four conditions (discussed in more detail in a later section). Furthermore, this work investigates a new domain – media consumption – a novel interactive form of media.

Experimental Setup

The data used in this paper was consensually collected during a user experience study of CAKE (Cook-Along Kitchen Experience)[4], an interactive cookery show that allows users to select and cook recipes that are subsequently presented at a pace matching that of the users, producing a personalised experience. The study aimed

[4] http://www.bbc.co.uk/rd/projects/cake

to evaluate the user experience of CAKE and to determine whether it better supports people while cooking compared to a traditional, non-adaptive cook-along video. One hundred participants were recruited from an internal staff newsletter, 95 of whom collected the ingredients required to cook, and 78 of whom completed the post-cooking questionnaire (indicating that they had completed the experience). As part of the study, there were two participant groups, one using CAKE and the other using a webpage containing a linear video sequence with written instructions. Furthermore, each participant was assigned a recipe type, either following a single recipe or multiple recipes. Nineteen people used CAKE with a single recipe, and 18 with multiple recipes; 25 people used a single linear video recipe and 16 multiple linear recipes. Of the participants, 11 experienced technical problems, such as videos not buffering or loading, which left 25 participants completing the linear single recipe and 14 completing each of the other three. Before the start of the study, participants reported a keen interest in cooking and were eager to take part in the study. The study required participants to cook a recipe in a place and at a time of their own choosing, by following the instructions in their videos; this was supplemented by a pre-test questionnaire completed in advance of ingredient collection and a post-test questionnaire completed shortly after cooking. The aim of the questionnaires was to measure how effective CAKE was in supporting the participants. The results showed that participants felt CAKE was more supportive compared to a traditional cook-along video.

The low-level interaction data collected during the study contained various types of mouse and keyboard events, recorded when a participant interacted with an interface in any of the formats. The aim of the current work is to explore whether there are behavioural differences between participants using the experience to support a cooking activity, and participants exploring, previewing or reviewing the experience. Data collected while the experience was being tested, in advance of deployment for the user experience study, are thus contrasted with the live data collected during the active period of the study. During the testing phase, 3,089 distinct low-level events were recorded, along with 26,717 low-level events when the study was live. The datasets contain actions such as mouse movements, scrolling, mouse over, key down, key up, mouse up, and mouse down.

Methodology

In this section, a description of the sequence of methods applied to the collected low-level event data is presented. The behavioural extraction is described, followed by the processes used during the exploration phase, in which features to be used for classification were selected. The final stage was the process of training a classifier to recognise the differences between testing and live data. The results from the exploration and classification predictions are presented and discussed in the following *Results* section.

The data is collected in the form of distinct user events. These events vary in form, as the information collected about each event depends on the type of event. For example, the data collected when a scroll is detected is different to when a mouse movement event occurs. This variability in events adds to the complexity of moving from fine-grained data to descriptive summary statistics. The method presented in [4, 10] is used to extract the statistics. This classifies the low-level data into micro-behaviours exhibited by the users. The types of metrics that are extracted include the session length, total number of events, total distance travelled, active time, and idle time.

A vital aspect of extraction of micro-behaviours is the process of detecting multiple sessions recorded by the same participant. A time-between events threshold of 40 minutes is used, as recommended in [4, 14], i.e. if there is a period between two events, produced by the same participant, of equal to or greater than 40 minutes, then the first event signals the end of a session while the second signifies the start of another. This enables detection and exploration of different behaviours by the same participant in the same context to find which of their multiple sessions is the ‘real’ one, i.e. when they cooked along with the study, through the training of a classifier.

The exploration of the extracted behaviours is performed to identify which features of the data to include in the training of the classifier. Due to the nature of the user study and the length of the tasks that the participant was required to perform, a relatively small amount of interaction data was collected compared with [4], for example only a small amount of valuable scrolling behaviour data was recorded with most values being zero. As a result, the scrolling behaviour extracted is removed from the feature set used to train the classifier, a process that is discussed later, and the focus of the analysis is on the rest of the mouse- and keyboard-related data. It should be noted that the CAKE experience has been designed to be a full screen service where no scrolling is necessary.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Principal Component Analysis (PCA) [15], a dimensionality reduction algorithm, and K-Means [16], a clustering algorithm, were applied to the selected behavioural features. This step helps to identify natural clusterings of data points and underlying correlations. The underpinning methodology followed was Knowledge Discovery in Databases (KDD) [17], which has the following steps: data selection, pre-processing, transformation, application of data mining algorithms, and interpretation of results. PCA was used as part of the pre-processing and transformation steps to provide insight into how behavioural features are correlated, while K-Means (application of data mining algorithms) was used as a descriptive measure of the data to identify clusters in the data.

The first step in training a classifier to differentiate between testing and live data was to scale the data (the transformation step). All the data were standardised by rescaling the features to have the properties of a standard normal distribution (a mean of zero and a standard deviation of one). This step is crucial as it prevents dominance of features with greater variances than others. For example, the duration of sessions has a more substantial variance due to their diverse nature, whereas the counts of events captured in a session have comparatively smaller variance which could be due to all participants having enforced interactions with the experience (clicking the 'next' button to move onto the following stage).

For classifier training, a combination of testing and live data is used. The live data contains participants that have only recorded a single session, while multiple session users are removed and used as the prediction dataset. An SVM classifier [18] is chosen because of its ability to perform well with the type of data being used [2, 5, 7].

Results

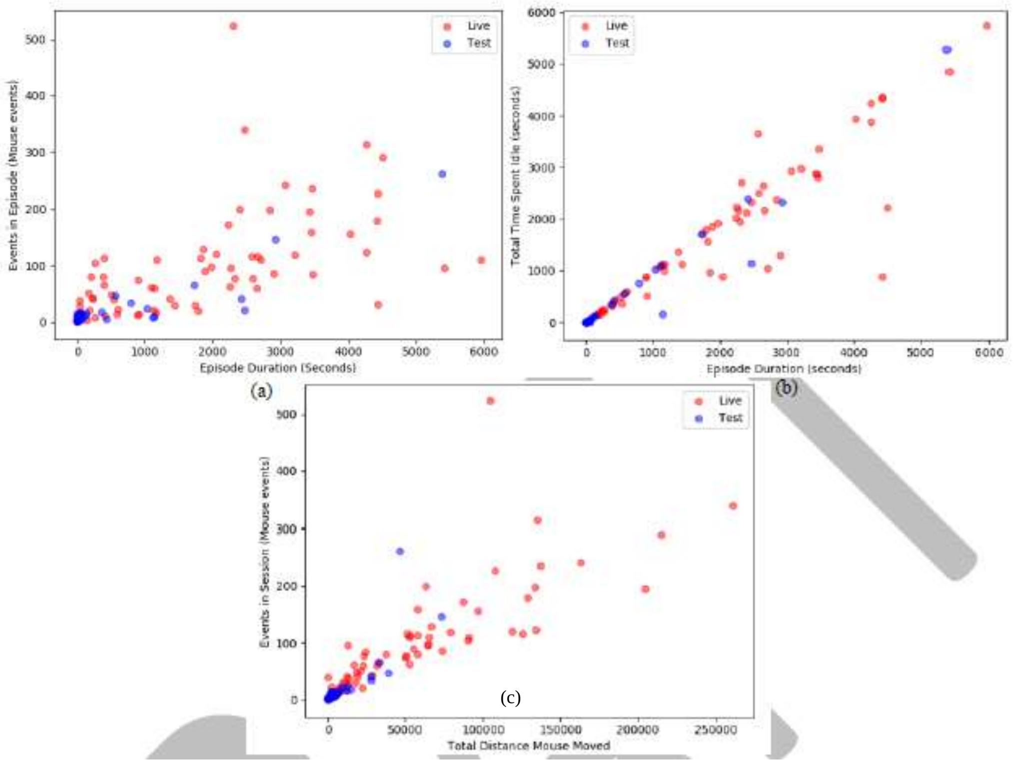


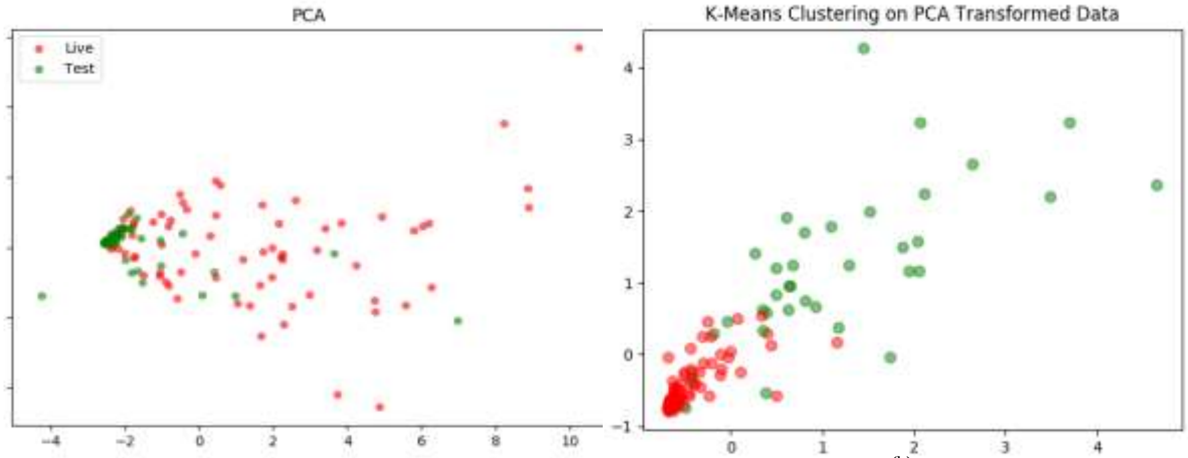
Figure 1. Behavioural feature plots: (a) episode duration in seconds vs. number of mouse events recorded in an episode, (b) episode duration in seconds vs. total time spent idle, (c) total distance the mouse moved in an episode vs. number of recorded events in an episode.

The plots in Figure 1 consist of extracted behavioural features compared before they are subsequently scaled for the classifier training phase. Figure 1(a) shows the number of events recorded in a session (*y*-axis) plotted against the duration (*x*-axis). Visually this indicates a strong clustering of testing data towards the lower end of both the *x* and *y* axis, suggesting that testing behaviour tends to have a short and concise session duration with fewer total events. The plotted live data appear more randomly spread out. However, there are some live data points that

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

cluster in the same area as the testing data, suggesting some participants may be exploring, previewing or reviewing the experience. A similar trend is found in Figure 1(b), which shows a strong clustering of testing data, and the sessions when considered in this way, are highly linear with most session time being idle (no events captured). This makes sense as in both study formats the participants spend the most of their time watching videos which require no interaction. Figure 1(c) further demonstrates a difference in behaviours between testing and live sessions, with the testing data points being considerably more tightly clustered when compared to Figure 1(a).

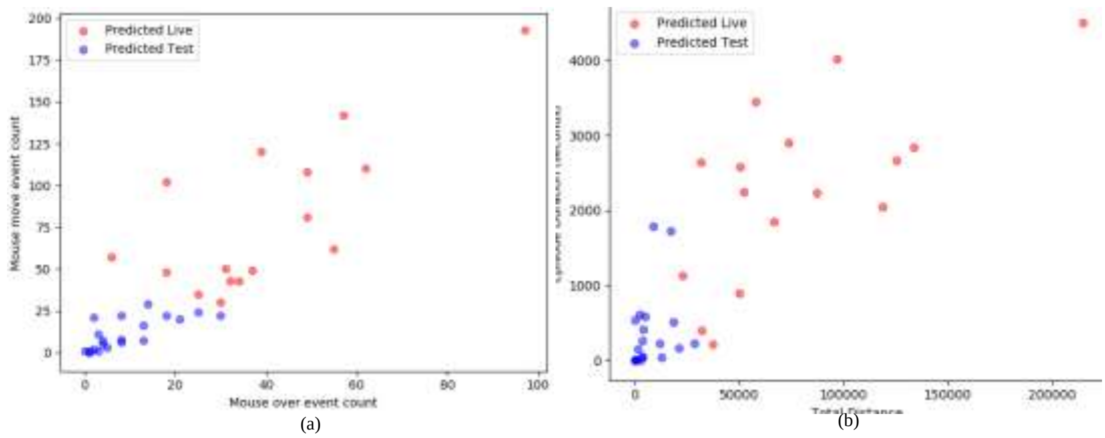
The plots in Figure 1 are indicative of differences in the behaviours exhibited in testing data when compared to live data. While this may make intuitive sense, the critical aspect is that live data points are clustered in the same space as the testing data, potentially suggesting that participants are exhibiting similar behaviour.



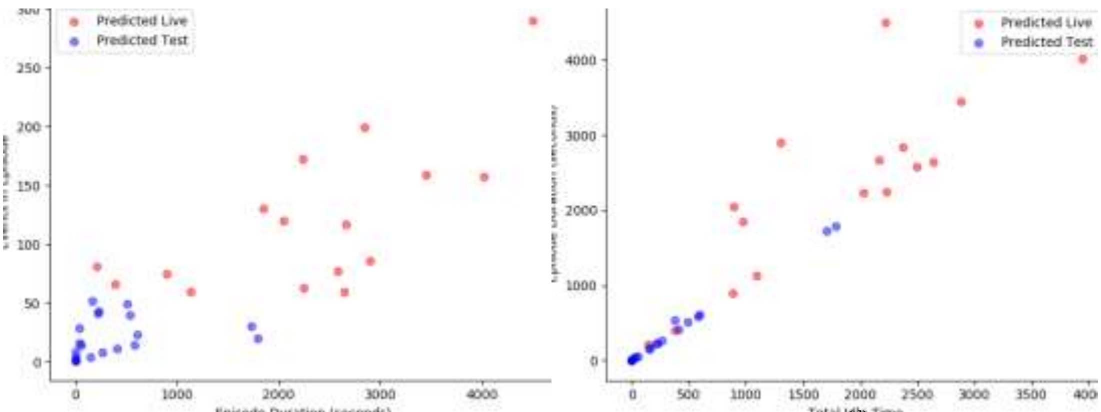
(a) (b)

Figure 2. (a) Result of applying PCA to the extracted behavioural features and (b) Result of applying K-Means clustering to the PCA data.

To reduce dimensionality of the behavioural features, PCA is applied to the data. PCA provides a summarization of features, can help to identify clusters within the data and hence helps to build hypotheses. In this case, it reduces the data into two principal components. The results, shown in Figure 2(a), demonstrate a strong clustering of the reduced testing data in a single point. This is like Figures 1(a), 1(b), and 1(c), suggesting that all testing features exhibit similar behaviour, whereas the live data are more spread out but show a tendency to cluster to the same point as the testing data. K-Means is then applied to the PCA-transformed data. The result is shown in Figure 2(b): one cluster is densely clustered at a single point while the other is more spread out. As previously discussed in the Methodology section, the live data are split into two subsets: one containing participants with a single session and the other containing participants with multiple sessions. The classifier analyses multi-session participants and predicts whether they exhibit behaviour like that of testing or live data points. Of the participants, 13 recorded multiple sessions with a mean of 2.76 sessions each (standard deviation = 1.16).



(a) (b)



(c) (d)

Figure 3. Plots demonstrating the behaviours of participants with multiple sessions and their predicted categories.

Figure 3 shows four plots that present behavioural features exhibited by participants with multiple sessions and their predicted category. The plots demonstrate that some participants exhibit behaviours during the live phase of the trial that the classifier identified as 'testing'. They also show how closely related some of the extracted behaviours are to actual testing data, strongly suggesting that these are in fact exploratory sessions.

Conclusion & Future Work

This paper presents an exploratory investigation of low-level interaction data collected during a user study of a novel interactive cookery experience. The aim of the investigation was to ascertain if characteristics of the data could differentiate behaviours exhibited during the testing or exploratory interactions with the experience from those exhibited when users were using the experience to support their cooking. In a more general sense, this can be characterised as differentiating exploration and testing of an interactive tool from its actual use.

A methodology is used that involves the process of synthesising the interaction data, the extraction of behaviours, the initial exploration of the behaviours, and the training of a classifier to differentiate between testing and live data. Once the model is created, participants that recorded multiple sessions during the trial are then classified to find their predicted category: test or live. The selected behavioural features have been analysed, and the results demonstrate that there are differences in the types of behaviours exhibited during the testing phase and the live phase. Both PCA and K-Means were applied to the extracted behaviours, and this has shown that the testing behaviours are consistent with the live data, being more random while still centring around a similar point as the testing data. Finally, the predicted categories of the multiple session users were presented. What this initial work and the preliminary results presented has shown is that participants in the user study exhibit behaviours indicative of reviewing, exploring, or previewing behaviour. Furthermore, it confirms that this behaviour can be successfully detected as discussed in the related literature.

This work is a first step in understanding behaviour within an interactive-media context using low-level interaction data, and it leaves many questions remaining. While detecting participants that exhibit 'test' behaviour in the live session has proven to be successful, understanding what those participants are doing in those sessions is an important next step. Exploring the ordering of sessions, produced by participants with multiple sessions, may provide further insight and additional information about the participants' overall experience. Furthermore, matching results to an overall rating of the experience is a crucial future step as it allows weight to be added to particular ratings along with the wider implications of the feedback results from pilots. Collecting a wider range of data may also provide further insights into the behavioural characteristics of participants. Finally, creating a tool that can generalise to the context but is able to analyse and extract behaviours from low-level interaction data could have significant implications for understanding users and how they consume media.

Acknowledgements

Jonathan Carlton would like to acknowledge the EPSRC and the BBC for iCASE PhD funding support (16000156).

References

1. Everitt, R., *et al.* (2003). Java-based internet biometric authentication system. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**: 1166-1172.
2. Shen, C., *et al*. (2017). Pattern-growth based mining mouse-interaction behaviour for an active user authentication system. *IEEE Transactions on Dependable and Secure Computing* **99**.
3. Gledson, A., *et al*. (2016). Combining mouse and keyboard events with higher level desktop actions to detect mild cognitive impairment. *IEEE International Conference on Healthcare Informatics* (Chicago, IL), 139-145.
4. Apaolaza, A., *et al*. (2015). Longitudinal analysis of low-level web interaction through micro behaviours. *Proceedings of the 26th ACM Conference on Hypertext and Social Media* (Guzelyurt, Northern Cyprus), 337-340.
5. Brown, E., *et al.* (2014). Finding Waldo: learning about users from their interactions. *IEEE Transactions on Visualisation and Computer Graphics* **20**: 1663-1672.
6. Thomas, P. (2014). Using interaction data to explain difficulty navigating online. *ACM Transactions on the Web* **8**: 1-41.
7. Fu, Y. E, *et al*. (2017). Your mouse reveals your next activity: towards predicting user intention from mouse interaction. *IEEE 41st Annual Computer Software and Applications Conference* (Turin, Italy), 869-874.
8. Guo, Q., *et al*. (2008). Exploring mouse movements for inferring query intent. *Proceedings of the 31st Annual International ACM Conference on Research and Development in Informatics* (Singapore), 707-708.
9. Guo, Q., *et al.* (2010). Ready to buy or just browsing? Detecting web search goals from interaction data. *Proceedings of the 33rd International ACM Conference on Research and Development in Information Retrieval* (Geneva, Switzerland), 130-137.
10. Apaolaza, A., *et al.* (2013). Understanding users in the wild. *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility* (Rio de Janeiro, Brazil), 13:1-13:4.
11. Vigo, M., Harper, S. (2013). Coping tactics employed by visually disabled users on the web. *International Journal of Human-Computer Studies* **71**: 1013-1025.
12. Arbelatiz, O., *et al.* (2016). User modelling in a social network for cognitively disabled people. *Journal of the Association for Information Science and Technology* **67**: 305-317.
13. Kaminsky, R., *et al*. (2008). Identifying game players with mouse biometrics.
14. He, D., Goker, A. (2000). Detecting session boundaries from web user logs. *Proceedings of the BCS-IRSG 22nd Annual Colloquium on Information Retrieval Research*, 57-66.
15. Tipping, M., Bishop, C. (1999). Probabilistic principle component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61**: 611-622.
16. Arthur, D., Vassilvitskii, S. (2007). k-means++: the advantages of careful seeding. *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 1027-1035.
17. Fayyad, U., *et al*. (1996). Advances in knowledge discovery and data mining.
    Cortes, C., Vapnik, V. (1995). Support-vector networks. *Machine Learning* **20**: 273-297.

Using Decision Tree Learning as an Interpretable Model for Predicting Candidate Guide Dog Success

Zach Cleghern[1], Margaret Gruen[2], David Roberts[1]

1 Department of Computer Science, North Carolina State University; 2 College of Veterinary Medicine, North Carolina State University

Training guide dogs is a time and resource intensive process that requires a copious amount of skilled professional and volunteer labor. Even among the best programs, many dogs are released from their training programs. The highest cost in producing guide dogs occurs during the professional training and placement. Selecting dogs for training is a crucial task and guide dog schools can benefit from both an increase in accuracy of their selection and the speed at which dogs can be screened out of the program. We present a method using decision trees to predict the future success or failure of dogs in a guide dog program based on existing data sources from a guide dog school. We achieved 60.6% accuracy on predictions in the test set compared to a failure rate of about 50% at this stage in dog training. Decision trees are easily interpretable; thus evaluators can benefit not only from the model's prediction but can also examine which features are used to determine success or failure.

Introduction

The process of training dogs as service animals is highly resource intensive. The monetary and time demands vary by type of work, but some estimates show the cost for an organization to raise and train a service dog may be $20,000 to $50,000 per dog[1]. Unfortunately, many dogs that enter training do not succeed as guide dogs, representing a major loss to organizations. Great effort has been put forward toward attempts to identify successful dogs early in life, allowing for enhanced selection and refinement of breeding. The majority of these have focused on behavioral and health traits. However, several studies have also evaluated maternal care and early life experiences[2][3]. Evaluation and screening takes place at several points including neonatal, puppy-raising and training phases, including assessments as puppies, adolescent, end of training, and results following placement. At Guiding Eyes, observations from most assessments are scored using the Behavior Checklist. While some associations with success have been found, particularly at older ages, the predictive value of puppy testing has remained relatively low because environment plays a key role.

Guiding Eyes for the Blind (Guiding Eyes), a large, not-for-profit guide dog organization in the United States, maintains records of results for these assessments from all their dogs. At Guiding Eyes, socialization begins at 1 week and observations of socialization events are recorded. A formal puppy assessment is conducted at around 7.5 weeks of age. About 11% of puppies tested are provided to other organizations that train service dogs, 16% are adopted as pets and the remaining 73% are raised as potential guide dogs by volunteer puppy raisers, who raise them in their homes, teach basic obedience and social skills and provide exposure to a variety of environments and experiences. Puppies are seen weekly then twice monthly by staff who also conduct formal assesments at 4 and 10 months of age. Their next assessment, the In-For-Training (IFT) test, is performed at a Guiding Eyes school when the puppy is about 16 months old. Following the IFT test, dogs may enter into training as a guide dog, enter training for other work (detection or emotional support), or be released as pets. This data corpus provides an opportunity to employ machine learning to assist in decisions about which dogs to invest training resources in.

Machine learning can identify subtle, otherwise difficult to discover patterns present in data. By incorporating data from both successful and unsuccessful dogs ("success" here means the dog was successfully placed with a handler and/or became a successful breeding dog), machine learning methods can detect patterns that are reflective of dog performance in ways that can be difficult for humans, even experts, to pick up on. Then, the results can be used to better understand early determinants of future success.

We chose decision trees due to their interpretability, which can be used to find features that are highly predictive of training outcome. The induction of decision trees[4] for classification is a standard method[5] in the realm of supervised machine learning. In this case the dependent variable, or "class", is a label indicating whether a dog is

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

successful in the program or not. We make predictions with decision trees by examining a set of *features*, or attributes, about a specific dog in the data set. A decision tree is a graph structure like a flowchart consisting of *nodes*, which represent a test on a condition, and connections between the nodes called *edges*, which represent the result of the test. By starting at the *root node*, or first decision, one can follow the edges to arrive at a *leaf node*, containing the prediction produced by the tree. The nodes between the root and leaf are *internal nodes*, which represent tests on specific attributes. For example, an integer attribute X and a node in the tree might test "X <= 0." When traversing the tree at this node, the expression is evaluated using the attribute X of the current data sample (e.g. the data associated with a single dog) and the edge corresponding to the result is selected.

In machine learning, an algorithm, such as CART[6] (Classification and Regression Trees- what we used), constructs decision trees by analyzing a training set of data samples. Induction of the tree happens incrementally beginning with the root node. Using some measure of impurity or information gain, such as Gini Impurity[7], the features are analyzed for the best possible split according to the measure. A feature and threshold value are chosen that splits the trainind data accordingly. Then each split is examined, creating new nodes until a stopping condition is met, such as maximum tree depth. Maximum tree depth is a number specified outside of the learning process, a *hyperparameter*. Other hyperparameters in decision tree learning are splitting criterion (how to measure each feature to select the next split), minimum samples in a split to add a new node, and maximum number of features to consider for a single tree.

Using the induced tree, we can classify a data sample by starting at the root node and testing the features at each node, following the associated path down the tree until a leaf node is reached. Once this happens, we have our prediction of the sample's class. Because each node explicitly states the attribute and threshold used in the test, this process is easily interpretable by humans allowing non-experts to use decision trees without computer interaction and also understand which features were relevant in the decision. By viewing the nodes, we can observe which features provide the most information about how to classify any sample. See Figure 1 for an example portion of a decision tree with 3 nodes shown. The first feature tested is an item from the Behavior Checklist involving the dog's confidence; the number of samples indicates how many data samples were in the portion of the data that this node is concerned with.

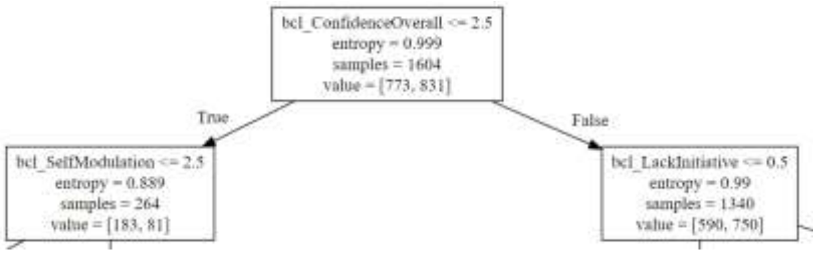


Figure 1. Example portion of decision tree with 3 nodes, each using Gini Impurity[7] splitting criterion.

The goal of this study is to show how machine learning, using existing behavioral data, can be employed to predict which dogs are more likely to succeed as guide dogs. Our objectives were to use decision tree learning with existing data collected by Guiding Eyes, including the Behavioral Checklist and the CBARQ test, and then to test the developed model using an unseen data set. If successful, our model would allow early predictors of success to be identified and used to increase accuracy of early career sorting and identifying which dogs are most appropriate for training. Ultimately, this would benefit the guide dog schools, the people they assist, and the dogs themselves. It is worth noting that our efforts are not the first to specifically address the problem of predicting success of guide dogs with advanced analytics and machine learning. Yim et al.[8] also used decision trees to predict guide dog success, but we are extending and improving on this work by restricting the time period to enable earlier predictions than previously available; ultimately this allows guide dog schools to filter out candidate dogs earlier in life and conserve valuable resources.

Data

Data were available for 1561 Labrador retrievers (765 males; 796 females) produced by Guiding Eyes over several years. Of these, 789 (50.5%) were successful in the guide dog program. Our definition of "success" here means that the dog was successfully placed with a person who is blind or visually impaired or possessed exceptional qualities and was selected as a breeding dog. The data used for the machine learning task included basic information about the dog (sex and breed) and the results of two common forms of evaluations, described here.

CBARQ

The CBARQ[9] is a 101 item questionnaire used to evaluate a dog's behavior and temperament. This questionnaire was developed by Serpell and Hsu, and has been validated for use in dogs[10]. It is commonly used in canine research [11][12], and comprises data that many guide dogs schools, in addition to Guiding Eyes, already collect[13]. The data we used from this questionnaire was collected when dogs were 12 months of age.

Behavior Checklist

Also developed by Serpell, the Behavior Checklist is a scoring system where observers can score up to 52 aspects of behavior such as excitability, fearful behavior (such as traffic or noise fear), emotional response to potentially stressful situations, and body sensitivity. Like the CBARQ questionnaire, the Behavior Checklist is a commonly used scoring system for scoring the behavior of potential guide dogs, which makes the test a good source of data if we want models that are useful for any guide dog school. The Behavior Checklist we utilized was collected at the In-For-Training evaluation.

Combined with the basic information about the dog (sex and a categorical breed code), these information sources result in 197 individual features. These features were tested for correlation with outcome in order to derive a restricted set of significant features for testing. Twenty-seven features had statistically significant ($p < 0.05$) correlation with success. These significant correlations in the 12 month CBARQ and Behavior Checklist data are shown in Tables 1 and 2, respectively. In the CBARQ questionnaire[9][10], the questions referred to in Table 1 are as follows:

4. Q10: [Displays aggressive behavior] when approached directly by an unfamiliar adult while being walked/exercised on a leash.
5. Q11: [Displays aggressive behavior] when approached directly by an unfamiliar child while being walked/exercised on a leash.
6. Q12: [Displays aggressive behavior] toward unfamiliar persons approaching the dog while s/he is in your car (at the gas station, for example).
7. Q52: [Displays fearful behavior] when having his/her feet toweled by a member of the household.
8. Q85: Nervous or frightened on stairs.

Table 1. CBARQ features in the data set with a statistically significant correlation with guide dog success.

Feature	Correlation	P
CBarq-Q10	-0.056	0.0269
CBarq-Q11	-0.063	0.0133
CBarq-Q12	-0.054	0.0341
CBarq-Q52	-0.05	0.0478
CBarq-Q85	-0.063	0.0131

Table 2. Behavior Checklist features in the data set with a statistically significant correlation with guide dog success.

Feature Name	Correlation	P Value	Feature Name	Correlation	P Value
BCL-	-0.052	0.0383	BCL-DogAggression	-0.196	$5.15*10^{-}$
BCL-FearDogs	0.053	0.0376	BCL-	-0.196	$5.96*10^{-}$
BCL-	0.118	$3.14*10^{-}$	BCL-	-0.14	$2.66*10^{-}$

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

BCL-	-0.051	0.0429	BCL-	0.051	0.0429
BCL-Footfall	-0.052	0.04	BCL-HandlerDogTeam	-0.263	$3.97*10^{-}$
BCL-	-0.147	$5.24*10^{-}$	BCL-PressuresDog	-0.185	$1.69*10^{-}$
BCL-	-0.177	$1.99*10^{-}$	BCL-KennelsPoorly	-0.055	0.0286
BCL-	0.052	0.0419	BCL-KennelAnxiety	-0.195	$8.94*10^{-}$
BCL-Olfactory	0.141	$2.32*10^{-}$	BCL-ChewsValuables	-0.184	$2.54*10^{-}$
BCL-Scavenges	0.17	$1.56*10^{-}$	BCL-Housebreaking	0.204	$4.45*10^{-}$
BCL-	-0.177	$1.97*10^{-}$	ComparisonRating	-0.057	0.0245

Building the Decision Tree Model

Decision trees are easy to understand by humans, which make them a suitable model for our prediction problem in which professionals evaluating dogs will want to know why a model chooses a particular prediction. We used the Python library scikit-learn[14], an easy to use machine learning package, specifically the CART algorithm[6]. We did not require much preprocessing of our data aside from reformatting some information to be usable by the scikit-learn library. For efficiency of the model building process, we removed any features (such as dog breed, since only Labrador Retriever data was submitted) whose values were equal for every dog in the training set. To evaluate the quality of the learned models, we divided the data into training and test sets, with a split of 80% in the training set and 20% in the test set, which was only used once the hyperparameters were optimized and we had trained the final model.

The hyperparameters to the induction algorithm that we optimized were maximum depth, splitting criterion (in scikit-learn, either the Gini Impurity[7] or Information Gain[7]), and minimum number of samples on which to split an internal node and also on a leaf node. To find the best set of values, we used grid search[15]. This is a simple method for finding sets of hyperparameter values that exhaustively searches the combinations of values in a search space by applying the learning algorithm and comparing the results on validation sets. Although more optimal approaches exist[15], grid search was sufficient for our small search space and the process did not take more than 5 minutes for even the full feature set. For each set of possible values, the algorithm further divides the training set into training and validation sets using 3-fold cross validation[16], which calculates a mean accuracy over the validation sets. The hyperparameters were chosen from the best scoring mean accuracies. We then used those values to fit a tree model on the entire training set. Finally, we evaluated the quality of the resulting decision tree on the previously unused test set using three metrics: accuracy, precision, and recall. (we did this for several subsets of the full feature set). *Precision* is the fraction of true positives (successful dogs that were also predicted to be successful) divided by the true and false positives (the number of times the model said the dog would be successful). *Recall* is the fraction of true positives divided by the number of successful dogs in the test set.

Results

The best set of hyperparameters for the full feature set were found to be: maximum depth of 5, entropy (information gain) splitting criterion, 6 as the minimum number of samples for a leaf node and 9 as the minimum samples for an internal node. We also tested the effectiveness of just the set of statistically significant features previously shown in Tables 1 and 2.

Table 3. Best cross-validation scores and test set accuracy, precision, and recall for each feature subset. Note that “breed” was actually removed as every dog in the data set had the same breed code.

Feature Set	Cross-validated Mean	Test Set Accuracy	Test Set Precision	Test Set Recall
Basic Dog Info (sex, breed)	0.5152	0.4856	0.5161	0.4819
BCL	0.638	0.608	0.62	0.698

CBarq	0.5521	0.5144	0.5507	0.4578
Info, BCL, CBARQ	0.635	0.606	0.620	0.684
Statistically significant	0.648	0.596	0.698	0.698

Table 3 shows the metrics obtained for several subsets of features. Allowing the learning algorithm to choose only from the set of features across all data sources whose correlation with success was statistically significant ($p < 0.05$) resulted in only slightly worse accuracy than the full feature set, but actually outperformed in terms of precision and recall. The full feature set achieved a 60.6% accuracy on unseen data, but since dogs at this point in time have a 50% failure rate this shows promise for the decision tree learning approach. Several subsets achieved recall scores of almost 70%. This indicates the decision tree models were effective at identifying the successful dogs (though false positives brought down the overall accuracy).

By using a decision tree *regressor* instead of a classifier, we were also able to learn decision trees that output a number between 0 and 1 and can be interpreted as a probabilistic prediction. To evaluate the quality of these predictions, we calculated the mean absolute value of the error, where *error* is the difference between the predicted output and the actual class. We treated success as the real number 1.0 and failure as 0. The results of this are shown in Table 4.

Table 4. The mean absolute error for each feature set when converting the prediction task to a regression problem.

Feature Set	Mean Absolute Error
Basic Dog Info (sex,	0.500
BCL	0.434
CBarq	0.488
Info, BCL, CBARQ	0.472
Statistically significant	0.464

Discussion

Using decision tree learning, we were able to successfully predict the future success or failure of potential guide dogs with 60.8% accuracy, 62% precision, and 68.4% recall on the full feature set. Since the eventual failure rate of dogs at this time point (the IFT test) is close to 50%, this is suggestive of an opportunity for cost savings for guide dog programs, especially if this accuracy score can be improved upon. Incorporating larger data sets from various guide dog schools could improve this accuracy. By examining sources of behavioral data that already exist, decision tree classifiers can extract patterns for guide dog schools to utilize in evaluating dogs for training. Including only features with significant correlations to success code was much faster (about 4 minutes and 4 seconds for the full feature set; 1 minute and 9.5 seconds for the significant features), but this is unlikely to be an issue in this domain.

Understanding what influences the ability of dogs to work in various roles, behaviorally or otherwise, has been a significant and often multidisciplinary area of research. There are studies that analyze what factors can tell us about the future behavior of dogs, such as the work by Weiss[17] in which behavioral tests were used to aid in the selection of shelter dogs of service dog training. The behavioral tests include assessments of reaction to touch, walking, fetching, and interacting with strangers. They then used regression analysis to predict the dog's performance over 5 weeks of training. Slabbert and Odendaal found in a two year longitudinal study several behavioral tests whose correlation with future success as police dogs was statistically significant. Some of these tests were able to be administered as young as 8 weeks of age[18]. Byrne et al.[19] used instrumented dog toys with sensors to develop a logistic model tree for classifying the eventual outcome of dogs in a service dog organization (Canine Companions for Independence) and were able to achieve an accuracy of 87.5%, which by their calculations could save $70,000 by identifying dogs that will likely fail the program.

Arata, Momozawa, and Takeuchi[20] studied behavioral factors that may predict success or failure in guide dog programs with the use of a questionnaire. They were able to assess that distraction, sensitivity, and docility are important behavioral factors that can predict a dog's future success. They found that distraction could predict a dog's success with 80.6% accuracy. Batt et al.[21] evaluated the ability of several behavioral tests to predict guide dog success and found several behavioral tests and physical characteristics that were predictive of success. Their results obtained at 14 months of age were more accurate than their results obtained at either 6 months or when the dog completed training. Using the CBARQ questionnaire, Duffy and Serpell found that tests administered at 6 and 12 months of age were useful for predicting success in guide and service dog programs, but their predictive value varied across organization[13].

A benefit of our approach is that we used data sets that are already commonly collected and did not implement our own tests. Our method is also able to provide a "line of reasoning" for its predictions by producing a path from root node to leaf node so that evaluators can know why a prediction was made. Probabilistic predictions may also prove useful by identifying dogs who may have high (but not exceptionally high) probability of success. This is useful for at least two reasons. Guide dog schools will be ultimately using their own discretion on the decision whether to train each dog and a probabilistic machine learning prediction may be more helpful in that regard. A second reason is these dogs may have specific issues that need to be addressed and their success could hinge on those issues.

Future Work

There are many avenues to explore in using machine learning to predict the future capabilities of guide dogs. Recent studies highlighting the importance of maternal care also open a new avenue for data that may refine our model[2][3]. Furthermore, Guiding Eyes collects a variety of text-based data associated with each dog at certain steps in its development such as training notes and comments recorded for various tests. Convolutional neural networks (CNNs) are well-suited to incorporating this data into a machine learning model alongside numeric features. Neural networks may be able to improve the accuracy but in contrast with decision trees can be very difficult to extract meaning from. In addition, IBM Watson[22] services for natural language processing may also be able to augment machine learning prediction of dog outcomes and also to tackle similar problems, such as optimal pairings of puppies with raisers.

Ethical Statement

This work utilized historical data provided by a not-for-profit organization and did not involve new data collection.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1329738. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. This work was also supported in part by a faculty award from IBM. The authors would like to give an enthusiastic thanks to Guiding Eyes for the Blind for their support as well. The views expressed in this paper are the authors' own, and do not necessarily reflect the views of IBM or Guiding Eyes.

References

1. Berns GS, Brooks AM, Spivak M, Levy K (2017). Functional MRI in awake dogs predicts suitability for assistance work. *Scientific Reports 2017*; **7**:43704. doi:10.1038/srep43704
2. Bray EE, Sammel MD, Cheney DL, Serpell JA, Seyfarth RM (2017). Characterizing early maternal style in a population of guide dogs. Frontiers in Psychology; 2017 doi: 10.3389/fpsyg.2017.00175
3. Bray EE, Sammel MD, Cheney DL, Serpell JA, Seyfarth RM. Effects of maternal investment, temperament, and cognition on guide dog success (2017). Proceedings of the National Academy of Science; doi:10.1073/pnas.1704303114.
4. Quinlan, J. Ross (1986). Induction of decision trees. *Machine learning* **1.1**: 81-106.

5. Weiss, Sholom M., and Ioannis Kapouleas (1990). An empirical comparison of pattern recognition, neural nets and machine learning classification methods. *Readings in machine learning*. 177-183.
6. Breiman, Leo; Friedman, J. H.; Olshen, R. A.; Stone, C. J. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software
7. Fayyad, Usama M., and Keki B. Irani (1992). The attribute selection problem in decision tree generation. *AAAI*.
8. Yim, Brendan, Su, Kun, Siahpolo, Soraya, Tseng, Chris (August 2, 2017). Guiding Eyes IBM Watson Project. *Personal correspondence with the authors*.
9. Segurson, Sheila A., James A. Serpell, and Benjamin L. Hart (2005). Evaluation of a behavioral assessment questionnaire for use in the characterization of behavioral problems of dogs relinquished to animal shelters. *Journal of the American Veterinary Medical Association* **227.11**: 1755-1761.
10. Hsu, Yuying, and James A. Serpell (2003). Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs. *Journal of the American Veterinary Medical Association* **223.9**: 1293-1300
11. Marshall-Pescini, Sarah, et al (2008). Does training make you smarter? The effects of training on dogs' performance (Canis familiaris) in a problem solving task. *Behavioural processes* **78.3**: 449-454.
12. Serpell, James A., and Yuying A. Hsu (2005). Effects of breed, sex, and neuter status on trainability in dogs. *Anthrozoös* **18.3**: 196-207.
13. Duffy, Deborah L., and James A. Serpell (2012). Predictive validity of a method for evaluating temperament in young guide and service dogs. *Applied Animal Behaviour Science* **138.1**:99-109.
14. Pedregosa, Fabian, et al (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12.Oct (2011)**: 2825-2830.
15. Bergstra, James, and Yoshua Bengio (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research* **13.Feb**: 281-305.
16. Kohavi, Ron (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai*. **Vol. 14. No. 2. 1995.**
17. Weiss, Emily (2002). Selecting shelter dogs for service dog training. *Journal of Applied Animal Welfare Science* **5.1**: 43-62.
18. Slabbert, J. M., and J. S. J. Odendaal (1999). Early prediction of adult police dog efficiency—a longitudinal study. *Applied Animal Behaviour Science* **64.4**: 269-288.
19. Byrne, Ceara, et al (2018). Predicting the Suitability of Service Animals Using Instrumented Dog Toys. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **1.4**: 127.
20. Arata, Sayaka, et al (2010). Important behavioral traits for predicting guide dog qualification. *Journal of Veterinary Medical Science* **72.5**:539-545.
21. Batt, Lara S., et al (2008). Factors associated with success in guide dog training. *Journal of Veterinary Behavior: Clinical Applications and Research* **3.4**:143-151.
22. High, Rob (2012). The era of cognitive systems: An inside look at IBM Watson and how it works. *IBM Corporation, Redbooks*.

Unfolding concealed equine pain behaviors with remote video technology

Britt Coles[1], Pia Haubro Andersen[1] and Linnéa Birgitsdotter[2]

1. Swedish University of Agricultural Sciences; 2. Distriktsveterinärerna Rimbo; britt.coles@slu.se

As the horse is a prey animal, previous research has suggested that horses may alter [1] or conceal [2] their behaviour if humans are present. Concurrently, veterinarians and researchers have a great interest in effective and timely equine behavioural pain assessment [3]. Subsequently, the aim of this observational unmatched paired cross-section study was to explore the possible variation of equine pain behaviours with and without an observer present in a clinical setting.

Twenty-one horses, at two different equine clinics, were video filmed from a wide-shot perspective from a camera mounted in the corner of the stall - using the same remotely operated technology with and without an observer present - before and after (20-26 hours post) orthopaedic surgery. From the subsequent 200 films, with an individual duration of 2-5 minutes, representative clips of 30 seconds showing the horse exhibiting distinct pain behaviour were chosen and randomized for a blind modified Equine Pain Scale (EPS) scoring by two experienced veterinarians. The EPS is a composite weighted pain scale that evaluates eight different equine behaviours [4] as well as facial expression [5]. For this study, the scale was modified to exclude three categories: the equine facial expression category as well as the two human-equine interaction behaviour categories.

A subsequent statistical analysis of these modified total pain scores using paired t-test test illustrated a significant (t=2,53, p = 0,020) mean increase of ≥ 2 points when comparing pain behaviours with and without an observer present after surgery. No such difference was noted before surgery. The noted post-operative increase in pain scores without an observer present was categorized by extrovert so-called gross pain behaviours (Fischer's exact test, p=0,0448). In the EPS, such behaviours are categorized by clearly visible physical behaviours such as kicking, excessive head movements, repeated stretching of the body, pawing, tail swishing, and playing with the mouth.

The results from this small study calls for larger representative studies of various equine populations to further explore the extent of possible hidden behaviours. Thus addressing an important welfare issue. In this endeavour, remote video technology could be a key ally in the daily veterinary therapeutic clinical work. In the future, remote video technology could include automated machine learning components for an automated decoding of equine pain behaviours and facial expression to help unfold and alert to such hidden equine pain behaviours. The early discovery of such pain behaviours is cardinal to optimal diagnostic and treatment procedures.

References

1. Ashley, F.H., Waterman-Pearson, A.E., Whay, H.R. (2005). Behavioural assessment of pain in horses and donkeys: application to clinical practice and future studies. *Equine Vet. J*. **37**: 565–575.

2. Coles, B.A. (2016). No pain, more gain? Evaluating pain alleviation post equine orthopedic surgery using subjective and objective measurements. *Master thesis: Uppsala, Dept. of Clinical Sciences, Swedish University of Agricultural Sciences.*

3. Bussières, G., Jacques, C., Lainay, O., Beauchamp, G., Leblond, A., Cadoré, J.-L., Desmaizières, L.-M., Cuvelliez, S.G., Troncy, E. (2008). Development of a composite orthopaedic pain scale in horses. *Res. Vet. Sci*. **85**: 294–306.

4. Gleerup, K.B., Lindegaard, C. (2015). Recognition and quantification of pain in horses: A tutorial review. *Equine Vet Educ* **28**: 47–57.

5. Gleerup, K.B., Forkman, B., Lindegaard, C., Andersen, P.H. (2015). An equine pain face. *Vet. Anaesth. Analg*. **42**: 103–114.

Functional near-infrared spectroscopy: beyond measuring workload

N.R. de Joux, E. Foderingham, D. Harris and L. O'Rourke

Department of Psychology, University of Nottingham, Nottingham, United Kingdom. (neil.dejoux@nottingham.ac.uk)

Functional near-infrared spectroscopy (fNIRS) is a tool used to measure oxygenated and deoxygenation hemoglobin in the brain, typical in the prefrontal cortex. In recent years, the fNIRS has gained traction in applied research as a measure of cognitive or mental workload during tasks. Changes in blood oxygenation are often attributed to differences in workload. There is, however, substantial evidence that attributing blood oxygenation changes to workload is not entirely accurate. The current paper presents an overview of a number of studies, both previously published and unpublished, where cerebral blood oxygenation is due to differences in task stimuli and visual processing requirements, which are not related to task workload. The findings suggest that researchers and practitioners using fNIRS should consider various aspects of task requirements beyond workload.

Introduction

Functional near-infrared spectroscopy (fNIRS) has gained increased use in the last 10 years in applied fields of human research, typically being used as a measure of cognitive or mental workload. Indeed, it is documented in previous research that there is a relationship between blood oxygenation and increases in task workload [1, 2]. It appears, however, that changes in blood oxygenation are not entirely due to workload demands. For example, changes in blood oxygenation using fNIRS have been found when comparing taste perception [3], between different age groups [4], and when comparing single task performance against dual task performance [1]. Unfortunately, a common trend when using fNIRS in applied settings is to ignore these potential factors, and solely attribute oxygenation changes to the impact of cognitive workload. By simply describing any changes in oxygenation as related to workload, users of fNIRS may be ignoring some of the underlying causes of these changes. The current paper presents a number of studies where changes in blood oxygenation were due to visual processing demands, rather than workload, during a series of sustained attention tasks.

Sustained attention tasks, also known as vigilance tasks, are tasks which require an individual to continuously monitor stimuli for an extended period of time and respond appropriately whenever a unique target object is presented [5]. Typically, vigilance tasks involve a monitoring period of 30 to 40 minutes, which allows researchers to observe differences in trends over time. These types of paradigms are extremely useful for observing the cumulative effect of tasks, as although a participant may be able to perform well in a task when they have adequate cognitive resources at their disposal, their performance on that same task will decline as they are unable to maintain that level of output [6]. These paradigms are also a useful mid-point between pure laboratory research and more applied settings. The principles of sustained attention paradigms have been used to investigate differences in performance in areas such as flight simulation and air traffic control [7, 8].

A number of trends have emerged from the literature on sustained attention, both in terms of performance and physiological response. The first major trend is that as time on task increases, performance on that task declines. This decline in performance is characterized by a slowing in reaction times and a decrease in accuracy, be that through increased misses and false alarms, or decreased hits and correct rejections [9]. The second major trend, and of more interest to main subject of the current paper, is that brain activity and blood oxygenation are right hemisphere lateralized. That is, the right hemisphere is typically more active than the left hemisphere during these tasks. This is a trend that has been found using fNIRS, but has also been found using a range of other neuroimaging techniques (for an overview, see Helton et al., 2010).

While this trend of right hemisphere lateralization does appear to be quite common during vigilance tasks, there are a number of documented cases during which deviant findings occur [11, 12]. In these studies, there is an increase in bilateral oxygenation, rather than one hemisphere being dominant. This is important, as an increase in bilateral activation means an increase in overall activation, a point that some researchers may use to infer an increase in workload. The cause of this bilateral activation is debated, however both task difficulty and arousing

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

stimuli have been shown to have an effect [10, 13]. Another factor that influences hemispheric lateralization is the configurative properties of the stimuli themselves [14, 15, 16, 17], and is the main focus for the current paper.

The reason why the configurative properties of stimuli results in differences in hemispheric activation during vigilance tasks is due to the underlying mechanisms that govern sustained attention and those that govern basic visual perception. Basic visual stimuli are composed of both local and global components, where the global component refers to the overall image, while the local component refers to the smaller parts that make up the overall image (see Figure 1 for examples). When investigating local-global processing from a solely perceptual perspective (i.e., without the requirement for sustained attention), the processing of global features results in greater right hemisphere activation, while the processing of local features results in greater left hemisphere activation [18]. As mentioned above, pure sustained attention tasks result in greater right hemisphere activation, which is due to the processes required for sustained attention being right hemisphere lateralized [19,20]. When employing local-global feature discrimination during vigilance tasks, a local vigilance task then requires greater left and right hemisphere activation, while a global vigilance task places a much greater processing load on the right hemisphere. This is a finding that has been established previously using tympanic membrane temperature [17] as well as using fNIRS [14, 15].

This previous fNIRS research [14] compared hemispheric activation during a vigilance task that required either local or global feature discrimination (for more detailed information on this, see the methods section). During this previous research, participants were placed into either a local or a global feature discrimination group for the duration of a vigilance task. As stated above, local discrimination during vigilance tasks should elicit greater bilateral activation over time, while global discrimination should result in higher right hemisphere activation. Shown in Figure 1, the hypothesized results were found, with significant hemispheric differences being found which corresponded directly to the expected areas of activation during the tasks. In both objective and subjective terms, there was nothing to indicate in the performance data that the local group resulted in any greater amount of ‘mental workload’, with superior performance found in the local condition, and no subjective differences.

The research presented in the current paper forms an extension of this previous research, and further explores how differences in oxygenation are a result of factors other than ‘cognitive workload’.

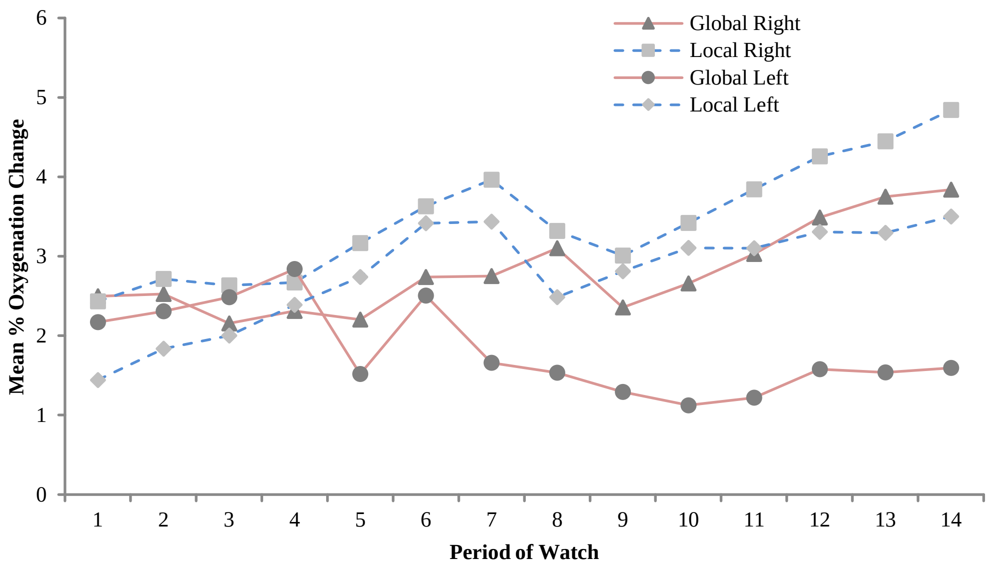


Figure 1. Mean percentage oxygenation change from baseline for left and right hemispheres in Experiment 1.

Methods

The results presented here are from two studies, which are an extension of the previous investigation mentioned in the introduction [14]. Both of these experiments have full ethical approval from the University of Nottingham

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

School of Psychology ethics committee. There are a number of characteristics that are shared across all of the experiments presented in this paper. The main shared characteristic is the stimuli used, which are basic Navon objects [21]. Shown in Figure 2, Navon objects are pictures where a number of smaller shapes are combined two form a larger, coherent shape. In order to induce local or global discrimination, participants were required to attend to one particular feature of the target object that was absent from the distracter objects. In all experiments, the target object was a large square composed of smaller circles. For the local discrimination groups, the distracter objects were a square made of triangles, and a diamond made of triangles. This meant that the unique feature of the target object was the local component (small circles). For the global discrimination groups, the distracter objects were a diamond made of triangles, and a diamond made of circles. This mean that the unique feature of the target object was the global component (large square). Three different font sizes which were used for each shape (12pt, 14pt and 16pt), meaning that the overall stimuli size varied between 50mm by 70mm, and 60mm by 95mm.

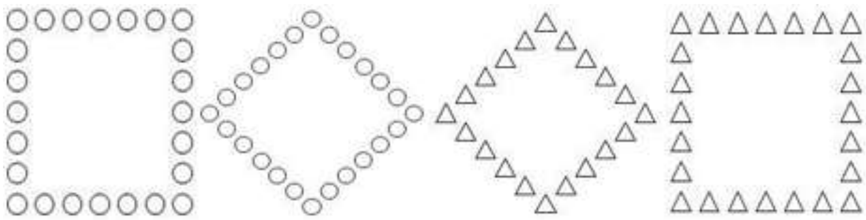


Figure 2. Examples of local-global stimuli. The left-most image was the target image in all conditions.

In the previously presented research, the fNIRS device used was the Nonin Near-Infrared Cerebral Oximeter. This is a total saturation index (TSI) fNIRS. The sensors for these two experiments were placed at the Fp1 and Fp2 positions (using standard 10/20 EEG placement), which correspond to the left and right prefrontal cortex. For the research presented here, the device used was the BIOPAC 100A, which is a continuous wave (CW) fNIRS. While this device offers a number of optic channels, the channels used here were those that were placed at the Fp1 and Fp2 positions. This was chosen for both comparison with previous research, as well as the practical implications of using a band sensor rather than individual sensors (as is the case with the Nonin device).

Forty-five participants completed experiment 1. For this experiment, participants were assigned to either a local, global, or no feature discrimination group where they were presented with all of the distracters. This meant that those in the no discrimination group did not have a unique component of the target object to identify. During this task, participants were presented with a brief display of the stimulus (250ms), before being presented with a black mask for 900ms. Distracter and target stimuli were presented in a randomized order, with a target probability of 0.14 and a distracter probability of 0.86. Participants completed this task continuously for 28 minutes and 12 seconds.

Sixty participants completed experiment 2. The stimulus presentation and length of the task was the same as used in Experiment 1. In this experiment, however, four separate groups were used; global traditional formatted task, local traditional formatted task, global SART (sustained attention to response task), and local SART. For the traditional formatted task (TFT) groups, participants were required to respond (press the space bar) only whenever they saw the target. For the SART tasks, participants were required to press the space bar whenever they saw a distracter target, and withhold their response whenever they saw a target. Whereas TFT vigilance tasks predominantly result in activation of areas of the brain associated with sustained attention, the SART evokes additional activation in areas of the brain associated with inhibition and motor control. These different areas of activation should result in differences in oxygenation.

Results

For all of the results, oxygenation was assessed in terms of change from baseline readings. For experiment 1, a 3 (condition: global, local or no discrimination) by 2 (hemisphere: left vs right) by 7 (period of watch) repeated measures ANOVA with orthogonal polynomial contrasts was performed. There was a significant time by condition linear effect, $F(2, 42) = 3.38$, $p = .043$, $\eta p^2 = .139$. Shown in Figure 3, the local discrimination shows a significant decrease in oxygenation over periods of watch. In contrast, the global and no discrimination conditions show little change over time, and show similar trends regardless of absolute level. Further, a simple contrast comparing the local group with the combined global and no discrimination group showed a significant difference, $F(2, 42) = 4.93$, $p = .012$, $\eta p^2 = .190$.

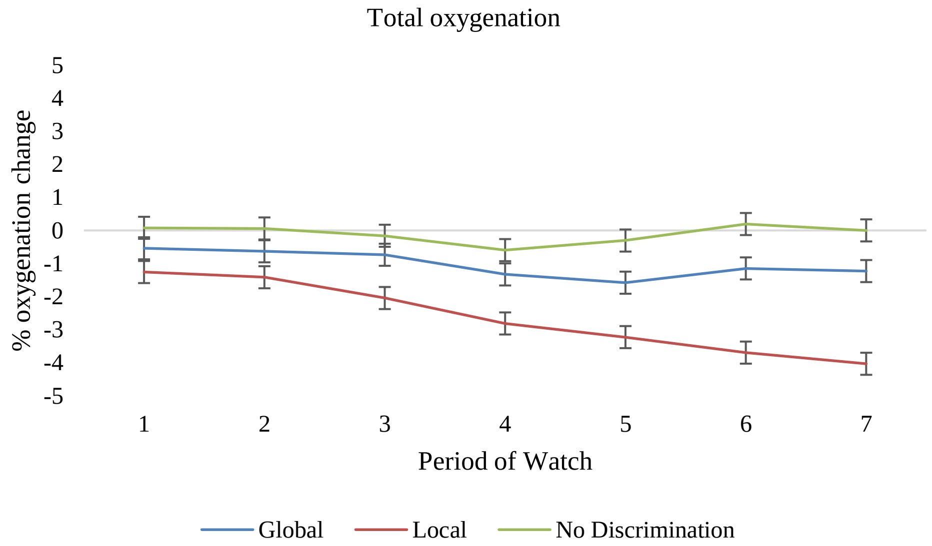


Figure 3. Mean percentage oxygenation change from baseline in Experiment 1.

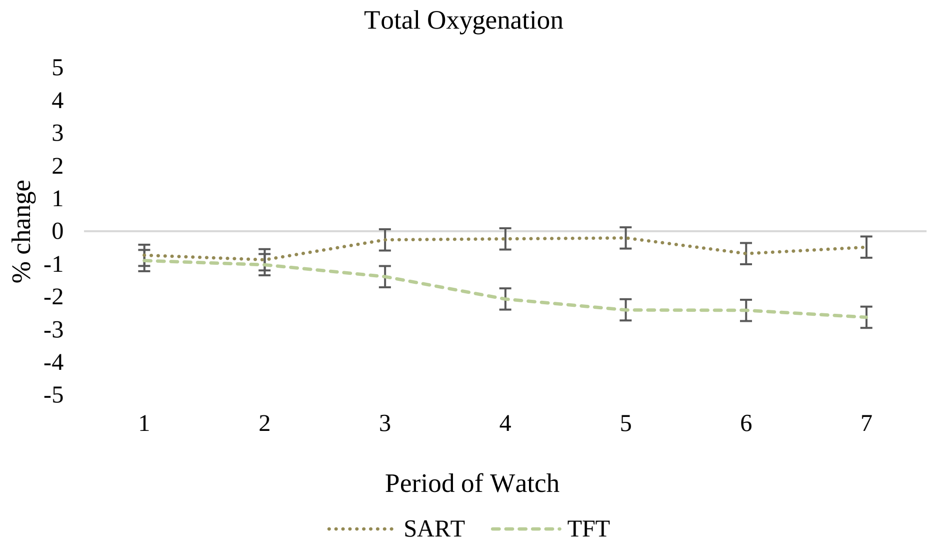


Figure 4. Total oxygenation over 7 periods of watch for TFT and SART response types. Error bars represent standard error.

For experiment 2, a 2 (group: local vs global) by 2 (response type: SART vs TFT) by 2 (hemisphere: left vs right) by 7 (period of watch) repeated measures ANOVA with orthogonal polynomial contrasts was performed on the oxygenation data. A significant period of watch by response type linear trend was found, $F(1, 54) = 7.93$, $p = .007$, $\eta p^2 = .128$, with the SART task and the TFT task showing different trends in regards to overall oxygenation (i.e., regardless of hemisphere). Overall oxygenation for the SART task shows little change from baseline, while the TFT task shows a sustained decrease from baseline over periods of watch (see Figure 4). A significant period of watch by response type by group linear trend was observed, $F(1, 54) = 4.90$, $p = .031$, $\eta p^2 = .083$. Shown in Figure 5, local and global discrimination groups show different oxygenation trends over time depending on the response type. For global discrimination, both TFT and SART groups show a slight decrease, although stay at the same relative level throughout the vigil. For the local discrimination groups, local TFT shows a steady decline in oxygenation over the vigil, whereas local SART shows an increase over the periods of watch.

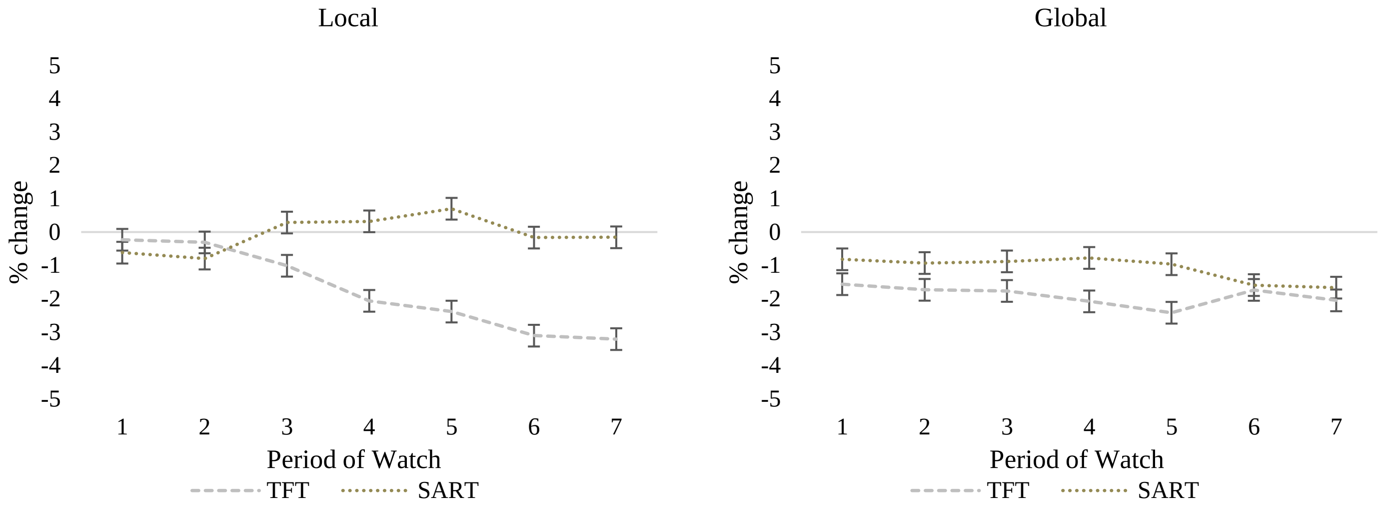


Figure 5. Total oxygenation over 7 periods of watch for local and global groups within the TFT and SART response types. Error bars represent standard error.

Discussion

Results in experiment 1 showed significant differences in overall oxygenation, with the local discrimination group showing a steady decline in oxygenation over periods of watch while the global and no discrimination groups showed no decline. Further, the global and no discrimination groups behaved similarly in terms of trends, with no significant differences between them. This is perhaps to be expected, as a no discrimination condition is likely to be closer to the brains "default" perceptual discrimination state, and previous research does show that people more readily respond to global stimuli initially [21]. Again, the local group resulted in faster reaction times compared to global and no discrimination conditions (which were extremely similar to each other). This finding mimicked that of previous research [14]. It should be noted that in this case, the fNIRS device used was a continuous wave fNIRS rather than a tissue saturation index fNIRS as used in the previous research. As such, it is likely that in this case a decrease in oxygenation represents higher activation [22]. Performance-wise however, this higher activation does not seem to result in any performance or workload deficits. If higher activation was indicative of higher workload requirements, then there should be a steeper vigilance decrement or decline. If this were the case, then the changes in oxygenation could more readily be attributed to workload. As this is not the case, then the changes in oxygenation should be attributed to differences in processing requirements alone.

The results of experiment 2 showed a number of key significant differences in overall oxygenation. First, there was a significant difference when comparing the TFT groups combined and the SART groups combined, with the TFT groups showing a decline in oxygenation over time, while the SART groups showed a slight increase. Furthermore, there were significant differences when comparing oxygenation trends in the local and global groups within their TFT and SART response formats. For the global discrimination conditions, both SART and TFT versions showed a slight decrease in oxygenation over time, with no significant differences between them. For the local discrimination conditions, the SART version of the task resulted in an increase in oxygenation over time, while the TFT version resulted in a decrease in oxygenation over time. The differences found in Experiment 2 are attributed to two key factors. The first is that differences in brain activation that are evoked by local and global feature discrimination, as shown in Experiment 1 and previous research. The second key factor is the type of response required (TFT or SART). The TFT is the more traditional sustained attention task, and requires the activation of a response from a resting state. The literature is clear that TFT results in greater activation in the prefrontal cortex [23], which is why a decrease is shown in Figure 4. The SART however requires that a participant

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

actively withholds a response that, over time, becomes a pre-programmed motor response. This has been shown to activate separate brain areas [2, 24], as opposed to the prefrontal cortex. As all of these experiments only observed oxygenation changes in the Fp1 and Fp2 positions (i.e., the prefrontal cortex), this increase activation was likely unable to be observed. This results in a lack of decrease in oxygenation in this case.

The combined findings from the three experiments described in the current paper provide an insight into how subtle changes in stimuli can result in differences in brain activation and cerebral oxygenation. The requirement to engage in a slightly different type of perceptual processing has been shown to have significant effects on oxygenation. Similarly, the type of response required has been shown to have a significant impact in experiment 2. While the current experiments are simple lab-based experiments, they do bring in to question the mechanisms behind oxygenation. With no corresponding differences in performance, and in fact the conditions with higher activation showing better performance, the results suggest that practitioners of fNIRS should show some caution when attributing changes in oxygenation to "increased mental workload" alone. That is not to say that there are not some effects that can be attributed to workload changes, however it is clear that not all changes are due to this. In order to obtain a more accurate account of oxygenation changes, users of fNIRS may need to determine the processing requirements of the tasks they are employing fNIRS devices in, and linking those oxygenation changes to activation in the corresponding neural networks.

References:

1. McKendrick R, Parasuraman R, Ayaz H. Wearable functional near infrared spectroscopy (fNIRS) and transcranial direct current stimulation (tDCS): expanding vistas for neurocognitive augmentation. Front Syst Neurosci. 2015;9:27. doi:10.3389/fnsys.2015.00027.
2. Rodrigo A, Domenico SI Di, Ayaz H, Gulrajani S, Lam J, Ruocco AC. Differentiating functions of the lateral and medial prefrontal cortex in motor response inhibition. Neuroimage. 2014;85:423-431. doi:10.1016/j.neuroimage.2013.01.059.
3. Bembich S, Lanzara C, Clarici A, et al. Individual Differences in Prefrontal Cortex Activity during Perception of Bitter Taste Using fNIRS Methodology. Chem Senses. 2010;35(9):801-812. doi:10.1093/chemse/bjq080.
4. Malcolm BR, Foxe JJ, Butler JS, De Sanctis P. The aging brain shows less flexible reallocation of cognitive resources during dual-task walking: A mobile brain/body imaging (MoBI) study. Neuroimage. 2015;117:230-242. doi:10.1016/j.neuroimage.2015.05.028.
5. Warm JS. Sustained Attention in Human Performance. Wiley, 1984.
6. Warm JS, Matthews G, Parasuraman R. Cerebral hemodynamics and vigilance performance. Mil Psychol. 2009; 21(Suppl 1):S75-S100. doi: 10.1080/08995600802554706.
7. Harrison J, Izzetoğlu K, Ayaz H, et al. Cognitive workload and learning assessment during the implementation of a next-generation air traffic control technology using functional near-infrared spectroscopy. IEEE Trans Human-Machine Syst. 2014;44(4):429-440. doi:10.1109/THMS.2014.2319822.
8. Gateau T, Durantin G, Lancelot F, Scannella S, Dehais F. Real-Time State Estimation in a Flight Simulator Using fNIRS. Chao L, ed. PLoS One. 2015;10(3):e0121279. doi:10.1371/journal.pone.0121279.
9. Davies D, Parasuraman R. The Psychology of Vigilance. Academic Press; 1982. https://books.google.co.uk/books/about/The_psychology_of_vigilance.html?id=1Kh9AAAAMAAJ.
10. Helton WS, Warm JS, Tripp LD, Matthews G, Parasuraman R, Hancock PA. Cerebral lateralization of vigilance: a function of task difficulty. Neuropsychologia. 2010;48(6):1683-1688. doi:10.1016/j.neuropsychologia.2010.02.014.
11. Shaw TH, Warm JS, Finomore V, et al. Effects of sensory modality on cerebral blood flow velocity during vigilance. Neurosci Lett. 2009;461(3):207-211. doi:10.1016/j.neulet.2009.06.008.
12. Shaw TH, Nguyen C, Satterfield K, Ramirez R, McKnight PE. Cerebral hemovelocity reveals differential resource allocation strategies for extraverts and introverts during vigilance. Exp brain Res. November 2015. doi:10.1007/s00221-015-4481-8.
13. Ossowski U, Malinen S, Helton WS. The effects of emotional stimuli on target detection: indirect and direct resource costs. Conscious Cogn. 2011;20(4):1649-1658. doi:10.1016/j.concog.2011.08.015.
14. de Joux NR, Russell PN, Helton WS. A functional near-infrared spectroscopy study of sustained attention to local and global target features. Brain Cogn. 2013;81(3):370-375. doi:10.1016/j.bandc.2012.12.003.

15. de Joux NR, Wilson K, Russell PN, Helton WS. The effects of a transition between local and global processing on vigilance performance. J Clin Exp Neuropsychol. 2015;37(8):888-898. doi:10.1080/13803395.2015.1068744.
16. de Joux NR, Wilson KM, Russell PN, Finkbeiner KM, Helton WS. A functional near-infrared spectroscopy study of the effects of configural properties on sustained attention. Neuropsychologia. 2017;94:106-117. doi:10.1016/j.neuropsychologia.2016.12.001.
17. Helton WS, Hayrynen L, Schaeffer D. Sustained attention to local and global target features is different: performance and tympanic membrane temperature. Brain Cogn. 2009;71(1):9-13. doi:10.1016/j.bandc.2009.03.001.
18. Martinez A, Moses CAP, Frank L, Buxton R, Wong E, Stiles J. Hemispheric asymmetries in global and local processing : evidence from fMRI. 1997;8(7):1685-1689.
19. Kimchi R. Primacy of wholistic processing and global/local paradigm: A critical review. Psychol Bull. 1992;112(1):24-38. doi:10.1037//0033-2909.112.1.24.
20. Kimchi R, Amishav R, Sulitzeanu-Kenan A. Gender differences in global-local perception? Evidence from orientation and shape judgments. Acta Psychol (Amst). 2009;130(1):64-71. doi:10.1016/j.actpsy.2008.10.002.
21. Navon D. Forest before trees: The precedence of global features in visual perception. Cogn Psychol. 1977;9(3):353-383. doi:10.1016/0010-0285(77)90012-3.
22. Hitchcock EM, Warm JS, Matthews G, et al. Automation cueing modulates cerebral blood flow and vigilance in a simulated air traffic control task. Theor Issues Ergon Sci. 2003;4(1-2):89-112. doi:10.1080/14639220210159726.
23. Kinomura S, Larsson J, Gulyás B, Roland PE. Activation by Attention of the Human Reticular Formation and Thalamic Intralaminar Nuclei. Science (80-). 1996;271(5248). http://science.sciencemag.org/content/271/5248/512. Accessed June 22, 2017.
24. Swick D, Ashley V, Turken AU. Left inferior frontal gyrus is critical for response inhibition. BMC Neurosci. 2008;9(1):102. doi:10.1186/1471-2202-9-102.

Crowdsourcing a Corpus of Eye Tracking Data on Web Pages: A Methodology

Sukru Eraslan[1], Yeliz Yesilada[1] and Simon Harper[2]

1. Middle East Technical University, Northern Cyprus Campus, 99738 Kalkanlı, Güzelyurt, Mersin 10, Turkey; {seraslan, yyeliz}@metu.edu.tr; 2. School of Computer Science, University of Manchester, Manchester, United Kingdom. simon.harper@manchester.ac.uk

Introduction

Eye tracking studies are increasingly used for understanding how people interact with web pages [1]. However, these studies are typically costly and time-consuming [2,3,4]. Some of the main reasons are the difficulty in finding participants, the cost of equipment and the need for a separate session for each participant. Due to these reasons, eye tracking studies are typically conducted with a small number of users. Therefore, the statistical analysis of the collected data tends to be underpowered which possibly causes a failure to detect a significant difference (Type II Error) [5]. One of the solutions for this problem is to allow researchers to crowdsource eye tracking data from other researchers and combine them in an appropriate way to have a large amount of data for their studies, if possible. In this paper, we propose a methodology for creating a corpus of eye tracking data on web pages. This corpus will facilitate crowdsourcing eye tracking data collected on web pages.

Researchers can analyse eye movements of users on web pages to discover which areas are attractive or distractive, how many times and how long these areas are visited and in which order. Based on their findings, they can propose a set of design guidelines or further process web pages to improve users' experiences. For example, an eye tracking study conducted by [6] on web forms illustrates that when error messages are displayed on the right side of erroneous input fields, they are noticed fastest by users. Another eye tracking study conducted by [7] on a web page with a number of portlets suggests that when there are multiple portlets on a web page, important ones should be placed on the upper-left or top of the page to minimise searching time. There are also some eye tracking studies which investigate differences between groups, such as people with and without autism [8].

Figure 1 visualises eye movements of a particular user on the home page of the AVG website by using a gaze plot. This visualisation technique uses circles to illustrate fixations where the eyes of users become relatively stable. When one fixation is longer than another one, it is illustrated with a larger circle. Each circle is also numbered to show the sequence of fixations.

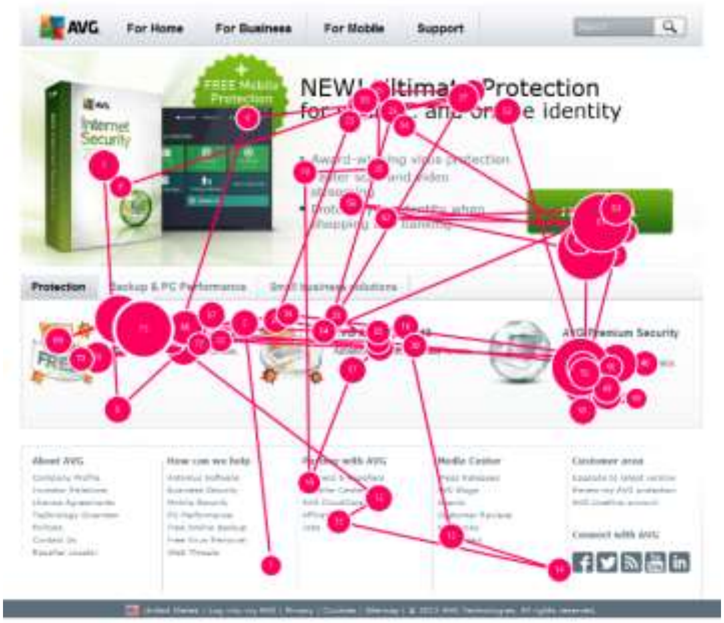


Figure 1. A gaze plot on the home page of the AVG website

There is, unfortunately, no specialised public repository or corpus of eye tracking data on web pages where researchers can upload their data or crowdsource some data to answer their research questions, test their hypotheses and/or validate their proposed algorithms. This situation may be one of the reasons why researchers typically do

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

not make their eye tracking data available to other researchers. When they do not share their eye tracking data, their findings cannot be reproduced and/or re-validated later by other researchers [9]. In addition, no benchmark can be created for other researchers to validate their algorithms. For example, when a researcher proposes an algorithm and wants to compare it with an existing algorithm whose validation data is not publicly available, s/he needs to collect new eye tracking data or try to find other public previously collected eye tracking data for a comparison purpose. For example, when [10] developed a scanpath analysis algorithm, they could not find public appropriate eye tracking data to compare their algorithm with other existing algorithms. Therefore, they needed to conduct a new eye tracking study. However, in some other established fields, such as Natural Language Processing (NLP)[5] and Medical Image Research[6], there are corpuses where other researchers can use to develop and evaluate their systems.

By creating a public shared corpus as proposed in this paper, researchers will be able to add new eye tracking data, access previously collected eye tracking data and combine some eye tracking data whenever possible. When researchers find appropriate eye tracking data in the corpus for their studies, they can replicate the study to collect more eye tracking data for achieving higher power in their statistical analysis. The corpus will be organised in a way that researchers can obtain all the required information about eye tracking data (equipment, participants, web pages and tasks) and filter eye tracking data based on different features, such as visual complexities of web pages.

A Corpus of Eye Tracking Data on Web Pages

To create a methodology for a possible public and shared corpus of eye tracking data on web pages, we have used a structured review methodology to investigate what kinds of eye tracking studies have been carried out in the literature in terms of their equipment, participants, pages, tasks and metrics. We have focussed on the academic papers published in the Journal of Eye Movement Research (JEMR)[7] and the Proceedings of ACM Symposium on Eye Tracking Research & Applications (ETRA)[8] as these are main venues for publication on eye tracking research. After analysing these sources in depth, we have found 20 relevant academic papers (9 journal articles [6,11,12,13,14,15,16,3,17] + 11 conference papers [18,19,20,21,22,23,24,25,26,7,2]). These papers show that different kinds of eye tracking studies have been conducted on web pages. The corpus should be able to store all the essential details of these eye tracking studies and make them usable.

Equipment: Different equipment has been used to record eye movements of participants on web pages. In particular, some of eye trackers used are as follows: Tobii TX300 [18], Tobii 1750 [23,24], EMR-HM8 [19], ISCAN RK-426 [22,25], ASL Model 504 [26] and SMI EyeLink [7]. These eye trackers capture data points with various sampling rates. For example, the sampling rate of Tobii TX300 is 300 Hz whereas the sampling rate of Tobii 1750 is 60 Hz, even though these eye trackers are products of the same company. Specifically, Tobii TX300 captures 300 data points per second whereas Tobii 1750 captures 60 data points per second. These data points are then clustered to create fixations [27]. Eye trackers can also have various accuracy rates. Therefore, if a particular eye tracking study is conducted by using different eye trackers, it would be a problem to combine eye tracking data later. Some of eye trackers are integrated into monitors. For example, Tobii TX300 is integrated into a 28” monitor whereas Tobii 1750 is integrated into a 17” monitor. There are also portable eye trackers which are not integrated into any monitors by default. For these kinds of eye trackers, researchers should use an external monitor. In particular, [7] used SMI EyeLink with a 17” external monitor. Thus, screen size and resolution should be taken into consideration for combining eye tracking data. Therefore, a public corpus should include at least the following details for the equipment:

- Eye Tracker – its brand and model should be provided, such as Tobii TX300.
- Sampling Rate – such as 60 Hz.
- Accuracy Rate – such as 0.5 degree.
- Screen Size – such as 17”.
- Screen Resolution – such as 1280 x 1024.
-

[5] An example of corpus in NLP: https://nlp.stanford.edu/projects/snli/
[6] A list of open-access medical image repositories: http://www.aylward.org/notes/open-access-medical-image-repositories
[7] https://www.bop.unibe.ch/jemr
[8] https://dl.acm.org/event.cfm?id=RE110

Participants: The behaviours of different groups of users on web pages have been investigated by analysing their eye movements. For example, an eye tracking study conducted by [21] aimed to investigate the influence of web-browsing experience on web-viewing behaviour. They divided their participants into two groups where the comparatively heavy internet users who browse and use various portal sites were located in the first group and the comparatively conservative users who use a regular portal site for fixed purposes were located in the second group. Their findings suggest that web-browsing experience influences web-viewing behaviour. Another eye tracking study conducted by [22] aimed to investigate the effects of age differences in visual search on web pages. In their study, they divided their participants into different age groups (children, teenagers, younger adults and older adults). According to their analysis, the groups generated based on the similarity of their eye movements were not consistent with the age groups.
The number of participants included in these eye tracking studies has also been varied. Figure 2 illustrates the number of participants included in some eye tracking studies which have been conducted on web pages. In particular, [23] conducted their eye tracking study with 11 participants (four males + seven females) whereas [26] conducted their eye tracking study with 30 participants (17 males and 13 females) who were students from different majors (such as, business, engineering and psychology). Unfortunately, there is no consistency among the number of participants. One of the main reasons would be the difficulty of conducting eye tracking studies with a large number of users as discussed in the Introduction section.

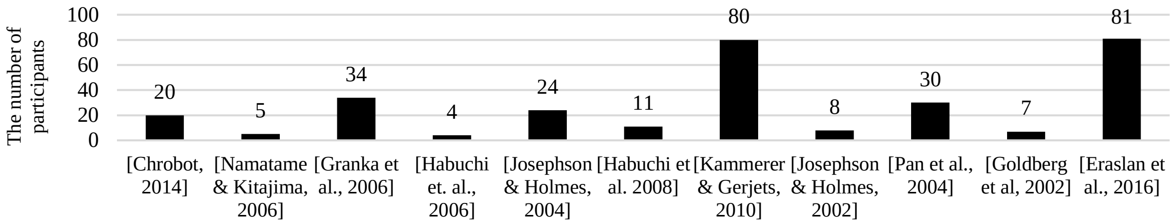


Figure 2. The number of participants included in some eye tracking studies conducted on the web

The corpus should store the total number of participants for each eye tracking study. If there are specific groups in an eye tracking study, then each group should have a unique identifier and the details of these groups should be included. For the participants and their demographics, the corpus should include at least the following details:

- Age – which can also be an age group (such as 18-24) or date of birth.
- Gender – which can be either male or female.
- Education Level – which can be a high school degree, associate's degree, bachelor's degree, master's degree, doctoral degree, etc.
- Education Background – which can be business, computer engineering, electrical engineering and psychology, etc.
- Web-browsing experience – which can be reported based on the information retrieved from the participants by using a five-point Likert scale.
- Group identifier(s) – which shows the group of the participant if s/he belongs to a specific group.
- Familiarity with the web page(s) used – which can be reported based on the information retrieved from the participants by using a five-point Likert scale.

To maintain confidentiality and anonymity for the participants, there should be no information that can reveal their identities. This criterion can be checked and validated by an administrator or moderator of the corpus before making the uploaded data public.

Pages: Different types of web pages have been used in eye tracking studies. For example, [25] conducted an eye tracking study on three different web pages which were from the portal, news and advertisement websites. The portal page consisted of many links, the news page was comprised of a lot of headings and sub-headings and the advertisement page contained many colourful visual items. In another eye tracking study conducted by [26], one popular website from each of the following categories was assigned to each participant: Business (jobtrack, macromedia & w3), Search (google, yahoo & cnet), News (cnn, msn & netscape) and Shopping (amazon & ebay). For each website, its home page and one of its internal representative pages were used. In addition, an eye tracking study conducted by [2] was performed on six web pages which had different levels of visual complexities determined by the ViCRAM (Visual Complexity Rankings and Accessibility Metrics) tool [28]. One of the main reasons was to achieve more generalizable results.

There were also some eye tracking studies that aimed to investigate how people interact with a specific web page. For example, [24] conducted an eye tracking study on search engine results pages (SERPs). They investigated how eye movements of users are affected when search results are presented in different ways (list and grid layouts). Furthermore, some researchers designed experimental web pages based on their purposes and conducted eye tracking studies on these pages [19,23,6,7]. Based on our current review, as illustrated in Figure 3, real web pages have been more frequently used instead of experimental web pages and there were more eye tracking studies with experimental pages in the past years.

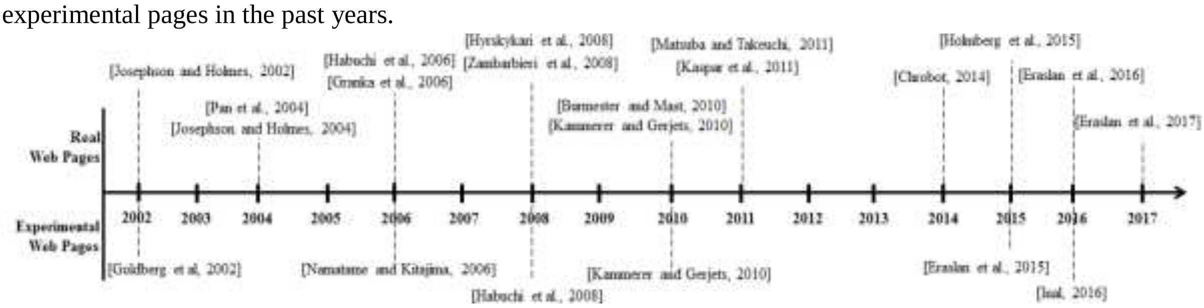


Figure 3. Real vs. experimental web pages used in eye tracking studies published in the JEMR and ETRA symposium

Regarding the web pages, the corpus should include at least the following details or features:

- Popularity – which can be measured by the reports provided by Alexa.com[9].
- Links – which can be coverage of links (as a percentage) or the number of links.
- Visuals – which can be coverage of visuals (as a percentage) or the number of visuals.
- Heading and subheadings – which can be coverage of heading and sub-headings (as a percentage) or the number of heading and sub-headings.
- Category – which can be determined by using the categories provided by Alexa.com[10].
- Visual Complexity – which can be measured by the ViCRAM tool [28].

Tasks: In some eye tracking studies, the participants have been asked to freely browse web pages for a specific amount of time, such as 30 seconds [26] and 15 seconds [25]. In other eye tracking studies, the participants have been asked to complete specific tasks. For example, [22] asked their participants to search for a "sign up" link on each page whereas [23] asked their participants to find out an article about a specific item in a simulated encyclopaedia in 130 seconds. Another example is from [24] who asked the participants to seek information in two sets of SERPs about two competing therapies (pros/cons) for Bechterew's disease (4 minutes for each set). As can be seen from these examples, different kinds of tasks have been performed by the participants in eye tracking studies. Thus, these tasks should be categorised to make them more searchable and usable by other researchers.

For the corpus, we suggest categorising tasks based on G. Marchionini’s Search Activities Model [29]. In his model, there are three main categories which are as follows: Lookup (fact retrieval, known item search, navigation, transaction, verification, question answering), Learn (knowledge acquisition, comprehension/interpretation, comparison, aggregation/integration, socialize) and Investigate (accretion analysis, exclusion/negation, synthesis, evaluation, discovery, planning/forecasting, transformation). Table 1 shows an example of task categorisation.

Task	Reference	Category
To freely browse web pages	[26]	Investigate
To search for a "sign up" link on each page	[22]	Lookup
To seek information in two sets of SERPs about two competing therapies for Bechterew's disease	[24]	Learn

Table 1. An example of task categorisation based on G. Marchionini’s Search Activities Model [29]

The corpus should store at least the following information for the tasks:

[9] https://www.alexa.com/topsites
[10] https://www.alexa.com/topsites/category

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

- Task Definition – what will be asked to the participants?
- Duration – how many seconds or minutes will be given for each task?
- Category – which can be lookup, learn or investigate.

Metrics: The eye tracking data collected on the web has been analysed by using various metrics based on the objective of the studies. For example, [25] and [13] aimed to test the scanpath theory [30] on the web, and therefore they analysed the scanpaths of their participants. However, as the objective of [21] was not related to scanpaths, they did not analyse the scanpaths of their participants. Instead, they divided the pages into six blocks and calculated the average number of fixations for each block for each group of participants. In contrast, a study conducted by [26] included both the analysis of scanpaths and other eye tracking metrics, such as mean fixation duration, saccade rates, etc. Even though an eye tracking study is conducted for a specific purpose, it can be used by other researchers for other purposes. Specifically, an eye tracking study might be conducted for only scanpath analysis but other researchers can use the study for in-depth fixation analysis.
The corpus should include all the fundamental information required to generate all these necessary metrics for analysis of eye tracking data. Since different algorithms can be used for clustering data points to identify fixations [27], it would be better to provide information about data points instead of fixations. Therefore, the corpus should include at least the following details of eye tracking data:

- Data point timestamp and duration
- Data point x and y coordinates
- Relevant Stimuli Name

When these details are made available to other researchers, they can generate necessary metrics for their studies. For example, they can use these details to generate total fixation duration and fixation count for a web page or a particular area on a web page.

Examples of Use Cases
An eye tracking study was conducted on the home pages of six popular websites by [10] to evaluate their algorithm. The study was then replicated with other 41 participants to further evaluate the algorithm and investigate the effects of the number of participants on the algorithm [2,3,31]. Therefore, a large amount of eye tracking data was collected with 81 participants. When the proposed corpus is implemented and the eye tracking data is uploaded with all necessary details, other researchers will be able to access and use the data as secondary data for their studies. If they have the required equipment, they can also replicate the study with more participants as all necessary information will be available to them.
By storing eye tracking data in an organised way and making them centralised, other researchers will be able to search for the most relevant data for their studies. For example, if researchers want to investigate how users interact with a particular type of web pages instead of specific web pages to reach more generic conclusions, they can search for the eye tracking data collected on that type of web pages. If the corpus is used by many researchers, then it will be possible to crowdsource a large amount of eye tracking data, and therefore researchers will not need a considerable amount of time and budget to conduct a new eye tracking study with many participants. The corpus will also be able to allow the external validation of proposed algorithms and replication of previous findings.

Concluding Remarks and Future Work
Eye tracking has commonly been used for understanding how people use web pages [1]. As eye tracking researchers, we typically experience difficulties in finding public eye tracking data for our studies. Of course, we can conduct our eye tracking studies to collect data but unfortunately these studies are costly and time-consuming. Our overall objective is to create a public corpus of eye tracking data on web pages which will allow researchers to make their eye tracking data publicly available in an organised way and access other previously collected eye tracking data for their studies. To make the corpus more usable by researchers, all essential information should be provided for each eye tracking study to allow researchers to determine whether or not these studies are suitable for their purposes. Therefore, we have firstly reviewed the literature to investigate what kinds of eye tracking studies have been conducted and which details should be provided when eye tracking data will be made publicly available. This paper summarises our initial list of principles. Although we suggest to include all of these details for each eye tracking study in the corpus, some of them can be defined as optional properties such that researchers will not

have to spend additional time to prepare all of these details. However, when all of these details are provided, eye tracking data will be more valuable and usable for other researchers.

The proposed methodology here has not been finalised. In this paper, we have focused on two specific venues for publications but of course the review could be extended. We are also planning to look at other venues, such as the Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)[11] and the Proceedings of the International World Wide Web Conference (WWW)[12]. We are also planning to conduct interviews with eye tracking researchers about the methodology and further improve it based on their suggestions. We may identify other information which should be provided when eye tracking data is uploaded to the corpus. When the methodology becomes ready, we are planning to implement the corpus and make it available to some researchers for testing. This will allow us to detect any possible usability problems and give some ideas for further improvement. Although we are currently planning to host and maintain the corpus by ourselves, it may also be hosted and maintained by a particular publisher in the long run.

This corpus will make useful contributions to eye tracking research on the web. It may lead researchers to make their eye tracking data publicly available, thus improving the overall reproducibility of their findings. Other researchers can also benefit from public eye tracking data. In particular, they can use public data for iterative development of their algorithms. When researchers do not share their eye tracking data, benchmarks cannot be created. Therefore, researchers cannot find any standard or point of reference which they can use for assessing their proposed algorithms. This corpus would also be beneficial for data scientists who want to access and use public eye tracking data with different types of data for their analyses.

References

1 Ehmke, Claudia and Wilson, Stephanie. Identifying Web Usability Problems from Eye-tracking Data. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI.But Not As We Know It - Volume 1* (UK, Swinton 2007), British Computer Society, 119-128.

2 Eraslan, Sukru, Yesilada, Yeliz, and Harper, Simon. Eye Tracking Scanpath Analysis on Web Pages: How Many Users? In *Proceedings of the Ninth Biennial ACM Symposium on ETRA* (New York, NY, USA 2016), ACM, 103-110.

3 Eraslan, Sukru, Yesilada, Yeliz, and Harper, Simon. Less users more confidence: How AOIs don't affect scanpath trend analysis. *Journal of Eye Movement Research*, 10, 4 (2017), article 6.

4 Pernice, Kara and Nielsen, Jakob. *How to Conduct Eyetracking Studies*. Nielsen Norman Group, 2009.

5 Gravetter, Frederick J. and Wallnau, Larry B. *Statistics for Behavioral Sciences*. Wadsworth Publishing, 2008.

6 Inal, Yavuz. User-Friendly Locations of Error Messages in Web Forms: An Eye Tracking Study. *Journal of Eye Movement Research*, 9, 5 (2016), article 1.

7 Goldberg, Joseph H., Stimson, Mark J., Lewenstein, Marion, Scott, Neil, and Wichansky, Anna M. Eye Tracking in Web Search Tasks: Design Implications. In *Proceedings of the 2002 Symposium on ETRA* (New York, NY, USA 2002), ACM, 51-58.

8 Eraslan, Sukru, Yaneva, Victoria, Yesilada, Yeliz, and Harper, Simon. Do Web Users with Autism Experience Barriers When Searching for Information Within Web Pages? In *Proceedings of the 14th W4A Conference* (New York, NY, USA 2017), ACM, 20:1-20:4.

9 Collberg, Christian and Proebsting, Todd A. Repeatability in Computer Systems Research. *Commun. ACM*, 59 (Feb. 2016), 62-69.

10 Eraslan, Sukru, Yesilada, Yeliz, and Harper, Simon. Scanpath Trend Analysis on Web Pages: Clustering Eye Tracking Scanpaths. *ACM Trans. Web*, 10 (Nov. 2016), 20:1-20:35.

11 Holmberg, Nils, Holmqvist, Kenneth, and Sandberg, Helena. Children's attention to online adverts is related to low-level saliency factors and individual level of gaze control. *Journal of Eye Movement Research*, 8, 2 (2015), article 2.

[11] https://dl.acm.org/event.cfm?id=RE151

[12] https://dl.acm.org/event.cfm?id=RE334

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

12 Zambarbieri, Daniela, Carniglia, Elena, and Robino, Carlo. Eye Tracking Analysis in Reading Online Newspapers. *Journal of Eye Movement Research*, 2, 4 (2008), article 7.

13 Burmester, Michael and Mast, Marcus. Repeated Web Page Visits and the Scanpath Theory: A Recurrent Pattern Detection Approach. *Journal of Eye Movement Research*, 3, 4 (2010), article 5.

14 Hyrskykari, Aulikki, Ovaska, Saila, Majaranta, Pävi, Räihä, Kari-Jouko, and Lehtinen, Merja. Gaze Path Stimulation in Retrospective Think-Aloud. *Journal of Eye Movement Research*, 2, 4 (2008), article 5.

15 Matsuda, Noriyuki and Takeuchi, Haruhiko. Joint Analysis of Static and Dynamic Importance in the Eye-Tracking Records of Web Page Readers. *Journal of Eye Movement Research*, 4, 1 (2011), article 5.

16 Kaspar, Kai, Ollermann, Frank, and Hamborg, Kai-Christoph. Time-Dependent Changes In Viewing Behavior On Similarly Structured Web Pages. *Journal of Eye Movement Research*, 4, 2 (2011), article 4.

17 Eraslan, Sukru, Yesilada, Yeliz, and Harper, Simon. Eye tracking scanpath analysis techniques on web pages: A survey, evaluation and comparison. *Journal of Eye Movement Research*, 9, 1 (2015), article 2.

18 Chrobot, Nina. The Role of Processing Fluency in Online Consumer Behavior: Evaluating Fluency by Tracking Eye Movements. In *Proceedings of the Symposium on ETRA* (New York, NY, USA 2014), ACM, 387-388.

19 Namatame, Miki and Kitajima, Muneo. Improving Web Usability for the Hard-of-hearing. In *Proceedings of the 2006 Symposium on ETRA* (New York, NY, USA 2006), ACM, 39-39.

20 Granka, Laura, Hembrooke, Helene, and Gay, Geri. Location Location Location: Viewing Patterns on WWW Pages. In *Proceedings of the 2006 Symposium on ETRA* (New York, NY, USA 2006), ACM, 43-43.

21 Habuchi, Yoshiko, Takeuchi, Haruhiko, and Kitajima, Muneo. The Influence of Web Browsing Experience on Web-viewing Behavior. In *Proceedings of the 2006 Symposium on ETRA* (New York, NY, USA 2006), ACM, 47-47.

22 Josephson, Sheree and Holmes, Michael E. Age Differences in Visual Search for Information on Web Pages. In *Proceedings of the 2004 Symposium on ETRA* (New York, NY, USA 2004), ACM, 62-62.

23 Habuchi, Yoshiko, Kitajima, Muneo, and Takeuchi, Haruhiko. Comparison of Eye Movements in Searching for Easy-to-find and Hard-to-find Information in a Hierarchically Organized Information Structure. In *Proceedings of the 2008 Symposium on ETRA* (New York, NY, USA 2008), ACM, 131-134.

24 Kammerer, Yvonne and Gerjets, Peter. How the Interface Design Influences Users' Spontaneous Trustworthiness Evaluations of Web Search Results: Comparing a List and a Grid Interface. In *Proceedings of the 2010 Symposium on ETRA* (New York, NY, USA 2010), ACM, 299-306.

25 Josephson, Sheree and Holmes, Michael E. Visual Attention to Repeated Internet Images: Testing the Scanpath Theory on the World Wide Web. In *Proceedings of the 2002 Symposium on ETRA* (New York, NY, USA 2002), ACM, 43-49.

26 Pan, Bing, Hembrooke, Helene A., Gay, Geri K., Granka, Laura A., Feusner, Matthew K., and Newman, Jill K. The Determinants of Web Page Viewing Behavior: An Eye-tracking Study. In *Proceedings of the 2004 Symposium on ETRA* (New York, NY, USA 2004), ACM, 147-154.

27 Salvucci, Dario D. and Goldberg, Joseph H. Identifying Fixations and Saccades in Eye-tracking Protocols. In *Proceedings of the 2000 Symposium on ETRA* (New York, NY, USA 2000), ACM, 71-78.

28 Michailidou, Eleni, Harper, Simon, and Bechhofer, Sean. Visual Complexity and Aesthetic Perception of Web Pages. In *Proceedings of the 26th Annual ACM International Conference on Design of Communication* (New York, NY, USA 2008), ACM, 215-224.

29 Marchionini, Gary. Exploratory Search: From Finding to Understanding. *Commun. ACM*, 49 (Apr. 2006), 41-46.

30 Noton, David and Stark, Lawrence. Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision research*, 11 (1971), 929-IN8.

31 Eraslan, Sukru, Yesilada, Yeliz, and Harper, Simon. Trends in Eye Tracking Scanpaths: Segmentation Effect? In *Proceedings of the 27th ACM Conference on Hypertext and Social Media* (New York, NY, USA 2016), ACM, 15-25.

Measuring Navigation Performance in Serious Games

Chris Ferguson, Herre van Oostendorp, Geert-Jan Giezeman, Sam de Redelijkheid, Egon L. van den Broek

Department of Information and Computing Sciences, Utrecht University, Utrecht, the Netherlands.
c.ferguson@uu.nl

Abstract: In serious games, in-game analytics is a major focus, as it is beneficial to see how a player is learning throughout the game. Different control schemes, such as ones often used in Virtual Reality (VR) games, open the door to a variety of existing analytical techniques. This article compares three measures adopted from two distinct areas to determine player's navigation behaviour, namely: task-finding lostness, gathering lostness and sequence similarity. Results on 13 children playing a serious Virtual Reality (VR) game, show both resemblances and differences among the three measures. Moreover, each of the measures show their constraints when applied in a serious VR game context. If anything, the current article illustrates, once more, how complex the analysis of player behaviour is and the need for the absent ground truth measure.

Introduction

Particularly in serious games, in-game analytics is a major focus, as it is beneficial to see how a player is performing in the game. Different control schemes, such as ones often used in Virtual Reality (VR) games, namely node-based movement, open the door to a variety of existing analytical techniques. Some of which can be adopted from other areas of computing, that could possibly be used to evaluate a player's performance within the game [11,12]. These are different to dedicated in-game analytics, which usually focus on heatmaps [1] or identifying switching behaviours, and the number of game objects (e.g., videos, resources, and locations) accessed per time unit [2].

This article compares three measures from two different areas of computing to determine player's (dis)orientation: task-finding lostness, gathering lostness and similarity sequence. The two lostness measures were originally introduced to analyse navigation in hypertext systems [5]. However, we hypothesize that these can be used for serious game navigation as well, as navigation is a key issue in the learning process. Additionally, a sequence similarity measure is included [6] that has been successful in a range of distinct domains including bioinformatics and music analysis. As such, this article contributes to methods and techniques in game playing analytics.

Distance measures

Lostness

The lostness measure [4] was created as a method for measuring hypertext usability in terms of task performance. The starting point for this was the belief of the authors that 'measures based on time and errors seem inappropriate for hypertext systems which, by their nature, encourage exploration and browsing'. Instead, a better approach is to assess task performance in terms of the effectiveness in how users find information and the degree in which they become lost in the information search whilst looking for this information [5].

Lostness [4] is defined by the number of information items inspected compared with the number of items that need to be inspected to locate the requested information:

$$\sqrt{\left(\frac{N}{S} - 1\right)^2 + \left(\frac{R}{N} - 1\right)^2},$$

with *R*, *S*, and *N* being respectively the minimum number of nodes that need to be visited to complete a task, the total number of nodes visited whilst searching, and the number of different nodes visited whilst searching.

Table 1. Comparison between old scope and new scope.

Initial (Hypertext) Scope	New (Game) Scope
Clicking a link to move to a new web page	Moving to another node within the environment

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Inspecting a web page containing information	Picking up an item
Finding information on a web page	Examining an item

Gathering versus Fact-Finding Lostness

Based on the identification of two types of activities, gathering and fact-finding, two lostness measures were constructed based on these. The gathering lostness measure involves calculating the minimum steps necessary to complete the whole task whereas fact-finding lostness calculates the lostness between each of the objectives in the task.

The difference between these two measures is that the gathering lostness is based on the number of steps identified manually imputed based on watching a perfect playthrough but, contrarily, fact-finding lostness is more automated. Instead, the spatial graph created for node-based movement, which defines which movement nodes are accessible from each node, is traversed and distance between each objective and the one accessed previously is calculated using a breath first search and is matched to the players path. This provides the minimum number of steps. This gives the opportunity to investigate if lostness is more representative if it is calculated per each objective or the whole task.

Needleman-Wunsch Similarity

Musaline [6] is a C++ library for alignment of sequences, which was initially developed for use with musical sequences; but, can also be used for comparing other sequences. It makes use of the Needleman-Wunsch algorithm [7]. In contrast to the lostness measure, 100% shows a perfect match, whereas 0% shows no match at all. To figure out how lost a player is, the path followed by the player is compared with the optimal path. While the lostness measure looks at path length and the number of unique nodes visited, the Needleman-Wunsch algorithm looks at two other aspects, namely:

- Identity, the visited nodes should be identical to nodes in the optimal path
- Order, the nodes should be visited in the same order as in the optimal path.

Given an optimal path and a player path, the Needleman-Wunsch algorithm efficiently searches through all possible alignments between both paths. An alignment is a list of nodes, where each node represents either a node in the optimal path, the player path or both, such that for each alignment we can reconstruct the user path by taking all nodes representing the user path and the optimal path by taking all nodes representing the optimal path.

A node in an alignment is considered correct if it represents both the player path and the optimal path, otherwise it is considered incorrect. We assign a cost to all incorrect steps and assigning a reward to each correct step. The similarity of an alignment is defined by summing over the costs and rewards for each step. Finally, the similarity of the paths is defined by taking the maximum similarity of all.

The cost and reward function chosen for the Chantry was based on several different user paths in the game and is 1 for a correct step and -0.35 for an incorrect step. Then we divide the similarity value by the number of steps in the optimal path and clamp it at 0 to create a value between 0 and 1. The values are chosen to match the gathering similarity values as closely as possible.

Weighted Measures

Simply taking the average of the lostness or similarity values does not suffice to obtain an exact overview of the player performance over the full game. The relatively short paths can be prone to noise and, due to the diminishing effect of extra steps, a player that gets completely lost on one or two locations may have a lower average lostness than someone who never gets completely lost; but, instead takes a side step on some otherwise perfect paths. Conversely, player performance could be overstated as low levels of lostness on simpler tasks may mask higher levels of lostness in more complicated tasks. Therefore, the complexity of tasks must be taken into account and weighted.

Although, for gathering lostness and sequence similarity, the values are simply weighted depending on the number of objectives needed to be completed in each gathering task, a different approach is taken for the fact-finding lostness. As well as calculating the lostness value for each objective, as the player also needs to move from the end of the objective to the start of another, the steps between the objectives are also taken into consideration. Therefore, this provides a weighted fact-finding lostness measure for the whole game:

$$\sqrt{\left(\frac{\sum_{o=1}^{n} N_o}{\sum_{o=1}^{n} S_o}-1\right)^2+\left(\frac{\sum_{o=1}^{n} R_o}{\sum_{o=1}^{n} N_o}-1\right)^2},$$

with n being the number of consecutive objective pairs in the player's path through the game and R_o, N_o, and S_o are respectively the required steps, total steps taken, and unique steps taken for the o^{th} objective pair.

Method

A total of 13 children (6 boys) aged 13-18, were recruited. Their VR experience varied substantially.

A PlayStation VR headset and over-ear headphones were used to secure the player's immersion with the game. A PlayStation controller was used to interact with the game.

Game play behavior was recorded time stamped, in the background. Consequently, it was possible to analyse the paths taken by the players and, hence, calculate the lostness and sequence similarity measures. Participants were given 30 minutes to play through the game at their own pace, learning about the story and completing tasks.

"The Chantry", an educational environmental narrative game for the PlayStation VR platform, was used as testbed. It tells the story of dr Edward Jenner and his invention of vaccination against the smallpox virus. To progress through the game, the player needs to explore the house of Jenner, finding out information about a particular story topic before moving on. Players interact with objectives that contain story information that must be found in order to continue. The full task is best described as a 'gathering activity' as the target information is spread out over different areas and has to be combined, whereas, finding each of the objectives present in the list can be referred to as more simple, local 'fact-finding activities' as all of the information is located in a specific place [3].

Results

The weighted lostness and sequence similarity values were investigated for all participants, to see how well they match each other. For this, the inverse of the lostness measures were used so that the perfect score represented by both measures would be 100%. So, a high (inverse) lostness score means that the participant exhibits a perfect search. The outcomes are reported in Table 2.

Overall, participants were not overly effective at the tasks, with all measures being below 54%. The difference between the gathering lostness and sequence similarity means is quite small (1.67), though fact-finding lostness is notably higher than both of these measures (9.47 higher than gathering lostness and 11.14 higher than sequence similarity). Moreover, compared to both the inverse gathering lostness and sequence similarity, the fact-finding measure gives somewhat different results. Inverse gathering lostness and sequence similarity (Pearson correlation .93, $p < .0001$) are almost equivalent, which was expected as the cost function of the sequence similarity was fitted to match the lostness value.

The data show that, across all measures, males performed better than females. This could be explained by males having more VR experience (mean: 2.33) compared to the females (mean: 1.71). However, these results could be contributed to gender differences in visual-spatial ability [8,9].

Table 2. Results for both inverse (weighted) lostness and the sequence similarity measures (in %).

	Gathering Lostness	Fact-Finding Lostness	Sequence Similarity
Total mean (SD)	44.07 (19.13)	53.54 (11.19)	42.40 (19.07)
Female mean (SD)	41.19 (22.38)	49.12 (10.63)	41.24 (23.43)
Male mean (SD)	47.42 (15.89)	58.69 (10.26)	43.74 (14.49)

Discussion

Altogether, three known analytic measures were introduced to a new domain: serious VR games. The measures indicate similar but, in parallel, distinct player behaviour. Additional studies must be carried out in order to identify which of these lostness measures is more accurate, which may also involve a fact-finding sequence similarity measure.

We signalled that each of the three measures has its constraints. For the lostness measures, the minimum steps necessary to complete a task must be constant. If the minimum value can unexpectedly change; then, an incorrect value will be returned. In addition, the path for lostness must be completely linear. Due to the nature of the measure, which is designed to punish node revisits, even if the minimum steps are correct, the user will be identified as lost. This is shown in the current study where one of the tasks returns a lostness value of 0.32, even when it has been completed perfectly. The sequence similarity allows multiple optimal orders, which is not always wanted. This is shown in the current study, where most tasks had 1-4 perfect paths, going up to 18 in one case.

One of the most illuminating results was that the mean task performance over most lostness measures was quite low, indicating a high level of disorientation in participants. This backs up that, as games are complex environments, players need to learn how to control the game and how it conveys the instructional material, before this material can be learned [10]. Therefore, to improve task performance, players should play a short tutorial to familiarise themselves with the environment, gameplay controls and mechanics, particularly the node movement system and picking up and using items.

Three distance measures were introduced to the domain of serious VR games. Each showed its behaviour and own constraints. If anything, they illustrated how complex the analysis of player behaviour is and the need for the absent ground truth measure. Results indicate the need for both basic research on intuitive distance measures in line with human behavior and their validation in ecological valid settings, such as serious VR games.

Acknowledgments

This work was supported by the REVEAL (Realising Education through Virtual Environments and Augmented Locations) Project, project number: 732599, funded under the European's Union's Horizon 2020 Research and Innovation Programme. Innovative Action (IA) H2020-ICT-2016-1. The experiment performed was carried out on the grounds of Sheffield Hallam University and approved by the University Ethics Committee (SHUREC).

References

1. Drachen, A., Schubert, M. (2013). Spatial Game Analytics. In: Seif El-Nasr M., Drachen A., Canossa A. (Eds.) *Game Analytics*. London, UK: Springer.
2. Westera, W., Nadolski, R., Hummel, H. (2014). Learning Analytics in Serious Gaming: Uncovering the Hidden Treasury of Game Log Files. In: A. de Gloria (Ed.), *Games and Learning Alliance. GALA 2013*. Lecture Notes in Computer Science, vol. 8605. Cham, Switzerland: Springer.
3. Puerta Melguizo, M.C., Vidya, U., van Oostendorp, H. (2012). Seeking information online: the influence of menu type, navigation path complexity and spatial ability on information gathering tasks. *Behaviour & Information Technology* **31**(1), 59-70.
4. Smith, P. (1996). Towards a practical measure of hypertext usability. *Interacting with Computers* **8**(4), 365-381.
5. Gwizdka, J., Spence, I. (2007). Implicit measures of lostness and success in web navigation. *Proceedings of the 31th European Conference on Cognitive Ergonomics (ECCE2013)*, Article no 17. New York: ACM.
6. Utrecht University (2012). Musaline: Musaline library. [online] Projects.science.uu.nl. Available at: https://www.projects.science.uu.nl/monochord/musaline/index.html [Last accessed 24 Apr. 2018].
7. Needleman, S.B., Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* **48**(3), 443-453.
8. Meijer, F., van den Broek, E.L. (2010). Representing 3D virtual objects: Evidence for an interaction between visuo-spatial ability and the nature of exploration. *Vision Research* **50**(6), 600–635.
9. Voyer, D., Voyer, S., Bryden, M.P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. Psychological Bulletin **117**(2), 250-270.
10. Wouters, P., van Oostendorp, H. (2013). A Meta-analytic Review of the Role of Instructional Support in Game-based Learning. *Computers & Education* **60**(1), 412-425.
11. van den Broek, E.L., Kuijpers, R. (2014). Game Over! On measurement and optimization of presence. In A.J. Spink, E.L. van den Broek, L.W.S. Loijens, M. Woloszynowska-Fraser, L.P.J.J. Noldus (Eds.), *Proceedings of Measuring Behavior 2014: 9th International Conference on Methods and Techniques in Behavioral Research*, p. 43–47. August 27–29, Wageningen, The Netherlands.
12. Sharma, K., Castellini, C., Stulp, F., van den Broek, E.L. (2018). Continuous, real-time emotion annotation: A novel joystick-based analysis framework. *IEEE Transactions on Affective Computing* **9**. DOI: 10.1109/TAFFC.2017.2772882. [in press]

Apathy's forms and neural basis in behavioral variant of frontotemporal dementia

Johan Ferrand-Verdejo[1], Bénédicte Batrancourt[1], Delphine Tanguy[1, 3], Rafaella 'Lara' Migliaccio and Richard Levy[1, 2]

1 Inserm U 1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMR S 1127, Institut du Cerveau et de la Moelle épinière, Paris, France. johan.ferrandverdejo@icm-institute.org 2 AP-HP, Hôpital de la Salpêtrière, Behavioural Neuropsychiatry Unit, Paris, France ; 3 Ecole Normale Supérieure, Paris Saclay, Cachan, France

Introduction

Apathy is a neurological syndrome present in many neurological diseases and in particular in the behavioral variant of fronto-temporal dementia (bvFTD). Apathy can be defined as a quantitative reduction of voluntary goal-directed behavior [1]. Its clinical evaluation remains subjectives as it relies on questionnaires in which sharp differences in responses can be observed between patients and caregivers [2]. Further, the underlying mechanisms and neural correlates of apathy are poorly understood. As consequence, management of apathy is still limited.

Differentiating the underlying mechanisms of apathy is crucial to propose more efficient treatments. Three different mechanisms have been postulated so far: motivational, cognitive and self-activation deficit, involving distinct regions of prefrontal cortex and basal ganglia [1, 3]. The present study is aimed at providing new and objective tools to characterize apathy, its different underlying mechanisms and to identify its neural basis, in patients affected by bvFTD.

Material & Methods

We assess apathy in bvFTD patients (n=7, mean age=59.14, mean Starkstein=15) and healthy subjects (n=4, mean age=63.25, mean Starkstein=8) using the Starkstein apathy scale [4].

The first goal was to determine the mechanisms (cognition and/or motivation) responsible for apathy in a population of bvFTD patients by using three different tools:

First, the Dimensional Apathy Scale **(DAS**) [5], a questionnaire assessing three dimensions of apathy: Executive, Emotional and Initiation. The first two subscales are used here to study cognition and motivation respectively.

Second, our team has developed a novel tool to assess apathy in semi-ecological condition (**ECOCAPTURE**, ClinicalTrials.gov/NCT02496312) [6]. The observation takes place in a waiting-room; instructions are to get comfortable and to enjoy the room. The experience consists of two successive phases in order to discriminate i) spontaneously (self-) generated (free phase) and ii) externally guided behaviors (guided phase). We focus on a few variables (the ratio of standing/walking and time-activity during the first 5 minutes of the free phase and the guided phase) allowing us to focus on another underlying pathophysiological mechanism: self-initiation, but it is also possible to analyse behavior to study the motivational and cognitive dimensions.

Thirs, we have also designed an experimental computer task, called **ICM_APATHY_TASKS**. This task is divided into two exercises with lower vs. higher difficulty, to study cognitive aspects of apathy. Moreover, each exercise is performed twice, by modifying the associated reward (lower or higher), in order to study the motivational dimensions.

The second goal was to identify the neural basis of different apathy forms in the bvFTD patients using **structural** Magnetic Resonance Imagery (**MRI**).

Results

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

To date, ECOCAPTURE has included seven bvFTD patients and four healthy controls. We will expect to enroll one additional patients and three healthy controls, for a total of eight patients and seven healthy controls, by the end of May 2018. At that time, we will then be able to present preliminary results of this study. Inclusion is still in progress for a total of 30 patients and 30 healthy controls, by the end of June 2019.

Ethical statement

We have obtained formal consent from the local ethical research committee: CPP, Comité de Protection des Personnes Sud-Méditerranée I, Région Provence Alpes Côte d'Azur, France. ECOCAPTURE, ID RCB: 2017-A00416-47, PI: Pr. Richard LEVY

Funding

Study Funded by the program "Investissements d'avenir" ANR-10-IAIHU-06; the foundation "Fondation pour la recherche médicale", FRM, frm.org; the ENEDIS company, Enedis - Tour Enedis, 34 place des Corolles - 92079 Paris La Défense - enedis.fr

References

1. Levy, R., and B. Dubois. "Apathy and the Functional Anatomy of the Prefrontal Cortex–Basal Ganglia Circuits." *Cerebral Cortex* 16, no. 7 (January 2006): 916–28.
2. Starkstein, Sergio E., Gustavo Petracca, Erán Chemerinski, and Janus Kremer. "Syndromic Validity of Apathy in Alzheimer's Disease." *American Journal of Psychiatry* 158, no. 6 (June 1, 2001): 872–77.
3. Brown, R.G. and G. Pluck. "Negative Symptoms: The 'pathology' of Motivation and Goal-Directed Behaviour." *Trends in Neurosciences* 23, no. 9 (September 1, 2000): 412–17.
4. Starkstein SE, Migliorelli R, Manes F, Teson A, Petracca G, Chemerinski E, et al. "The prevalence and clinical correlates of apathy and irritability in Alzheimer's disease". Eur J Neurol (1995); 2: 540-546.
5. Radakovic, R., and S. Abrahams. "Developing a New Apathy Measurement Scale: Dimensional Apathy Scale." *Psychiatry Research* 219, no. 3 (November 30, 2014): 658–63.
6. Lecouturier K., B. Batrancourt, J. Ferrand-Verdejo, C. Azuar and R. Levy. Unpublished.

Building empathy: Preparation of professionals and peers for an optimization of mediation

Pedro Ferreira Alves and Tâmara Rodrigues

Portuguese Association of Relational-Historic Psychology; psicologia.relacionalhistorica@gmail.com

The compromise between teaching and learning does not arise spontaneously, and therefore the Rita Leal School methodology allows the creation and recreation, in a significant way, of interactive and systematic cooperation and relational moments adequate to the evolution of each child. The practical aim of the Rita Leal School intervention lies on the development of the relational basis necessary to promote the basic confidence of the child, which is indispensable to the motivation to learn. All activities have been filmed with the objective of optimizing the efficacy of the intervention. These sessions have then been supervised in group meetings, to promote the attentive and relational quality of peers and professionals.

At the end of each session with the children and adolescents, peers and professionals were asked to film their testimony of that day's session. Testimonies were filmed because emotions are greatly expressed through non-verbal communication.

The analysis of these testimonies was then made with The Observer XT 12 software, together with an investigation model of qualitative content, the Analysis of Nuclei of Signification (Aguiar, Soares and Machado, 2015). This method intends to give an understanding of the meanings which constitute the content of discourse of this sample through nuclei of signification. 108 individual reports were analyzed. The verbalization of the experiences produced three principal nuclei of signification. A total of 38 discriminators were defined, organized in 3 main categories, which are "Self-appraisal" (10 indicators); "Hetero-evaluation" (11 indicators) and the description of ASD individuals, in turn subdivided into two categories: "Dialogic Relationship" (12 indicators) and "Cooperative Learning" (5 indicators).

We observed the reflexive-formative potential of autoscopy and heteroscopy with clinical supervision resorting to video feedback (French Socio-Historic Ergonomy). The process of auto and hetero analysis had as result a reflection of professional practice, with the intent of rendering the activity more scientific and the creation of individualized observation habits of each child, allowing a more personalized intervention. The purpose was to create an always more emphatical, more decentered team in which each team member (re)builds their own professional identity.

Reference

Aguiar, W. M. J., Soares, J. R., Machado, V. C. (2015). Núcleos de significação: Uma proposta histórico-dialética de apreensão das significações [Nuclei of signification: A historical-dialectical proposal of apprehension of meanings]. *Cadernos de Pesquisa* **45**(155), 56-75. doi: 10.1590/198053142818

The LocoWhisk Arena: A novel behavioural set-up that quantifies rodent exploration and locomotion behaviours

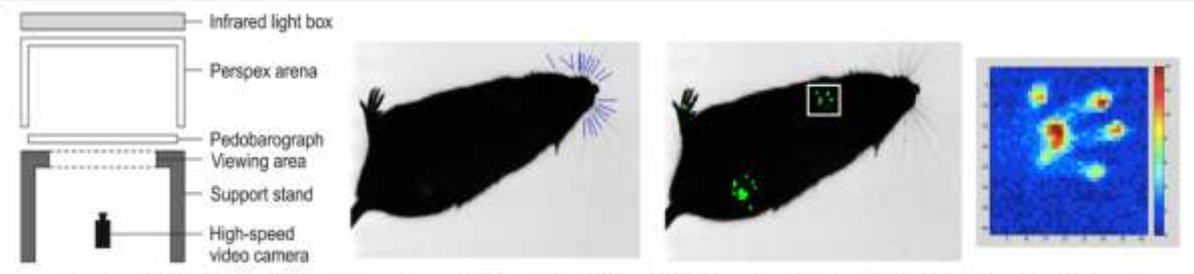
Robyn A Grant

Division of Biology & Conservation Ecology, Manchester Metropolitan University, robyn.grant@mmu.ac.uk

Quantitatively assessing behaviour to measure animal health is challenging because there is a lack of unobtrusive behavioural models to refine animal testing procedures. Our novel arena and software solution (the LocoWhisk arena) simultaneously measures locomotion behaviours and whisker movements, providing robust and quantitative measures of rodent behaviour for basic research or drug discovery. The LocoWhisk system is a new, portable behavioural set-up that incorporates both gait analysis (using a pedobarograph) and whisker movements (using high-speed video camera and infrared light source).

In this presentation, we will demonstrate the highly-quantitative, quick and automated nature of the LocoWhisk arena. We will show that using the LocoWhisk arena provides robust measures of motor control, and can give insights in to small mammal behavior that have never been documented and quantified before. Indeed, we will demonstrate that small quadrupedal mammals appear to use their whiskers to tactually guide safe foot positioning. We show that forepaw placement always falls within the whisker field of small mammals, and that forepaw width is always smaller than whisker span. Insights from the LocoWhisk arena has allowed us to propose that guiding locomotion might be a common function in whisker touch sensing in all small, quadrupedal mammals. We suggest that simultaneosly measuring both whisker movements and locomotion can provide robust measures of motor control.

All experiments were conducted with approval of the local ethics committee at Manchester Metropolitan University.



LocoWhisk: A novel behavioural arena that quantifies rodent exploration and locomotion behaviours

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

From BAP to DogBAP – Adapting a Human Body Movement Coding System for Use in Dogs

A. Huber[1,2], N. Dael[3], C. Caeiro[2,4], H. Würbel[1], D. Mills[2] , S. Riemer[1]

1 Division of Animal Welfare, University of Bern, Bern, Switzerland. annika.huber@vetsuisse.unibe.ch
2 School of Life Sciences, University of Lincoln, Lincoln, UK.
3 Institute of Psychology, University of Lausanne, Lausanne, Switzerland.
4 School of Psychology, University of Lincoln, Lincoln, UK.

Measuring behaviour in both humans and nonhuman animals requires robust methods to obtain reliable and replicable results, important prerequisites for the scientific quality of a study. Methods used for behavioural coding can differ on several levels: a) the level of subtlety of the behaviours measured [1], ranging from more holistic behaviours such as classifying the facial or bodily expression as a whole to the movement of single muscles (see e.g. [2]), b) the quality of behavioural descriptions used [1], or c) the level of possible subjective inference by the coder [3]. All these factors potentially affect the reliability and replicability of the results. In human research, methods have been developed for coding subtle behaviours with precisely defined movements, following a standardized coding protocol. These methods satisfy the scientific criteria of objectivity and reliability and they facilitate transparency, replicability and comparisons of results across studies [3]. For the face, the Facial Action Coding System (FACS [4]) is a prominent example. FACS is an anatomically-based coding system for measuring subtle facial muscle movements. Aiming to maximize reliable and objective coding, it is considered as the gold standard for coding human facial expressions [5]. FACS was initially developed for humans but has been adapted to several nonhuman species including both primates and non-primates (see e.g. [6, 7]).

In contrast, methods comparable to FACS for the coding of subtle expressions of the human body were lacking – until the Body Action and Posture Coding System (BAP [3]) was developed. Like FACS, BAP is anatomically-based and provides a standardized protocol for coding with systematic descriptions of subtle human body movements [3]. More specifically, BAP allows a highly objective and reliable coding of postures and actions of a human's neck/head, trunk, arms, and limbs by integrating information from three different levels: (1) The anatomical level (*different articulations of the codeable body parts*). (2) The form level (*direction and orientation of movement along the three human body axes*). (3) The functional level (*communicative and self-regulatory functions*). Since its development, the BAP coding method has been successfully applied in a range of studies across different research topics for measuring subtle body movements of humans (see e.g. [8, 9, 10]).

In animals, the use of methods comparable to the ones used in human research are considered to be promising. This has recently been proposed for investigating canine emotions [11]. Dogs are increasingly popular in research [12], and with their adaptations to living in human societies [13], they provide a unique model species for comparative investigations with humans ([13, 14]). Using the FACS method, very recently a comparative study revealed new insights into subtle facial expressions of humans and dogs in emotion-inducing situations [5]. So far, to the authors' knowledge, an anatomically-based method for systematically coding subtle body movements comparable to the human BAP system has been lacking in canine science. Nevertheless, a BAP coding system for dogs could greatly improve the objectivity and reliability of dog behaviour coding, supporting both transparency of analysis as well as reproducibility of study findings by other researchers. Furthermore, it would allow researchers to conduct comparative studies of human and dog bodily expressions.

Here, we present the Dog Body Action and Posture Coding system (DogBAP) – the first adaptation of the BAP coding system for a nonhuman species. DogBAP is a system that – like the human BAP – is anatomically-based, allowing the coding of postures and actions of the dog´s head, neck, trunk, limbs, and tail, following a standardized coding protocol. Movements of these body parts are defined in comparison to a dog's standard position according to two integrated levels: (1) The anatomical level (*anatomical articulation of the codeable body parts*), and (2) the form level (*form of the movement; described by using anatomical terms of direction with reference to three orthogonal axes of a dog's body (*ventrodorsal, craniocaudal, transversal*)*). To give an example, the head

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

articulation "Lateroflexion right" is defined as "Movement of the head in a lateral right direction around the ventrodorsal axis relative to the anatomical standard position". The DogBAP codes were developed by studying possible movements of the codeable body parts along the three orthogonal body axes according to a dog's anatomy. Afterwards, the DogBAP coding was first tested with a selection of 10 dog pictures showing 10 individuals in different bodily alignments. Subsequently, DogBAP was used with a selection of videos of freely moving dogs, collected during a laboratory-based experiment on emotional expressions of dogs (Ethical approval for this experiment was given by the University of Lincoln, College of Science Research Ethics Committee, UID CoSREC304). In this experiment, food was used to induce positive (via food anticipation after the dogs have learned to anticipate a food reward in a certain location of the experimental room) and negative (via food frustration when the dogs could see but not access the food reward for 60 s) emotional states.

After outlining the most relevant features of the human BAP coding system, we will compare and discuss the current version of DogBAP with its main developmental milestones, its evaluation in terms of reliability, possible future steps and potential areas of application both in research as well as in our daily practical handling of dogs.

References:

1. Benjamini, Y., Lipkind, D., Horev, G., Fonio, E., Kafkafi, N., Golani, I. (2010). Ten ways to improve quality of descriptions of whole-animal movement. *Neuroscience and Biobehavioral Reviews* **34**: 1351-1365.
2. Parr, L.A., Waller, B.M., Vick, S.J., Bard, K.A. (2007). Classifying chimpanzee facial expressions using muscle action. *Emotion* **7**: 172-181.
3. Dael, N., Mortillaro, M., Scherer, K.R. (2012). The body action and posture coding system (BAP): development and reliability. *Journal of Nonverbal Behavior* **36**: 97-121.
4. Ekman, P., Friesen, W., Hager, J. (2002). *Facial Action Coding System (FACS): manual and Investigator's Guide*. Salt Lake City, UT: Research Nexus.
5. Caeiro, C., Guo, K., Mills, D. (2017). Dogs and humans respond to emotionally competent stimuli by producing different facial actions. *Scientific Reports* **7**: 15525. doi: 10.1038/s41598-017-15091-4.
6. Caeiro, C.C., Waller, B.M., Zimmermann, E., Burrows, A.M., Davila-Ross, M. (2013). OrangFACS: A muscle-based facial movement coding system for orangutans (Pongo spp.). *International Journal of Primatology* **34:** 115–129.
7. Wathan, J., Burrows, A.M., Waller, B.M., McComb, K. (2015). EquiFACS: The Equine Facial Action Coding System. *PLOS ONE* **10**: e0131738.
8. Dael, N., Mortillaro, M., Scherer, K.R. (2012). Emotion expression in body action and posture. *Emotion* **12**: 1085-1101.
9. Walsh, J., Eccleston C., Keogh, E. (2014). Pain communication through body posture: The development and validation of a stimulus set. *Pain* **155**: 2282-2290.
10. Velloso, E., Bulling, A., Gellersen, H. (2013). AutoBAP: Automatic coding of body action and posture units from wearable sensors. *Proceedings - Humaine Association 5th International Conference on Affective Computing and Intelligent Interaction, ACII* (Geneva, 2-5 Sept 2013): 135-140.
11. Kujala, M. (2017). Canine emotions as seen through human social cognition. *Animal Sentience* **13**: 1-34.
12. Bensky, M.K., Gosling, S.D., Sinn, D.L. (2013). The world from a dog´s point of view: a review and synthesis of dog cognition research. *Advances in the Study of Behavior* **45**: 209-406.
13. Miklósi, Á., Topál, J., Csányi, V. (2004). Comparative social cognition: what can dogs teach us? *Animal Behaviour* **67**: 995-1004.
14. Topál, J., Miklósi, Á., Gácsi, M., Dóka, A., Pongrácz, P., Kubinyi, E., Vilmos, C. (2009). The dog as a model for understanding human social behavior. *Advances in the Study of Behavior* **39**: 71-116.

Applying camera trapping as a novel method for conservation/behaviour research of arboreal mammals

M.C. Kaizer[1], T.G. Alvim[2], C.L. Novaes[2] and R.J. Young[1]

1 School of Environment and Life Sciences, University of Salford, Manchester, United Kingdom. marikaizer@hotmail.com; 2 Caparaó Muriqui Project, Rede Ecodiversa para Conservação da Biodiversidade, Brazil.

Introduction

Half of the world's biodiversity is in the canopies of tropical forests [1], including a great richness of mammals and a large proportion of primates [2]. Most of arboreal mammals are threatened of extinction, and an understanding of their behaviour is a key component for the success of the conservation initiatives. However, quantifying the behaviour of wild animals represents significant challenges, especially for rare or elusive species in remote or dense forest habitats. Remote sensing camera-traps are increasingly applied in wildlife research, expanding the borders of traditional behavior ecological inference and conservation strategies, and allowing sampling of wild population at large temporal and spatial scales [3]. Although camera trap has been widely applied for terrestrial mammals monitoring, the technique has rarely been used for arboreal species [4,5]. The aim of this study is to present data on the methodology and effectiveness of canopy camera trapping for arboreal mammal's conservation behaviour research.

Methods

The study site was the Caparaó National Park, one of the most important remnants of Atlantic Forest, in southeastern Brazil (20°37' and 20°19' S; 41°43' and 41°55' W). At a potential aerial route (i.e., tree branches) through the canopy, we placed Bushnell 119774C camera traps with infrared LED and triggered by a motion sensor in 17 sampling locations at a mean height of 11m. We used rapid ascent/descent (RAD) and single rope (SRT) techniques to climb the trees [6], and to maximize detection probability, positioned the cameras to monitor the potential crossing points to capture animals. We set the cameras to work continuously, 24 hours a day, and to take two pictures followed by a 30-second video per trigger, with less than one second between photographs, five seconds between photographs and video, and an interval of one second between triggers. The photograph size was 8MP, and the HD video size of 1280 x 720 pixels per frame.

Results

After six months of data collection, we obtained over 32,000 records including photos and 30-second videos. Ten species of medium and large-sized mammals were recorded including the critically endangered primate Northern Muriqui (*Brachyteles hypoxanthus*). Other species of small-sized mammals, and reptile and bird species were also recorded. Analysis based on time-stamped camera data are useful to describe the temporal patterns of the species, including species' activity patterns and niche partitioning. The requirement for specialized climbing techniques is one of the logistic limitations of the canopy camera trapping, as well as the large potential for false triggers due the wind and non-target stimulus (e.g. leaves).

Conclusions

Our results suggest that canopy camera trapping can provide reliable data of behavioural indicators such activity repertories and daily activity patterns of a diversity of arboreal mammals' species. Animals do not change their activity patterns to avoid the cameras, indicating that canopy cameras are relatively non-invasive research method and have high applicability for conservation behavior research of arboreal animals in dense rainforest environments.

Ethical statement

This research adhered to the Brazilian legal requirements and to the American Society of Mammalogists (ASM).

References

1. Lowman, M.D., Devy, M.S., Ganesh, T. (2013). Treetops at risk: challenges of global canopy ecology and conservation. Annu. Rev. Ecol. Sys 27: 55-81.
2. Kays, R., Alyson, A. (2001). Arboreal tropical forest vertebrates: current knowledge and research trends. In: Linsenmair et al., (eds) Tropical forest canopies: ecology and management. Springer: 109-120.
3. Marvin, D.C., et al., (2016). Integrating technologies for scalable ecology and conservation. Global Conservation and Ecology 7: 262-275.
4. Gregory, T., Carrasco, R.F., Deichmann, J., Kolowski, J., Alonso, A. (2014). Arboreal camera trapping: taking a proven method to new heights. Methods in Ecology and Evolution 5(5): 443-451.
5. Bowler, M.T., Tobler, M.W., Endress, B.A., Gilmore, M.P., Anderson, M.J. (2016). Estimating mammalian species richness and occupancy in tropical forest canopies with arboreal camera traps. Remote Sensing in Ecology and Conservation: 1-12.
6. Maher, J. (2010). The Roof of the Rainforest: ITEC Manual for Canopy Access Techniques. The Tree Climber's Coalition and the Institute for Tropical Ecology and Conservation. Atlanta, Georgia.

Longitudinal multivariate measurement of emerging individuality

Gerd Kempermann[1], Anna Grzyb[1], Sara Zocher[1], Susan Schilling[1] and Andreas Brandmaier[2]

1. German Center for Neurodgenerative Diseases (DZNE) Dresden 2. Max Planck Institute for Human Development; gerd.kempermann@dzne.de

Gene x environment interactions are at the core of both developmental processes and plasticity across the lifespan. Historically, the experimental paradigm of environmental enrichment / enriched environments (ENR) has been used, primarily in rodents but by extension also in other species, to experimentally capture this interaction. The key principle is that inbred mice or rats are exposed to an laboratory environment that is richer in social and inanimate stimuli than the normal laboratory housing.

This research has mostly been concerned with assessing group differences in ENR conditions in comparison to standard laboratory housing. By and large, ENR exerts strong effects on a large variety of variables, especially the brain. But ENR induces also in increase in variance, so that the animals in the ENR become more diverse and different from each other than in the control conditions [1,2]. This means that individuality emerges, even if genes and environment are held constant. In addition, the correlation patterns among the variables fundamentally change. This effect is particularly prominent in the brain and seems essentially absent for example for metabolic parameters. The individualization of brain plasticity can to some extent be explained by variance in behavioral trajectories.

Neurobiological research on the emergence of individuality thus calls for a novel type of enriched environment, which allows the longitudinal assessment of behavioral parameters of large number of mice in one cage system as well as multivariate analysis tools. A technical solution to this challenge, which is presented here, is the INDIVIDUALITY cage, a system of 70 standard cages in one rack that are linked by connector tubes to allow free roaming across all cages for the entire population of mice held in this system (see Figure 1). Each connector tube is equipped with a ring antenna. The mice are tagged with RFID transponders, so that a longitudinal assessment of motility is possible. The data can be analyzed in various ways. A parameter that has proven to be extremely useful in this context, is roaming entropy (RE), which is essentially the probability that a mouse can be found at a given place at a given time and thus a measure of territorial coverage and exploration. The system allows also some indirect conclusions about social structure and its dynamics. Combined with a large set of cross-sectional mid- and endpoints measurements, this system allows a unique and unprecedented assessment of emerging variance and individualization in the ENR paradigm.

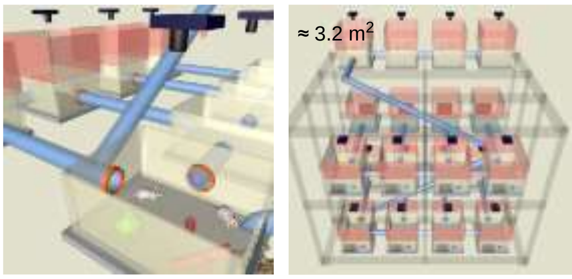


Figure 1. Schematic rendering of the INDIVIDUALITY cage system, in which up to 70 standard cages are connected by tubes (blue) that are quipped with RFID ring antennas (orange rings) that allow tracking the mice via the RFID transponders implanted under their neck felt.

1. Freund J, Brandmaier AM, Lewejohann L, Kirste I, Kritzler M, Kruger A, et al. Association between exploratory activity and social individuality in genetically identical mice living in the same enriched environment. Neuroscience. IBRO; 2015;309: 140–152. doi:10.1016/j.neuroscience.2015.05.027

2. Freund J, Brandmaier AM, Lewejohann L, Kirste I, Kritzler M, Kruger A, et al. Emergence of Individuality in Genetically Identical Mice. Science. 2013;340: 756–759. doi:10.1126/science.1235294

The Effect Of Wing Tears On The Flight Behaviour Of Common Pipistrelle Bats *(Pipistrellus Pipistrellus)*

Rana Khayat, Kirsty J. Shaw, Louise M. Melling & Robyn Grant

Manchester Metropolitan University, Manchester, United Kingdom, rana-osama.s.khayat@stu.mmu.ac.uk; robyn.grant@mmu.ac.uk

Bats are in the order Chiroptera and are the only mammals capable of flight, due to them having large flapping wings [1]. However, these wings are also thin, which causes them to be very prone to tearing [2]. This study developed a method to investigate the effect of these tears on the flight behaviour of common pipistrelle Bats (*Pipistrellus pipistrellus*). Bats, both injured and healthy, were filmed using a high-speed camera (phantom camera, MIRO_M110) and the video recordings were used to track bat wing movements [Figure 1] using the Manual Whisker Annotator software (MWA) [3]. The data gathered from the MWA software included the following variables for both the left and right wings: maximum angle, minimum angle, mean angle, amplitude-RMS (Root Mean Square Amplitude) and frequency, as well as the body orientation. Tears significantly affected wing movement during flight, however the body orientation was affected even more. Future methods to measure the effect of wing tears will focus on measuring aspects of the body, rather than tracking the wing movements. This study had ethical approval from the local committee at Manchester Metropolitan University.

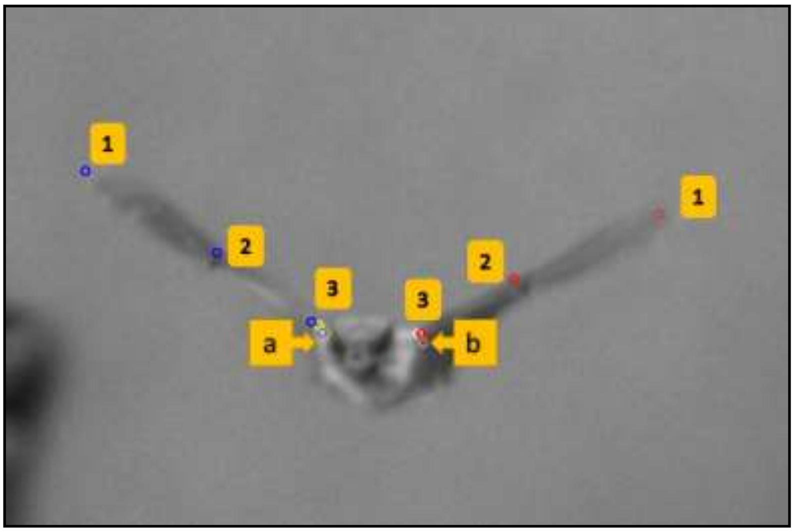


Figure 4. The tracking points for the bat movements. The two points at the middle of the bat (a and b) are for the body orientation. The three points on each wing are for tracking the wing movement.

References

1. Voigt, C., Borrisov, I. and Voigt-Heucke, S. (2012) 'Terrestrial locomotion imposes high metabolic requirements on bats'. *Journal of Experimental Biology*, 215(24) pp.4340-4344.
2. Ceballos-Vasquez, A., Caldwell, J. and Faure, P. (2014). ‘Seasonal and reproductive effects on wound healing in the flight membranes of captive big brown bats’. *Biology Open*, 4(1), pp.95-103.
3. Hewitt, B., Yap, M. and Grant, R. (2016). ‘Manual Whisker Annotator (MWA): A Modular Open-Source Tool’. *Journal of Open Research Software*, 4 (16), p.p. 1-7.

Left Alone – Effects of Acute and Prolonged Social Isolation on the Behavioral Profile in Adult Male and Female Wistar Rats

S. Lundberg, H. Le & E. Roman

Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden. stina.lundberg@farmbio.uu.se

Introduction

The rat is a social animal and should therefore be housed in groups as often as possible to promote the expression of natural behaviors for increased animal welfare. Social isolation (SI) is a known stressor for rats, but different lengths of SI are used in scientific studies for a variety of reasons, from being an integrated part of the research question to a necessary mean for certain methods. The effect of SI is often overlooked and it is not clear how different lengths of isolation affect behavior in the adult rat. This study uses the Multivariate concentric square field™ (MCSF) test, a multi-domain behavioral test developed to study rodents under semi-naturalistic conditions, to investigate the effect of acute and prolonged SI in adult male and female rats.

Methods

Wistar rats (RccHan:WI, Envigo, Horst, the Netherlands) were subjected to acute (30 minutes, n=12/sex) or prolonged (approximately 3 weeks, n=12/sex) SI before testing in the MCSF test. As control condition (n=12/sex), and before the acute isolation, animals were group-housed in groups of three. All experimental procedures were performed during the dark period and animals had *ad libitum* access to food and water in the home cage. The experimental protocol was approved by the Uppsala Animal Ethical Committee and was consistent with Swedish and European legislations.

The MCSF (see Figure 1) [1] is a 100×100 cm arena, divided by inner walls to a Center (70×70 cm) and surrounding Corridors on three sides and an elevated and illuminated stainless-steel construction on the fourth. One corner contains an enclosed shelter and in the next corner is the raised Hurdle with a hole-board mounted with photocells to detect nose pokes. The stainless-steel construction is subdivided into three separate zones; the Slope, which is the access-way from the Corridor; the Bridge Entrance, the first body length of the level portion; and the Bridge. In the middle of the Center, a Central Circle (CTRCI, 25 cm in diameter) is added during analysis.

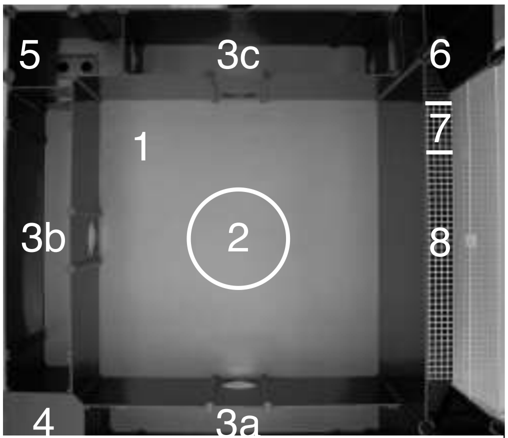


Figure 5. The MCSF arena with defined zones: 1) center, 2) central circle, 3a-c) corridors, 4) dark corner room (shelter), 5) hurdle with hole-board, 6) slope, 7) bridge entrance, 8) bridge.

When starting the test, an animal is placed in the Center, facing the wall separating the Center and Bridge. The animal is then allowed to freely explore the arena for 20 minutes. The test is videotaped and analyzed with EthoVision XT 12.0 (Noldus Inc., Wageningen, the Netherlands). Automatic tracking is used to obtain latency (s), frequency and duration (s) to all zones except the Slope, Bridge Entrance and Bridge, as well as distance (cm) and mean velocity (cm/s) for the whole arena and by Center, CTRCI and total Corridors. Manual scoring is used to score latency, frequency and duration in the zones excepted from automatic tracking, as well as frequency of rearing, grooming and stretched attend posture. Automatic tracking and manual scoring is used in combination for the zone parameters due to the high illumination on the bridge compared to the rest of the arena. The number of urinations, defecations and nose pokes in the Hurdle is noted after completion of each trial.

The parameters from the MCSF test is analyzed directly as obtained and in the form of the trend analysis [2], a rank-order procedure that groups parameters into functional behavioral categories (general activity, exploratory activity, risk assessment, risk taking and shelter seeking). The

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

trend analysis enables interpretation of the performance in the MCSF with an ethological and functional perspective giving rise to the behavioral profile.

Results and conclusions

As seen in Figure 2, acute SI did not lead to any differences in the behavioral profile compared to the controls in either sex. However, prolonged SI caused more pronounced differences, especially in risk-taking behavior which was decreased compared to controls in both sexes and relative acute SI in females. Female rats exhibited fewer SI-dependent differences than males and had higher general and exploratory activities than males in general.

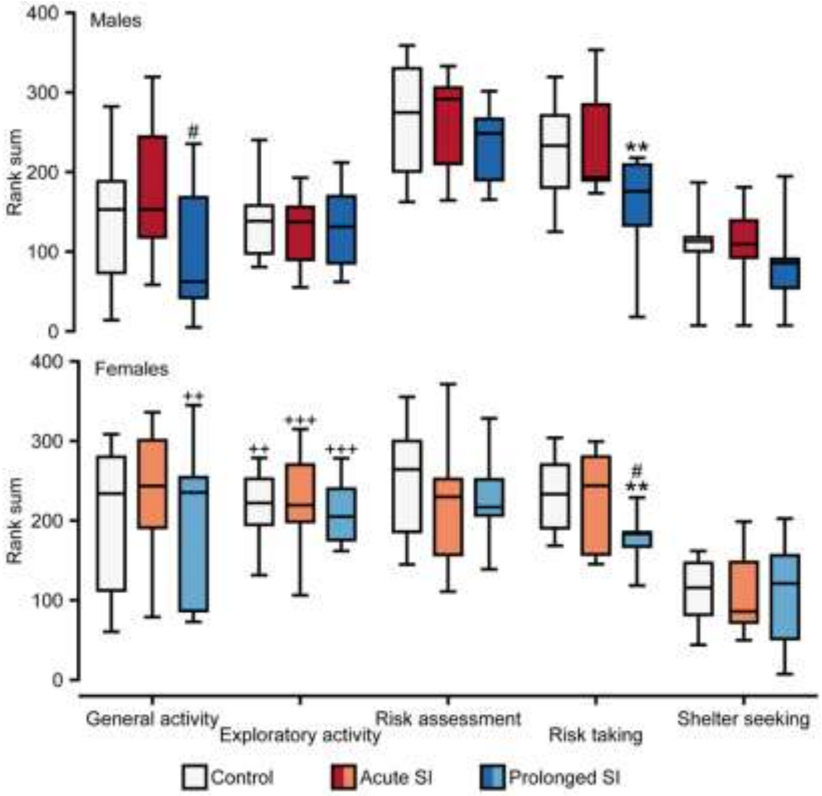


Figure 6. Box plot of the functional behavioral categories of the trend analysis used for behavioral profiling in the MCSF test in male and female adult rats subjected to acute or prolonged SI or group-housing as control condition.

** $p<0.01$ relative to controls; # <0.05 relative to acute SI; ++<0.01, +++<0.001 relative to males, *post hoc* Mann-Whitney U-test.

In preclinical animal research, it is important to consider the impact of all experimental procedures, not just on the posed research question, but on the animals' general welfare. Increased attention to animal welfare and added ethological perspectives in research influence study design and result interpretation and leads to improved validity of the research performed.

References

1. Meyerson BJ, Augustsson H, Berg M, Roman E (2006). The Concentric Square Field: a multivariate test arena for analysis of explorative strategies. *Behavioral Brain Research* **168**:100–113.

2. Meyerson BJ, Jurek B, Roman E (2013). A rank-order procedure applied to an ethoexperimental behavior model - the Multivariate Concentric Square Field (MCSF) test. *Journal of Behavioral and Brain Sciences* **2013**:350–361.

A Model of Cooperation of Two Rats in the Interactive Rat Touch Screen Chamber

I.S. Matiulko[1], E.P. Murtazina[1], B.V. Zhuravlev[1]

1P. K. Anokhin Research Institute of Normal Physiology, Moscow, Russian Federation. Irinamatulko@gmail.com

Different experimental models are applied by researchers to study cooperation in animals. In the present study we designed a model of cooperation in which both animals in a pair had to touch a stimulus light (conditioned signal) within a short period of time. The model was implemented by using the Rat Touch Screen Chamber (Lafayette Instrument, USA) that is controlled through Lafayette Instrument's versatile ABET II hardware and Whisker's Control system (https://lafayetteneuroscience.com/products/rat-touch-screen-chamber). On this platform, we created our own applications (Schedule Designer) that allowed to display a stimulus, to detect the behavioral acts (response to the stimulus), and to analyze the data obtained.

Introduction

The study of behavioral and neurophysiological mechanisms that underlie social relationships in animals is an issue of current importance in the physiology of higher nervous activity of mammals and birds, especially in the populations of social animals. One of the most complex forms of social behavior is cooperative interaction. Ethologists usually study animal cooperative behavior by observing the processes of cooperative hunting [13, 9], cooperative brood care [18], and cooperation in defense against predators [5]; for example, cooperative rearing was found in Slender–tailed Meerkats [3]. Thus, the ability of animals within the population to cooperate may strengthen the relations among the subjects and increase their chances of survival. To study cooperative behavior, researchers developed different instrumental models where animals in pairs have to cooperate in order to receive positive reinforcement. These models were applied to chimpanzees [1, 4, 12], elephants [17], dogs and wolves [14, 16], dolphins [3], and birds [11]. Since cooperation is a complex behavioral action, different neurophysiological and social parameters such as the cognitive abilities of these animals [15] and their hierarchical status [10] contribute to the effectiveness of cooperation.

Cooperation in rats and the role of ultrasound communication in this process were studied in the experiments of Gavrilov V.V. [6]. He used the model of operant foraging behavior where rats in pairs, after a set of individual learning sessions, learnt to press a lever simultaneously for a food reward. The results of the experiment demonstrate different patterns of electrical activity of the brain when animals perform individual actions and when they exhibit cooperative behaviors. What is more, specific electric potentials in the brain arose during ultrasound communications. These data indicate that rats are capable of learning and performing cooperative actions and that ultrasound may mediate communication when rats exhibit cooperative behavior [7,8].

The aim of this study was to develop and to test the model of learning cooperative behavior in rats and to reveal the factors that may affect the learning process and its effectiveness.

Materials and methods

Experimental animals

Male Wistar rats (190 ± 10 g body weight at the start of the study), two months of age, were used in the study. Animals were housed in the groups (5-6 per cage) under standard laboratory conditions of a 12:12-hour light/dark cycle with room temperature maintained between 20°C and 22°C. The animals had ad libitum access to food and water. All the procedures used in this study were performed in strict accordance with the Directive 2010/63/EU on the protection of animals used for scientific purposes and approved by the official Ethical Committee in Biomedical Research appointed by the P. K. Anokhin Research Institute of Normal Physiology.

Apparatus

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

The Rat Touch Screen Chamber (Lafayette Instrument, USA) was used in the study to design the experimental model of cooperation in rats (Fig. 1). The chamber is equipped with a modular panel with two troughs for food or/and water (pellet feeders) on one side and a touch sensitive display (touch screen) for stimulus presentation (stimulus lights) on the other side. This interactive environment provides an opportunity for cognitive assessment in animals by enabling the researchers to display a prearranged signal on the screen and, at the same time, allowing the animals to execute operant feeding or drinking behavior in order to achieve food or/and water reward. The size of the chamber where animal can explore the territory (animal work space) is 30.5×26.0 ×20.0 cm with the monitor viewing area (effective display area) of 10.25” W × 7.9” H where the stimulus light is displayed.

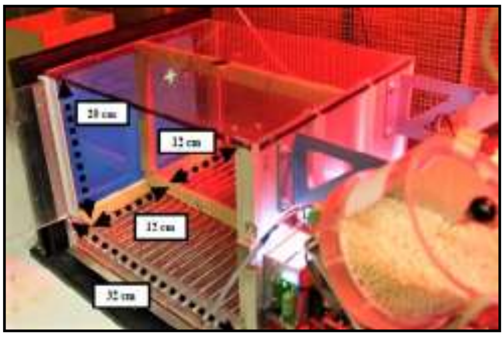


Figure 1. Interactive Rat Touch Screen Chamber (30.5×26.0 ×20.0) divided into two equal parts by the metal mesh. Each part of the chamber is adjusted for one animal and equipped with a pellet feeder and interactive sensory monitor.

Based on the ABET II hardware (Lafayette Instrument's versatile ABET II and Whisker's Control system) platform that controls the chamber, we developed and launched the software application that allows to set the time for operations, run automated food and water supply, detect different types of actions that animals perform, such as touching the screen within the stimulus light area, opening the door of the food trough, and so on. Automated pellet and water dispensers outside the chamber provide water and food supply (water, 0.5 ml, and one food pellet, 45 mg) into the food trough that are on the side opposite to the monitor (thirty centimeters apart to observe and record the rats` behavior, a video camera (Logitech C270 WER HD 960-000635) was placed above the cage.

For the sessions of learning cooperation by two rats for earning food reward, the chamber was divided into two equal parts by a metal mesh, so that each part of the space contained a food trough and a half of the sensory display. The metal mesh did not impede communication between two animals through the visual, tactile, olfactory, and sound afferentations.

Procedure

The experimental procedure was comprised of three parts:

- Determination of the hierarchical status of each animal (once in two weeks).
- Individual learning of rats to earn food and water reward (20 sessions, twice a week).
- Cooperation learning in pairs of animals (14 sessions, twice a week).

1. **Determination of the hierarchical status**

Before and throughout the sessions of individual and cooperative learning we evaluated the ranks of the animals in each group (hierarchical status) according to the dominancy that was based on the competition for access to water in the drinking vessel after 24 hours of water deprivation. After water deprivation, rats of one group were placed into the empty cage surrounded by the transparent plastic walls for 30 minutes. A drinking vessel with water was attached to the plastic wall and was fixed at some distance from the floor of the cage so that a rat had to climb on the platform (plastic cube, 10×10 ×10 cm) to reach the bottle. Since only one rat could fit the space of the platform, animals had to compete with each other, thereby demonstrating dominance over other animals in the group. We recorded animals` behavior during each of the 30-minute sessions and then calculated the total number of behavioral acts when a rat was drinking while standing on the platform and total duration of drinking for each animal. To determine the hierarchical status of animals we compelled the sociometric matrices in which the ranks were calculated as the sum of the victories divided by the sum of the defeats for each rat. The victory/defeat ratio reflected the dominance of individual animal relative to the other animals of the same group.

2. **Individual learning**

Before learning cooperation, rats were trained individually to acquire the instrumental skill of earning a food reward after applying a light touch to the conditioned signal (stimulus light) on the certain area of the sensitive screen. Rats were deprived of food for 24 hours prior to the individual learning session. Light, sound, and other stimuli in the environment were minimized during the experiments. When a rat touched one of the stimulus lights on the screen (green cross on one side and blue circle on the other side), control software ran automated dispensers for food supply (food pellet, 45 mg) to the opposite feeding window that is on the same side of the chamber as the stimulus light is. When pellet dispenser was activated, a short sound from the side of the automated food dispenser signified pellet entry into a food trough; thus, a rat could recognize the moment when the food pellet was delivered. Duration of the first learning session was 1 hour, of the second session – 40 minutes, of the third one – 30 minutes, and then 20 minutes for each of the remaining sessions.

ABET II hardware system contains a set of programs for planning, collection, and analysis of the experimental data. There programs allow to provide a chronological sequence of output signals (displaying of the stimulus lights on the screen, running automated food and water supply) and input signals, such as quantitative measure of behavioral acts (touching the image of the stimulus, feeding and drinking, opening and closing the doors of the food trough). For the sessions of individual training we performed quantitative measure of the time parameters of behavioral acts and calculated latent periods between them. Cognitive abilities of the animals were evaluated according to the learning rate throughout the sessions and the total number of effective eating and drinking patterns during each session.

3. Cooperation

After the sessions of individual training, rats of one group were placed in pairs into two separate parts of the experimental Rat Touch Screen Chamber for acquisition of a skill of cooperation for earning the food reward. The chamber was the same as the one used in the experiment on individual learning and was modified and adjusted according to the conditions required for cooperation: the metal mesh divided the chamber into two parts; each part contained a half of the screen and one feeder (Fig. 1). Animals were paired randomly; once formed, pairs remained unchanged throughout the experiment. Since cooperation implies behavioral synchronization in touching the stimulus light and consequent earning food reward by two rats, two types of events were detected during the session (Fig. 2): 1) effective cooperation achieved by the coordinate touch to the stimulus light on each part of the screen by two rats within the period of 5 seconds (latent period) followed by delivering two food pellets; 2) no cooperation detected when the second rat touched the image of the conditioned signal with delay (incoordinate behavior), i.e. with the latent period of more than 5 seconds («delay») or when the second rat did not make operant response by touching the stimulus light within the 10-second period after the first rat («no response»).

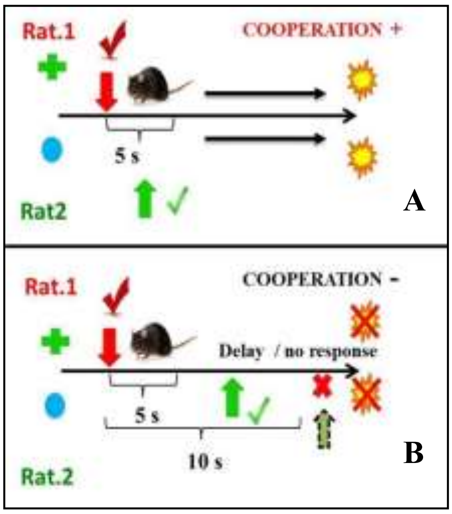
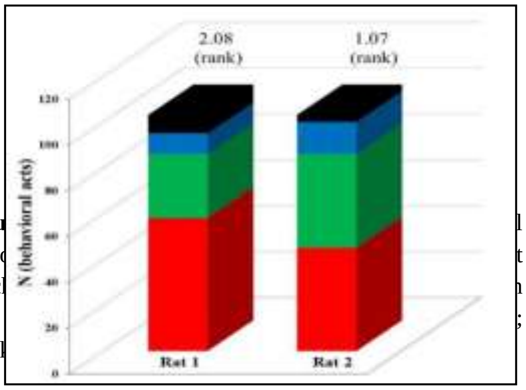


Figure 2. Time scheme of the experiment on cooperation between two rats in a pair for earning food reward. A – effective cooperation; B - no cooperation.

For each experimental trial we recorded the temporal dynamics of the events: time of the stimulus light display, time when a rat touched the conditioned signal and looked inside the food trough, and time when the food pellets were delivered. Characteristics of the behavioral patterns were calculated and analyzed for each pair of animals and included the total numbers of effective coordinated acts of cooperation, incoordinate acts, and failures of cooperation. The following parameters were calculated individually for each rat: the number (N1) and latent periods (LP1) of events when a



Figu[…] l
resp[…] t
touc[…] n
(t < […] ;
blac[…]

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

rat touched the stimulus light first; the number of events when a rat touched the stimulus light after the first rat (N2, «cooperation») with the latent period of up to 5 seconds (LP2 ≤ 5 s, «delayed response»); the number of touches with the latent period of more than 5 seconds (LP3 > 5 s); the number of events when the second rat did not touch the conditioned signal after the first rat`s touch (N4, «no response») (Fig. 3).

Analysis of the proportions between different behavioral patterns for animals in each experimental pair during the trials of cooperation allowed to identify the leaders and the followers within a pair and to reflect on the general effectiveness of cooperation among animals of one group. In most cases, leadership (dominance) detected when an animal touched the stimulus light first, was characteristic for a rat with the higher rank determined by the victory/defeat ratios in the competitions for access to water. Nevertheless, effectiveness of cooperation also depended on the response of the second animal: cooperation was effective when the second animal recognized the action of the first one and responded by touching the stimulus light within 5 seconds. The average latent period of the coordinate response (touching the image of the conditioned signal after the first animal) was 3,7±0,4 seconds.

Data analysis
Data processing and analysis were performed using ABET II Analysis and MS Excel software.

Conclusion

The results of the experiment on implementation of the model of cooperation in rats shown that this model allows to effectively and precisely determine rats` individual abilities, including cognitive abilities, required for cooperation with other animals from the group and, together with additional analysis and observations, to reveal the factors that may affect cooperative abilities.

Reference

1. Bullinger, A. F., Melis, A. P., Tomasello, M. (2014). Chimpanzees (Pan troglodytes) instrumentally help but do not communicate in a mutualistic cooperative task. *Journal of Comparative Psychology* **128(3)**, 251–260.
2. Clutton–Brock, T. H., Brotherton, P. N. M., O’Riain, M. J., Griffin, A. S., Gaynor, D., Kansky, R., McIlrath, G. M. (2001). Contributions to cooperative rearing in meerkats. *Animal Behaviour* **61(4)**, 705–710.
3. Daura–Jorge, F. G., Cantor, M., Ingram, S. N., Lusseau, D., Simões–Lopes, P. C. (2012). The structure of a bottlenose dolphin society is coupled to a unique foraging cooperation with artisanal fishermen. *Biology Letters* **8(5)**, 702–705.
4. Duguid, S., Wyman, E., Bullinger, A. F., Herfurth–Majstorovic, K., Tomasello, M. (2014). Coordination strategies of chimpanzees and human children in a Stag Hunt game. *Proceedings of the Royal Society B: Biological Sciences* **281(1796)**, 2014.
5. Garay, J. (2009). Cooperation in defence against a predator. *Journal of Theoretical Biology* **257(1)**, 45–51.
6. Gavrilov, V.V. (2016). Ultrasound communication in rats during the operant foraging behavior. (in Russian). *7th International Conference on Cognitive Science*, 216–217.
7. Gavrilov, V.V., Arutunove, C.P. (2013). Formation of the individual experience in the model of cooperative behavior in rats. (in Russian). *Evolutionary and comparative psychology in Russia: traditions and perspectives*, 106–112.
8. Gavrilov, V.V., Kosyakov, N.N. (2015). Brain activity in cooperating rats. (in Russian). *Collection of the scientific papers: Neuroscience for medicine and psychology. 11th International Interdisciplinary Congress*, 122–123.
9. Gazda, S. K., Connor, R. C., Edgar, R. K., Cox, F. (2005). A division of labour with role specialization in group–hunting bottlenose dolphins (Tursiops truncatus) off Cedar Key, Florida. *Proceedings. Biological Sciences* **272(1559)**, 135–40.
10. Grewal, J. K. (2012). Cooperation versus Dominance Hierarchies in Animal Groups. *A thesis submitted for the degree of MSc in Mathematical Modelling and Scientific Computing*. P. 7–11.
11. Heaney, M., Gray, R. D., Taylor, A. H. (2017). Keas Perform Similarly to Chimpanzees and Elephants when Solving Collaborative Tasks. *PLoS ONE* **12(2)**, e0169799.

12. Hirata, S., Fuwa, K. (2007). Chimpanzees (Pan troglodytes) learn to act with other individuals in a cooperative task. *Primates* **48**, 13–21.
13. Lührs, M.–L., Dammhahn, M. (2010). An unusual case of cooperative hunting in a solitary carnivore. *Journal of Ethology* **28(2)**, 379–383.
14. Marshall–Pescini, S., Jonas, F. L. Schwarz, Kostelnik, I., Virányi, Z., Range, F. Importance of a species' socioecology: Wolves outperform dogs in a conspecific cooperation task. *PNAS 2017* **114 (44)**, 11793–11798.
15. McNally, L., Brown, S. P., Jackson, A. L. (2012). Cooperation and the evolution of intelligence. *Proceedings. Biological Sciences* **279(1740)**, 3027–3034.
16. Ostojić, L., Clayton, N.S. (2014). Behavioural coordination of dogs in a cooperative problem–solving task with a conspecific and a human partner. *Anim Cogn.* **17(2)**, 445–459.
17. Plotnik, J. M., Lair, R., Suphachoksahakun, W. de Waal, F. B. (2011). Elephants know when they need a helping trunk in a cooperative task. *Proc. Natl. Acad. Sci. USA* **108**, 5116–5121.
18. Stuart, A. M. (1969). Social Behavior and Communication. *Biology of Termites* **1,** 193–232.

Towards Arousal Sensing With High Fidelity Detection of Visual Focal Attention

Oludamilare Matthews, Markel Vigo and Simon Harper

Department of Computer Science, University of Manchester, Manchester, United Kingdom.
oludamilare.matthews@manchester.ac.uk, markel.vigo@manchester.ac.uk, simon.harper@manchester.ac.uk.

Measuring emotions objectively in human-computer interaction is complicated because it involves selecting the appropriate detection mechanism, developing the computational technique to analyse it, and evaluating the correctness of the technique. Existing solutions are either manual, multimodal (requiring multiple sensors), have latency in their response, or do not provide context to why the participant experiences certain emotions. Our approach is an algorithm that measures physiological arousal through analysis of pupillary response and relating it to participants' focal attention. We built our algorithm through a data-driven approach by doing a secondary analysis of two independent datasets. The algorithm works by sensing an increase in arousal through peak detection of the pupil size and compounding the magnitude of this increase with the time spent on the corresponding area of interest. Our contribution to affect detection is that our detection mechanism is unimodal (eye-tracking) and unobtrusive, yet senses arousal with added information about participants' focal attention when they experienced a certain measure of arousal. In its preliminary evaluation, we used Stroop's effect to elicit multiple states of arousal on 19 participants. We found a moderate correlation, $r(76)=.61$, $p<.01$ between the induced states of arousal and our algorithm's computed level of arousal. This result suggests that our algorithm could be used to complement existing research methods in usability, UX and studies of visual research behaviour. The committee for ethics at the University of Manchester approved this study.

Introduction

Affect detection is well researched, and several results in the literature report high accuracy in detection. The challenge lies in integrating these solutions to fit real-world applications. This challenge arises mainly due to the affect detection technique used. These techniques can be categorised into three: self-reported, behavioural/physical and physiological.

Due to their simplicity, **self-reported** (diaries, questionnaires, surveys) methods are the most common techniques used to gain emotional feedback from subjects. The major problem with self-reported techniques is its susceptibility to response biases including acquiescence [1], demand characteristics [2], extreme responding [3] and social desirability [4]. This method of data collection is labouring to the subjects, which means the information that can be extracted is limited and may not be in real-time with the emotional experience. Another category of affect detection is by observing **behavioural** and **physical** features. These include direct approaches such as facial expressions, body gestures, voice prosody and less direct methods such as, through natural language processing, text mining, keyboard, and mouse dynamics. These approaches are more similar to the way human beings observe each other to detect emotions. The problem with transforming this system of emotion detection onto computational techniques is that it may not always be applicable because it depends majorly on the type of interaction. For instance, researchers do not always have the opportunity to capture text, voice or gestures. If this method is introduced into an experimental procedure where it would otherwise be unnatural, it will be disruptive to the participant's normal behaviour thereby introducing a confounding emotion. The third one, **physiological** means of affect detection makes use of changes in the autonomic activities of the body. Physiological means include the skin's electrical properties (electro-dermal activities - EDA) [5], brain activity monitoring (electroencephalogram - EEG) [6], [7], measuring the muscle's electrical activities (electromyography - EMG) [8]. Several factors limit the applicability of these mechanisms like the cost of purchasing the sensors, the skills to set them up, latency and sensitivity of their response [9], [10], and their obtrusiveness to natural behaviour. Contrary to the challenges of the aforementioned techniques, pupillary responses are physiological responses that can be detected by observing the changes in the size of the pupil (contraction and dilation) [11], [12]. Pupillary responses are known indicators of the autonomic activities in the nervous system, especially physiological arousal [13]. The sympathetic nervous system, which enables the body to respond to perceived danger is activated upon an increase in arousal. On the Pleasure-Arousal-Dominance (PAD) emotional scale, arousal represents the intensity of the emotion [14]. Amongst other physiological changes such as increased heart rate, pulse rate, arousal also causes the pupils to dilate. Because there is a direct correlation between pupillary response and arousal, the pupil sizes can be measured through eye trackers and analysed appropriately to sense arousal [15]. In visual researches, eye trackers can be used to capture the size of the pupil upwards of 50 times per second (50Hz). Eye trackers also detect fixations, which are prolonged visual gaze on a single location. Contrary to self-reported emotions, pupil dilation is not easily prone to bias because it is a measure of autonomic activities, which are non-deliberate. It is unobtrusive and adds further context to understand the participant's behaviour because we can use the same device that captures pupil dilation to capture the user's focal attention. Capturing gaze behaviour makes our approach unique because if we know where they are looking at on the screen, and the temporal aspects of this (i.e., how long they looked at it), we would have contextual information that can be used to understand people's behaviour.

Proposed solution

Our proposed approach is an algorithm that analyses pupillary response to extract levels of arousal and relate these measures to the area of their visual context, i.e. what they are looking at. This algorithm was built using a data-driven approach through the secondary analysis of two independent datasets from studies in different domains to ensure that it is not over-fitted to a specific application. Furthermore, we used expert knowledge from the conductors of the original experiments and self-reported feedback from the participants as our ground truth for the formative phase in developing the algorithm.

Methodology

The algorithm was formed with two different sets of eye tracking data; hereafter named, Study 1 and Study 2. Study 1 contained static images of ECG scans while Study 2 was collected during participants' interaction with a knowledge engineering tool called Protégé [16].

Study 1

The original aim of this experiment was to study the visual behaviours of medical experts as they interpret ECG scans to improve the accuracy of ECG interpretation.

There were 31 participants - 23 (74%) female and 8 (26%) male. Majority of these participants (74.2%) were cardiac physiologists/technicians and students of cardiac physiology while the remaining 25.8% were of other professions including nurses, doctors, and students. Students make up 12.9% of participants. All participants had some training on ECG interpretation, although varying in level and experience. The stimuli presented to the participants were ECG scans (18 of them) in random order, without time limit. Further details about the initial study can be found elsewhere [17].

For our analysis, we selected two ECG stimuli, both having 12 leads. Those two stimuli were selected because they had the closest number of correct as incorrect interpretations so that the analysis is balanced between these groups for statistical validity. Each stimulus was split into different AOIs that correspond to each of the 12 ECG leads, using the eye tracking software see Figure 2.

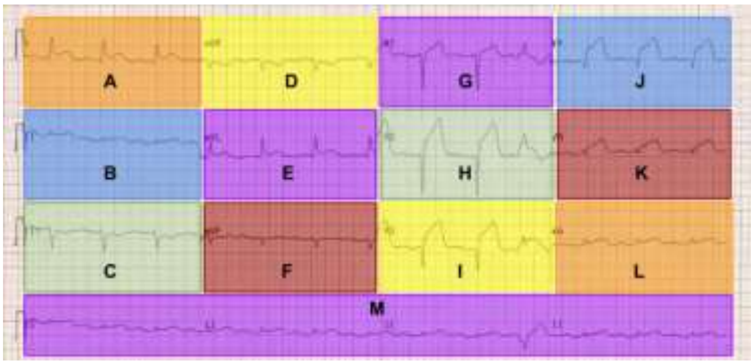


Figure 2: Areas of interest overlaid on a 12-lead ECG

This makes it possible to relate areas where participants gazed at, with a specific lead on the ECG. At the end of the eye tracking session with subjects, they were asked to report their level of arousal in simpler terms such as stress, fear, anxiety, pressure.

Eye trackers capture data at a frequency of 50Hz, i.e. 50 records per second. The dataset for a participant viewing a single stimulus over a 30 second dwell time will contain thousands of record which is much more than the anticipated frequency of change in arousal. The relevance of analyzing this data for our algorithm was to use data-driven techniques to find out the optimal data aggregation size and technique that most accurately detects when the participants felt an increase in arousal. The ground truth (quasi) we used here was the participant's self-reported arousal, and the medical expert's annotation of their thought process regarding where they looked at that informed their decision making for interpreting the ECG scan.

We explored two different fixed-size windowing techniques: simple moving averages and non-overlapping windows). We aggregated both using three aggregate sizes of 5, 50 and 100 records. Recall, data was collected at 50Hz, so these are equivalent to 0.1ms, 1ms and 2ms respectively. Our hypothesis was that: Participants will experience an increase in arousal when they gaze at salient leads of the ECG (for instance, H, I and J represented salient AOIs in one of the ECG stimuli). To determine the optimal setting (windowing technique and aggregate size), we randomly extracted participant's data and applied the algorithm under different settings. We used statistical specificity (recall) to select the best setting. The specificity was computed by:

$$\text{Specificity} = \frac{\text{Correctly detected peaks (True Positive)}}{\text{correctly detected peaks (True Positive)} + \text{ incorrectly rejected peaks (False Negative)}}$$

This was a suitable approach because our ground truth only stipulates some instances (not all instances) where we expect increased arousal. Also, we are not interested in using the algorithm to detect the true negative rate (i.e. when there should not be a peak). However, to reduce false positives, we selected the setting with the highest accuracy but fewest peaks. We selected the fixed point aggregate over moving average for the reason that moving averages return the same number of records as the input. These high-frequency data will return high values in specificity, but there will be many false positives. Also, it will be difficult to identify what the cause of the peak is because the aggregation is too fine. We selected 50 records (1s) because the specificity was 100% compared to 83% by five records (0.5s). Although 100 records (2s) also returned 100% specificity, it was difficult to identify the most fixated AOI during a particular window because there were many ties and the likelihood to miss out an event will be higher in larger groupings.

Study 2

The original aim of this experiment was to evaluate the impact of a plug-in that was built to support the users in authoring and managing ontology.

Participants in this study were required to have expertise in natural language processing and ontologies. From the data, nineteen (19) participants' eye tracking data were captured; fourteen (14) male, fifteen (15) female, between twenty-two (22) and fifty-seven (57) years of age (mean 33.28). Participants interacted with Protege, version 5.0 which is an open source ontology engineering tool developed by University of Manchester and Stanford University. Protégé was pre-configured with the inference inspector and a custom developed plugin – Protégé Survey Tool (PST). For this study, the user interface was segmented into 5 AOI's - view, progress, scenario, question and action; Figure 3 shows this.

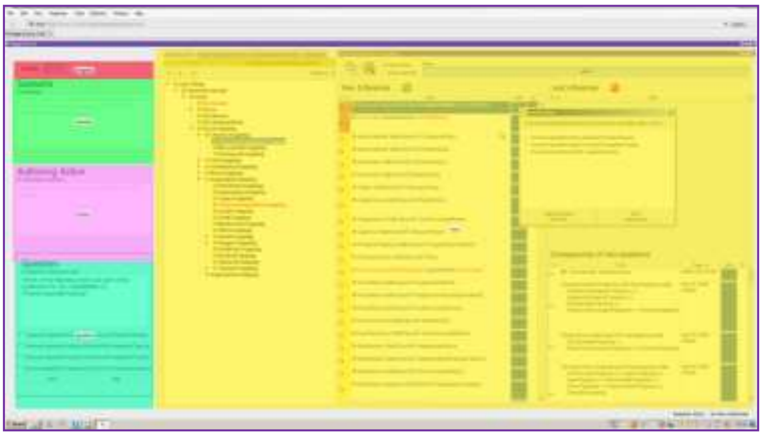


Figure 3: Protégé user interface with Areas of Interest overlaid.
See [16] for more information about the study.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

This dataset poses a unique challenge in that it contains user interaction and the data spans for a longer time compared to study 1.

The relevance of analysing this dataset to developing our algorithm was first to decide on a technique to split the data into chunks (respecting temporal order) and secondly, to relate areas on the screen with arousal levels during user interaction. We explored several techniques including fixed-size quantisation, clustering and changepoint detection. Changepoint detection is implementable in non-real time and real-time data. It works by detecting the likelihood of change in the statistical properties of a dataset and the likelihood that the change takes place at a certain record in the dataset. By specifying a certain property of the interaction such as interaction event, or several other properties if desired as input, we can deploy the univariate or multivariate implementation of the Bayesian changepoint algorithm. The flexibility and robustness of this technique informed our decision to use this algorithm to split the data wherever the average dwell time per stimulus exceeds 5 minutes.

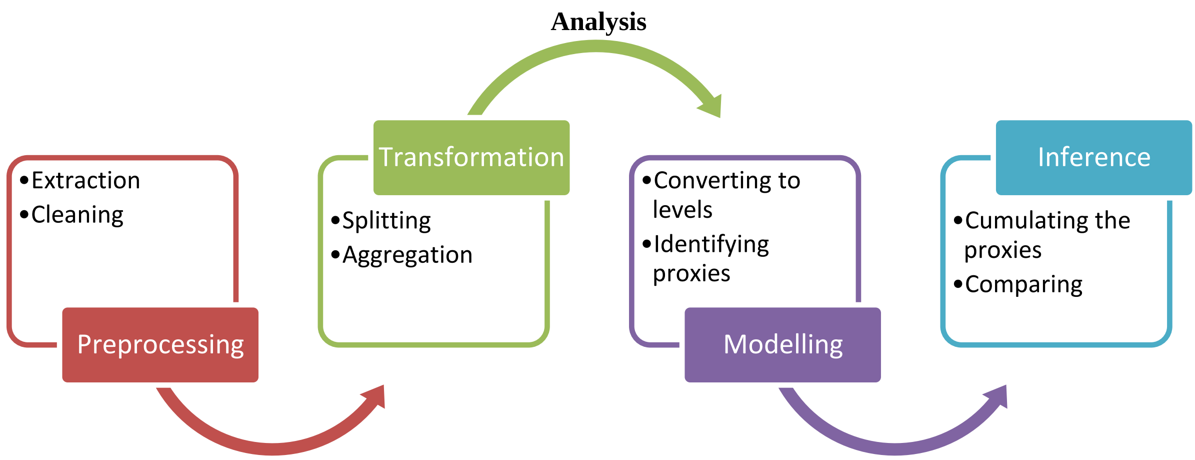


Figure 4: Execution flow

The execution as illustrated in Figure 4, starts with preprocessing - extracting the desired variables (Participant, Recording Timestamp, Fixated AOI, Right Pupil, Left Pupil, and Interaction Event) from the eye tracker. The data is cleaned using linear interpolation to replace missing pupil values and values not within pupil range (2-8mm). Next, we transform the data. If the data for a participant for a stimulus spans more than 300 seconds, we split it into chunks by passing the interaction timing as a parameter into the splitting algorithm (Bayesian Changepoint) by Johannes Kulick (MIT 2014). The Bayesian Changepoint algorithm computes the probability that there is a change and the probability that the change occurs at a particular record/data point. We then set a filter so that if the likelihood of change is greater than our desired threshold, we split the data at that point. The algorithm was implemented on stored data, but solutions exist for real-time change detection, albeit less accurate. After transformation, we select each chunk and perform aggregation using fixed non-overlapping windows also known as tumbling windows. For each window, we also compute the most fixated AOI. The next stage is modelling each window to detect an increase in pupil dilation. Firstly, we transform the pupil diameters to discrete levels. The steps for doing this are listed below.

Get the average pupil size (x_w) for each window (w)

$$\bar{x}_w = \frac{\sum_{i=1}^{n} x_i}{n}$$

Get the 'pupil span' by subtracting the minimum from the maximum average pupil size for each window.

$$\text{Span} = \bar{x}_{max} - \bar{x}_{min} \text{ ; where } \bar{x}_{max} = \max(\bar{x}_1, \ldots \bar{x}_w) \text{ and } \bar{x}_{max} = \max(\bar{x}_1, \ldots \bar{x}_w)$$

To compute the unit of increment (δ) for a participant, divide the span in step 2 by the degree of freedom (desired number of levels subtracted by 1).

$$\delta = \frac{\text{Span}}{\text{df}}; \text{ where d f } = \text{ No. Of ArousalLevels - 1}$$

This value (δ) is the magnitude of arousal to change from one level of arousal to another.

Get the median (μ) of pupil dilations for the participant under this stimulus. Median is preferred here to reduce the influence of outliers.

For each window, subtract the median (μ) in step 4 from its mean ($\bar{x}_w$) from to get the change in arousal for that window (δ_w).

$\delta_w = x_w - \mu$

The arousal level for each window (l_w) is gotten by dividing the change in arousal for each window (δ_w) from step 5 by the unit of change computed in step 3 (δ).

$$l_w = \frac{\delta_w}{\delta}$$

Finally, the most frequently fixated Area of Interest (AOI) for each window is selected. If there is more than one most frequently fixated AOI, the most recent one is selected, and if there is no fixated AOI within the window, the most frequently fixated AOI from the preceding window is selected. If there is none from the preceding window (s), null value is returned.

Our new vector contains a representative arousal level for each widnow (l_w). Next, we detect peak (p), similar to the local maxima in signal processing. In our peak detection function, we defined a peak to be a window where the preceding window has a lower aggregated level, but the immediate window after it has either equal or greater aggregated pupil level. Next, we identify other interesting indices relative to the peak, illustrated in Figure 5.

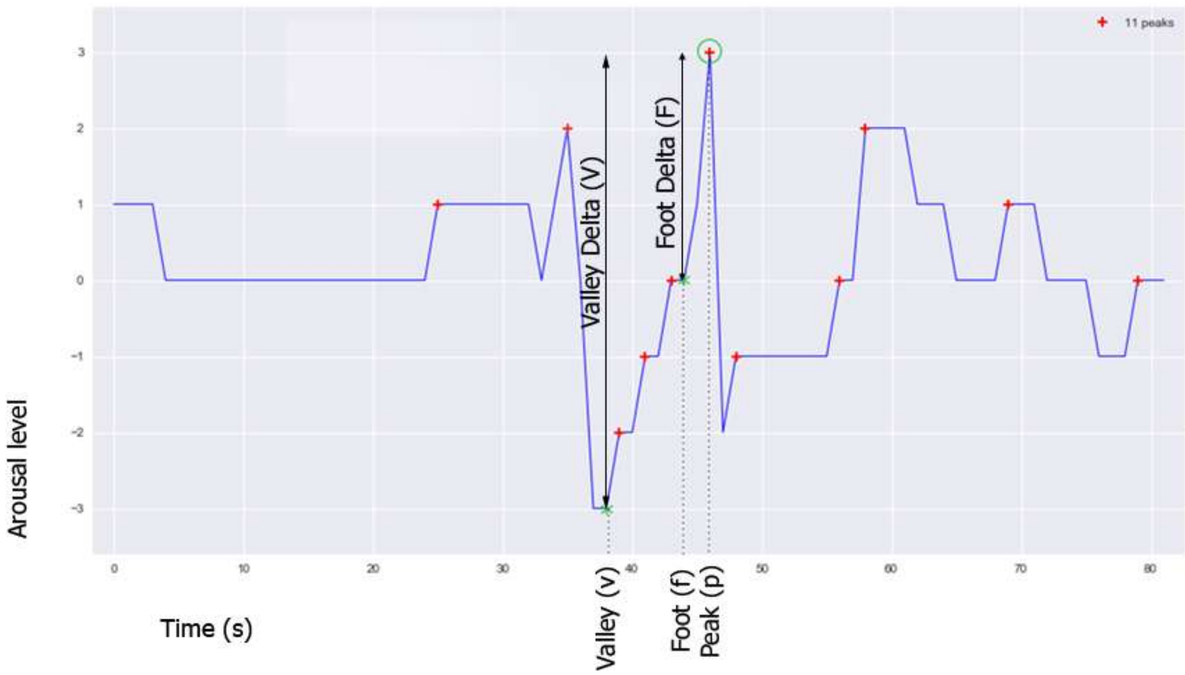


Figure 5: Graph of arousal level against time.

We hypothesize that these proxies will indicate the magnitude of change in pupil dilation, the user's visual focus (fixated AOI) during and before experiencing such magnitude of change, and what the total duration of the user on that particular area is. These measures are defined thus:

9. The immediate window before the peak (Foot - f): The user's visual context at this point may contribute to why there was an increase in arousal in the next window.
10. The change in level between the peak and f (Foot delta - F): The magnitude of change in pupil dilation between the foot and the peak is indicative of how intense the increase in arousal is.
11. The most fixated AOI at this point (Foot AOI – a(f)): The particular area on the screen the user mostly fixated on may be the cause of the increase.
12. The lowest level before a peak (Valley - v): The user was normal before this point. At this point, the user started to experience an increase in arousal.
13. The most frequently fixated AOI at the valley (Valley AOI – a(v)): The particular area on the screen the user mostly fixated on. This may be the reason that the user started experiencing a steady increase in arousal before the peak.
14. The change in level between the peak and v (Valley delta - V): This is the absolute magnitude of change the user experienced between the valley and the peak.
15. Total fixation duration for each area of interest, t (a): This is the total duration spent on a particular area of interest. Fixation duration is often associated with attention. The longer they spend on a particular area, the

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

more they are exposed to the effect caused by that area, therefore the more they are likely to be impacted by the arousal intensity of that area until they are habituated (reduced sensitivity to the stimulus) then, the impact begins to reduce.

To make inferences, the first step is to compute a cumulative score of arousal for each area of interest, for each stimulus, for each participant. The cumulative arousal level experienced by a participant due to an area of interest on a stimulus is given by:

$$A(a) = (\sum F + \sum V)\ t(a).$$

All values are linearly normalised between 1 and 2, per participant.

We hypothesise that this score gives a cumulative value for the level of arousal experienced by a user, after fixating on an area of interest (a) for a particular duration (t(a)).

Preliminary evaluation & discussion

We evaluated the accuracy of the algorithm using Stroop's effect to induce two levels of arousal (normal and high). Stroop's effect is a psychological effect where participants experience a decrease in cognitive efficiency measured by accuracy and response time when they are distracted by incorrectly named objects **[**18**]**. This study was approved by the University of Manchester, school of computer science committee on ethics.

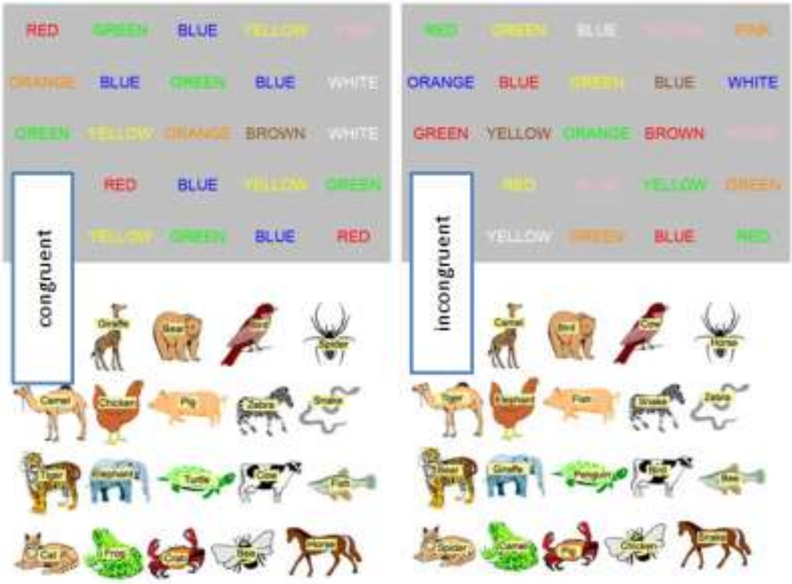


Figure 6: Stimuli for stroops effect.

Participants (19) were tasked to name congruent (correctly labelled) and incongruent (wrongly labelled) objects (animals and colours) displayed on the screen, see Figure 6. Our ground truth enables us to have two levels of arousal as seen in Table 1. The aim is to compare the output of our algorithm with these levels.

Stimuli		Expected Arousal level
Animals	Congruent	High
	Incongruent	Low
Colours	Congruent	High
	Incongruent	Low

Table 1: Stimuli and expected levels of arousal

Using point biserial correlation, we found that there was a moderate correlation, $r(76)=.61$, $p<.01$ between the expected arousal level and our algorithm's arousal level, A(a).

Breaking this down by stimuli, we found there to be a low correlation, $r(76)=.35$, $p<.01$ for animal naming between the expected arousal level and our algorithm's arousal level, A(a) while there was a high correlation, $r(76)=.82$, $p<.01$ for colour naming.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

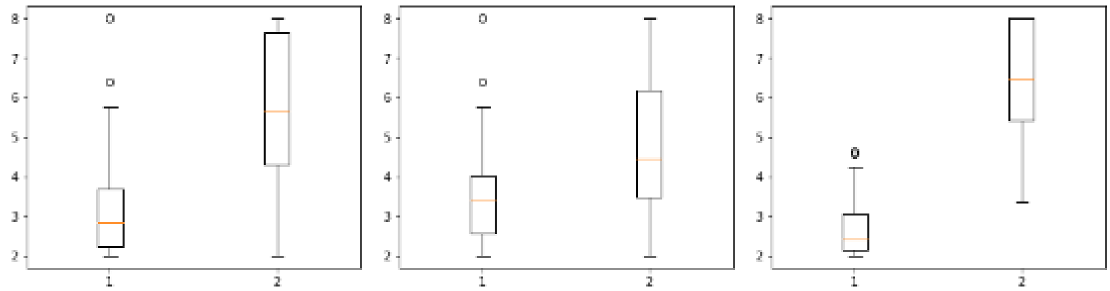


Figure 7: Boxplot of the expected arousal level against the algorithm's Arousal level comparing congruent (1) and incongruent (2) states

As seen from Figure 7, the algorithm was able to discriminate between high arousal and low arousal induced by colour naming compared to animal naming. Some participants verbally said they were able to look at the animals without seeing the text which made it less confusing than the coloured text. Overall, this result shows promise.

Future work

A more comprehensive evaluation is being planned to examine the accuracy of the algorithm. The next phase is to perform a more robust evaluation using stimuli that are rated and elicit a more varied emotional response. While Stroop's effect elicits two states of arousal: normal and high, the next study is designed to induce multiple levels of arousal unto subjects. The visual stimulus is from the International Affective Picture System's (IAPS) database [19], [20]. This database contains pictures of different affective ratings classified using the PAD (Pleasure-Arousal-Dominance) scale [21], [22]. We will select pictures that do not show violence or erotic behaviours in order not to violate the ethical comfort of the subjects. Nevertheless, we will select pictures of varying arousal ratings to test the ability of our arousal in recognising this. Stroop's test is purely cognitive hence, the valence level (pleasantness or unpleasantness of the stimuli) is kept constant. The next study will test our algorithm when users are presented with stimuli that have varied valence rating. It will be interesting to find out the impact of valence on fixation duration and how this impacts the accuracy of the algorithm.

Conclusion

In our quest to develop a solution that relates visual contents with how users feel while they view them, we have analysed independent datasets from two different domains and used our findings to model this in the form of an algorithm. We tested the accuracy of the algorithm in detecting two states of arousal induced by Stroop's effect and found a moderate correlation, $r(76)=.61$, $p<.01$ between the induced arousal and the output of our algorithm. So far, our solution shows promise, and upon further evaluation, we anticipate that our algorithm will complement existing research method to study visual behaviour. Furthermore, our solution is lightweight, non-intrusive therefore when cameras with adequate fidelity for pupillometry become more accessible (e.g., eye-tracking web cameras), the algorithm can be deployed in naturalistic settings to sense physiological arousal for adaptive and recommender systems.

References

[1] C. E. Ross and J. Mirowsky, "Socially-desirable response and acquiescence in a cross-cultural survey of mental health," *J. Health Soc. Behav.*, pp. 189–197, 1984.

[2] A. L. Nichols and J. K. Maner, "The good-subject effect: Investigating participant demand characteristics," *J. Gen. Psychol.*, vol. 135, no. 2, pp. 151–166, 2008.

[3] C. H. Hui and H. C. Triandis, "Effects of culture and response format on extreme response style," *J. Cross. Cult. Psychol.*, vol. 20, no. 3, pp. 296–309, 1989.

[4] R. J. Fisher, "Social desirability bias and the validity of indirect questioning," *J. Consum. Res.*, vol. 20, no. 2, pp. 303–315, 1993.

[5] Y. Ayzenberg, J. Hernandez Rivera, and R. Picard, "Feel," *Proc. 2012 ACM Annu. Conf. Ext. Abstr. Hum. Factors Comput. Syst. Ext. Abstr. - CHI EA '12*, p. 2357, 2012.

[6] Z. Gao and S. Wang, "Emotion Recognition from EEG Signals using Hierarchical Bayesian Network with Privileged Information," pp. 579–582, 2015.

[7] C. A. Torres-valencia and M. A. Alvarez, "Comparative analysis of physiological signals and Electroencephalogram (EEG) for multimodal emotion recognition using generative models," 2014.

[8] M. Soleymani, G. Chanel, J. J. M. Kierkels, and T. Pun, "Affective ranking of movie scenes using physiological signals and content analysis," *Proceeding 2nd ACM Work. Multimed. Semant. - MS '08*, p. 32, 2008.

[9] S. K. Yeom, H. Il Suk, and S. W. Lee, "Person authentication from neural activity of face-specific visual self-representation," *Pattern Recognit.*, vol. 46, no. 4, pp. 1159–1169, 2013.

[10] D. Hutchison and J. C. Mitchell, *Lecture Notes in Computer Science*. .

[11] M. P. Janisse, "Pupil Size, Affect and Exposure Frequency," *Soc. Behav. Personal. an Int. J.*, vol. 2, no. 2, pp. 125–146, 1974.

[12] T. Partala, M. Jokiniemi, and V. Surakka, "Pupillary responses to emotionally provocative stimuli," *Proc. 2000 Symp. Eye Track. Res. Appl.*, pp. 123–129, 2000.

[13] M. M. Bradley, L. Miccoli, M. A. Escrig, and P. J. Lang, "The pupil as a measure of emotional arousal and autonomic activation," *Psychophysiology*, vol. 45, no. 4, pp. 602–607, 2008.

[14] S. Gobron, J. Ahn, G. Paltoglou, M. Thelwall, and D. Thalmann, "From sentence to emotion: A real-time three-dimensional graphics metaphor of emotions extracted from text," *Vis. Comput.*, vol. 26, no. 6–8, pp. 505–519, 2010.

[15] C. Aracena, S. Basterrech, V. Snášel, and J. Velásquez, "Neural Networks for Emotion Recognition Based on Eye Tracking Data."

[16] N. Matentzoglu, M. Vigo, C. Jay, and R. Stevens, "Making Entailment Set Changes Explicit Improves the Understanding of Consequences of Ontology Authoring Actions," in *Knowledge Engineering and Knowledge Management: 20th International Conference, EKAW 2016, Bologna, Italy, November 19-23, 2016, Proceedings*, E. Blomqvist, P. Ciancarini, F. Poggi, and F. Vitali, Eds. Cham: Springer International Publishing, 2016, pp. 432–446.

[17] A. Davies, "Data Driven Analysis of Visual Behaviour for Electrocardiogram Interpretation : Technical Report (initial study)," no. May, 2016.

[18] F. Bousefsaf, C. Maaoui, and a. Pruski, "Remote detection of mental workload changes using cardiac parameters assessed with a low-cost webcam," *Comput. Biol. Med.*, vol. 53, pp. 1–10, 2014.

[19] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (IAPS): Technical manual and affective ratings," *NIMH Cent. Study Emot. Atten.*, pp. 39–58, 1997.

[20] P. J. Lang, "International affective picture system (IAPS): Affective ratings of pictures and instruction manual," *Tech. Rep.*, 2005.

[21] R. a. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Trans. Affect. Comput.*, vol. 1, no. 1, pp. 18–37, 2010.

[22] M. Pilarczyk and T. Wieczorek, "How to measure emotions ? Latest research , future possibilities .," pp. 1–8.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

A Novel Behavioural Test For Detecting Concealed Face Recognition

A.E. Millen and P.J.B. Hancock

Psychology, University of Stirling, Stirling, UK. ailsa.millen@stir.ac.uk

Introduction

Imagine you are being interviewed by police in connection with a serious crime. They wish to connect you with other suspects and show you a picture of one of your accomplices. You deny knowing them but can you avoid giving the game away? The aim of our research is to use a novel behavioural test to detect such concealed recognition using blurred faces. Our approach is based on knowledge that humans are experts in familiar face recognition, thus are less dependent on image quality for familiar faces than genuinely unfamiliar faces [1, 2].

We present images for recognition in a sequence, gradually making them less blurred. Figure 1 shows two example sequences. Working from left to right, it should be easier to decide that you do know the first face, than to decide that you (probably) do not know the second one. The idea, then, is that when presented with such a sequence a suspect will recognize, and then deny knowing, a familiar face more quickly than they should, had the face been genuinely unfamiliar.

Figure 1. Photographs of faces are systematically de-blurred from very blurred (Gaussian blur level 30 on the left) to clearer resolutions (blur level 20 on the far right). To illustrate the effect, here we show differences in 2 levels of blur between each image. In the experiment, the difference between each subsequent blur level was 1. Each blur level remained on the screen for 500ms ±500. Photo credits: (Top) President Donald J Trump by White House is licensed under CC BY 2.0. (Bottom) Unfamiliar face image sourced from Psychological Image Collection at Stirling (PICS: http://pics.stir.ac.uk/) with full permission for research purposes, reports, journals and conferences.

Method

In a concealed recognition test, 26 participants responded to photographs of familiar and unfamiliar faces according to three task instructions: deny knowledge of one familiar face, honestly identify a second person, and correctly reject genuinely unknown faces (known as the Concealed Information Test 3-Stimulus Protocol [3]) while we systematically reduced the blur level of each face from extremely blurred to full quality until a yes/no response was made. Button presses marked the degree of image blur at response. All participants completed two test blocks in a fixed order: 1) with simple test instructions to conceal knowledge, 2) with instructions to 'beat the test' after an explanation of how the test worked.

For analyses, we compared the blur level at which no responses were made to familiar faces (lies), compared to no responses to genuinely unfamiliar faces (truths) using a 2 (familiar, unfamiliar) x 2 (test 1, test 2) repeated measures analysis of variance. Interactions were investigated with paired sample t-tests.

Results and Conclusion

Results revealed that familiar faces were recognised and responded to earlier in the blur sequence than genuinely unknown faces, despite explicit denial of knowledge. The data confirm that superior recognition of familiar faces caused participants to reject them sooner than they would had they been truly unknown. This recognition advantage was significant in both conditions, demonstrating a robustness to both test experience and strategies to evade detection. This finding shows how we can exploit knowledge of differences in familiar and unfamiliar face processing to develop practical tools to expose concealed recognition. We plan to further develop this method to optimise robustness to countermeasure strategies and establish a rigorous test methodology for potential field use. Our novel approach presents a practical tool that might be useful for police officers seeking to gather intelligence on criminal associations.

Ethical Statement

The research was approved by Stirling University General University Ethics Panel (GUEP 219).

References

1. Burton, M. A., Wilson, S., Cowan, M. & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. Psychological Science, **10**(3), 243-248. https://doi.org/10.1111/1467-9280.00144.
2. Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy ("lie detection") with event-related brain potentials. Psychophysiology, **28**(5), 531-547. https://doi.org/10.1111/j.1469-8986.1991.tb01990.x.
3. Schwaninger A., Lobmaier J.S., Collishaw S.M. (2002) Role of Featural and Configural Information in Familiar and Unfamiliar Face Recognition. In: Bülthoff H.H., Wallraven C., Lee SW., Poggio T.A. (eds) Biologically Motivated Computer Vision. BMCV 2002. Lecture Notes in Computer Science, vol 2525. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-36181-2_64.

How do Sea lions detect texture using their whiskers?

Alyx Milne, Mollie Taylor, Robyn A Grant

Division of Biology & Conservation Ecology, Manchester Metropolitan University, robyn.grant@mmu.ac.uk

Pinnipeds (seals, sea lions and walruses) have the most sensitive whiskers of all mammals. Sea lions, specifically, have the longest whiskers and have been shown to control them during active sensorimotor tasks, such as ball-balancing [1]. Sea lions are able to use their whiskers to discriminate between different textures, shapes and sizes of objects, as precisely as a human fingertips, but how they can do this is still not understood. Humans engage in active touch sensing and make task specific movements to maximize sensory information; for example, lateral movements are associated with texture discrimination [2]. This presentation introduces a novel training and filming set-up that has been built in order to measure whisker touch sensing behaviours in California sea lions, to assess how sea lions can discriminate between different textures using only their whiskers.

All procedures were approved by local ethics committees at Blackpool Zoo and Manchester Metropolitan University. Sea lions were trained to discriminate between different textured rubber fish stimuli, with and without a blindfold. They were filmed using an underwater Go Pro (Hero 5) camera. The video was then analysed using custom behavioural software, the Manual Whisker Annotator (MWA) [3], which manually tracks head and whisker movements during stimulus exploration. This method gives highly quantitative and precise movement measurements of whisker amplitude, asymmetry and angular position, as well as head orientation. We present here the process of training and the set-up design, which was built to optimize whisker and head measurements. We present our first data on how sea lions use their whiskers and head movements to judge texture.

Figure 1: Example video footage from the test set-up. The sea lion is exploring the stimuli with its whiskers to find the correctly textured fish.

References

1. Milne, A. and Grant R.A. (2014). Characterisation of whisker control in the California sea lion (*Zalophus Californiaus*) during a complex, dynamic sensorimotor task. *Journal of Comparative Physiology A,*
2. Gibson, J.J. (1962). Observation on Active Touch. Psychological Research, 6(69), p.p. 477-490.
3. Hewitt, B., Yap, M. and Grant, R. (2016). 'Manual Whisker Annotator (MWA): A Modular Open-Source Tool'. *Journal of Open Research Software,* 4 (16), p.p. 1-7.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

The Hexagon - A New Tool for Multiple Simultaneous Demonstrations in Social Learning

Magdalena Monier[1], Sabine Nöbel[1], Anne-Cécile Dagaeff[1], Laurent Polizzi[2], Guillaume Isabel[3] and Etienne Danchin[1]

1. Laboratoire Evolution et Diversité Biologique, CNRS UMR5174, 2. CNRS/IRSAMC/LCAR, UMR 5589, 3. Centre de Recherches sur la Cognition Animale (CRCA), Centre de Biologie Intégrative (CBI), CNRS UMR 5169.

Introduction & Methods

The Hexagon is a new tool for multiple simultaneous demonstrations in behavioural experiments of social learning. We are studying mate-copying, a special form of social learning, in which individuals use the mate choice of conspecifics as source of information revealing potential partners' quality or compatibility. It leads females to either mate preferentially with the specific male they saw being chosen by another female (individual-based copying), or with a male showing similar characteristics as the male they saw being chosen by another female (trait-based copying) [1]. This behavioural pattern was documented in many vertebrate species, and in one invertebrate species: *Drosophila melanogaster* [2-4].

The experimental design to test mate-copying in *D. melanogaster* involves a demonstration during which one demonstrator female copulates with one of two males, one dusted in green, the other in pink powders, followed by a mate-choice test during which the observer female encounters two virgin males, one pink and one green, and can choose to mate with one of them. Several studies [3, 4] showed that females can perform trait-based mate-copying of artificially-coloured male phenotypes (green versus pink), after watching only a single live demonstration.

However, in the wild, drosophila often live in dense populations and aggregate on rotting pieces of fruits. Females can thus witness multiple and simultaneous mate choice demonstrations and can also be affected by the population sex-ratio [5], which determines the availability of a mating partner. We can expect females to adapt their choosiness to the population sex-ratio, and possibly, to mate-copy more or less often depending on the context. In order to address these questions and measure mate-copying in a more natural situation of multiple simultaneous demonstrations, we created a hexagonal device, called the Hexagon.

The Hexagon is composed of a central arena devoted to the observers, and six peripheral compartments devoted to the demonstrators. Each peripheral compartment can be seen from the central arena, as they are separated by a glass partition. Each peripheral chamber can host a copulating couple and a single male of the opposite colour, providing a demonstration to the observers placed on the central arena (see figure).

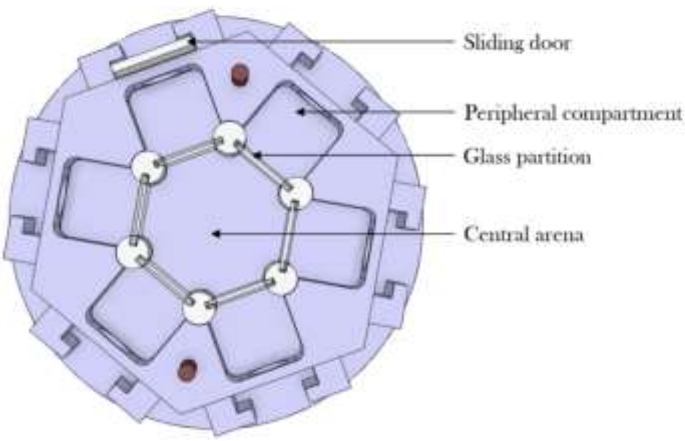


Figure: The Hexagon device allows observer females placed in the central arena to witness up to six simultaneous mate-choice demonstrations, occurring in the peripheral compartments [4]. The device "Hexagon" can be purchased from Toulouse Tech Transfer and Paul Sabatier University through a Material Transfer Agreement.

As the hexagon allows us to choose the number of observer females and demonstrator trios (and thus group size and sex-ratio), we created a gradient of sex-ratio during demonstrations of mate-copying experiments by changing the number of females observing from the central arena six simultaneous demonstrations unfolding in the six peripheral compartments of our hexagonal device.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Results & Discussion

We found that experimental male:female sex-ratio during demonstrations did not affect mate-copying indices, but positively affected the proportion of both males courting the female during mate-choice tests, as well as male courtship duration. Thus, as predicted, the sex-ratio affected female choosiness positively, and drosophila females seem to have evolved a mate-copying ability independently of sex-ratio, and a capacity to adapt their choosiness to male availability.

Conclusion & Outlook

Addressing the effect of sex-ratio and group size on mate-copying is only one example for the range of application of the Hexagon. Our device can be used to study many questions related to mate-copying or more generally visual-based social learning in fruit flies. For instance, it has been used in a mate-copying context to demonstrate conformism, i.e., that flies show an exaggerated tendency to copy the majority [6]. We were able to show that flies can detect majorities as subtle as 60% and learned to prefer males of the most commonly chosen colour as efficiently whatever the majority during the demonstrations (range from 100% down to only 60%). The Hexagon can also be used to detect the ability to count (or to estimate numeric values), to study the behavioural response to the commonness of a phenotype (by using two male phenotypes which proportions differ from 50% each) or of a specific situation such as the presence of stressors, in a mate-copying context. Finally, this device can be used for a diversity of behavioural experiments involving observers and demonstrators (i.e. situations of social learning) as long as vision only is needed, like mate choice or laying site preference.

With its multiple usage options the Hexagon can greatly help in gaining insight into cognition and social learning.

Ethical statement

Our study involved a population of *Drosophila melanogaster* that have been maintained exclusively under laboratory conditions for hundreds of generations. The experiments contained behavioural observations of *D. melanogaster*, which required no ethical approval and complied with French laws regarding animal welfare. We kept the number of flies used in this study as small as possible. We handled flies by gentle aspiration without anaesthesia to minimize damage and discomfort. After the experiments individuals were euthanized in a freezer.

References

1. Bowers, R.I., Place, S.S., Todd, P.M., Penke, L., Asendorpf, J.B. (2012). Generalization in mate-choice copying in humans. *Behavioral Ecology* **23**, 112-124.
2. Mery F, Varela SAM, Danchin E, Blanchard S, Parejo D, Coolen I, Wagner RH (2009) Public versus personal information for mate-copying in an invertebrate. *Current Biology* **19**, 730-734.
3. Dagaeff A-C, Pocheville A, Nöbel S, Loyau AL, Isabel G, Danchin E (2016) Drosophila mate copying correlates with atmospheric pressure in a speed learning situation. *Animal Behaviour* **121**, 163-174.
4. Monier, M., Nöbel, S., Isabel, G., Danchin, E. (2018) Effects of a sex ratio gradient on female mate-copying and choosiness in *Drosophila melanogaster*. *Current Zoology*.
5. Passos, C., Tassino, B., Reyes, F., Rosenthal, G.G. (2014). Seasonal variation in female mate choice and operational sex ratio in wild populations of an annual fish *Austrolebias reicherti*. *Plos One* **9**, e101649.
6. Danchin É, Nöbel S, Pocheville A, Dagaeff AC, Demay L, Alphand M, Ranty-Roby S, Allain M, Monier M, Gazagne E, Isabel G. Fruit fly social learning allows the emergence of cultural traditions in mate-choice. Submitted.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

The models of competitive and cooperative activities in dyads applied by using biological feedback obtained from the electromyographic signals

E.P. Murtazina[1], N.Yu. Trifonova[1], B.V. Zhuravlev[1], I.S. Matiulko[1], A. Yu. Gorovaya[1]

1P. K. Anokhin Research Institute of Normal Physiology, Moscow, Russian Federation.E.murtazina@nphys.ru

In our research, we used and developed methods, a survey protocol, and computer models of individual and joint activities of two subjects on the basis of physiological feedback from their electromyographic (EMG) signals. The model of competitive activity is based on the rivalry of participants to maintain the height of the column displayed on the screen (training "Columns") in the target range as long as it possible by changing the muscular tension of the flexors of the hand. The telemetric recordings were performed by using the "Kolibri" sensors (Neurotech, Taganrog, Russia). The coordinated cooperative activity was modeled on the basis of the same training "Column". The subjects had to keep the height of one column in the target range while looking at the integral indicator which was set in accordance with the EMG signals from both subjects in the pair. The developed protocol allows to carry out a comparative psychophysiological analysis of individual and joint activities in the various social contexts. During the processes of both individual tasks and trainings in dyads, we performed synchronous recordings of the heart rate variability of two participants, and also conducted synchronous telemetry 16 (2x8) - channel EEG recording (EEG-hyperscanning).

Introduction

The studies of psychophysiological features and neurophysiological mechanisms of social relationships between people in the family, cultural, educational, and professional activities in the human community are the issues of importance for many scientists. Different researches and surveys have been launched to study the relationships between a mother and a child [10, 12, 15], singers and musicians [16, 18], performers and the audience [1], circus jugglers [9]. Some authors have also studied the psychophysiological mechanisms which underlie interaction between a doctor and a patient [7]. The neurobiological basis of the cooperative social behavior has been investigated by using economic games which allowed to reveal the direct, indirect, and generalized reciprocity and altruism in individuals [17].

Nowadays, the researchers use wide variety of sensorimotor tests in which two or more subjects have to simultaneously perform different tasks in different contexts: individual, competitive or cooperative ones [13, 19]. As an example, there are the studies on the role of negative and positive feedback in the effectiveness of cooperation between the individuals [5, 6]. The hyperscanning technique is also frequently used to reveal and to analyze the correlation between the activities of different brain structures in two people during their intersubjective social interaction [3, 4, 8, 18]. For the present, the coherent activity of the brain structures is shown in civil aviation pilots during their joined activity on the training vehicles [21]. It may be also assumed that the method of hyperscanning may help study the emotional component of interaction between some individuals when they listen to the music [1]. In other studies, the researchers using the method of magnetoencephalogram hyperscaning have found that the synchronous activation and correlation between the brain activities in two subjects in kinematic tasks depend on the role of individuals during the interaction: leadership or fellowship [11, 20]. It is believed that that determination of the neural mechanisms of intersubjective coherence of the brain activity would allow to use these mechanisms as the neuromarkers of the quality of the social interaction in daily life and professional activities [14]. However, in many studies, the researches use different and often disparate models of individual and joint activities. Taking into account the abovementioned, the aim of this study was to develop and approbate the computer–based models of competitive and cooperative interactions between two subjects and to perform a comparative study of the psychological and neurophysiological mechanisms of competition and effective cooperation.

The biofeedback systems are used for the psychophysiological training and correction of various functional problems for the patients with wide spectrum of diseases affecting the neural, locomotor and cardiovascular systems of the body, as well as in case of psychoemotional problems. BFB training is used for correction,

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

treatment, and training of various conditions. It can be used in the field of psychology to treat insomnia, to improve attention concentration, relaxation and rest. In the field of neurology and in rehabilitation, it can be used for the patients with ICP, heavy head traumas, as well as for those suffering from brain attacks, and injuries of the peripheral nervous system. It can also be used in the field of cardiology to regulate the blood pressure, especially when its elevation is caused by the high stress level.

Methods

The complex developed included the synchronous recording of electrocardiograms in the standard lead in two subjects by using the hardware-software "Varikard" system (LLC "Ramena", Ryazan, Russia). The standard values of the heart rate variability (HRV) were calculated in accordance with the "International Standard for Registration, Physiological Interpretation and Clinical Use of Cardiac Variability, 1996". The analysis of the HRV parameters was carried out at all the stages of the survey, starting from the background state of operative rest, then with individual and joint activities for each subject separately. In addition, a cross-correlation analysis was performed to compare the dynamic of changes in the HRV parameters of the two subjects.

The complex also included synchronous 8-channel telemetry electroencephalogram registration in two subjects by using the MP150 + BioNomadix system (BIOPAC Systems, United States) in the following areas of the cerebral cortex of both hemispheres: frontal, central, parietal, and visual. EEG hyperscanning provided subsequent cross-correlation analysis of the changes in spectral-coherent characteristics between brain activities in two subjects. Coherence and spectral capacities in the cerebral cortex regions were calculated and then compared in the following EEG frequency ranges: theta, alpha 1 and 2, beta 1 and 2 rhythms. Synchronous ECG and EEG registrations allowed to perform the cross-correlation analysis of the characteristics of HRV and EEG in both of the subjects (Fig. 1).

Figure 1. Illustration of the telemetry encephalograms and electrocardiograms recording in two participants.

Models

The modelling of the individual learning and competitive and cooperative behavior in the dyads were performed by using the "BOS–Kinesis" program (Neurotech, Taganrog, Russia). This software and hardware complex is based on the telemetric recording of the physiological signals and consequent appliance of the characteristics of these signals as a biofeedback (BF) in a variety of trainings. Registration of the electromyographic signals (EMG) was performed with the "Kolibri" sensors that were attached onto the skin of the hands in the area above the flexor muscles of the leading hand of a participant (Fig. 2). This device offers the wireless data transmission technology and incredibly versatile Kolibri sensors. The main advantage of the wireless registration is the minimum level of motion artifacts which allow to be applied to any position of the patient's body, including free movement around the premise. The basic version of the device supports up to four simultaneously working sensors. Thus, the complex can be used for the BFB training with four parallel channels. Technical characteristics are as follows: potential registration type – bipolar, working frequency of the wireless interface - 2,4 GHz, voltage range - from 0 to 200 mV. All the sensors have independent controls and the user is able to change the modality, the frequency range, and the amplitude range for each of the using the application. There are several ways of sensor fixation for various BFB modalities: for EMG, ECG, and breathing - disposable adhesive electrodes with a special band; for EEG - a special band (in a form of a headband) with a system of electrodes. The wireless complex is used in conjunction with the software for BFB Kinesis system.

The "BOS–Kinesis" program provides an apparatus for the online mathematical data processing of the EMG signal which is converted into the dynamic of changes in the total spectral power of the EMG. In our study, the values of the spectrum power were determined by the changing height of the column displayed on the screen so that the participant can see the result of his actions. Five–second calibration of the EMG signal was performed at the beginning of each training session during which the participant was instructed to make the maximum wrist flexion with the sensor installed on it.

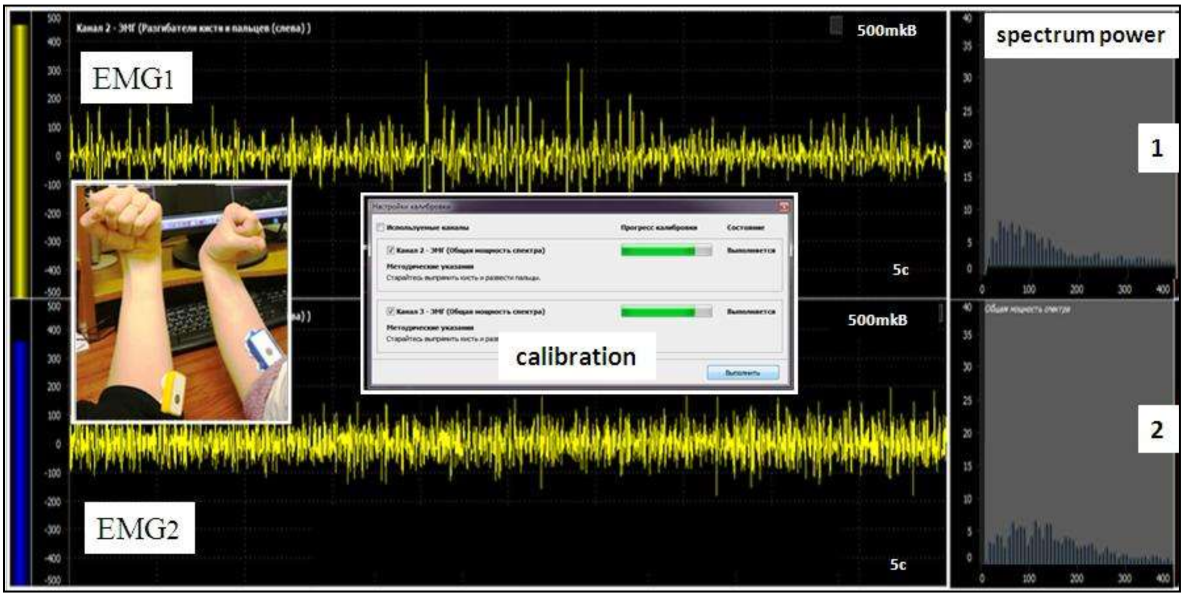


Figure 2. Illustration of the initial stage of the examination: telemetry recording of the electromyograms of two participants (EMG 1 and EMG2) by using the "Kolibri" sensors for calibration and calculation of the averages spectral power.

The average level of the EMG spectral power signal obtained during the calibration determined the average level of the column, as well as the upper and lower thresholds, and the ranges according to which the effectiveness of the BF EMG training is calculated: "Fail, below the line" – [0%; 20%]; "Good, below the line" – [20%; 40%]; "Excellent" – [50%; 10%]; "Good, above the line" – [60%; 80%]; "Fail, above the line" – [80%; 100%].The experimenter had the opportunity to change the thresholds and the ranges of evaluation of the BF EMG training effectiveness. The participants had to maintain the height of the column within the middle range which had green color by changing the contractile force of the flexors. The green color of the column signified high effectiveness of the action («Excellent»). When the height of the column

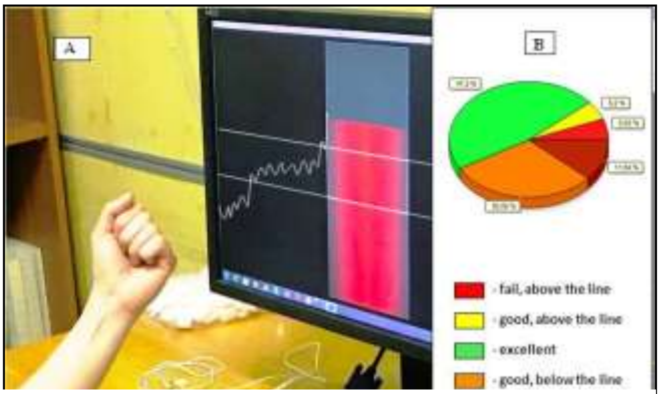


Figure 3. The stage of individual examination (A) and the ratio diagram of the periods in which the height of the column falls within the specified ranges (B).

approached the threshold values, the bar`s color turned to yellow (the result is "Good"); when the column height deviated significantly from the target range and was higher or lower of this range (the result is "Fail"), the bar`s color turned into red (Fig.3.A). After an automated processing of the primary data of the BF EMG training, the program calculated the percentages of the durations of the periods during which the height of the column fell into the "Excellent", "Good, above the line", "Fail, above the line", "Good, below the line", and "Fail, below the line" ranges and then calculated their maximum and average durations (Fig.3.B.).

At the first stage of the examinations, two subjects were trained individually (3 sessions, 3 min each). After that, the participants had to perform a BF EMG training together in a pair in the competitive context. Before the session began, they were instructed to follow a competitive type of the test. A person who acquired a higher score of effectiveness was considered as the winner. During the training session, the participants were sitting in front of one monitor which showed two bars each of which reflected the effectiveness of the corresponding person. Similar to the stage of the individual training, the parameters of performance were obtained for each participant in a pair during competition.

At the next stage of the testing procedure, we studied the psychophysiological processes during the cooperative activity of two subjects by using the same BF EMG training methodology. The same as in a previously described stage, the subjects sat next to each other in front of one monitor. However, unlike the previous stages, the subjects were instructed to cooperate their actions in order to regulate the height of one common column by changing the muscle tension. The height of the column was determined by the dynamic pattern of the resultant calculated by the individual parameters of the participants (Fig. 3). Before conducting the test on cooperation, the procedure of calibration of the EMG signal was performed for both subjects in order to correctly evaluate the contribution of each subject to the result.

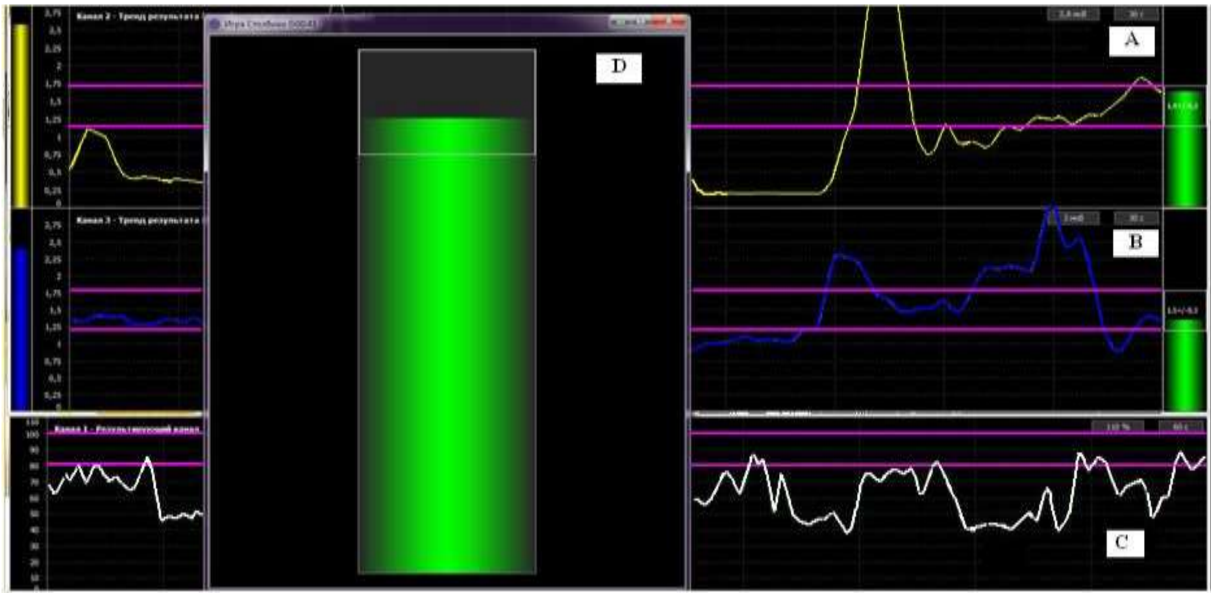


Figure 4. A stage of the joint activity: trends of individual (A, B) and resulting dyadic (C) spectrum power; D – window of the common column.

The evaluation of the cooperative activity was performed under two conditions: 1) when the subjects saw the changes in both common and individual columns; 2) when they could only see the common column without feedback about the success of maintaining the height of the individual columns; this situation was considered as so-called blind cooperation based on integrated joint performance. After the sessions of cooperative activity, individual performance indicators, as well as integral indicators of the dynamics of changes in the resulting height of the common column were calculated and analyzed.

In this study models were tested on healthy male and female volunteers (20–35 years old) who were paired up and knew each other. No material reward was implied for individuals for the participation in the experiment since it

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

was supported by their own interest. The tests were performed after informing the subjects about the purposes of the research. The participants were introduced to the principles of Biomedical and Neuro Ethics and gave their voluntary consent to be examined and to take part in the surveys [2].

The results of the preliminary examinations of eight pairs of subjects shown that the characteristics of the BF EMG training in the social context significantly differ from an individual activity. The ratios of the periods of the muscular overload to the periods of insufficient tension reflect the levels of competitive and cooperative motivation. Thus, the model of cooperation allowed to determine the leader and the follower within the dyads, as well as to identify some pairs that were incapable of effective cooperation.

Ethical statement

All the experimental procedures were performed in accordance with the Additional Protocol to the Convention on Human Rights and Biomedicine, concerning Biomedical Research [2] and approved by the official Ethical Committee in Biomedical Research of the P. K. Anokhin Research Institute of Normal Physiology.

References

1. Acquadro, M.A., Congedo, M., De Riddeer, D. (2016). Music Performance As an Experimental Approach to Hyperscanning Studies. *Front Hum Neurosci*. **25,** 10–242.
2. Additional Protocol to the Convention on Human Rights and Biomedicine, concerning Biomedical Research. Done at: Strasbourg, Date enacted: 2005–01–25.
3. Astolfi, L., Toppi, J., De VicoFallani, F. (2011). Imaging the Social Brain by Simultaneous Hyperscanning During Subject Interaction. *IEEE Intell Syst.***26(5),**38–45.
4. Babiloni, F., Astolfi, L. (2014). Social Neuroscience and Hyperscanning Techniques: Past, Present and Future. *Neuroscience and Biobehavioral Reviews***0,** 76–93.
5. Balconi, M., Vanutelli, M.E. (2017). When cooperation was efficient or inefficient. Functional near–infrared spectroscopy evidence. *Front. Syst. Neurosci.***11**, 26.
6.Balconi, M., Gatti, L., Vanutelli, M.E. (2018). When cooperation goes wrong: brain and behavioural correlates of ineffective joint strategies in dyads. *Int J Neurosci*.**128(2),** 155–166.
7.Benedetti, F. (2013). Placebo and the New Physiology of the Doctor–Patient Relationship. *Physiological Reviews***93(3),** 1207–1246.
8.Dumas, G., Nadel J., Soussignan, R., Martinerie, J., Garnero, L. (2010). Inter–brain synchronization during social interaction. *PLoS One* **5(8)**, e12166.
9.Filho, E., Bertollo, M., Tamburro, G., Schinaia, L., Chatel–Goldman, J., di Fronso, S., Comani, S. (2016). Hyperbrain features of team mental models within a juggling paradigm: a proof of concept. *PeerJ***4,** e2457.
10.Jin, Z., Zhang, X., Han, Z. R. (2017). Parental Emotion Socialization and Child Psychological Adjustment among Chinese Urban Families: Mediation through Child Emotion Regulation and Moderation through Dyadic Collaboration. *Front Psychol***8,**2198.
11.Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R. (2004).Intersubject synchronization of cortical activity during natural vision. *Science* **303(5664),** 1634–1640.
12.Hollenstein, T., Tighe, A.B., Lougheed, J.P. (2017). Emotional development in the context of mother–child relationships. *CurrOpin Psychol.***17,**140–144.
13.Hu, Y., Li, X., Pan, Y., Cheng, X. (2017). Brain–to–brain synchronization across two persons predicts mutual prosociality. *Social Cognitive and Affective Neuroscience***12(12),** 1835–1844.
14.Koike T, Tanabe HC, Sadato N. Hyperscanning neuroimaging technique to reveal the "two–in–one" system in social interactions. Neurosci Res. 2015 Jan; 90:25–32.

15.McQuade J.D., Breaux R.P. Parent emotion socialization and pre–adolescent's social and emotional adjustment: Moderating effects of autonomic nervous system reactivity. Biol Psychol. 2017 Dec;130:67–76. doi: 10.1016/j.biopsycho.2017.10.007.
16.Novembre, G., Sammler, D., Keller, P. E. (2016). Neural alpha oscillations index the balance between self–other integration and segregation in real–time joint action. *Neuropsychologia* **89,** 414–425.
17. Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., Kilts, C. (2002). A neural basis for social cooperation. *Neuron* **35(2),** 395–405.
18. Sänger, J., Müller, V., Lindenberger, U. (2013). Directionality in hyperbrain networks discriminates between leaders and followers in guitar duets. *Frontiers in Human Neuroscience* **7,** 234.
19. Stolk, A., Noordzij, M. L., Verhagen, L., Volman, I., Schoffelen, J.M., Oostenveld, R. Toni, I. (2014). Cerebral coherence between communicators marks the emergence of meaning. *Proceedings of the National Academy of Sciences of the United States of America* **111(51),** 18183–18188
20. Scholkmann, F., Holper, L., Wolf, U., Wolf, M. (2013). A new methodical approach in neuroscience: assessing inter–personal brain coupling using functional near–infrared imaging (fNIRI) hyperscanning. *Front Hum Neurosci*. **7,** 813.
21.Toppi, J., Borghini, G., Petti, M. (2016). Investigating Cooperative Behavior in Ecological Settings: An EEG Hyperscanning Study. *PLoS One* **11(4),** e0154236.

Comparing farmers' perceptions of the severity of pig aggression seen in video clips against actual severity

Rachel S. E. Peden[1, 2*], Irene Camerlink[1], Laura A. Boyle[3] & Simon P. Turner[1]

1Animal and Veterinary Sciences Research Group, SRUC, Edinburgh, UK, Rachel.Peden@SRUC.ac.uk; 2The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, UK; 3Teagasc, Pig Development Department, Animal & Grassland Research and Innovation Centre, Fermoy, Ireland

The uptake of animal welfare research in commercial practice is highly dependent upon whether the farmer perceives it as a problem that needs to be addressed. Our objective is to assess farmer perceptions of the longstanging animal welfare issue of aggression between pigs. Mixing unfamiliar pigs is common practice in pig husbandry creating high levels of aggression as dominance relationships are established. Aggression threatens pig welfare through exhaustion, injury and stress. These physical costs can also threaten farm productivity. A large body of literature exists and a number of management strategies to reduce aggression in intensive farming systems have been identified, however, few have been adopted in practice (1). A recent survey of 167 UK pig farmers revealed that the majority did not perceive aggression as a problem (2). We employ a novel method to investigate whether farmers underestimate the physical impact of the issue; specifically, farmers were asked for their perception of aggression seen in video clips and their response is compared to actual measures of severity.

92 UK pig farmers were recruited at six seperate farmers' group meetings. They viewed four 20 s video clips of pigs either during (clip 1) or immediately after (clips 2-3) a fight and answered a paper based survey employing visual analogue scales (VAS). Control groups of non-farmers with variable experience of working with pigs completed an amended version of the survey (35 agricultural and 26 animal science students and 10 veterinarians). Video footage was obtained from a seperate research project whereby 168 growing pigs were video recorded in aggressive dyadic encounters. For each pig measures of skin lesions and blood lactate were taken pre- and post-fight to indicate relative change as a result of aggression. Skin lesions indicate the severity of aggression and blood lactate gives a measure of physical fatigue. A stepwise selection process was adopted to choose each video clip. Based on the severity of skin lesions and blood lactate we identified from the entire dataset the encounters whereby both pigs obtained high (upper quartile), medium (interquartile range) or low (lower quartile) severity measures. In each meeting the clips were displayed in a different order. Clip 1 was always first and acted as a practice and a common start point, but was not for analysis. Following each clip, farmers were asked to place a downward line through each of the 100mm VAS at a point they felt best represented: i) how much of a negative emotional reaction they had; ii) how exhausting they believe the fight was; and iii) how severe they believed the fight was. Results are curently under analysis and will be presented at the conference. Farmers' scores will be compared against the scores of the control groups and the scores of all groups will be compared against the objective measures. Analysis will focus on whether farmers underestimate the physical impact of aggression. We will highlight the importance of understanding human behaviour in resolving animal welfare problems. This study received internal ethical approval from the Human Ethical Review Committee at the University of Edinburgh. Animal experimentation was approved by SRUC's Animal Ethics Committee and the U.K. Government Home Office ensuring compliance with EC Directive 86/609/EEC.

References

1. Peden RSE, Turner AI, Boyle LA, Camerlink I. The translation of animal welfare research into practice: the case of mixing aggression between pigs. Applied Animal Behaviour Science. 2018 in press.
2. Camerlink I, Turner SP. Farmers' perceptions of aggression between growing pigs. Applied Animal Behaviour Science. 2017; 192C 42-7.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Possibilities, Constraints and Limitations of Image-based Animal Tracking in Natural Environments

B. Risse[1], M. Mangan[2], B. Webb[3]

[1]**Faculty of Mathematics and Computer Science, University of Münster, Münster, Germany (b.risse@wwu.de)**
[2]**School of Computer Science, University of Lincoln, Lincoln, United Kingdom (mmangan@lincoln.ac.uk)**
[3]**School of Informatics, University of Edinburgh, Edinburgh, United Kingdom (bwebb@inf.ed.ac.uk)**

Introduction

Recent advances in automatic image-based tracking of individually behaving animals have enabled the collection of rich datasets underpinning breakthroughs in biological and medical research [1]. Driven by parallel advances in imaging technology, computational power and computer vision algorithms, tracking systems have been implemented for behaviour analysis of organisms ranging from tiny insects (e.g. *Drosophila melanogaster*) to larger vertebrates (e.g. mice) [2]. Yet, such systems have been largely developed for controlled laboratory conditions and struggle to generalise to real-world situations [3] preventing tracking of animals in their natural environments [1]. The importance of in-field behavioural analysis has been highlighted by a variety of studies, ranging from the influence of fertilisers on navigation capabilities [4] over the impact of factorised farming on animals regulating greenhouse gas emissions [5] to the threats of light pollution on biodiversity in general [6]. Most behavioural quantifications for these animals are still done manually so that the need for novel automatic methodologies is emphasised in multiple publications [1], [7]-[9].

Challenges of in-field tracking include but are not limited to potentially small and varying animal sizes; changing animal appearances; clutter and occlusions; limited number of recordings; varying illumination and shadows; and an unknown and theoretically huge region of interest [1]. To account for this challenges a robust detection and tracking algorithm is required which relies on as few constraints as possible. For example, a freely moving camera is required to capture animal paths in arbitrarily sized environments. Also machine learning-based detection mechanisms limit the applicability since they require a significant amount of training data and the animal has to feature a reasonable resolution to allow discriminative correlation [11]. Generating training data is however very time consuming and might be unavailable given a limited number of recordings. Finally, long trajectories extracted from a moving camera will inevitably suffer from drift and thus error accumulation [12].

An ideal image-based tracking system for biologists should be applicable to a diversity of unmarked species, function within diverse complex environments and experimental conditions. Furthermore, it should be mostly automated, simple to use, and inexpensive [1]. From an end-users point of view it would be desirable if only a monocular freely moving hand-held or done-operated camera would be required to record the movement of the animal. Based on the resultant recordings the tracking system should output behavioural features without the need of any other information like inertial measurement unit readings (e.g. accelerometers). Moreover, camera calibration should not be a requirement to enable processing of already existing videos.

In this work we will discuss some of our steps towards such a desired visual tracking system for animals in natural environments. For our real world experiments we chose ants (*Cataglyphis velox*) as a model organism for mainly three reasons: (1) These desert ants are highly researched to understand navigational capabilities of insects; (2) a freely moving camera is required to cope for their potentially huge foraging distances; and (3) this scenario cannot be solved using state-of-the-art methods (mainly due to the small and rapidly changing appearance of the target in combination with the cluttered background [10]).

Methods

We addressed the above listed challenges using a novel tracking strategy posing only four constraints: (1) The animal has to move in more than 50% of the frames and the motion has to be roughly equally distributed over the video sequence; (2) the frame rate has to be fast enough to ensure relatively small displacements between consecutive frames; (3) the background has to have enough distinctive texture to allow feature-based image warping; and (4) the imaging plane of the camera should be roughly parallel to the ground (i.e. birds eye view) and the distance to the ground should not vary strongly. Note constraint (3) and (4) are only required given a moving camera. Furthermore, a violation of the last constrain will not prevent our algorithm from tracking but will

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

result in scaling issues in the final trajectory (see bellow). More importantly these constraints are independent from the animals' size and appearance as well as the appearance of the environment and do not not require any labelling or learning. The resultant system is robust against varying illumination, shadows, clutter and occlusions, and can be directly applied to already existing videos recorded using a stationary or moving camera (e.g. hand-held, drone operated).

In general, our animal tracking algorithm consists of two major processing steps called detection and tracking which can both operate in two modes to account for a stationary or moving camera. Since the stationary mode is a special case of the moving camera mode we will explain the entire moving camera algorithm first.

As illustrated in Figure 1 the detection process can be split into three processes: We first extract the camera motion between consecutive frames (in the moving camera mode only) using ORB features [13] in combination with a robust estimator (i.e. RANSAC) to calculate the homography (i.e. full perspective transformation) between frames at $t+1$ and t (called $f(t+1)$ and $f(t)$; cf. constraint 3). Subsequently, we can warp the camera position at $f(t+1)$ on the position at $f(t)$ by using the homography resulting in a new virtual frame $\tilde{f}(t+1)$. The virtual frame at time $t+1$ appears to be *at the same location* as its reference frame $f(t)$ given the impression that the camera is stationary between these images. Therefore, frame differencing between $f(t)$ and $\tilde{f}(t+1)$ can be used to extract the remaining motion which consists of the animal motion and noise (cf. constraint 1). The difference image is a 2D heat map where bright pixel indicate the remaining motion. These cues are treated as observed variables which are combined with a smooth motion model (cf. constraint 2) to formulate an inference problem. Smooth motion is implemented by using small (i.e. low variance) 2D Gaussian distributions centred at the hypothetical position at time t and are called pairwise potentials. Both, the unaries and the pairwise potentials are combined in a so-called factor graph which is a graphical representation of the underlying inference problem. Animal detections are then calculated by extracting the maximum value over all unaries whereas consecutive unaries are smoothly connected by the pairwise potentials. This is done globally optimal by using the max sum algorithm (for details of the detection algorithm see [14]).

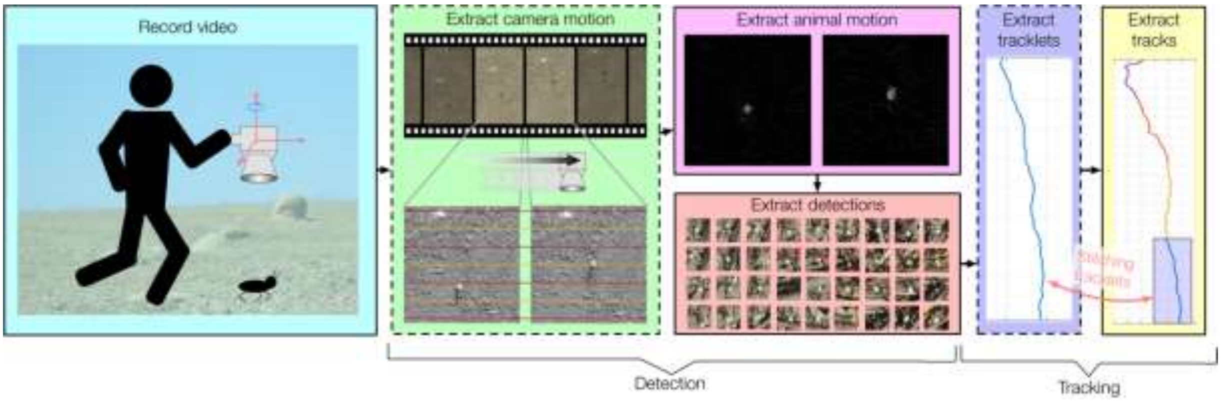


Figure 1: Overview of the imaging and tracking system. From left to right: Recordings are extracted using a hand-held camera. The algorithm determines the camera motion and extracts the remaining animal motion (i.e. unaries). Detections are calculated using global optimisation [14]. Subsequently, these detections are used to generate tracklets which are finally stitched to generate overall trajectories (red arrow between "Extract tracklets" and "Extract track"; blue tracklet is the first segment of the extracted track). Dashed boxes indicate processing steps necessary in the moving camera mode only.

Note that the resultant detections only specify 2D animal positions in the respective frame coordinate system. In order to extract movement trajectories, the detections have to be warped relative to a given reference frame (i.e. fixed camera position) using the homographies again. However, the concatenated use of consecutive homographies over long sequences will inevitably accumulate an error and result in tracks with a decreasing accuracy over time [12]. Therefore, we separate the global track into multiple overlapping tracklets: each tracklet only warps a fixed number of points (e.g. 200 detections) relative to the first tracklet position (instead of warping all detection relative to the first observation). Each tracklet overlaps by 50% with the subsequent tracklet. The overlap is used to identify the Euclidean transformation which has only 3 degrees of freedom: 2 for translation and 1 for rotation (see Figure 1 right: individual tracklets are colour coded). As a result, error accumulation within this fragment is limited to the user-specified tracklet length.

The tracklet stitching algorithm utilises projective geometry and makes use of constraint 4: Assuming continuous birds eye recordings, the overlaps of the two tracklets can only differ by rotation, translation and scaling (i.e. by a

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

similarity transformation). Therefore, we estimate the similarity transformation between the overlaps using the RANSAC paradigm. Subsequently scaling is removed from the resultant transformation to get rotation and translation only. This results in a Euclidean transformation and is necessary to prevent collapsing trajectories: visual camera tracking and the resultant perspective warping with scaling will inevitably result in a scale drift over longer periods [12]. By removing the scaling each tracklet is scaled relative to its respective reference frame and not influenced by possible inaccuracies of distant homographies. Thus, scaling errors can be roughly controlled during recording by ensuring a constant distance between the camera and the ground (cf. constraint 4).
In the stationary camera mode neither the camera motion extraction (and removal) nor the tracklet stitching is necessary. Instead, we directly estimate the remaining motion between consecutive frames $f(t)$ and $f(t+1)$ to extract the detections [14] and use these detections in the fixed camera coordinate system as the final trajectory. Steps necessary in the moving camera mode are indicated as dashed boxes in Figure 1.

Tracking System Prototype & Preliminary Results

To collect data, allow manual interactions and to get first insights into the tracking performance in natural environments, we implemented a rudimentary graphical user interface (GUI) shown in Figure 2. The GUI is used to initialise all necessary parameters and to inspect the video. It is also used to visualise the tracking results: Potential detections are indicated by a red circle and temporally warped trajectories are projected onto the image plane (yellow line in Figure 2). The GUI also allows manual user interaction to correct wrong detections and integrate this information into the optimisation framework [14]. To evaluate the influence of tracking and post-processing parameters a preference panels can be unfolded if necessary (Figure 2 right). An exemplary movie illustrating the overlays generated by the GUI can be found here [16].
To identify the minimal set of constraints specified and to perform a first test we used a hand-held camera (Samsung HDC-SD900; resolution: 1920x1080 pixel @ 60 fps) to record desert ants in their natural environment. Extracting ground truth data is very complicated for small and freely behaving animals in the wild since direct positional measurements cannot be done [8]. Furthermore, most existing tracking approaches identify targets on visual appearance features (e.g. using correlation filters) and do not extract camera-motion compensated trajectories so that a direct comparison is not feasible [10]. To get a first qualitative impression we attached a differential GPS sensor (DGPS; Leica System 1200) to the camera. The DGPS system has centimetre accuracy but can only operate with a low temporal resolution (2 Hz in our case). Since the camera is hovering approximately over the ant DGPS does not exactly specify the animals' position. However, we ensured that the camera is always roughly above the animal by attaching a laser targeting device next to the lens to enable centred recordings (cf. four red laser pointers visible in Figure 2 left).

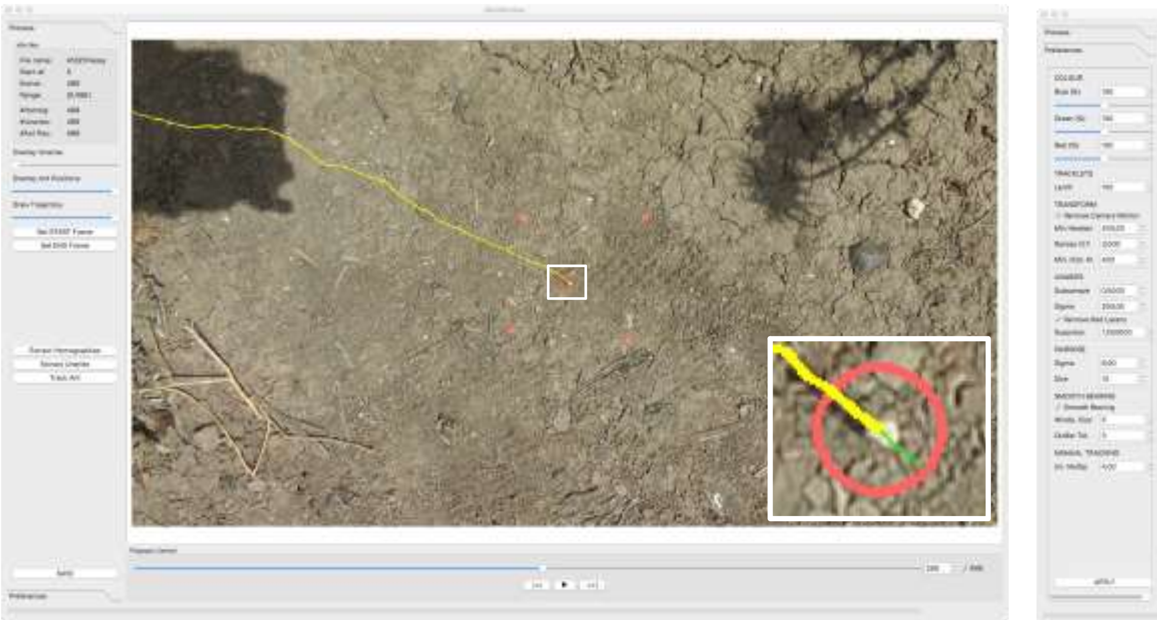


Figure 2: Interactive GUI to collect tracking data. Left: Main window to inspect meta information, initialise tracking, review the video and plot / interact with the results. A close up of the tracked animal is given in the bottom right (white box). Right: Preference panel to control all relevant parameters and additional post processing steps.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

We used the DGPS track and the calculated animal track to identify major flaws and to evaluate the temporal warping accuracy. In Figure 3 left three trajectories are given: in the top the DGPS track of the camera is shown in blue (unit: meters). The middle trajectory (red) shows the animal trajectory generated by warping all positions relative to the first camera position (i.e. first frame used). The bottom-most trajectory is generated by stitching shorter tracklets as described above. The tracklets are colour coded (200 frames per tracklet). Note that both, the non-stitched and stitched animal trajectory is given in pixels and that the GPS and video tracks are not spatially or temporally registered. For comparison we manually rotated the the GPS track by 30°.

Obviously the stitched trajectory shares more similarity with the DGPS path than the non-stitched track. The difference in appearance is mainly caused by the spatial collapsing. As indicated by the axes scaling the non-stitched animal track is one order of magnitude shorter than stitched track. The first 200 positions in the non-stitched track are however identical to the stitched positional estimates since exactly the same transformations are applied to these points in both cases (highlighted in yellow in Figure 3 left). However, in the subsequent frames scaling drift results in constantly decreasing distances between consecutive position estimates. Therefore, the entire trajectory collapses into an x-axis length of approximately 1200 pixel (from x-axis coordinates 2500 to 3700 in Figure 3 mid; highlighted in violet). In contrast, Euclidean transformations of short tracklets does not change the scaling so that the stitched trajectory does not collapse: Each of the 200 frames tracklets have a comparable length (cf. coloured tracklets in Figure 3 bottom).

To compare the DGPS and the stitched trajectory directly, we scaled the dimensions of the animal track according to the camera trajectory. As shown in Figure 3 right the blue DGPS camera path and the stitched animal path are oriented in the same direction throughout the entire sequence indicating no rotational stitching errors. Furthermore, the overall shape appears meaningful and reflects the difference between camera and animal motion. However, the DGPS track appears smother which is partially caused by the lower sampling rate (2Hz vs. 60 fps) but also due to jitter within the animal detections. We note that no spatial or temporal calibration / synchronisation was used so that the results shown here can only reveal a correct tendency and cannot be used for in-depth quantitative evaluations.

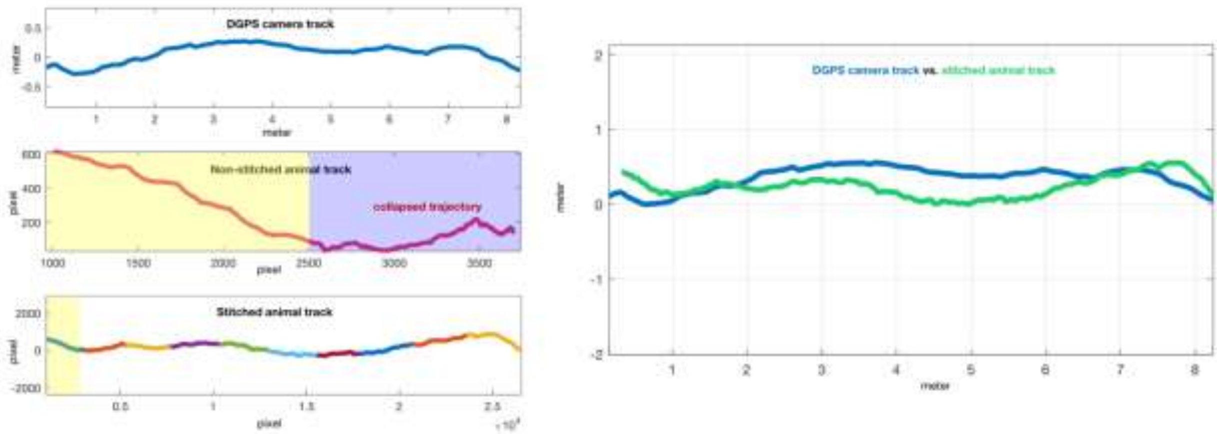


Figure 3: Preliminary trajectory evaluation. Left: The DGPS track of the camera (top) compared to the non-stitched animal trajectory (mid) and the stitched animal trajectory (bottom; 200 frame tracklets coloured). Corresponding segments of the non-stitched and stitched trajectory are highlighted in yellow and the collapsed segment in the non-stitched track is indicated in violet. Note that DGPS measurements are given in meters whereas the animal trajectories are given in pixel. Right: Direct comparison of the DGPS camera trajectory (blue) and the scaled stitched animal trajectory (green).

Discussion

Image-based tracking of animals in their natural environment is a challenging and as yet unsolved problem [1]. In particular, the complex appearance of the fore- and background as well as frequent disturbances prevent the usage of classical techniques developed for pre-defined laboratory situations. In order to gain insights into almost unconstrained and thus flexible in-field animal tracking we designed a tracking prototype capable of extraction positional quantities given moving or stationary camera. Our algorithm utilises only four basic constraints, is not limited by the appearance or resolution of the target and does not require any training. As a consequence, the

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

algorithm is not limited to animals or natural environments and can be applied to all kinds of moving objects and scenes (an examples of correctly detected artificial objects in urban environments is given in [14]).
Since tracking results are difficult to interpret (especially in situations in which a moving camera is used) we implemented a rudimentary GUI to generate and inspect trajectories manually (c.f. [16]). Whereas a reliable detection of in-frame positions is possible in many different situations [14] a reliable camera-motion compensated trajectory extraction is still an open challenge: A comparison with differential GPS reveals that advanced trajectory generation is required to cope for the drift which will inevitably occur in case of visual camera tracking [12]. Furthermore, our preliminary study does not claim to produce reliable animal tracks in real-world coordinates. For example, failed tracklet stitching can result in incorrect heading directions and thus wrong trajectories after this mistake. Moreover, the resultant trajectory is in camera coordinates and only relative to the reference frames so that additional scaling routines are required for meaningful interpretations.
Despite the remaining challenges described, our study indicates that tracking of even tiny animals like ants in complex environments is possible using only one uncalibrated camera. Such a straight-forward system can be generalised for tracking all kinds of species without the need of adding additional sensors. This is particularly important for the insect community since it has been shown that state of the art approaches using telemetry have an impact on behaviour [15] and can only be used to study 0.3% of all terrestrial animal species [8]. Therefore, additional studies and efforts are necessary to enable flexible and easy-to-use tracking of animals in their natural environments.

References

1. A. I. Dell, J. A. Bender, K. Branson, I. D. Couzin, G. G. de Polavieja, L. P. J. J. Noldus, A. Pérez-Escudero, Pietro Perona, A. D. Straw, M. Wikelski, and U. Brose, “Automated image-based tracking and its application in ecology,” *Trends in Ecology & Evolution*, vol. 29, no. 7, pp. 417–428, Jul. 2014.
2. S. E. R. Egnor and K. Branson, “Computational Analysis of Behavior,” *Annu. Rev. Neurosci.*, vol. 39, no. 1, pp. 217–236, Jul. 2016.
3. A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, “Visual Tracking - An Experimental Survey.,” *IEEE Trans Pattern Anal Mach Intell*, 2014.
4. J. Fischer, T. Müller, A.-K. Spatz, U. Greggers, B. Grünewald, and R. Menzel, “Neonicotinoids Interfere with Specific Components of Navigation in Honeybees,” *PloS ONE*, vol. 9, no. 3, pp. e91364–10, Mar. 2014.
5. T. J. Hammer, N. Fierer, B. Hardwick, A. Simojoki, E. Slade, J. Taponen, H. Viljanen, and T. Roslin, “Treating cattle with antibiotics affects greenhouse gas emissions, and microbiota in dung and dung beetles.,” *Proc. Biol. Sci.*, vol. 283, no. 1831, pp. 20160150–7, May 2016.
6. F. Hölker, C. Wolter, E. K. Perkin, and K. Tockner, “Light pollution as a biodiversity threat,” *Trends in Ecology & Evolution*, vol. 25, no. 12, pp. 681–682, Dec. 2010.
7. R. Kays, M. C. Crofoot, W. Jetz, and M. Wikelski, “Terrestrial animal tracking as an eye on life and planet.,” *Science*, vol. 348, no. 6240, Jun. 2015.
8. W. D. Kissling, “Animal telemetry: Follow the insects,” *Science*, vol. 349, no. 6248, pp. 597–597, Aug. 2015.
9. R. Dirzo, H. S. Young, M. Galetti, G. Ceballos, N. J. B. Isaac, and B. Collen, “Defaunation in the Anthropocene.,” *Science*, vol. 345, no. 6195, pp. 401–406, Jul. 2014.
10. M. Kristan et al., “The Visual Object Tracking VOT2016 Challenge Results.,” *ECCV Workshops*, 2016.
11. M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, “Convolutional Features for Correlation Filter Based Visual Tracking.,” *ICCV Workshops*, pp. 621–629, 2015.
12. H. Strasdat, J. M. M. Montiel, and A. J. Davison, “Scale Drift-Aware Large Scale Monocular SLAM.,” *Robotics Science and Systems*, 2010.
13. E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An efficient alternative to SIFT or SURF,” presented at the 2011 International Conference on Computer Vision, 2011, pp. 2564–2571.
14. B. Risse, M. Mangan, L. Del Pero, and B. Webb, “Visual Tracking of Small Animals in Cluttered Natural Environments Using a Freely Moving Camera,” *ICCV , Workshop on Visual Wildlife Monitoring*, pp. 1–10, Nov. 2017.
15. T. McIntyre, “Animal telemetry: Tagging effects,” *Science*, vol. 349, no. 6248, pp. 596–597, Aug. 2015.
16. Exemplary movie demonstrating our algorithm: https://www.dropbox.com/s/3vm16tisxd2atgw/Ant7R8_linux_croped.mp4?dl=0

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Mobile-phone based data-collection methods for social studies: case studies from Estonia

Siiri Silm and Rein Ahas

Department of Geography, University of Tartu, Tartu, Estonia. siiri.silm@ut.ee

The presentation aims to give an overview of using indicators collected through mobile-phone tracking for research into human behaviour. I will introduce two types of data. Firstly, passive mobile-phone tracking data collected from the log files of mobile network operators, which give a good overview of the phone use and spatio-temporal movement of large populations [1]. The most wide-spread passive data sets are Call Detail Record (CDR) data, collected from phone use logs. Mobile network operators' logs also collect data communication (DDR), Location Update (LU), antenna connection (signalling, bulk data) and other data. Different studies have used different types of data, depending on the system and practice of the mobile operator. The advantages of passive mobile-phone tracking data include large populations and long periods, which enable receiving detailed information on movement. The disadvantage of such data, however, is complicated access to data due to privacy and business secret requirements. Another issue is the unspecific geographical location: data carry antenna coordinates, which hinders exact downscaling on the house or street canyon level. Using CDR-type data in behavioural studies presumes generation of indicators based on these data. These indicators may be generated for spatial behaviour, temporal behaviour and interaction. In the presentation, I will introduce the options for generating the indicators and give examples of studies on social segregation and time use [2].

The second dataset described is active mobile-phone tracking, which is carried out along with active phone location inquiries [3]. Currently, most inquiries from smart-phones are made with special software; these may be initiated by the owner without the need to seek approval from the operator or any third party for the collection of data. The advantage of tracking smart-phones is high geographical precision (due to GPS), exact temporal resolution, and the possibility to collect additional information on movement, phone logs and Internet use records. Most importantly, however, it allows sending inquiries to the respondent via special software or messages and making location-specific inquiries. Smart-phones collect longitudinal data, i.e., research periods with contractual customers are normally extensive. These data are also used in various panel studies. The disadvantage of that method is the relatively small number of respondents, because their involvement is expensive and complicated [4,5]. At the same time, the received information is of very high quality, which places the privacy and data protection issues first. In case of data collection based on smart-phones, it is also necessary to generate indicators descriptive of the respondent's behaviour from raw data, which is often a complicated task. These may be generated based on spatial behaviour, temporal behaviour and interactions; additional information on these indicators may be collected with questionnaires or interviews. In my presentation, I will introduce these indicators and their use in urban geography studies and individual behaviour studies [6]. In cooperation with Estonian Genome Centre, we are studying the behaviour of 500 respondents by employing, in addition to mobile-phone data, a genetic dataset and a personality test.

References

1. Shoval, N., Ahas, R. (2016). The use of tracking technologies in tourism research: the first decade. *Tourism Geographies* **18** (5), 587-606.
2. Silm, S., Ahas, R., Mooses,V. (2017). Are younger age groups less segregated? Measuring ethnic segregation in activity spaces using mobile phone data, *Journal of Ethnic and Migration Studies* (Published online October 2017). http://www.tandfonline.com/doi/full/10.1080/1369183X.2017.1400425
3. Poom, A., Ahas, R. (2016). How Does the Environmental Load of Household Consumption Depend on Residential Location? *Sustainability* **2016**, 8(9), 799; doi:10.3390/su8090799 http://www.mdpi.com/2071-1050/8/9/799.
4. Birenboim, A., Shoval, N. (2016). Mobility research in the age of the smartphone. *Annals of the American Association of Geographers* **106** (2), 283-291.

5. Ahas, R. (2010). Mobile positioning in mobility studies. In: Mobile Methods (Büscher, M., Urry, J., Witchger K eds.) Routledge, pp. 183-199.
6. Eurostat (2014). Eurostat Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics, Consolidated Report of Eurostat Contract No 30501.2012.001-2012.452, 31p. http://ec.europa.eu/eurostat/web/tourism/methodology/projects-and-studies

Use of wearable device to reduce musculoskeletal risk behaviours

R.E. Sims[1], M. Hart[2] amd A. Vasina[2]

[1]University of Derby, Derby, UK. R.Sims3@derby.ac.uk, [2]Soter Analytics, Croydon, UK. Info@soteranalytics.com

Back injuries account for 38% of all work-related musculoskeletal disorders reported to the Health and Safety Executive in 2016/17, representing 194,000 workers [1]. Reducing these injuries would have a significant positive effect on the demands on employers and National Health Service. It is proposed that most health and safety attention needs to be focused on the effects of sustained and repetitive movements, and impact of the physical requirements of the roles [1]. Most musculoskeletal injuries do not occur due to a single event, in the majority of cases repeated or sustained positions and activities over many hours, day after day, make the musculoskeletal system more prone to injury [2]. A number of injuries might be prevented if the 'dangerous' moving pattern is noticed and changed in time.

Soter Analytic's SoterSpine sensor uses accelerometers and gyroscopes to continuously capture human movement data (including movement and force on the body), identify at-risk movements and notify the worker. Workers received immediate feedback through the colour of light shown on the sensor, as well as daily reports via smartphone on their movements during the day. An app, Anna, has been designed and is being tested alongside the sensor feedback alone. This app will provide wearers with information and guidance aimed at reducing risky postures and behaviours.

The current project is investigating the effectiveness of the SoterSpine sensor with and without the Anna app, to see if behaviour can be changed during a 6 month time period. The study is being conducted by Soter Analytics and the University of Derby, with funding from Innovate UK. Ethical approval from University of Derby Human Sciences Research Ethics Committee will be granted before data collection starts.

Participants will be assigned to three groups: Device+app (participants can see the light on the sensor and also have access to the Anna app), Device-only (participants can see the light on the sensor), and Control (participants wear the sensor but receive no feedback). A total of 150 participants (workers from the same company, aged 18-65yrs) will be recruited across the three groups, and all will complete a questionnaire about current pain locations and intensities, and current work risks (bend, lift, reach, twist). Participants will then wear the sensors for 4 workings weeks, with data being collected continually during the working day on their postures and movements. After 4 weeks participants will again be asked to answer questions about current pain locations and intensities and current work risks, and the same questions will be asked 3 months after the sensor-wearing period has ended. Participant reports of movements and pain will be compared with the data from the sensors to verify whether the risks identified by the sensors correlate with the perceived risks and experienced pain (or not) of participants. It is of interest to see whether reported behaviours match those detected by the sensors, and if any change in behaviour is sustained to three-months after the sensor was worn. Validation of the device from earlier studies and findings of this project to date, and any issues arising will be presented.

References

1. Health and Safety Executive (2017) Work-related Musculoskeletal Disorders (WRMSDs) Statistics in Great Britain 2017, HSE, < http://www.hse.gov.uk/statistics/causdis/musculoskeletal/msd.pdf > Accessed 31 January 2018.

2. Punnett, L., Wegman, D.H. (2004) Work-related musculoskeletal disorders: the epidemiologic evidence and the debate. *Journal of Electromyography & Kinesiology* **14(1)**, 13-23.

Context-Aware Experience Sampling Method to Understand Human Behavior in a Smart City: a Case Study

A. Smets, B. Lievens and R. D'Hauwers

imec-SMIT, Vrije Universiteit Brussel, Brussels, Belgium. annelien.smets@imec.be

The study of human behavior is inherently complex. A wide range of social-psychological behavior models identify external elements as one of the factors influencing human behavior. Consequently, researchers require methods to study behavior within its real-life environment, capturing external factors that influence the behavior. Today's new technologies and their large-scale deployment offer new opportunities to implement these methods. In this paper, context-aware experience sampling is presented as a method to study human behavior in its context and capture the influence of external factors. This method, part of a behavior change research methodology, is first explained and consequently demonstrated by means of a case study in a large-scale IoT living lab for smart city services in Antwerp (Belgium).

Introduction

Human behaviors are inherently complex and understanding them is one of the hardest challenges in social sciences. Numerous research disciplines study human behavior and from a social-psychological perspective different behavior models have been put forward identifying the underlying factors that influence human behavior [2]. We can distinguish internal, merely psychological and individual determinants and external determinants [6]. The latter could be for example social situations, institutional contexts or cultural norms [20, 22]. Conducting behavior research is challenged by the influence of external determinants such as time, location and social settings (other influences, see [20]). When studying human behavior, it is therefore necessary to take these external factors into account. The term context can be used as an overarching term to describe the external factors influencing behavior. Despite its acknowledged importance in the study of behavior, there are only a limited amount of methods that aim to understand human behavior on a large scale while incorporating the context. These methods either rely on objective measures of behavior (e.g. web log analysis) and thereby neglect the dimension of human experience and motivations; or they use self-reporting methods (e.g. diary-studies) that are context-independent and therefore more susceptible to biases (e.g. recall bias).

In this paper, we aim to contribute to the study of behavior by illustrating the potential of smart cities in behavioral research and more specifically in capturing the context of behavior. According to Townsend [25], "smart cities are places where information technology is combined with infrastructure, architecture, everyday objects and even our bodies to address social, economic and environmental problems". This increased digitalization could enable behavioral research by surpassing some boundaries of current methods such as limited resources and time. As argued in Spagnoli et al. [23], digital technologies make it possible to (1) capture data in real-life settings, (2) regain control over data by capturing the context of behavior, and (3) analyze large sets of information through continuous measurement. Moreover, the study of behavior is highly relevant in smart cities. Human behaviors play an important role in dealing with urban challenges [18, 19] and their interaction with technologies highly affects the smart city. Therefore, understanding human behavior is crucial in designing and implementing smart city services. In this paper, we present an explorative research design for behavior understanding that incorporates the context of behavior. More specifically, we describe a method, mainly relying on context-aware experience sampling, whose implementation can be improved and facilitated through the characteristics of a smart city. By describing a use case that applies this method to study human behavior in a smart city, we highlight the expansion of research opportunities that arise through smart cities. This results in a research design that extents current practices by (1) validating behavioral assumptions in a real-life context and (2) providing feedback loops to the user using their real-life data. In the remainder of this paper, we first present existing behavioral models and explain the concept of behavior understanding. Subsequently, we discuss the weakness of existing behavior understanding methods to capture context and position context-aware experience sampling as a viable solution. Finally, we illustrate this method by describing a case study within City of Things in Antwerp (Belgium), a large-scale Internet of Things (IoT) living lab and testbed for smart city services. In describing the case study, we mainly focus on the

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

methodology and how we put it into practice. The paper ends with our main learnings and some recommendations for future research.

Context in measuring behavior

Studying human behavior and its interaction within an urban environment requires a tailored approach that takes into account different contextual factors due to the many influences in a city. Hence, a framework for behavior analysis in a city context should capture a range of internal mechanisms (psychological and physical) and the external environment (physical and social environment). To this end, the Modular Behavioral Analysis Approach (MBAA), allows to conceptualize, implement and evaluate behavior change interventions in a smart city (see Figure 1) [7, 23]. Michie et al. [16] mention insufficient attention is generally given to the analysis and the understanding of behavior as a starting point of behavior change interventions. The MBAA, however, recognizes the importance of understanding (current) behavior in the broader context and foresees a specific step for mapping the current behavior and an identification of behavioral determinants, i.e. factors that influence the behavior. In this paper, we will focus on this activity of understanding behavior.

Figure 1. Modular Behavioral Analysis Approach (adapted from [23]).

The importance of context

To understand current practices of behavior understanding, we conducted a literature review consisting of 17 scientific studies published between 2014 and 2017. We observed that the majority of these studies rely on (semi-)structured interviews, which almost never took place during, or in the context of, the studied behavior. The latter might be a potential drawback of these studies because concerns have been raised which call into question the validity of this kind of self-reporting outside the context [27]. Research suggests that people are unable to accurately reconstruct their behavior and experiences [10]. Li [12] argues that there are two main sources for this inability: incomplete knowledge and biases. Additionally, contextual information supports people to become aware of factors that influence their behavior [13]. This insight is frequently being used in health research on physical activity or eating behaviors [4, 13, 24]. To achieve this contextual enrichment, these studies make use of different sensors capturing physiological parameters, often combined into a wearable. The use of these new technologies also provides the opportunity to accurately capture the actual behavior and therefore overcome recall bias. This approach is supported by literature on design for reflection, which describes different ways technology can support people to reflect on their own behavior [3]. First, technology can be used to record the behavior or experience itself in order to provide a record of events that can be looked at again. Second, it can augment this information by capturing data that otherwise might not be consciously observed. For example, exposing people to their physical activity (i.e. step count) helps them to reconstruct their daily activities [13]. Considered the importance of behavior to cities, the authors investigated other methodologies, facilitated by technologies in a smart city, that aim to improve current behavior understanding methodologies by including context.

Context-aware experience sampling method

A method that is often used to overcome the lack of context, is contextual inquiry. This is a user-centered design method that focuses on observing and interviewing people while they are carrying out a particular task in the appropriate context [5]. Due to its design, contextual inquiries are highly reliable and result in more detailed information compared to surveys [26]. However, in contrast to the advantage of surveys being scalable and relatively cheap, contextual inquiries are highly resource-intensive. A combination of the scalability of surveys and the contextual richness of contextual inquiries would provide researchers a method to obtain data with a higher external validity [21]. The experience sampling method (ESM) is an established method that combines these two advantages. The essence of ESM is studying experiences in the moment and thereby capture the experience in the context [29]. Originally, the method was designed to question people about their experience in random occasions and has been primarily used for time-use research [8, 10]. This has been implemented by the use of beepers and paper diaries, although more recently, technological advancements have made researchers use more often mobile devices as they have become a part of daily life [21]. Numerous studies using ESM throughout different domains

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

acknowledge its value by being less susceptible to recall errors than other self-report methods [1, 10, 29]. Despite its strength to capture experiences in the context, ESM still lacks the ability to capture the context itself. To this end, the concept of a context-aware experience sampling method (CA-ESM) has emerged [8]. It improves ESM "by using sensing technologies to automatically detect events that can trigger sampling and thereby data collection" [17]. Intrinsically smart cities provide an optimal infrastructure to deploy CA-ESM since urban IoT platforms enable integrating a large number of heterogeneous context data streams originating from different end systems [28], such as weather data and traffic information. These data streams can be used to map and understand the context of behavior in a more thorough way, thereby reducing the recall basis and self-reporting weaknesses of current methodologies within behavioral understanding research. To include context when studying behavior in a city, we propose a research design that complements current practices by adding two additional research steps: (1) validate behavioral assumptions in a real-life context and (2) feedback loops to the users using their real-life data.

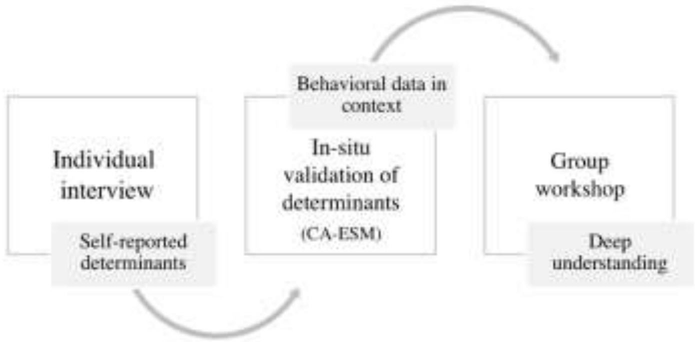


Figure 2. Proposed process to understand behavior.

The proposed research design consists of three steps (see Figure 2). First, we conduct individual interviews to elicit users' thoughts. In contrast to other studies, we use the interviews to discover *assumptions* about behavioral determinants. In the second step, we validate these assumptions (i.e. self-reported determinants) in a real-life context using CA-ESM. This consists of mapping people's actual behavior and studying the contextual influences. We focus on both the status of the determinant (e.g. "How is the weather?") and the influence of the determinant (e.g. "Did the weather influence your decision?"). The final step focuses on gathering a deep understanding of the determinants. This is accomplished during a group workshop where the participants are being confronted with their actual behavioral data, enriched with contextual information. According to the literature on context, providing this data would support the reflection on one's own behavior [12, 13].

Case study: understanding commuting behavior

To test and validate this use of CA-ESM, a pilot study was conducted. The scope of this pilot was to investigate how the method could be used to understand the behavioral determinants of people commuting to work. The pilot was conducted in October 2017 with five citizens from the city of Antwerp (Belgium). The researchers chose a small group because the goal was to get insights in the behavior understanding process, not to obtain an understanding of the behavior with statistical significance levels. During the next two months, the research steps as illustrated in Figure 2 were executed. Every participant signed an informed consent form by which they allowed the researchers to capture and use their personal data for the research purposes. Throughout the experiment, the participants could also contact a team of experts, in case they had any questions related to their participation. This project has been part of City of Things, a smart city living lab and testbed located in Antwerp (Belgium) [11].

Step 1: Self-reported determinants

The first step was conducted similarly to most of the research in behavior understanding: by means of individual semi-structured interviews. This initial interview was necessary to elicit the participants' assumptions about their current behavior that would be validated in the next step. The structure of the interview was based on the COM-B model, which is an existing model that is used to identify behavioral determinants [15,16]. The participants received a number of cards, each containing one abstract keyword representing a determinant from the COM-B model. Some examples: physical strength, self-confidence, time, location, money, identity, culture and belief. During the interview, the participants structurally went through the cards, explaining which cards had an impact on their commuting behavior. Specific attention was given to describing the meaning of the determinants. For

example, when the participant indicated that weather influenced his behavior, the researchers tried to clarify whether this meant either light rain, a thunderstorm or snow.
Finally, the participants were asked to rank their selected determinants (i.e. cards) from most important to less important. It was clear that this exercise was easy for external determinants (such as environmental and social factors). The participants considered those as rational determinants. The internal, more psychological determinants (e.g. self-confidence, beliefs) were perceived as more difficult to grasp. Some of the participants spontaneously placed these cards along the side to indicate that these determinants are always influencing their behavior in a slumbering way.

Step 2: Behavioral data in context

Following the individual interviews, a two-week during test was launched to validate the participants' assumptions, by means of CA-ESM. However, at the time of the pilot not all technological components were in place and therefore we used the Wizard of Oz method [9] to simulate the CA-ESM. To this end, the participants were asked to install an application (Life360 [14]) on their smartphone that allowed the researchers to track the location of the users in real-time during their commute and communicate with the participants. This research set-up allowed us to capture three types of data (see Figure 3): an objective measurement of (a) the participants' actual (commuting) behaviors and (b) the context in which these occurred. By means of the messaging function within the application, the researchers could obtain (c) additional feedback from the participants about what influenced their commuting behavior.

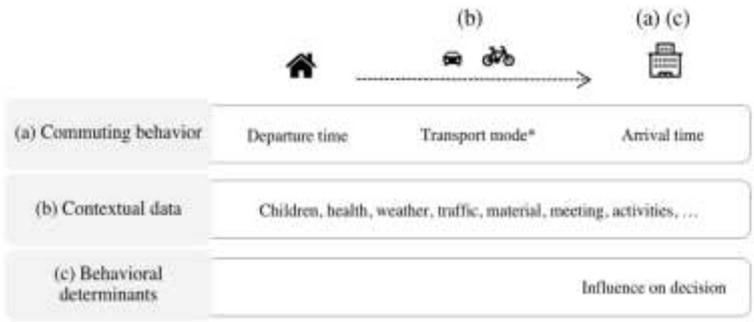


Figure 3. Data captured during in-situ validation of determinants.

(a) Commuting behavior: the participants' commuting behavior resulted in information about their departure and arrival time and the route they chose, as described in Figure 3(a). Ideally, the CA-ESM application would automatically track the transportation mode (bike, car, public transport, ...), but this was not possible with the given application* and the researchers had to collect this information through the participants. The application warned the researchers when one of the participants arrived at work. This event triggered the researchers to question this participant about his/her commute. In this way, the participants were questioned about their commute right after it happened in order to prevent recall bias. The set of questions the participants received upon arrival at work was based on the individual interviews, for example statements like "when it rains, I take the car". We wanted to investigate how contextual data can be captured and how behavioral determinants can be validated with CA-ESM. Therefore, two different types of questions were asked in week 1 and week 2.
(b) Contextual data: during the first week, we wanted to identify the added value of including contextual information (see Figure 3(b)). Ideally, the researchers would possess contextual information (e.g. weather, traffic, etc.) captured by sensors throughout the city. However, in this case, we were not able to access such data directly and consequently had to gather this information ourselves. Accordingly, upon arrival at work, each participant received a customized questionnaire containing questions related to the status of the determinants that were identified during the interview (e.g. "Did it rain when you left for work?").
(c) Behavioral determinants: the second week focused on the influence of the context on the behavior. Each participant received the same questionnaire which no longer questioned the status of the determinant, but the influence of the determinant on the behavior (e.g. "Did the weather influence your transport choice?"). In this second week, we used identical questionnaires because we wanted to investigate if it was possible to use a unified approach for all participants. Based upon the observations during the first interviews, the researchers decided not to include some of the underlying, more psychological determinants in the daily questionnaires. The assumption is that these determinants are less subject to change and do not determine behavior at a specific moment, but are

rather explanatory determinants, which promote a certain habit or a default behavior. For example, if a person beliefs biking is better for the environment, he will try to take the bike most of the time except if a contextual factor prevents him to perform the preferred behavior, e.g. if it rains. The CA-ESM approach is especially suited to provide insights on this kind of contextual influences.

Step 3: Deep understanding
After the two-week trial, a workshop with all the participants was organized. The goal of this workshop was to gain a better understanding of the captured data and how the context influenced the behavior. To this end, the participants were confronted with their actual behavior during the two-week test period. More specifically, they received an overview of their transport mode complemented with the contextual information they had provided through the daily questions. Based on this information, the participants got the opportunity to discuss their behavior and comment on how they had experienced the experiment. During the workshop, three out of the five participants concluded that their actual behavior deviates from their assumptions raised during the first interview. The behavior of these participants was influenced by a larger set of determinants compared to the remaining two participants. This demonstrates the difficulty of assessing the influence of a complex set of determinants. Being provided with information about their behavior and the context, clearly helped them to better understand and explain the influence of the determinants.

Lessons learned
The ESM-CA pilot study allows us to draw some learnings concerning the in-situ validation of the determinants. First of all, due to the short term of the study, the type of data from week 1 (the *status* of the determinant) was insufficient to discover patterns in the data (e.g. someone only cycles when it does not rain). It was therefore necessary to also question the *influence* of the determinant on the transport choice. However, solely questioning the influence of the determinant without knowing the status of the determinant turned out to result in incomplete knowledge as well. It is therefore recommended to combine data about the influence of a determinant with a continuous data stream of contextual parameters obtained from a smart city platform (such as weather conditions). By combining these data types, the method becomes more robust with regard to self-reporting biases, because the researcher can cross-check the answers from the participants with the objective contextual data. Secondly, the study showed that a uniform questionnaire was not sufficient to get a good understanding of people's behavior; working with personalized questionnaires or providing a field for free-text input is required. Finally, this trial also led to some practical concerns related to the CA-ESM tool. More specifically, when choosing the CA-ESM tool, one should pay specific attention to the following requirements: real-time accessible data, integration with contextual data, user-friendly content management system, pre-scripted triggering rules and question sets that can be send automatically and to each participant individually.

Conclusion and further research
Behavior studies often use various ways of self-reporting methods to acquire an understanding of factors influencing one's behavior. However, these methods raise some challenges as they are susceptible to biases and often inadequately capture the context of the behavior, resulting in incomplete information. Especially within an urban context, this challenge becomes even bigger as the set of influencing factors is not only much larger, but also more complex. In our pilot, we observed a difference between the initial self-reported behavior and the actual measured behavior of participants. When confronting them with this difference we were able to gather a set of new insights, which we would not have discovered by only relying on the (traditional) interview.
In order to understand human behavior within this urban context, it is necessary to able to capture the contextual elements throughout the research activities. In this paper, we presented a research design relying on the context-aware experience sampling method (CA-ESM) to deal with this challenge. The CA-ESM augments current practices of understanding behavior by validating behavioral assumptions in a real-life setting. Moreover, the CA-ESM provides additional value by combining the strengths of qualitative and quantitative research: a large number of participants can be questioned in real-life, while capturing additional contextual information to enrich the data. Today's technological possibilities make it easier to collect, integrate and analyze these different datasets. Additionally, they allow the CA-ESM to be implemented on a large scale by automatically asking the right

questions to the right people based on contextual cues. In our pilot, we were able to interact with our different participants in a direct and individual way, based on the simulated CA-ESM tool.

According to the proposed research design, the CA-ESM has to be preceded by a process to elicit the behavioral assumptions. In our pilot, we used the COM-B model as a guidance to identify the behavioral determinants. This model has been very useful throughout the research. Nevertheless, we needed to grasp the proper meaning of the determinants as they were interpreted differently by each of the participants. Although these individual interviews are necessary, they limit the scalability of the research design and a purposeful sampling of the initial panel is required. However, in contrast to our pilot that had a small number of participants throughout the entire study, the second step of the research design could be easily up scaled by applying the CA-ESM to a large group of people. When applied to a larger sample, it is possible to assess the relevance of each of the determinants throughout the whole panel, enabling the development of user segments related to behavioral profiles.

Overall, the pilot demonstrates that the proposed approach offers a good framework to build an understanding of (current) human behaviors by capturing several contextual determinants. Since it is difficult for people to elicit these determinants adequately, the interaction with contextual data provides a more complete result that is less susceptible to biases compared to other approaches. Based on our experiences, we consider the CA-ESM as an instrument to collect data on behaviors and their context and should be applied within a broader research approach on measuring human behavior or behavior intervention design. Our approach can not only provide behavioral profiles, but also identify building blocks on which behavior interventions can be built on. Additionally, our approach would be suited to contribute to the validation of behavior interventions. Therefore, we argue for the use of CA-ESM during various phases of behavior research and especially in those steps where in-depth insights in the motivation and the context of the behavior is required.

References

1. Consolvo, S., & Walker, M. (2003). Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Computing* **2**(2): 24-31.
2. Darnton, A. (2008). Reference Report: An overview of behavior change models and their uses. Centre for sustainability, University of Westminster.
3. Fleck, R., & Fitzpatrick, G. (2010). Reflecting on reflection: framing a design landscape. *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*: 216-223.
4. Frost, J., & Smith, B. K. (2003). Visualizing health: imagery in diabetes education. *Proceedings of the 2003 conference on Designing for user experiences*: 1-14.
5. Hanington, B. & Martin, B. (2012). Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions. Rockport Publishers.
6. Heimlicha, J., & Ardoinb, N. (2008). Understanding behavior to understand behavior change: a literature review. *Environmental Education Research* **14**(3): 215–237.
7. Hoelck, K., McCrory, G., van der Graaf, S. (2017). A behavioral analysis approach for the 'City of Things'. *Proceedings of the 21the Conference of the Environmental and Sustainability Management Accounting Network (EMAN) Liège 2017*.
8. Intille, S. S., Rondoni, J., Kukla, C., Ancona, I., & Bao, L. (2003). A context-aware experience sampling tool. *CHI'03 extended abstracts on Human factors in computing systems*: 972-973.
9. Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Office Information Systems* **2**(1): 26–4.
10. Larson, R. & Csikszentmihalyi, M. (1983) The Experience Sampling Method. *New Directions for Methodology of Social & Behavioral Science* **15**: 41-56.
11. Latré, S., Leroux, P., Coenen, T., Braem, B., Ballon, P. & Demeester, P. (2016). City of Things: an integrated and multitechnology testbed for IoT smart city experiment. *Smart Cities Conference (ISC2), IEEE International*: 1-8.
12. Li, I. (2011). Personal informatics and context: using context to reveal factors that affect behavior. Doctoral dissertation, Carnegie Mellon University.
13. Li, I., Dey, A. K. & Forlizzi, J. (2012). Using context to reveal factors that affect physical activity. *ACM Transactions on Computer-Human Interaction (TOCHI)* **19**(1): 7.
14. Life360 (n.d.). <https://www.life360.com>. Accessed 10 January 2018.

15. Michie S., van Stralen M. & West R. (2011). The behavior Change wheel: A new method for characterising and designing behavior change interventions. *Implementation Science* **6**(1): 42.
16. Michie, S., Atkins, L. & West, R. (2014). The Behaviour Change Wheel: A Guide To Designing Interventions. Silverback Publishing: London.
17. MIT (2008). Context-aware experience sampling. <http://web.mit.edu/caesproject/index.htm> Accessed 15 November 2017.
18. Mostashari, A., Arnold, F., Maurer, M. & Wade, J. (2011). Citizens as sensors: The cognitive city paradigm. *8th International Conference & Expo on Emerging Technologies for a Smarter World (CEWIT)*: 1-5.
19. Oskamp, S. (2000). A sustainable future for humanity? How can psychology help?. *American Psychologist* **55**(5): 496-508.
20. Patrick A. (2009). Ecological validity in studies of security and human behaviour. <https://www.andrewpatrick.ca/cv/Andrew-Patrick-ecological-validity.pdf>. Accessed 10 January 2018.
21. Pejovic, V., Lathia, N., Mascolo, C. & Musolesi, M. (2016). Mobile-based experience sampling for behavior research. *Emotions and Personality in Personalized Services*: 141-161.
22. Shove, E. (2010). Beyond the ABC: climate change policy and theories of social change. *Environment and Planning A* **42**(6): 1273-1285.
23. Spagnoli, F., van der Graaf, S., Ballon, P. (2017). Changing behavior in the Digital Era: A Concrete Use Case from the Domain of Health. *World Academy of Science, Engineering and Technology International Journal of Psychological and Behavioral Sciences* **11**.
24. Ståhl, A., Höök, K., Svensson, M., Taylor, A. S. & Combetto, M. (2009). Experiencing the affective diary. *Personal and Ubiquitous Computing* **13**(5): 365-378.
25. Townsend, A. (2013). Smart cities: big data, civic hackers, and the quest for a new utopia. W.W. Norton & Co, New York.
26. Vacek, R. (2015). Contextual Inquiry. <https://www.slideshare.net/vacekrae/contextual-inquiry-how-ethnographic-research-can-impact-the-ux-of-your-website>. Accessed 21 November 2017.
27. Willems, E. (1969). Planning a rationale for naturalistic research. *Naturalistic viewpoints in psychological research*.
28. Zanella, A., Bui, N., Castellani, A., Vangelista, L. & Zorzi, M. (2014). Internet of things for smart cities. *IEEE Internet of Things journal* **1**(1): 22-32.
29. Zirkel, S., Garcia, J. A. & Murphy, M. C. (2015). Experience-sampling research methods and their potential for education research. *Educational Researcher* **44**(1): 7-16.

New tools to study nectar consumption in honey bees: testing demand law in honey bees

Michel B.C. Sokolowski

Department of Psychology, INSERM 24 (ERI 24), Université de Picardie

When the price of a good is increased, the economic demand law predicts that the consumption of the good should decrease [1]. To test such predictions with animals, researchers have suggested using Skinner boxes that allows to control the price of the good and to measure the consumption at the same time [2]. With a new design of lab operant conditioning chambers following Sokolowski and Abramson [3, 4], we tested the demand law with honey bees in closed economies (that is when the animals are consuming all their daily food in the conditioning chambers). The price has been defined as the amount of syrup a bee received following a response. In our protocol, the price increased from 0.5 responses/µl to 4 responses/µl. According to the demand law, we observed decreasing daily syrup consumption. At the same time, the behavioral output (that is the number of responses emitted by the animal to get the consumption level) increased with price (inelastic demand). To our knowledge, it is the first time the demand law has been successfully tested in an invertebrate.

Our study illustrates the usefulness of using Skinner boxes with invertebrates.

References:

1.Hursh, S. R., & Roma, P. G. (2016). Behavioral Economics and the Analysis of Consumption and Choice. Managerial and Decision Economics, 37, 224-238.

2.Rachlin, H., Green, L., Kagel, J. H., & Battalio, R. C. (1976). Economic Demand Theory and Psychological Studies of Choice11The chapter was jointly written by H. Rachlin and L. Green; the experiments described were done in collaboration with J. H. Kagel and R. C. Battalio. The work was partially supported by an NSF grant to H. Rachlin, NIH Fellowship to L. Green, and NSF grant to J. H. Kagel and R. C. Battalio. We are grateful to Dr. Peter Henderson who developed the computer program. In G. H. Bower (Ed.), Psychology of Learning and Motivation (Vol. 10, pp. 129-154): Academic Press.

3.Sokolowski, M.B.C., Abramson, C.I.A. (2010). From foraging to operant conditioning: A new computer-controlled Skinner box to study free-flying nectar gathering behavior in bees. Journal of Neuroscience Methods, 188, 235-242.

4.Sokolowski, M.B.C., Abramson, C.I.A. (2016). Operant Conditioning of the Honey Bee [Video file]. Retrieved from https://www.youtube.com/watch?v=2o5pSji_g3E&t=223s, Accessed 5 February 2018.

Behaviour Monitoring in Context: a Methodology for Measuring the Impact of the Human Factor on Smart Cities Big Data and Technologies

F. Spagnoli[1], S. van der Graaf[1], A. Smets[1]

(imec-SMIT, Vrije Universiteit Brussel, Brussels, Belgium, Francesca.spagnoli@imec.be; Shenja.vanderGraaf@imec.be; Annelien.Smets@imec.be)

This paper proposes a methodology to design large scale user experiments within the context of smart cities, supporting the measurement of objectively, contextualised and thus realistic user behaviour in real-time. There is a long tradition of studying human behaviour in many academic disciplines, including economics, sociology, clinical and social psychology; understanding what motivates humans and triggers them to perform certain activities, and what are the determinants of (changing) behaviour is central both for researchers and companies, as well as for policy makers to implement efficient public policies. While numerous theoretical approaches for diverse domains such as health, retail and environment have been developed, the methodological models guiding the evaluation of such studies are said to have started showing their limitations in the contemporary everyday context. More specifically, digitisation, the Information and Communication Technologies (ICT), the Internet of Things (IoT) connecting networks of devices, and new possibilities to collect and analyse massive amounts of data have made it possible to study behaviour from a realistic perspective, as never before. Digital technologies enable to (1) capture data in real-life settings, (2) regain control over information by capturing the context of behaviour, and (3) analyse huge sets of information through continuous measurement. Within this context, by exploiting the proposed framework, behavioural research can be conducted to address the issues of different domains, such as mobility, environment, health or energy, as it has especially been designed for a smart city context.

To this end, our paper, therefore, affords a model linking technologies and the so-called ‘human factor’ which strongly relies on defining a methodology for shaping behavioural profiles – referred to as ‘urban citizenship’ profiles - and related interventions, as well as the evaluation of side-effects on the creation of new business models and sustainability plans. This methodology will be particularly developed to support city governments and public institutions to effectively use large data sets generated in urban environments. Moreover, this integrated framework represents the need of considering both the personal determinants as well as the contextual ones as the latter are often recognized as important, yet, sufficient insights into how context can be taken into account are overall lacking. Thus, within the complex environment of a smart city, technology is generally seen as a facilitator for, particularly, behaviour change, even if often there is a de-synchronisation between tech development and context of use.

It is estimated that by 2050, 70% of the world population will live in cities [1]. Urban big data can develop relevant benefits for the smart cities [2]. Indeed, ICT has always played a primary role in mobilising urban innovation and supporting the creation of the “Smart Citizenship” concept [3]. Traditionally, smart cities use the IoT technologies and big data to improve its services and products, by taking advantage of co-creation processes, methodologies and tools [4]. It has been demonstrated that the use of urban big data is a fair tool for creating new business opportunities and markets in smart cities, especially in terms of service innovation and creation of new business models [5, 6].

What is the rationale behind so-called ‘smart cities’ then? The concept has been variously defined within the literature, on the one hand as urban places composed of “Everyware”, a context pervaded from ubiquitous computing to monitor urban data flows with the final scope of engaging with its citizens [7]; on the other hand, the term has been related to the ability to enact the development of knowledge economies, where the governance of an urban environment is triggered by the innovation and creativity of its smart citizens [8]. It is clear that smart cities can play an active role for enabling socio-economic progress for increasing the sustainability of urban contexts [8, 9]. Indeed, they have produced a relevant impact in different sectors, especially in mobility, health, energy and education. These social and economic transformations are even more enacted by the implementation of big data applications to improve the living standards of citizens [10]. According to Batty, big data in a smart city context have the potential to activate a change in every sector of the society and to generate a positive impact on the economy of all nations [11]. The fact that the current volume of big data is approximately 2.5 quintillion bytes [12] demonstrates that the potential is enormous. In cities, data analysis (and integration) is increasingly seen as the main source of promoting growth and wellbeing, thus improving the quality of life of citizens and the sustainability of cities.

In this view, smart cities are likely to be able to achieve a full urban sustainability, but only if they succeed to trigger or support a sustainable behaviour of their citizens, as well as drafting a sustainability plan, but only if they work collectively, involving all actors and citizens, who ultimately are also needed to be incentivised to change their current behaviour to reach smart city goals. This includes the human factor as a required element for changing behaviour. Urban sustainability means that both social, economic and environmental impacts are exploited, and this is only possible by considering the whole complex system of a smart city, as well as applying a participatory governance model including the "human factor" within this process.

The proposed methodology has been drafted by taking into account these projections in the near future and by collecting directly the challenges arising in smart cities. The methodology builds on an extensive literature review from psychology, health and behavioural economics, and provides a comprehensive approach, enabling a reduction in weaknesses apparent in current available theories used in research within the smart cities context [13]. As behaviour is intrinsic context-dependent, the main scope of this methodology is to map therefore behaviours into specific profiles, determine the value of behaviour, and collect a set of data derived from behaviour change interventions, which can be used for predicting the effects of behaviour. The final expected results to implement such methodology are to:

- Steer tailored interventions.
- Provide requirements for technological development stages.
- Make informed decisions for several points of interventions (value trade off).
- Deploy data-streams and real-time randomized trials in a cost-effective way through an enriched city dashboard
- Identify path dependencies influencing business modeling.
- Provide policy recommendations to municipalities, service and utility providers.
- Increase technological take-up and minimize adoption risks.
- Improve the sustainability of technological solutions.

The proposed model will be first applied to the "City of Things" within the city of Antwerp in Belgium, to be further refined and validated.

Ethics

Research protocol and procedure will be submitted for ethical review to the VUB Ethics Committee that includes a multidisciplinary panel of experts for ethical review of research in the humanities and the behavioral or social science research traditions.

References:

1. M. Lierow, B2City: the next wave of urban logistics, Oliver Wyman, available at http://www.supplychain247.com/paper/b2city_the_next_wave_of_urban_logistics, (last accessed on 24 July 2014), 2014.
2. R. Kitchin, The real-time city? Big data and smart urbanism, GeoJournal, 79(1), (2014) 1–14.

3. C. Martelli, A point of view on new education for smart citizenship, Future Internet MDPI, 9(1), (2017) 4.
4. Spagnoli, F., van der Graaf, S., Brynskov M. (Forthcoming). The Paradigm Shift of Living Labs in Service Co-Creation for Smart Cities: SynchroniCity Validation. Springer series: Lecture Notes in Information Systems and Organisation (LNISO). Volume: "Organizing in the digital economy. At the interface between social media, human behaviour and inclusion".
5. Y. Pan, Y. Tian, X. G. D. Liu, G. Hua, Urban big data and the development of city intelligence, Engineering Volume 2, Issue 2, June 2016, pp. 171-178, Elsevier, 2016.
6. G. Morelli, F. Spagnoli, Creative industries and big data: a business model for service innovation. In: S. Za, M. Drăgoicea, M. Cavallari (eds), Exploring Services Science. IESS 2017. Lecture Notes in Business Information Processing, Springer, Cham, 2017, vol. 279, p.156.
7. A. Greenfield, Everyware: the dawning age of ubiquitous computing, first edition, New Riders, Berkeley, 2006.
8. K. Kourtit, P. Nijkamp, D. Arribas-Bel, Smart cities perspective. A comparative European study by means of self-organizing maps, Innovation, 25(2), (2012) 229–246.
9. A. Townsend, Smart cities: big data, civic hackers, and the quest for a new utopia. W.W. Norton & Co, New York, 2013.
10. C. E. Jimenez, A. Solanas, F. Falcone, E-government interoperability: linking open and smart government, Computer, 47(10), (2014) 22-24.
11. M. Batty, Big data, smart cities and city planning. Dialogues in Human Geography, Sage, 3(3), (2013) 274–279.
12. J. M. Koh, M. Sak, H.X. Tan, H. Liang, F. Folianto, T. Quek, Efficient data retrieval for large-scale smart city applications through applied Bayesian inference, Proceedings of the 10th International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), Singapore, 2015, 1-6.
13. Spagnoli, F., van der Graaf, S. Ballon, P (2017), Changing Behaviour in the Digital Era: A Concrete Use Case from the Domain of Health, World Academy of Science, Engineering and Technology International Journal of Psychological and Behavioral Sciences Vol:11, No:12, 2017

Assessing disinhibition behaviour in behavioural variant frontotemporal dementia patients using ecological and cognitive tasks

Delphine Tanguy[1,3], Johan Ferrand-Verdejo[1], Richard Levy[1,2], Julie Socha[1], Raffaella 'Lara' Migliaccio[1] and Bénédicte Batrancourt[1]

1 Inserm U 1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMR S 1127, Institut du Cerveau et de la Moelle épinière, Paris, France. delphine.tanguy@icm-institute.org; 2 AP-HP, Hôpital de la Salpêtrière, Behavioural Neuropsychiatry Unit, Paris, France ; 3 École Normale Supérieure Paris-Saclay, Cachan, France.

Introduction

Behavioural variant frontotemporal dementia (bvFTD) is an early-onset neurodegenerative disease (the second most common form of dementia after Alzheimer's disease) associated with behavioural disturbances. These troubles result from a frontotemporal lobar neurodegeneration [1]. This behavioural variant of FTD is characterized by a progressive deterioration of personality, social conduct and cognition. In particular, cognitive and behavioural disinhibition is systematically reported in bvFTD patients and represents the core of current clinical criteria [2]. Disinhibition has been identified as one of the major cause for caregiver distress [3]. However, its assessment, characterization and neural correlates are poorly known. The aim of our study is two-folded: first, to identify a signature of disinhibition on a behavioural and cognitive plan in order to generate new clinical tools to assess this symptom, and second, to study the neuronal networks associated with disinhibition in bvFTD patients.

Project/method

We are studying cognitive and behavioural inhibition troubles in bvFTD using a population of eleven patients (Mini Mental State Evaluation >20 for a correct cognition) and nine control subjects, following two approaches:

A behavioural approach with an investigation in "real-life" situation (ecological task) ECOCAPTURE (http://clinicaltrials.gov/NCT02496312) is a tool developed by the laboratory, used here to assess in a quantitative way the disinhibition behaviour, in a "real-life" situation. The subject is in a waiting room during 45 minutes, instructions are to get comfortable and to enjoy the room. A questionnaire has to be filled, making the subject interacting with the environment. A video recording system allows us to encode and analyse some disinhibition behaviours (social inappropriate behaviour, loss of manner or decorum and impulsive, rash or careless actions). With all these data, the aim is to retain some metrics leading to a differentiation in disinhibition between patients and control subjects. In 2015, six bvFTD and six controls were explored by this tool, and results were already discussed in a few congresses. A multivariate power analysis showed that nine subjects per group were necessary and sufficient to get valid significant results.

A cognitive approach via a neuropsychological evaluation (cognitive task)

The Hayling test is a classical tool evaluating cognitive disinhibition with completion of sentences. Three scores will be available: completion time in part A (automatic condition), completion time in part B (inhibition condition) and errors number. An augmentation of completion time in part B and an augmentation of errors number are expected with patients.

These two approaches are complementary and provide a cognitive and behavioural signature (composite score) of inhibition troubles, leading to a new and objective diagnostic tool. This double assessment will allow us to define the whole profile of disinhibition. Furthermore, a validation study will be necessary for a clinical utilisation.

Results

At the time of this submission, five patients and three controls performed the Hayling test and no significant differences were found. However, a tendency appears for errors number (p=0.059, patients > controls). Regarding the behavioural task, including 11 patients and nine controls, the metrics of inappropriate behaviour show a significant difference (p = 0.044, patients > controls). Inclusion is still in progress for these two tasks, for a total of 30 patients and 30 healthy controls, by the end of June 2019.

Ethical statement

We have obtained formal consent from the local ethical research committee: CPP, Comité de Protection des Personnes Sud-Méditerranée I, Région Provence Alpes Côte d'Azur, France. ECOCAPTURE, ID RCB: 2017-A00416-47, PI: Pr. Richard LEVY

Fundings

Study Funded by the program "Investissements d'avenir" ANR-10-IAIHU-06; the foundation "Fondation pour la recherche médicale", FRM, frm.org; the ENEDIS company, Enedis - Tour Enedis, 34 place des Corolles - 92079 Paris La Défense - enedis.fr

References

1. Massimo et al. (2009). Neuroanatomy of apathy and disinhibition in frontotemporal lobar degeneration. Dement Geriatr Cogn Disord, 27:96–104.
2. Rascovsky et al. (2011). Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain, 134:2456–2477.*
3. O'Connor et al. (2017). Behavioral-variant frontotemporal dementia: Distinct phenotypes with unique functional profiles. *Neurology*, 89(6), 570-577.
4. Peters et al. (2006). Orbitofrontal dysfunction related to both apathy and disinhibition in frontotemporal dementia. *Dement Geriatr Cogn Disord, 21:373–379.*
5. Paholpak et al. (2016). Person-Based Versus Generalized Impulsivity Disinhibition in Frontotemporal Dementia and Alzheimer Disease. *Journal of geriatric psychiatry and neurology*, *29*(6), 344-351.
6. Lecouturier K., B. Batrancourt, J. Ferrand-Verdejo, C. Azuar and R. Levy. Unpublished.

Thermal Imaging as an Emerging Technique to Study Proximate Causes of Behaviour – A Review on Current Methods and Future Directions

H. Telkanranta[1, 2], E. Paul[1] and M. Mendl[1]

1 Bristol Veterinary School, University of Bristol, UK. helena@telkanranta@bristol.ac.uk 2 Faculty of Veterinary Science, University of Helsinki, Finland

Causal factors affecting behaviour fall into two categories. Ultimate causes, studied for example in behavioural ecology and evolutionary biology, involve the fitness of genotypes; they help explain why natural selection has favoured some behavioural strategies over others. Proximate causes, studied in the fields of cognition, motivation and emotion, involve individual-level perception, decision-making, value judgments and learning processes; they help explain why an individual is motivated to carry out particular behaviours at particular times. In furthering our understanding of proximate causes of behaviour, one way forward is the development of new methods to measure these internal processes of individuals, whether conscious or non-conscious. Subjective states cannot be measured directly, but substantial information can be gleaned about the processes of behavioural decision-making through measuring physiological processes linked to them. Ideally, such measures should be non-invasive and remote, in order not to affect the processes being measured.

One of the technologies increasingly utilised to this end is thermal imaging, also known as infrared thermography. It is considered to hold substantial promise as a research tool for the future, including automated monitoring of animals in laboratories, zoos and veterinary clinics, on farms and in the wild. Thermal imaging of individual animals allows for remote measurement of the distribution of temperatures across a surface, such as an animal's face or body, by an array of sensors detecting infrared radiation. These measurements are then converted to numerical temperature data, which often are visualised as heat map images (see Fig. 1 for an example). There is a wide range of different types of thermal cameras for different purposes and with different levels of image resolution. In most cameras used for scientific research, spatial resolution varies from 240 x 360 pixels to 768 x 1024 pixels, and temporal resolution varies from 8 to 200 images per second. The higher end of these allows for detailed measurements of free-moving animals at distances of a few metres to several tens of metres, depending on the size of the animal and the resolution required. Prices of thermal cameras have decreased substantially over the years and are expected to continue to do so, which has added to the interest in developing methods to use them in a variety of settings.

In the context of studying proximate causation of behaviour, the most common use of thermal imaging so far has been the study of the relationships between thermal physiology, ambient temperatures and behaviour (e.g. in laboratory rodents; farm animals such as cattle, pigs and poultry, and wild animals including both endothermic and ectothermic species). Recent years have also seen a rapid increase in research to develop thermal imaging methods to measure physiological processes that are directly linked to emotional states, such as activation of the sympathetic nervous system during emotional arousal, which causes peripheral vasoconstriction and therefore a measurable reduction in peripheral temperature. Research by several groups including ours has demonstrated substantial potential to utilise such findings in the development of new methods to measure perception, emotion and motivation in a wide range of species. However, the increased availability of thermal cameras and the aesthetically appealing nature of thermal images is a double-edged sword. While thermal imaging is a mature technology, its use in the development of new research methodology to measure physiological changes linked to causal processes is still at an early stage.

One of the essential requirements for successful further development is an increase in interdisciplinary collaboration between researchers with backgrounds in ethology, physiology and cognition as well as thermal physics. Among the methodological research needs, one of the most important involves quantifying how other physiological processes, such as digestion of recently eaten food, exercise and diurnal rhythm, affect surface temperature distribution in those parts of the face and body that are typically of interest in emotion studies. As an example, the ear pinna, in which peripheral vasoconstriction and the resulting temperature drop during emotional arousal have been found (e.g. in the dog and rabbit), also acts as part of the thermoregulatory

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

vasoconstriction/vasodilation system in several species, leading to the prediction that there are species-specific threshold values for low and high ambient temperatures that would mask emotion-related effects in ear pinna temperature. Another major area in which methodological research is still needed concerns the effects of the environment can have on the temperature parameters of interest. Air currents, radiated heat from nearby animals, reflected heat from nearby surfaces as well as residual heat from recent touch by other animals or recent direct sunlight, are all examples of external factors that often have substantial confounding effects in temperature readings recorded by the thermal camera. Systematic, interdisciplinary studies of these effects is therefore one of the essential components of developing reliable thermal imaging methods for measuring motivation for behaviour.

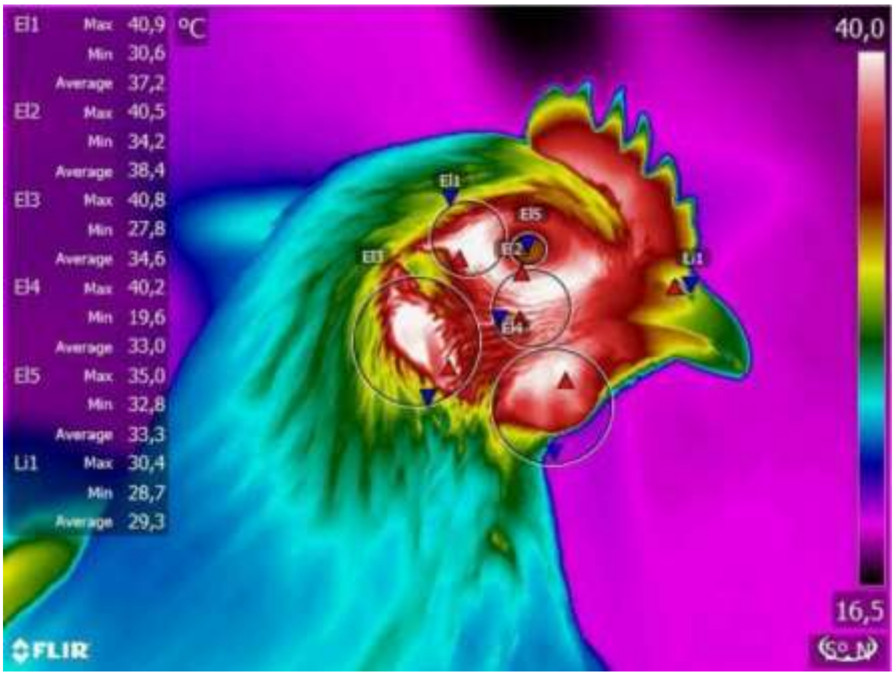


Fig. 1. An example of extracting temperature data from a thermal image. Data on the left includes the minimum, maximum and average temperatures in the regions of interest, and the temperature bar on the right shows which colours correspond to which surface temperatures in this visualisation. The compass reading in the lower right-hand corner records the orientation of the camera as the deviation from pointing north. From Telkanranta et al., in prep.

Sheep nocturnal grazing behavior using infrared point-of-view camera – a preliminary study

M. Terra-Braga[1]✉, M. I. Ferraz-de-Oliveira[2], J. A. Lopes-de-Castro[2], E. Sales-Baptista[2]✉

[1]**MSc Livestock Farming Systems, Montpellier SupAgro. Montpellier, France.**

[2]**Instituto de Ciências Agrárias e Ambientais Mediterrânicas (ICAAM), Department of Zootechny, University of Évora. Évora, Portugal.**; marina.terra-braga@supagro.fr; elsaba@uevora.pt

Introduction

Extensive livestock production systems rely largely on forage availability as feed source. Under free ranging conditions, grazing behavior develops circadian rhythmicity. Grazing activity occurs mainly during the day, normally with two important periods around sunrise and sunset [1]. Nevertheless, grazing patterns and herbage intake are constantly affected by environmental factors. In dry climates, such as Mediterranean, precipitation is typically scarce during the hot season. In addition, native vegetation in these areas is dominated by annual species. As a result, pasture becomes poor in quality along the growth cycle, which leads animals to graze for longer periods, including at night, to cover their nutritional needs [2]. Nocturnal grazing may also increase when air temperature during the day is high. In this case, ruminants tend to avoid foraging in the hottest part of the day [3]. For those reasons, the study of sheep grazing patterns and behavioral responses to environmental limitations is essential to achieve successful management practices while improving animal welfare.

The study of livestock foraging behavior is a multidisciplinary field that involves many levels of evaluation, from grazing patterns to forage selection. During the day, behavior recording is relatively simple through live observation. However, observations at night are restricted by limited sight in the dark. Besides, it is difficult to predict whether the presence of an observer may alter the normal behavior, unless we use recording equipment, such as accelerometers that measure grazing by detecting movements of the head [4]. Despite conveiyng information on grazing patterns, accelerometers cannot provide information on forage selection.

In the present study, we evaluate sheep grazing behavior through animal-borne cameras. Animal-borne video and environmental collection systems (AVED) have been used for studies on wild animals habitat and behavior for three decades [5]. Since then, technological development has made the equipment smaller and lighter, allowing the use on smaller species. Depending on the anatomy of the species and the aim of the study, the camera may be attached using different techniques (e.g. epoxy glue, harness or neck collar).

Besides, the weight of commercial cameras is not expected to interfere in the normal behavior. A study on sheep circadian rythm using GPS neck collars representing approximately 2.2% of their body weight found no significant differences between collared and no-collared (control) groups [6]. Finally, another advantage of this method against fixed cameras is the possibility of measuring behavior in large paddocks, since it is attached to the animal.

Animal-borne video has already been used by our team for studies in sheep grazing behavior, and the results were positively correlated with observations of flock behavior [7]. However, despite the accuracy of this method, the nocturnal camera has recorded only five hours of video footage. Thus, it was not possible to evaluate grazing behavior from dusk to dawn.

The aim of this preliminary study was to test a filming equipment to record grazing behavior from the sheep's point of view all through the night.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Materials and methods

The study was conducted in the region of Alentejo, Portugal, 12 km south-west of Évora, within the University of Évora's Mitra Experimental Station (38º31'44"N, 8º01'00"W). The climate is typically Mediterranean, with long dry summers and mild winters.

One black merino from a flock of 37 adult non-lactating ewes was selected to bear a camera. The flock was kept under continuous grazing in a 2.3 ha paddock of permanent native sward without any other feed supplement distribution.

We used one infrared night vision PatrolEyes® (PatrolEyes. Ada Township, Michigan USA; Figure 1B) attached to a neck collar to record from the animal's point of view. A protective case was manufactured with a transparent plastic soapbox and water-resistant with tape (see Figure 1A). In order to extend the recording, a portable power bank was plugged into the camera. In total, the filming equipment weighted 400.2 g, representing approximately 0.6 % of the sheep's body weight.

The recording was made in the night of May 18th to 19th, starting at 9:00 pm. Sunset was at 8:39 pm and sunrise at 6:17 am. Weather was clear with wind speed from 4.5 to 6 m/s. Lowest temperature was 8.2 ºC and highest was 22.6ºC on the 18th during the day.

A B

Figure 1. (A) Photograph of a sheep fitted filming equipment; (B) Photograph of PatrolEyes® infrared night vision camera.

The video footages were analyzed through the software Behavioral Observation Research Interactive Software (BORIS®; www.boris.unito.it). Screen captures are shown in Figure 2. Grazing duration and number of bouts were analyzed.

All the procedures used on animals were approved by the ICAAM board under a license from DGAV (Direcção Geral de Alimentação e Veterinária).

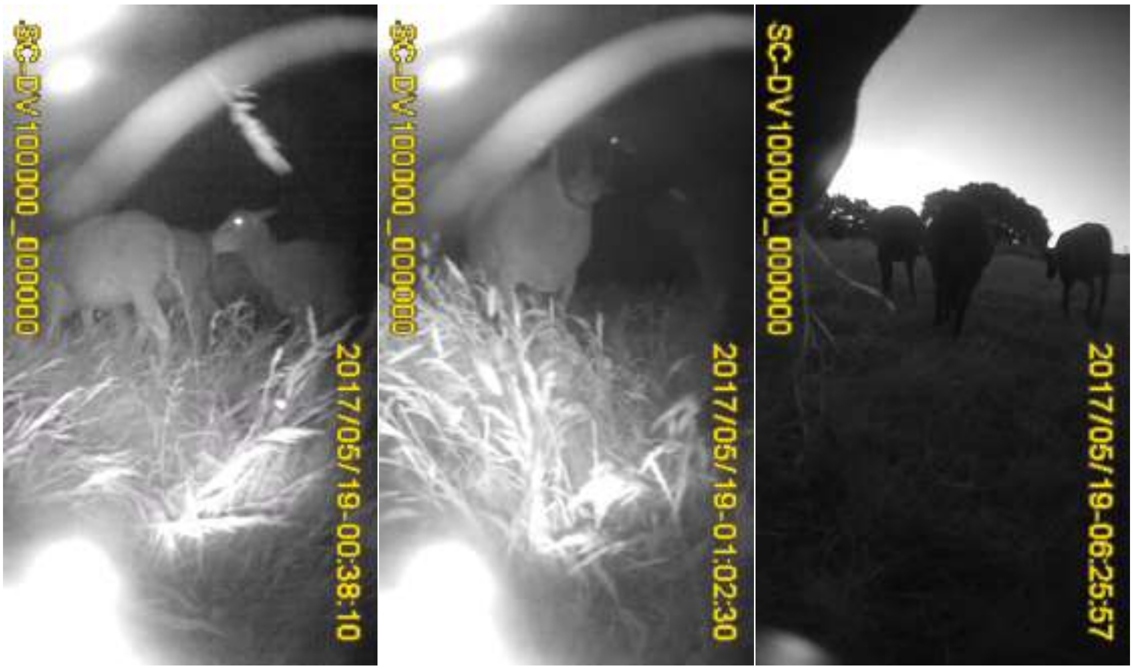


Figure 2. Screenshots of the video footage.

Results

The length of the video footage was of ten hours, from 9pm to 7pm. Total grazing and number of bouts are presented in Figure 3. The first hour, which corresponds from 9pm to 10pm, had the longest grazing duration (49 minutes and 54 seconds). Subsequently, grazing duration dropped to 11 minutes and 28 seconds in the second hour and zero in the third hour. There was a small peak of 14 minutes 39 seconds of grazing from midnight to 1am, followed by 3 minutes of grazing in the next hour. There was not any grazing from 2am to 5am, and a new peak of 12 minutes 36 seconds is observed from 6am to 7am, in the hour when the sun rose. Number of bouts followed the same pattern as grazing duration, the highest peak being at the first hour.

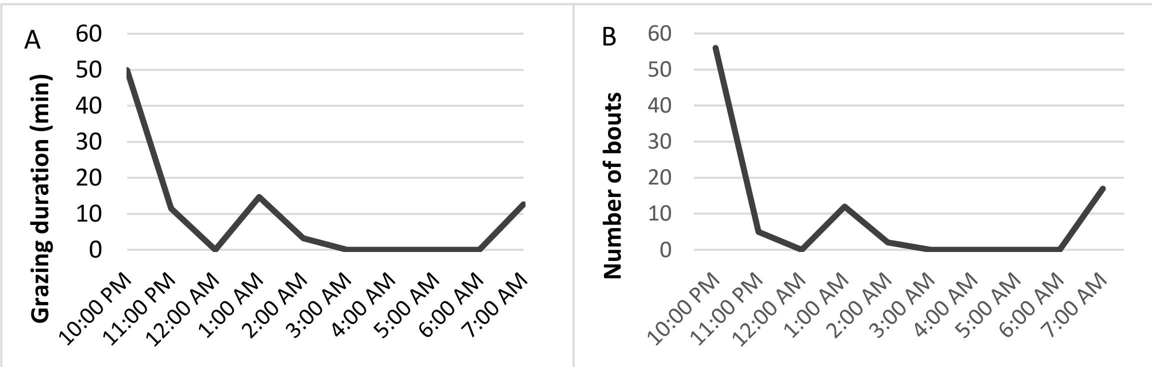


Figure 3. (A) Grazing duration per hour; (B) Number of bouts per hour

Discussion

Our filming equipment was able to record from the animal's point of view from dusk to dawn, permitting access to grazing, among other behaviors. Behavior studies through video footage allow a deep and concentrated analysis of the material. Furthermore, it spares researchers from long hours of behavior observation in the field and eliminates the observer effect.

In our preliminary study, we observed a grazing pattern that was very similar to previous results [8]. Although grazing was intense after sunset, most of the night it did not occur at all, except for midnight to 1am, when there was a small peak. Then, our ewe resumed grazing around dawn. Air temperature was mild in the day that preceded the evaluation, suggesting that this grazing activity is a regular pattern and not a response to heat stress during the day.

When the footage is in high definition, as it was in our trial, animal-borne video could also serve as a tool for studies on feed selection. A previous study on white-tailed deer evaluated grazing selection continuously using neck collar infrared cameras and were able to identify the plant species 84% of time [9]. Therefore, this method is an accurate and practical tool for studies on livestock grazing species, such as cattle, sheep, goats and horses. It is particularly useful in native swards where biodiversity is high.

Battery duration is still the main limitation for using animal borne cameras. This leads to frequent manipulation of the animals, apart from a need to have multiple devices since the batteries need to be charged. External power banks may therefore be a good solution to minimize human interference and to obtain longer footages, but they require manufacturing protective cases. There are many advantages of the animal-borne video method. However, further studies are necessary to reinforce its validity, reliability and feasibility.

References:

1. Van Soest, P.J. Nutritional ecology of the ruminant. 2.ed. Ithaca: Cornell University, 1994. 476p.

2. Manteca, X., & Smith, A. J. (1994). Effects of poor forage conditions on the behaviour of grazing ruminants. *Tropical Animal Health and Production*, *26*(3), 129–138. https://doi.org/10.1007/BF02241068

3. Arnold, G.W. (1984). Comparison of the time budgets and circadian patterns of maintenance activities in sheep, cattle and horses grouped together. *Applied Animal Behaviour Science*, *13*, 19–30. https://doi.org/0168-1591/84/$03.00

4. Alvarenga, F. A. P., Borges, I., Palkovič, L., Rodina, J., Oddy, V. H., & Dobos, R. C. (2016). Using a three-axis accelerometer to identify and classify sheep behaviour at pasture. Applied Animal Behaviour Science, 181, 91-99.

5. Moll, R.J., Millspaugh, J.J., Beringer, J., Sartwell, J., & He, Z. (2007). A new "view" of ecology and conservation through animal-borne video systems. *Trends in Ecology and Evolution*, *22*(12), 660–668. https://doi.org/10.1016/j.tree.2007.09.007

6. Hulbert, I.A.R., Wyllie, J.T.B., Waterhouse, A., French, J., & Mcnulty, D. (1998). A note on the circadian rhythm and feeding behaviour of sheep fitted with a lightweight GPS collar. *Applied Animal Behaviour Science*, *60*, 359–364.

7. Sales-Baptista, E.; Terra-Braga, M.; Castro, J.A.L.& Ferraz-de-Oliveira, M. I. (2017). Use of point of view cameras for welfare management in extensive grazing systems. *Proceedings of the IX International Congress on Silvopastoral Systems.* Manizales, 6-8 September 2017).

8. Colquhoun, I.R. (1970). *The grazing ecology of red deer and blackface sheep in Perthshire, Scotland*. PhD thesis, University of Edinburgh.

9. Lavelle, M.J., Blass, C.R., Fischer, J.W., Hygnstrom, S.E., Hewitt, D.G., & VerCauteren, K.C. (2015). Food habits of adult male white-tailed deer determined by camera collars. *Wildlife Society Bulletin*, *39*(3), 651–657. https://doi.org/10.1002/wsb.556

High accuracy indoor positioning system to monitor spatial use in dwellings

N. Van Loy[1], G.Verbeeck[2] and E. Knapen[3]

1 Department of Architecture and Arts, University of Hasselt, Hasselt, Belgium. Nick.vanloy@uhasselt.be; 2 Department of Architecture and Arts, University of Hasselt, Hasselt, Belgium. Griet.verbeeck@uhasselt.be; 3 Department of Architecture and Arts, University of Hasselt, Hasselt, Belgium. Elke.knapen @uhasselt.be

Introduction

In recent years, occupancy sensing in indoor environments has been used in a wide field of applications, such as smart control of HVAC-systems, detection of humans in emergency interventions [1], assisted living [2], searching for lost people or objects, shopping behaviour, industrial applications [3], player tracking in sports, games, etc.. Three different levels of occupancy sensing can be distinguished: occupancy detection, occupancy counting and occupancy tracking [4]. Occupancy detection provides information about whether someone is present in a specific room or not, but not about the number of occupants or the spatial use within a room. It can be done by several techniques, such as infrared-detection [5] and sound-detection. . Occupancy counting, e.g. by CO_2-measurements, gives more information about the number of occupants in a room [6, 7]. The highest level of occupancy information is obtained by occupancy tracking, which gives, besides the presence and number of occupants, the exact location of occupants within a room and can be measured by indoor localisation techniques based on e.g. Bluetooth, Wi-Fi [4] and ultra-wideband (UWB) [1, 8]. Indoor localisation techniques are in general more intrusive for residents because they need to wear a tag [4], although some of them can be build-in in residents' smartphones (Bluetooth, Wi-Fi).

Nowadays, indoor localisation systems are mostly applied in large buildings with large spaces, such as office spaces, airports, etc.. For monitoring spatial use patterns in dwellings [9], a very precise indoor localisation system is needed which is able to track the position of residents in small rooms with a lot of obstacles. In this paper, system requirements for indoor tracking in dwellings are derived and an overview and evaluation of commonly used techniques for occupancy sensing in buildings are provided.. Four commercially available, ultra-wideband based localisation systems are compared. In a case study analysis, one commercially available system, Pozyx™, which provides accurate positioning and motion information, is used for monitoring spatial use patterns of residents in their dwelling.

Requirements for accurate indoor tracking techniques in dwellings

System requirements are crucial when selecting a sensing technology that is able to monitor the moving position of multiple residents over time separately. Mautz [10] defined several criteria which can be used for the assessment and comparison of different occupancy sensing techniques. In order to obtain the most fitting technique, the specifications of the system need to match the predetermined requirements which depend on the application. When searching the most suitable system, the different requirements have to be weighed against each other which is not a straightforward method [10].

When monitoring the exact location of residents in dwellings, some boundary conditions and difficulties have to be taken into account. First of all, the monitoring system has to deal with a small, but very complex environment with lots of obstacles, e.g. indoor walls and furniture that can disturb the location signal. Furthermore, the system has to be accurate enough (min. horizontal accuracy 1 m, min. vertical accuracy 2m), because, due to the small size of dwellings, small errors in accuracy can lead to large errors in the results, such as locating a person incorrectly in an adjacent room. Privacy is another critical issue when tracking residents in their own dwelling. The monitoring is rather intensive for the residents and can be perceived as intrusive, which can lead to drop-out or changes in their normal behaviour. Therefore, the monitoring system has to be as compact as posssible and may not restrict their normal living patterns. Additionally, some extra practical requirements are added, such as the possibility to set up the whole system in a relative simple and quick way (around 2,5 hours) and the need for limited infrastructure/devices. Lastly, the cost of the detection system has also to be taken into account. Table 1 shows a summary of the requirements which are derived for occupant tracking within rooms of dwellings.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Table 1: Requirements for occupant tracking within rooms in dwellings.

CRITERIA	REQUIREMENTS
COVERAGE AREA (MEASUREMENT AREA/VOLUME)	Semi-large dwelling (2-3 floors)
HORIZONTAL ACCURACY (POSITION IN A ROOM)	<1m
VERTICAL ACCURACY (INDICATION OF THE FLOOR LEVEL)	<2m
LEVEL OF OCCUPANCY SENSING (DETECTION, COUNTING, TRACKING)	Tracking
OUTPUT DATA (SPATIO-TEMPORAL DATA)	Coordinates with timestamp
NUMBER OF USERS (NUMBER OF TAGS)	One family (1-5 tags)
PRIVACY (NATURE OF GATHERED DATA)	According to ethical regulations
INTRUSIVENESS (DISTURBING DAILY ACTIVITIES)	Low/Moderate
SIZE/WEIGHT (SIZE/WEIGHT OF MEASUREMENT UNIT)	Handheld
INSTALLATION COMPLEXITY (MAN-HOURS TO INSTALL THE SYSTEM)	<2,5hours
REQUIRED INFRASTRUCTURE (DEVICES/CABLES)	Moderate
COST (PRICE OF THE COMPLETE SYSTEM)	Moderate

Evaluation of occupancy sensing techniques

Occupancy sensing in indoor environments can be achieved by using different sensors such as infrared-detection, CO_2-concentration measurements, cameras, indoor localisation systems (based on radio signals), although not all of them provide complete information about the location within a room. In Table 2, commonly used systems for occupant sensing and indoor localisation are showed, including the evaluation of their specifications according to the predetermined requirements.

Category	**Techniques** / ***Requirements***	*Level of occupancy sensing*	*Horizontal accuracy*	*Vertical accuracy*	*Coverage area*	*Output data*	*Number of users*	*Privacy*	*Intrusiveness*	*Size/weight of the tag*	*Required infrastructure*	*Installation time*	*Cost*	*References*
Sound	Ultrasonic	Detection	mm	mm	Scalable	Distances	NA	High	Low	NA	High	High	High	[10, 11]
	Audible sound	Detection	Room	Room	Scalable	Presence	NA	Med	Low	NA	Med	Med	Low	[10, 11]
Concentration	CO_2-concentration	Counting	Room	Room	Scalable	Presence	NA	High	Low	NA	Med	Low	Med	[6, 7, 12]
Infrastructure	Pressure-sensors/computer activity	Detection	Zone	Zone	Scalable	Location	NA	High	Low	NA	High	High	High	[11, 13]
Image detection	Visual	Tracking	dm-mm	dm	Scalable	movie	High	Low	Low	NA	High	Med	High	[2, 10, 11, 14, 15]
	Infrared-detection	Detection	Zone	Zone	Scalable	Presence	NA	High	Low	NA	Med	Med	Low	[2, 10, 11]
Radio frequency signals	GPS	Tracking	25m	X	NA	Coordinates	High	Med	Med	Small	Low	Low	Low	[16]
	Wi-Fi	Tracking	3-10m	Floor	Scalable	Coordinates	High	Med	Med	Small	Med	Med	Low	[4, 10, 17, 18]
	Bluetooth	Tracking	1-3m	Floor	Scalable	Coordinates	High	Med	Med	Small	Med	Med	Med/high	[10, 19, 20]
	Ultra-wideband	Tracking	dm	dm	Scalable	Coordinates	Med	Med	High	Med	Med/high	Med	Med/high	[1, 3, 8, 10]
	RFID	Tracking	1-3m	Floor	Scalable	Coordinates	High	Med	Med	Small	Med	Med	Low	[10]

Table 2: Commonly used techniques for occupant sensing

If the aim is to investigate spatial use within rooms, the exact location of residents needs to be monitored and, therefore, some of the techniques of Table 2 need to be excluded. Ultrasonic sound-detection, audible sound-detection, infrared-detection and CO_2-concentration measurements only provide information on the level of occupancy detection or occupancy counting [3, 10]. These techniques can be valuable for other studies in which - the exact location of residents within a room is not needed, as they are in general cheaper and less intrusive to the residents. Infrastructure-based techniques, such as pressure tiles, can deliver fine-graded information about the location, depending on the number of sensors [13]. These techniques are most suitable to situations where occupant information is needed for a specific place such as a desk or a conference room [11]. Visual camera-detection can offer very detailed information about the location of occupants and even their activities and actions, but privacy concerns related to camera detection will be very high. Additionally, the image processing and required infrastructure will lead to a large cost [11].

Most techniques that can determine the exact location of residents are radio-frequency based, but there is a large difference in accuracy between those techniques. Most techniques require the resident to wear a tag (or smartphone) [4]. The most well-known localisation technique is GPS (global positioning system), which can reach an outdoor accuracy up to a few meter by using satellites. Once indoor, GPS cannot be used anymore because GPS-signals cannot penetrate solid walls and it will lead to inaccurate results [3]. Therefore, other radio-frequency based signals are commonly used which can work in an indoor environment, such as Wi-Fi, Bluetooth, Ultra-Wideband and RFID (Radio-frequency identification).

Two most commonly used measuring principles for positioning techniques are (multi)lateration and (multi)angulation. Lateration, trilateration and multilateration rely on measuring distances to calculate the position (Figure 1a). For two-dimensional environments, minimum three distances are needed, while, for three-dimensional environments, minimum four distances are needed. Lateration can be used in combination with algorithms to calculate the distance between the transmitter and receiver, such as:

- Received Signal Strength: RSS-based algorithms measure the signal strength of the received signals to estimate the distance between the transmitters and receivers. RSS is sensitive to NLOS-situations (not-line of sight) and multipath environments which leads to less accurate results in these environments and makes them less attractive for accurate indoor localisation [3, 10].
- Time Of Arrival: TOA-based algorithms estimate the distance by calculating the traveling time between the transmitters and receivers and divide them by the wave speed. Therefore, the clock between the transmitters and receivers has to be precisely synchronized [3, 10].
- Time Difference Of Arrival: TDOA-based algorithms measure the difference between the time of arrival measurements and has as advantage that only the transmitters have to be synchronized [3, 10].
- Two Way Ranging: TWR-based algorithms are measuring the traveling time from a transmitter to a receiver and back. In this way no time synchronisation is needed between the devices, which leads to lower complexity and lower cost. As drawback of TWR, measurements of different devices need to be managed sequentially to avoid interference [10].

Multi-angulation (Figure 1b) calculates the position by forming triangles from receivers wherefrom the location is known. Angulation can be used in combination with Angle Of Arrival (AOA) algorithms [1, 3, 10].

AOA-based algorithms measure the angles to the reference points which leads to higher complexity compared to other algorithms and they are more sensitive for errors which can lead to inaccurate results [3, 10].

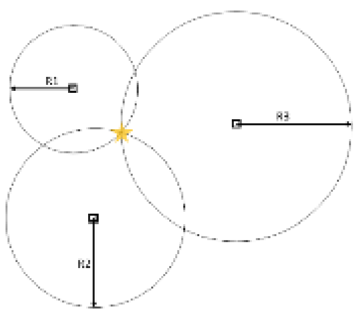


Figure 1a: Positioning by trilateration

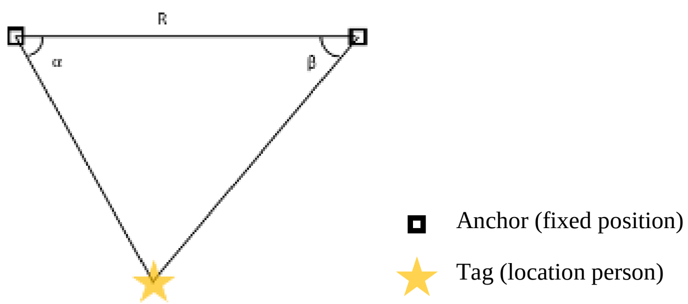


Figure 2b: Positioning by triangulation

Bluetooth-, Wi-Fi-, RFID-based localisation mostly rely on the RSS-algorithm which is more sensitive for obstructions such as internal walls [10] which leads to lower accuracy and is not high enough to get insights into the circulation within a room. Bluetooth-based systems mostly have very small and low-power tags or can be integrated into a smartphone [10, 20]. Wi-Fi-based systems has a wide coverage range between 50 and 100m and can use the existing network. RFID can also be used for proximity positioning where the position is equated to the

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

anchor point with the strongest signal strength. Hereby the accuracy depends on the density of the anchors and signal range. They can all localize a large number of occupants because no time management is needed. Nevertheless, RSS is less reliable for high accurate measurements in an indoor environment [3, 10].

Ultra-wideband (UWB) is a radio frequency technology for short range and high-bandwidth (larger than 500MHz) communication with less multipath interference and a signal that passes through walls and objects [3, 10]. UWB can be used for accurate indoor positioning by using different positioning methods, such as TOA, TDOA, TWR and RSS. The first three methods are more reliable for measuring distances in indoor environments because they measure time instead of signal strengths. UWB-based systems have also some drawbacks: They operate outside the licenced radio frequencies, causing that the power has to remain below 12dB (Europe) [3]. The cost of UWB system is mostly higher than similar systems which are using Bluetooth or Wi-Fi. However, UWB is a good technique for high accurate indoor localisation and has been used for this research to gain insights into the spatial use of residents within a room in a dwelling.

Comparison of commercially available ultra-wideband indoor positioning systems

System / ***Requirements***	*Level of occupancy sensing*	*Horizontal accuracy*	*Vertical accuracy*	*Coverage area*	*Output data*	*Number of users*	*Privacy*	*Size/weight of the tag*	*Required infrastructure*	*Installation time 4 anchors/1tag*	*Cost 4 anchors/1tag*	*Positioning algorithm*
Ubisense™	Tracking	15cm	15cm	Unlimited	Coordinates	>1000	Location	38x39mm, 25g (incl. battery)	Tags, anchors, timing cable	10’	High	AOA + TDOA
Eliko™	Tracking	30cm	30cm	scalable	Coordinates	10 (4Hz), 40 (1Hz)	Location	85x55x18mm, 54g (incl. battery)	Tags, anchors, master	30’	Medium (±€1800)	TWR
Open Rtls™	Tracking	30cm	30cm	Scalable	Coordinates	>1000	Location	75x50x17mm (incl. battery)	Tags, anchors	/	Medium	TDOA, TWR
Pozyx™	Tracking	10cm	30cm	Max 16 anchors	Coordinates, Free platform	Frequency/tags	Location	60x53 mm, 14g (excl. battery) 58x70x33mm (incl. battery, self-made)	Tags, anchors, master tag	30’	Low (±€600)	TWR

Table 3: Commercially available ultra-wideband positioning systems

Table 3 shows an overview of four different commercially available ultra-wideband based positioning systems which are compared according to the requirements which are discussed earlier. All four listed indoor localisation systems achieve the requirements, although they use different algorithms which have an influence on the number of tags and the required infrastructure.

Ubisense™ offers an indoor localisation system based on a combination of AOA and TDOA to calculate the position of a tag [21]. Therefore, the time between the different anchors has to be configured, which can be done by a timing cable between the anchors [22]. The system is able to cover a large area and can locate many tags because they can be measured individually and do not interfere with one another. The compact tag of Ubisense™ is the smallest compared to the other systems and has a long life battery, which has a positive effect on the intrusiveness for the residents. Eliko™ offers an indoor localisation system which is based on TWR, with the advantage that no time synchronisation is needed. Hence, multiple tags have to be located sequentially to prevent interference between the different tags. The tags of Eliko™ are larger with a rechargeable battery, but are still portable for users [23]. Open RTLS provides real-time location and can be synchronized for TDOA or TWR, in case of TDOA there is a wireless synchronisation. The tags of open RTLS™ are comparable with those from Eliko™. Pozyx™ is a commercially available system which offers an indoor localisation and motion sensing system (accelerometer, gyroscope, magnetometer, pressure sensor) which works according to TWR. Tags have to be managed by a master which also gathers and saves all the collected data. These tags are compatible with the open Arduino platform which make it able to collect raw data [24]. Pozyx™ offers a development kit without a case or battery and the kit cannot be handed over to the residents directly. The Pozyx™ system is used in the research project to monitor the circulation patterns of residents throughout the dwelling because of its open platform, development possibilities and its low cost. Practical implications and some preliminary results are shown in the next paragraph.

Spatial use monitoring in a dwelling using Pozyx™

To gain insights into the spatial use within rooms in Flemish dwellings and their seasonal variations, spatial use patterns, actions of residents and the indoor climate in the dwelling are monitored for nine consecutive days in each season. Pozyx™ is used as indoor localisation system. It consists of minimum four anchors (Figure 2_{left}) for three-dimensional positioning (lateration), one tag (Figure 2_{middle}) for each resident and one master (Figure 2_{right}) to control and readout the data of the tags carried by the residents. The master will send a signal sequentially to one of the tags carried by the residents. Where after, the resident's tag sends a signal with a time stamp to the anchors that send the signal back to the resident's tag (principle of two-way ranging). The resident's tag calculates its position and sends it to the master tag [24] which saves it on an SD-card. After this cycle, the master sends a signal to the next resident's tag, and this process is repeated continuously during the whole monitoring period when the residents are at home and awake. The number of anchors depends on the design and structure of the dwelling. Therefore, the accuracy has to be tested during each set up. Generally, four anchors are needed for each floor level but the presence of obstacles (walls) can result in the need for more anchors to obtain the same accuracy. The location of the anchors can be gathered by the anchors themselves or need to be measured manually and loaded into the program. Pozyx™ offers a development kit without standard battery, on-off switch or case. For carrying the tags, a case has been designed with the Pozyx tag, a 3700 mAh battery and a charger to charge the battery and to switch the tag on and off. According to the technical description, the tags should use 200mA which leads to a theoretical autonomy of 18,5 hours. Due to the conversion of 3.7V to 5V, a reduced autonomy of 12-14 hours is mostly obtained.

Figure 2: modules to locate multiple persons: anchor (left), tag (middle), mastertag (right)

One way to present the gathered data of the localisation system is as a heat map (Figure 3), which visualises the use intensity of places by a resident during the monitoring period. The figure shows that some areas are only partially used while others are not used at all. This heat map does not make a distinction between use patterns where the residents have been staying on one place or when they are just passing multiple times at a certain spot. To discover the difference between those, more research on 'stops', i.e. places where residents stay for a longer time, has to be done. Additionally, circulation patterns which visualize the effective walking line can be derived. Figure 4 shows the circulation pattern of one resident during a time frame of 23 minutes while coming home. The resident starts his tag at the door, passes the dining table, where after he goes to the kitchen, stay for a relatively long time at the extra work table and then goes upstairs for a couple of minutes. This figure shows where the residents has walked as well as where the resident has stayed for some minutes e.g. the work table and kitchen table.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

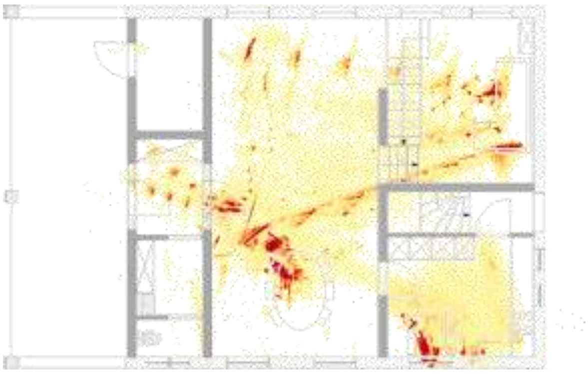


Figure 3: heat map of spatial use of one resident

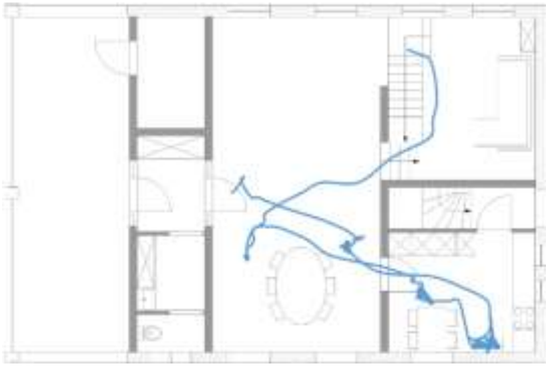


Figure 4:Circulation pattern of one resident during 23'

Conclusion

Many techniques can be used for occupancy sensing and can be used in a lot of research areas. However, for monitoring detailed spatial use patterns within rooms in dwellings, a more advanced monitoring technique is needed because most detection systems give only insights in the occupancy of a room as a whole. A literature review has shown that ultra-wideband has the most potential for accurate indoor positioning in dwellings. Characteristics of UWB, such as large bandwidth and short pulses, makes it able to penetrate obstructions, such as walls, more easily and therefore UWB is more suitable for complex indoor environments. UWB-based systems mostly use time measurements to calculate the location of a tag. When Two-Way Ranging is used, no time synchronisation is needed. Comparison of four ultra-wideband indoor positioning systems showed that they all meet the requirements and they all are highly accurate. For a research project on spatial use patterns in dwellings, the indoor localisation system Pozyx™ is used and some preliminary results are presented as a heat map and as a circulation pattern. In further research, a distinction between movements and stops will be made to get better insights into the areas where the residents are staying for a longer time instead of areas where residents pass multiple times.

Ethical statement

The case study has been submitted to the ethical committee of Hasselt University. Before the residents take part in the research, they are informed about the complete research process and about their right to stop their participation at any time. Each resident signs a document which guarantees that all the data will be processed confidentially and gives permission to use floorplans and pictures for the analysis and to publish results of the analysis.

References

1. Fuchs, C., Aschenbruck, N., Martini, P., and Wieneke, M. (2011) Indoor tracking for mission critical scenarios: A survey. *Pervasive and Mobile Computing*. **7**: 1-15.
2. Lymberopoulos, D., Bamis, A., Teixeira, T., and Savvides, A. (2008) BehaviorScope: Real-Time Remote Human Monitoring Using Sensor Networks 533-534
3. Alarifi, A., Al-Salman, A., Alsaleh, M., Alnafessah, A., Al-Hadhrami, S., Al-Ammar, M.A., and Al-Khalifa, H.S. (2016) Ultra Wideband Indoor Positioning Technologies: Analysis and Recent Advances. *Sensors*. **16**: 36.
4. Zou, H., Jiang, H., Yang, J., Xie, L., and Spanos, C. (2017) Non-intrusive occupancy sensing in commercial buildings. *Energy and Buildings*. **154**: 633-643.
5. Nguyen, T.A. and Aiello, M. (2013) Energy intelligent buildings based on user activity: A survey. *Energy and Buildings*. **56**: 244-257.
6. Calì, D., Matthes, P., Huchtemann, K., Streblow, R., and Müller, D. (2015) CO2 based occupancy detection algorithm: Experimental analysis and validation for office and residential buildings. *Building and Environment*. **86**: 39-49.
7. Dedesko, S., Stephens, B., Gilbert, J.A., and Siegel, J.A. (2015) Methods to assess human occupancy and occupant activity in hospital patient rooms. *Building and Environment*. **90**: 136-145.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

8. Haneda, K., Takada, J., and Kobayashi, T. (2005) On the Cluster Properties in UWB Spatio-Temporal Residential Measurement, National Institute of information and communications technology,Italy.
9. Van Loy, N., Bosserez, A., Verbeeck, G., and Knapen, E. (2018) Data collection methods for accurate spatial use within rooms *Windsor conference* (Windsor, 12 - 15 April),
10. Mautz, R. (2012) IndoorPositioning Technologies, Department of Civil, Environmental Geomatic Engineering, ETH Zurich,
11. Labeodan, T., Zeiler, W., Boxem, G., and Zhao, Y. (2015) Occupancy measurement in commercial office buildings for demand-driven control applications—A survey and detection system evaluation. *Energy and Buildings*. **93**: 303-314.
12. Ramos, T., Dedesko, S., Siegel, J.A., Gilbert, J.A., and Stephens, B. (2015) Spatial and temporal variations in indoor environmental conditions, human occupancy, and operational characteristics in a new hospital building. *PLoS One*. **10**: 24.
13. Orr, R.J. and Abowd, G.D. (2000) The smart floor: a mechanism for natural user identification and tracking *CHI '00 Extended Abstracts on Human Factors in Computing Systems* (The Hague, The Netherlands, 275-276
14. Carlorattiassociati Local Warming http://www.carloratti.com/project/local-warming/ Accessed 30 november 2016
15. Topouzi, M., Grunewald, P., Gershuny, J., and Harms, T. Everyday household practices and electricity use: early findings from a mixed-method approach to assessign demand flexibility *Behave 2016* (Coimbra, 8-9 september 2016), 15
16. Handcock, R.N., Swain, D.L., Bishop-Hurley, G.J., Patison, K.P., Wark, T., Valencia, P., Corke, P., and O'Neill, C.J. (2009) Monitoring Animal Behaviour and Environmental Interactions Using Wireless Sensor Networks, GPS Collars and Satellite Remote Sensing. *Sensors (Basel)*. **9**: 3586-603.
17. Carlorattiassociati Fondazione Agnelli http://www.carloratti.com/project/fondazione-agnelli/ Accessed 28 november 2016
18. Goga, N., Vasilateanu, A., Mihailescu, M., Guta, L., Bocicor, I., Bolea, L., and Stoi (2016) Evaluating indoor localization using WiFi for patient tracking *International Symposium on Fundamentals of Electrical Engineering* (Bucharest, 30 June- 2 July),
19. Cheng, H.T. and Zhuang, W. (2009) Bluetooth-Enabled In-Home Patient Monitoring System: Early Detection of Alzheimer's Disease. 15.
20. Blooloc https://www.blooloc.com/ Accessed 27 september 2017
21. Ubisense Ubisense https://ubisense.net/en/products/angleid
22. Ward, A. (2007) Ultra-wideband location system, User's manual.
23. Eliko Precise UWB-based real time location system https://www.eliko.ee/products/kio-rtls/ Accessed 18 Januari 2018
24. Pozyx Pozyx Accurate positioning https://www.pozyx.io/ Accessed 19 december 2017

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Self-confrontation with traces of one's activity: making meaning of technology-rich in-situ natural behavior

Mikael Wahlström[1], Deborah Forster[2] and Erwin Boer[3]

1. VTT Technical Research Centre of Finland; 2. Contextual Robotics Institute UC San Diego & Center for Equine Health UC Davis; 3. Delft University of Technology; mikael.wahlstrom@vtt.fi

With the explosion of technology platforms that track behavior, there is a growing emphasis on those platforms that feed-back to the 'user' (either immediately or after a delay) traces of their own activity. Confronting someone with video presentations of their own activity [1] emerged as an explicit autoconfrontation methodology in the ethnographic toolkit in the 1990s and continues to evolve [2]. Based on literature review and ongoing research and innovation projects, this oral presentation addresses two recent issues related to the application of self-confrontation methods: firstly, use of novel sensor networks, which are able to capture multiple modalities, and, secondly, the development of autonomous transportation systems through the use of this data that has been considered and used in self-confrontation sessions.

Based on the existing literature, two distinctive opportunities describe self-confrontation in contrast to usual recall-based data-collection methods: Firstly, the user is to be confronted with aspects of the activity that they do not remember or did not attend to at the time. The other opportunity is the so-called 'hyper recall' that sometimes ensues when aspects of activity are presented - users are able to recall and reflect on more than is recorded in the data. Third possibility, emphasized and discussed by us, is in presentation and dialogue regarding data on sensor-based modalities and regarding initial research findings made by the researchers. Self-confrontation facilities the reduction of the boundary between the study subject and researcher, that is, hypotheses on the metrics on the study subject's behavior can be generated and considered collaboratively.

Our past and current use cases come from two domains: surgical robotics [e.g, 4] as well as development of intelligent driver support [e.g., 3] and vehicle autonomy systems. We demonstrate this methodology in (1) the 'pre-design' phase, where it supports task analysis and the discovery phase of user-experience research, and (2) the 'validation / calibration of behavioral metrics' phase in simulated or prototype-testing environments - exposing mismatches between user-experience and automated or autonomous technology behavior. Recent work on autonomous vehicles (cars and ships) points to the value of this method in (3) the 'training/coaching and behavior change' phase that current ADAS (Advanced Driver Assistance Systems) design aspires to.

References:

1. Cahour, B., & Licoppe, C. (2010). Confrontations with traces of one's own activity. *Revue d'anthropologie des connaissances*, 4(2).
2. Clot, Y., Faïta, D., Fernandez, G., & Scheller, L. (2000). Entretiens en autoconfrontation croisée : une méthode en clinique de l'activité [Interviews in Crossed Self-confrontation : An Activity Clinic Method]. *Perspectives Interdisciplinaires sur le Travail et la Santé.* 2: 1–9.
3. Boer, E.R., Joyce, C.A., Forster, D., Chokshi, M., Mogilner, T., Garvey, E. and Hollan, J. (2005). Mining for meaning in driver's behavior: A tool for situated hypothesis generation and verification. In Noldus, L. P., Grieco, F., Loijens, L. W., & Zimmerman, P. H. (2005). Measuring behavior 2005. In *5th International Conference on Methods and Techniques in Behavioral Research, Wageningen, The Netherlands*.
4. Wahlström, M., Seppänen, L., Norros, L., Aaltonen, I., & Riikonen, J. (in press). Resilience through interpretive practice – A study of robotic surgery. Safety Science.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Mental Workload Measurement: A Maritime Simulation Study

R. Vreede and C.Guiking

Maritime Research Institute Netherlands (MARIN) Wageningen, The Netherlands. c.guiking@marin.nl

Introduction

There has been limited research on the workload of maritime operators, although the concept of mental workload (MW) is one of most widely used concepts in human factors. Understanding how MW influences performance is vital for system and maritime operation design, especially during the introduction of new technologies on board ships that potentially change the cognitive demands of operators.

The aim of this study was to investigate the applicability, sensitivity and selectivity of mental workload indices in a maritime simulator environment. This study focuses on measurement of mental workload of tugboat captains. Tugboat captains experience varying levels of mental workload, because of diversity in manoeuvres as well as changing environmental circumstances. Therefore, tugboat captains are an interesting group for studies on mental workload.

Study design

To assess the tug captain's mental workload two scenarios were designed. The first scenario is a racecourse, see Table 1, and the second scenario a hoisting operation, see Table 2. The racecourse consists of six common manoeuvres, which the tug captain had to complete under time pressure. Each tug captain performed the manoeuvres in the same order and had to sail to the next manoeuvre after finishing one. The racecourse was developed to induce low (sailing between the manoeuvres), medium and high mental workload (the manoeuvres) by changing the task load. These form three difficulty levels. For the development and classification of the manoeuvres, tugboat experts underwent and assessed the manoeuvres in a preliminary study. The second scenario is a maritime hoisting operation where the tugboat captain has to pick up personnel in a basket that was being lowered from a crane on an offshore platform. Two environmental conditions were created, the first condition without waves and wind (medium task load) and the second condition with increased waves and wind (high task load). The tug captains had to keep position in order to receive the basket with personnel. Sailing to and from the crane were seen as phases with low task load. The two environmental conditions were performed one after each other, including sailing back and forth to the starting point with and without waves and wind.

The scenarios were performed by the tugboat captains in a tug simulator at the Maritime Research Institute Netherlands (MARIN). MARIN is the largest independent maritime research institute in the world, with several advanced test facilities and simulators. The tug simulator included all relevant instruments, displays and controls to simulate a tug bridge. To this extent, a database was prepared including visual guidance for the human operator.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Table 1. Manoeuvre description of the racecourse scenario

Manoeuvre	Description	Expected task load
AB	Sailing A to B	Low
B	Zigzag	Medium
BC	Sailing B to C	Low
C	Moor at the quay frontally, depart backwards.	Medium
CD	Sailing C to D	Low
D	Pivot around the buoy.	Medium
DE	Sailing D to E	Low
E	Zigzag backwards	High
EF	Sailing E to F	Low
F	Zigzag backwards, smaller distance between buoys.	High
FG	Sailing F to G	Low
G	Navigate backwards into starting position, after the notification that starboard engine has failed.	High

Table 2. Chronological order of the hoisting scenario

Manoeuvre	Description	Expected task load
	From starting point towards the easy manoeuvre	Low
Hoisting in calm weather	Picking up personnel from the basket without any sea or swell	Medium
	Navigating back to starting position	Low
Hoisting in heavy weather	Picking up personnel from the basket with increased sea and swell	Hard
	Navigating back to starting position	Low

Participants

Twenty tugboat captains participated in this study. Two groups were created, a low experience group (about 5 years of experience) containing eight tugboat captains (μ = 37, SD = 9), and a high experience group (about 36 years of experience) with twelve tugboat captains (μ = 58, SD = 6). They all filled in an informed consent (IC) document. The research was approved by the ethical commission of the University of Twente in the Netherlands. All tug captains received financial compensation for their participation.

Materials and methods

Mental workload was measured using three categories of measurement instruments. Task performance (primary & secondary), subjective reports (e.g., surveys), and physiological measures.

Primary performance. The primary tasks performance criteria are unique in this study. Several parameters were selected for primary performance for specific manoeuvres. In the racecourse scenario zigzag manoeuvres in between buoys were conducted and an ideal line was calculated. This ideal racing line was calculated based on the derived track of the five fastest captains (without hitting the buoys). The cross track error (difference between ideal and realized track) served as primary task performance parameter. A lower standard deviation indicated a higher performance. For the hoisting scenario the mean distance to the basket with the personnel was calculated and served as primary task performance parameter.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Secondary task. The peripheral detection task (PDT) was used as secondary task in order to measure mental spare capacity as indication for the mental workload. The PDT device consists of a head mounted LED that emit a signal every 3-5 seconds to which the operator has to respond. The operator has to press the switch as quickly as possible, a reaction time of more than two seconds is deemed as a miss. The data was processed with R Studio, and averaged per manoeuvre for hit rate (%) and reaction time (ms). If visual demands of the primary task are high, PDT has proved to be a sensitive indicator of mental workload for car drivers and cyclists [1,2]. A more demanding primary task results in increased reaction time and more missed cues. The visual demands of a tugboat captain while manoeuvring in and near a port or while he is navigating close to an offshore platform are taxed heavily. The PDT (equipment shown in Figure 1) was therefore seen as a fitting measure of secondary performance.

Figure 1. PDT

The NASA-TLX and the Rating Scale Mental Effort (RSME) were selected as subjective self-assessment surveys. The NASA-TLX is a multidimensional self-assessment scale, which is used and has been validated in various domains, including the maritime industry [3,4]. An average of the subscales (mental demand, physical demand, temporal demand, performance, effort, and frustration) reflects an integrated measure of overall workload. The RSME is a unidimensional self-assessment survey recorded with by pen and paper during the run. The RSME contains a fifteen centimetre vertical line, is scored from 0 to 150, and has descriptions on non-equidistant intervals (i.e . Zijlstra [5]). The NASA-TLX was recorded after every run with pen and paper. An unweighted score was calculated. The RSME was administered in both scenarios.

Four physiological indices of mental workload were selected for this study:

1. Pupil dilation
2. Functional near infrared spectroscopy (FNIRS)
3. Electrodermal activity (EDA)
4. Electrocardiography (ECG).

Pupil dilation When the lighting of the environment stays constant, pupil size varies in a systematic manner with respect to physiological and psychological parameters [6]. This variation can be attributed to a change in cognitive demands [6]. A mobile eye tracker from SMI was used to measure pupil dilation. The glasses of the mobile eye tracking system had two small cameras mounted on the inside of the frame. These cameras recorded the pupil size. The collected data was post-processed using SMI's BeGaze software. From BeGaze, a text file was extracted that included a list of the pupil size as a function of time. The pupil dilation was collected with a frequency of 60 Hz. The pupil dilation data was averaged per manoeuvre.

Functional near-infrared spectroscopy (FNIRS). FNIRS utilizes certain wavelengths of light to measure the oxygenated and deoxygenated haemoglobin in the prefrontal cortex (PFC), which is related to the blood oxygen level dependent contrast, which is also applied in functional magnetic resolution imaging [7]. An increase in blood oxygen levels in the prefrontal cortex can be related to activation of the working memory and mental workload [6]. FNIRS is safe to use for an operator, relatively cheap, and simple to administer [8]. The FNIRS 100A-W hardware from Biopac was selected for this study. The data was collected using four light sensors. In between the sensors, a LED emitted light to assist the sensors while recording the signal. The sensors were attached just above the eyes of the operator. The wires that originate from the sensors led back to a recording network device which, in turn, had a wireless connection with a dedicated server. Before the collection of the data started, a baseline was established, which took no more than five seconds. After collecting the data it was post-processed. Three filters were applied to minimize the noise in the signal. The first filter minimized ambient noise that was left in the signal. The second filter was a finite impulse response (FIR) filter which minimized undesirable frequencies in the signal. The last filter minimized the amount of artefacts (e.g., unwanted movement of the sensor). After filtering, the average oxygenation could be calculated (de-oxygenated haemoglobin subtracted from oxygenated haemoglobin).

Electrodermal activity (EDA), which indicates the level of transpiration. The sweat glands are under control of the sympathetic autonomic nervous system (SANS). The electrodermal activity gives an indication of arousal and stress [9]. Mental workload is related to a decline in parasympathetic autonomic nervous system (PANS), and an activation of the (SANS) [10]. The EDA was measured with the Biopac EDA100C. This wireless device can be attached to a subject's wrist. Two electrodes are connected to the palm of the tugboat captain. The wireless transmitter sends the data to a dedicated server. The device was calibrated before the data collection started. The mean amplitude of the skin conductance response (SCR) prior the manoeuvre, was subtracted from the mean amplitude of the SCR measured during the manoeuvre.

Electrocardiography (ECG). Mental workload is related to arousal, and neural activity is preceded by an increase in metabolic demand, which might be the cause of an increased heart rate [11]. Heart rate variability (HRV) decreases with an increase of mental workload [11]. Heart activity with low frequencies (LF=0.05-0.15 Hz) reflects sympathetic activity with vagal modulation, whereas higher frequencies (HF=0.16-0.40 Hz) reflects parasympathetic activity [14]. An increase in the LF/HF ratio therefore might indicate an increase in workload of the tugboat captain [11]. The ECG data was collected using the Biopac ECG100C. A wireless device was attached to the abdomen of the tugboat captain with a Velcro band, this device administers the ECG and sends the data to a dedicated server. Three electrodes were placed on the tugboat captain. The first electrode on the right collarbone, the second electrode beneath the lowest rib on the right side, and the third electrode on the lowest rib on the left side. The electrodes were taped off to minimize the noise in the signal. The data was post processed with Biopac's AcqKnowledge software. Three parameters from the ECG were used for further analysis, namely the R-R peak interval (RRI), the ratio of LF HRV to HF HRV, and the root of the mean square difference of successive R-R intervals (RMSSD).

Procedure

Before the experiments, the participants filled in a demographic form. The experiments were explained to the tugboat captains. The physiological sensors were attached and subsequently the recording quality was checked. Tugboat captains were reminded of the objectives of the scenario. They familiarized themselves with the simulator, before performing the racecourse scenario. During each manoeuvre the RSME was conducted. Performance and physiological data were collected continuously during the run of the scenarios. At the end of the run the NASA-TLX was conducted. After completing the racecourse scenario the tugboat captains had a break. After the break the tugboat captains were interviewed about their experience with the simulator and the workload indices that were used. After the racecourse scenario the hoisting scenario was conducted according to the same procedure. Each experiment lasted approximately four hours for each tugboat captain.

Data analysis
Sensitivity of the workload indices was investigated with a repeated measures analysis of variance. The difficulty level functioned as within-subject factor, and tugboat experience (low and high) was the between-subjects factor. Simple contrasts were used. Post hoc tests were performed with the Bonferonni correction. The inter-correlations were calculated with Pearson's correlation coefficient.

Results
The instruments were tested for sensitivity with analysis of variance and selectivity by calculating the inter-correlations. Sensitivity testing for the primary performance parameters are described for their respective manoeuvres. The rest of the instruments were tested between the three difficulty levels (low, medium and high). Additionally, each specific manoeuvre (medium or high task load) was compared to low task load (set as baseline). A difficulty level is a grouping of manoeuvres with supposedly the same difficulty level. If the task load changes induced by the design of the scenario the indices should respond accordingly. Secondly, for the selectivity the inter-correlations were calculated with the data of the entire scenario (e.g., hoisting or racecourse), and per difficulty level within a scenario (e.g., low, medium, and hard difficulty) and per manoeuvre (B, C, D, E, F, & G). If there is a general mental workload response, the indices should correlate.

Data was missing because of malfunctioning or cumbersome software and hardware. One third of the data from the FNIRS, eye tracker, EDA, and ECG could not be retrieved. The collection of the data from the eye tracker could not be checked if it was functioning correctly during an experiment, which caused some of the missing data. The FNIRS, EDA, and ECG electrodes sometimes let loose, which caused additional loss of data.

Sensitivity results:
Primary performance. The frequency of the propeller use increased significantly from manoeuvre B to E, no other differences were found. Results for the hoisting scenario revealed no statistical significant difference for the used parameter: between the distance to the basket of the easy and hard phase in the hoisting operation.

Secondary performance. For the PDT hit rate significant results were found between each difficulty level in the racecourse scenario, where the higher difficulty level generates a lower hit rate. No significant differences were found for the PDT reaction time, or between-subjects and interaction effects. The PDT hit rate showed significant differences between the low task load and each manoeuvre , except for manoeuvre B. There were no statistically significant results for the PDT measurement in the hoisting operation.

Subjective reports. No statistically significant results were found for the NASA-TLX, for both scenarios. For the RSME statistical significant results were found between each difficulty level in the racecourse scenario, with a positive correlation between difficulty level and RMSE score. No between-subjects or interaction effects were found. Same results were found for manoeuvre level, except for the mean of manoeuvre C, which did not significantly differ from the low task load. Statistical significant results for the RSME in the hoisting operation were found, where a higher difficulty resulted in a higher score.

Physiological measures:

- No statistically significant results were found for the pupil dilation between each difficulty or manoeuvre and the low task load within the racecourse scenario. There were also no statistically significant results for each difficulty level in the hoisting scenario.
- For FNIRS measurements during the racecourse, the mean oxygen level of hard difficulty level significantly increased compared with medium difficulty level. No significant effects were found on the manoeuvre level. In the hoisting scenario a significant effect was found between the means of the low phase and medium phase, where the low phase showed more oxygenation. A significant reverse effect was found between the medium and hard phase of the hoisting scenario.
- The EDA measure did not show any significant results on difficulty level during the racecourse scenario, however the low task load showed significantly lower EDA compared to manoeuvre B and C. No significant results were found for the hoisting scenario.
- The RRI of the ECG showed significant results between the difficulty levels of the racecourse scenario, where an increase in difficulty resulted in an increase in RRI. Furthermore, manoeuvre E,

F, and G showed a higher RRI than B, C, and D, as expected. No significant results were found for the hoisting scenario.

- For RMSSD of the ECG no significant results were found on difficulty level for the racetrack scenario. On manoeuvre level, the RMSSD was significant lower for manoeuvre D compared to the baseline. No significant results were found for the hoisting operation.
- No significant results on difficulty and manoeuvre level were found for the LF/HF ratio of HRV for the racecourse and hoisting scenario.

Inter-correlations of mental workload indices: Tables for both scenarios of the inter-correlations for mental workload indices can be found in Appendix A. The top half of the tables are the correlations over the entire scenario, while the bottom half of the tables represents the three difficulty levels (low task load first, then medium and hard). A brief overview of the inter-correlations is described. Some indices are correlated, but the results are fractured.

For the racecourse scenario, PDT hit rate and PDT reaction time significant correlate during the whole racecourse scenario, between the difficulty levels and between four manoeuvres (low task load, B, D, G). The results show that an increase in reaction time relates to a decrease in hit rate. Of all the correlated indices, the PDT is often the common denominator. The PDT reaction time and hit rate correlates with oxygenation (on medium difficultly level and manoeuvre B). When oxygenation decreases, reaction times increase. The reaction time also correlates with RRI and RMSSD, both on the low (baseline) and medium difficulty level. Both indices have a negative correlation with reaction time, when RRI or RMSSD decreases, the reaction time increases. The PDT reaction time also has a positive correlation with EDA. Other indices such as EDA show a high negative correlation with pupil diameter (manoeuvre F). EDA also correlates negatively with RRI and RMSSD exclusive for high difficulty level. The LF/HF correlates negatively with RMSSD for all difficulty levels, but not for the individual manoeuvres. For the hoisting scenario, the PDT reaction time correlates negatively with pupil diameter, contrary to expectations. The PDT hit rate correlates negatively with RRI and RMSSD. The correlation between the hit rate and the RRI was according to expectations, the correlation between the hit rate and RMSSD contrary to expectations. All the significant correlations in the hoisting operation were found in the low task load level, the expected low task load.

Discussion

In this study several mental workload instruments were applied and combined with primary task performance criteria for specific scenarios. The instruments were tested for sensitivity, selectivity, and practicality. For the racecourse scenario, sensitivity was tested by comparing the means of three task difficulties (low, medium and high) within groups and between groups (two groups of experience, low and high). The same procedure was performed for the hoisting operation, for a high and medium task load condition (with and without wind and wave excitation on the tug). The selectivity was tested with analyses of the inter-correlations of the used workload indices. Finally, the practicality of the instruments were evaluated based on the entire procedure.

Sensitivity: Most workload indices proved not to be sensitive to the maritime scenarios as set-up in this study. Small effect sizes were found for all primary task performance parameters indicating that the task load differed are in line of the expected task load. The RSME was sensitive to the different difficulty levels, for both the racecourse and the hoisting scenario. The physiological instruments showed little sensitivity to the three manoeuvre difficulty levels of the racecourse scenario. The RRI of the ECG showed significant results between the difficulty levels of the racecourse scenario, where an increase in difficulty resulted in an increase in RRI, which can be explained by tugboat captains standing instead of sitting in a chair in the higher difficulty levels manoeuvres. Additionally, RMSSD was only significantly lower for manoeuvre D (Pivot around the buoy) compared to low difficulty level, indicating a higher mental workload. A higher HRV (indexed with RMSSD), is an indicator for lower mental workload. The oxygenation, pupil dilation, EDA & ECG lacked statistical power due to missing data and due a low number of participants in this study. In total 20 tugboat captains participated and were divided in two groups, 8 novices and 12 experts. The variance between and within the groups due to experience or physiological differences could have masked the effects on the used measurements. The racecourse manoeuvres all differed in design and therefore resulting in a high amount of factors that increased the variability. The higher

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

variability could have induced more diversity in the output of the used measurements and therefore missing the intended contrast of the manoeuvres.

Selectivity: Only the PDT and the ECG – heart rate showed significant results between the means of the three manoeuvre difficulty levels, which is in line with expectations. The selectivity of the workload indices were mixed with a low number of significant correlations. The lack of significant results on and inter-correlations may be partly explained by missing data or lack of contrast induced by the manoeuvre difficulty levels.

Practical points: The simulator data was logged using a dedicated server and the output files were analyzed using Matlab. The PDT was easy to use, but required significant post-processing in order to derive the actual hit rates and reaction times per manoeuvre. The switch of the PDT that had to be pressed to confirm that the tug captain observed the signal was operated with a switch on the tug captains left or right hand. This was sometimes hard to use, because the controls of the tugboat required both hands intensively.

The RSME was easy to administer. We conducted the RSME with pen and paper, but an app could enhance the practicality of administration. We choose to not create a weighted NASA-TLX, in order to save time during the evaluation procedure after the manoeuvres. In hindsight, it might have been better to use a weighted NASA-TLX, because of the differences in the scenarios. The racecourse scenario came with the instruction to complete the course as quickly as possible, while the hoisting scenario did not. The ECG & EDA apparatus was easy to use and analyze, but a disadvantage of the EDA was that it had to be measured on the palm of the hand, which coincidentally just like the PDT, could become cumbersome for the operator. A point of concern with the ECG was that the signal was noisy (with artefacts) for tugboat captains with a higher body mass when they moved. The artefacts were minimized by taping off the electrodes, in order to prevent movement of the wiring. The setup of the fNIRS apparatus took approximately five minutes. This is short compared to setting up an electroencephalography, which can take up to an hour to set up. There were some problems with environmental light causing noise in the signal. The software that was used to log the data (CobiSTUDIO) did not give any warnings if the data collection did not start properly.
The software for the post processing part (Fnirsoft) worked satisfactory. The eye-tracker system did not allow to continuously check if the data was collected. The mobile phone on which the software for the data collection ran, became increasingly hot after prolonged use in time. Batteries of the mobile phone had to be changed about every hour, which was impractical because some of the experiments lasted longer than an hour. The post-processing of the data required significant processing power, and took over 48 hours of continuous processing. Aside from those points, the software worked in accordance with expectations.

Conclusions:
Most importantly, the lack of results on sensitivity and selectivity could be the result of analysing data on manoeuvre level and not within a manoeuvre on a continuous timescale. By using means of all the indices on manoeuvre level, valuable data could have been missed. The results suggest that the used manoeuvres are not providing sufficient contrast. However the results of the RSME did reflect the predicted contrast, namely a higher RMSE score for the higher task load manoeuvres. The hoisting procedures did not provide the intended contrast of the two conditions (medium and high difficulty), which could be attributed to an inadequate procedure. The time for keeping the tugboat in position differed and therefore changing the conditions for each tugboat captain and in consequence changing the difficulty. The lack of significant outcomes could also result from a low number of participants. The PDT and RSME proved to be most valuable in this study, these instruments are reliable and relatively easy to use and analyse. The ECG – RMSSD could be useful to measure suppression of the PANS and therefore indicating a higher mental workload. Other measurements methods proved not to be useful to measure mental workload in this study. For this study the data was analysed with parameter means over manoeuvres, however for an inspection of effects within manoeuvres it is advised to analyse the data on a continuous time scale.

Acknowledgements
We would like to thank Noldus (Patrick Zimmerman, Tobias Heffelaar, and Anouk den Hamer) for helping us setting up the experiments and analysing the data.

References

1. Vlakveld W. P., Twisk, D., Christoph, M., Boele, M., Sikkema, R., Remy, R., & Schwab, A. L. (2015). Speed choice and mental workload of elderly cyclists on e-bikes in simple and complex traffic situations: A field experiment. *Accident Analysis & Prevention, 74,* 97-106. doi: 10.1016/j.aap.2014.10.018
2. Martens, M. H. & Van Winsum, W. (2000). Measuring distraction: The peripheral detection task. *Proceedings NHTSA: Internet Forum on the safety impact of driver distraction when using in-vehicle technologies*
3. Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 239-250). Amsterdam: Noord-Holland.
4. Rauffet, P., Chauvin, C., Nistico, C., & Judas. (2016). Analysis of submarine steering: effects of cognitive and perceptual–motor requirements on the mental workload and performance of helmsmen. *Cognition, Technology & Work, 18,* 657-672. doi: 10.1007/s10111-016-0384-0
5. Zijlstra, F.R.H. (1993). Efficiency in work behavior: A design approach for modern tools (Doctoral dissertation). Retrieved from Repository TU Delft (90-6275-918-1)
6. Rodriguez-Paras, C., Yang, S., & Ferris, T. K. (2016). Using pupillometry to indicate the cognitive redline. *Proceedings of the Human Factors and Ergonomics Society, USA, 60,* 685-685. doi: 10.1177/1541931213601157
7. Sato, H., Yahata, N., Funane, T., Takizawa, R., Katura, T., & Atsumori, H. (2013). A NIRS-fMRI investigation of prefrontal cortex activity during a working memory task. *Neuroimage 83,* 158–173. doi: 10.1016/j.neuroimage.2013.06.043
8. Ferrari, M. & Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *Neuroimage 63,* 921-935. doi: 10.1016/j.neuroimage.2012.03.049
9. Roth, W. T. (1983). A Comparison of P300 and Skin Conductance Response. *Advances in Psychology, 10,* 177-199. doi: 10.1016/S0166-4115(08)62039-7
10. Gawron, V. J., Schiflett, S. G., & Miller, J. C. (1989). Measures of in-flight workload. *Aviation Psychology,* 240-287
11. Hogervorst, M. A., Brouwer, A. –M., & Van Erp, J. B. F. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in Neuroscience, 8,* 1-14. doi: 10.3389/fnins.2014.00322

Analysis of monkey group behavior reveals "cheating"

Fan Xu

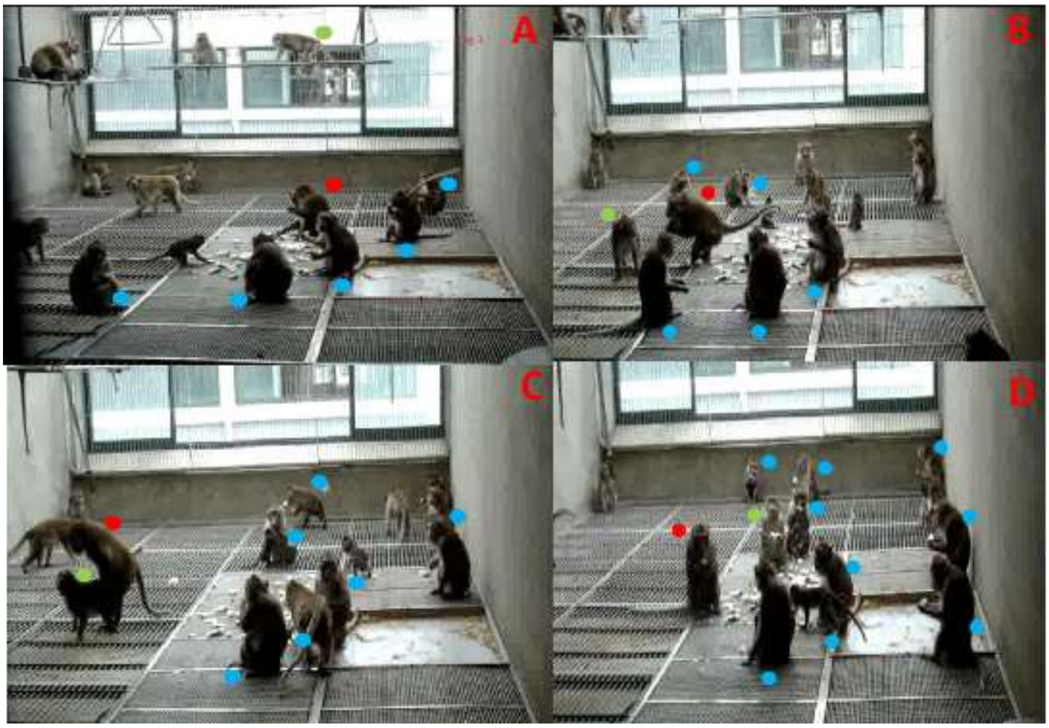
Chengdu Medical College, China

The law of the jungle plays a crucial role for subject survival. Food-taking sequences are dominated by social hierarchies based on relativly stable social relationships in a monkey colony. Shively reported that the social hierarchy is linearly in non-human primate communities. Namely, the stronger individuals, who can protect the group members, infants and weak ones, will enjoy the priority to take the food. Therefore, the food taking sequence is a good indicator to reflect the subject's social rank in a house feeding community. However, with traditional observation methodology, it is hard to record the food taking sequence in a group.
Objective: Identification and quantification of "cheating" behavior in macaque social hierarchy

Methods: We employed the focal observation method in a monkey group and recorded the video of group behaviors via high solution SONY DV in the Sichuan Meishan Monkey Base (License: SC 2012-0038) from Nov 2016 to May 2017.With these recordings, we then used the Noldus 10.0 XT to analyze the behaviors of individual monkeys in the group. Specifically, we marked the food take sequence of each subject manually.

Results: Food taking sequences are stable in a population. An interesting finding was that the subject adopts smart strategies to jump to the previously stable feeding sequence via a sexual exchange with the dominant monkey. As a result, the individual takes food as higher-class members do. Smart "cheating" helps the weaker subjects to survive, thus circumventing the "law of the jungle".

To understand the social roles played by individual subjects in the social hierarchy, it is necessary to distinguish their rank, specifically during the feeding period, i.e., which animal presented more submission and more aggressive behaviors respectively, as we have described in our previously research.



The dominant male monkey and higher ranking ones were had priority in taking the season fruit melon (see Fig A, upper left) with red and blue cycle marker. During this sequence, the target subject climbed down from the window and walked around the food disk (see Fig B with green cycle marker). This target subject seems to make eye contact with the dominant male monkey. As an agreement was approached, the dominant male monkey walked to the target subject and copulated with her (see Fig C red and green cycle marker). The target subject made a sharp call, as if she were saying "It's my turn" when copulation terminated (see Fig D). After this occurred, the target subject walked to the food disk and enjoy the food with higher rank ones peacefully.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

The Movement Ecology of Seabirds revealed with Unsupervised Behavioural Clustering

Zuzana Zajková [1, 2], José Manuel Reyes-González [1], Jacob González-Solís [1] and Frederic Bartumeus[2,3,4]

1 Institut de Recerca de la Biodiversitat (IRBio) and Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals (BEECA), Universitat de Barcelona, Barcelona, Spain. zuzulaz@gmail.com, josemanuel.delosreyes@gmail.com, jgsolis@ub.edu; 2 Centre d'Estudis Avançats de Blanes (CEAB-CSIC), Blanes, Spain. fbartu@ceab.csic.es; 3 Centre de Recerca Ecològica i Aplicacions Forestals (CREAF), Cerdanyola del Vallès, Spain; 4 Institució Català de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

Movement and behaviour of wild animals is attracting an increasing attention, from ecology and evolution to conservation and management fields. Technological advances in the field of *biologging* have led to an increase in types of devices to remotely monitor movement and behaviour of wild animals (e.g. GPS loggers, accelerometers, depth recorders, etc.) [1]. These devices can automatically record huge amounts of data over long periods and provide essential data to understand how the underlying motor system organization relates to the environment, a major goal in movement ecology. However, subsequent behavioural annotation of time-series usually requires an expert supervision, which is often overwhelming and may also bring some subjectivity to the analysis. To overcome these limitations, we can currently apply different automated machine learning algorithms on single- or multi-sensor data [2]. Automated approaches to behavioural annotation are particularly useful in organisms such as pelagic seabirds, which are highly mobile marine top-predators out of the human's sight for most of their live. To study their movement ecology and behaviour, geolocation-immersion loggers have been widely used. After 20 years since its advent, these loggers still remain as the most cost-effective tracking method and the unique sensors capable to record annual movements of pelagic seabirds while ensuring the welfare of tagged individuals. In addition to the positional data (2 low-resolution spatial locations per day), geolocation-immersion loggers also provide high-resolution (3s) wet-dry conductivity data, from which we can potentially infer year-round behaviour of seabirds at sea. However, this data has been typically used by calculating aggregated activity budgets (i.e. through averaging per unit of time -hour, day, month, etc.-), disregarding potentially valuable information contained within its inherent temporal structure that can be used for behavioural annotation. Either alone or combined with GPS devices, the structure of wet-dry temporal correlations (i.e. landings, take-offs, sit-and-wait foraging, drifting on water) of the immersion data can be directly used for behavioural annotation or to improve GPS-driven behavioural annotation. Here we present a new methodological approach where high-resolution wet-dry data is used as the key segmentation variable to identify behaviours. The approach is grounded on the t-Distributed Stochastic Neighbour Embedding (t-SNE) [3], a machine learning algorithm recently used for unsupervised behavioural clustering [4][5][6], but adapted to a new big data protocol (Garriga et al. in prep). As a proof of concept, we applied our approach on a small dataset, composed of 20 long foraging trips (duration ranged from 7 to 17 days) of a highly mobile pelagic seabird species, the Cory's shearwater, tracked simultaneously with 5 min resolution GPS loggers and high-resolution geolocation-immersion loggers. We applied a breakpoint algorithm on a cumulative sum of binary wet-dry (immersion) data series (wet 1, dry -1) to segment the trajectory and obtained various activity metrics, together with average velocities, turns and directional persistence at the segment level from GPS data. We ended up with high dimensional dataset involving 11 variables and 1953 data points. This analysis allowed us to identify elementary behavioural modes (movement clusters) (Fig. 1). We explored the sequences and structure of the behavioural clusters under different contexts, and related them to changes in environmental conditions in order to validate and interpret each cluster. Our classification can be used to dissect changes in movement patterns underlying different environmental contexts. We also show that geolocation-immersion loggers alone (without any GPS data association) collect relevant enough information to depict most of the behavioural landscape and clustering. This result may provide the required step for extending segmentation analyses and unsupervised clustering to year-round trips, opening new avenues to understand behavioural patterns over the entire annual cycle of marine species spending most of their lives out of the human's sight.

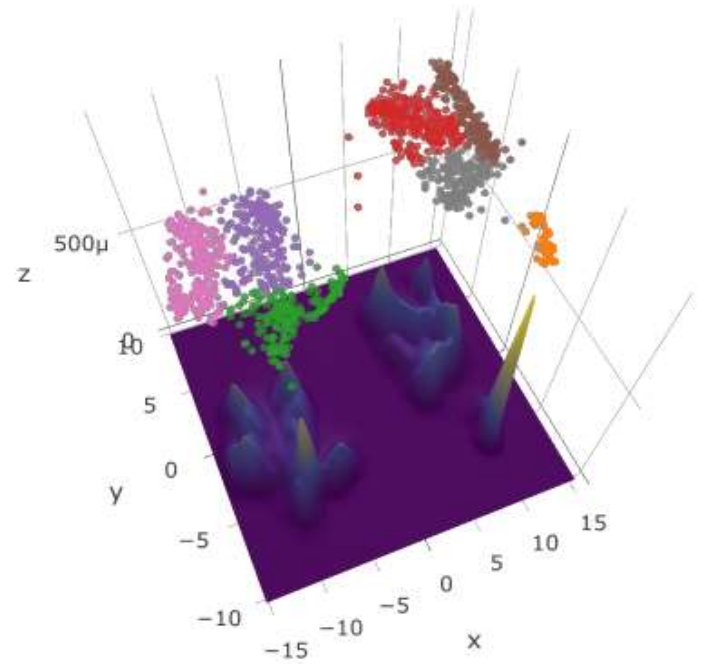


Figure 1: Movement behavioural landscape of Cory's Shearwater. A multi-dimensional dataset which consisted of 11 wet-dry derived metrics and 1953 data segments extracted from 20 foraging trips was used as an input to an adapted t-SNE algorithm. The bottom layer represents the estimated probability density function of all segments reduced to two dimensions (x, y) through t-SNE. The top layer shows behavioural clusters resulting from unsupervised clustering method (each point represents one segment and behavioural clusters are represented by a different colours).

Ethical statement

Model species for this study was Cory's Shearwater *Calonectris borealis*. All fieldwork was conducted in the Southwest of Gran Canaria, Canary Islands, Spain (27° 50' 29" N, 15° 47' 29" W) during July 2011 under licence to carry out research with this species approved by the regional committee for scientific capture (Consejería de Medio Ambiente del Cabildo de Gran Canaria, Canary Is.) and appropriate ringing license from Oficina de Especies Migratorias (Ministry for Agriculture, Fishing and the Environment, Spain). All animals were handled in strict accordance with guidelines for ethical research and good animal practices as defined by the current European legislation and the University of Barcelona Local Ethical Review Procedures. We selected 40 potential nests in the incubating period, from which we used 20 for deployment of loggers on parental adults. We equipped incubating birds with a GPS logger (Perthold Engineering, Germany) weighting ~ 20 g, previously adapted for waterproof with heatshrink tube. These loggers were attached on the birds back with Tesa© tape as commonly done in seabird at-sea tracking research projects (i.e. [7]). We also used a 2.4 g light logger geolocator (Biotrack Ltd ©) attached with a cable tie to a PVC ring which was deployed on the tarsus of the bird. Body mass of tagged birds ranged from 600 to 900 g so both devices amounted to 2.5 - 3.7 % of the birds' weight (below the deleterious recommended threshold of 3-5%; [8][9][10]). Birds were caught during changeover of the incubation bout to avoid time the egg being unattended. The data stored in both GPS and geolocator loggers cannot be broadcast, and thus the unit must be recovered to retrieve the data. The handling process took less than 15 min for each deployment/retrieval to minimize disturbance, and birds were immediately released to the entrance of their nest. Any bird was tagged for no more than 17 days and no more than two foraging trips. We measured hatching success and adult weight after recovery from nests where devices were deployed (N=25) and compared to other nests not used for tracking (N=15), and no differences were found, suggesting there was no significant short-term effects on tagged birds.

References

1. Ropert-Coudert, Y., Wilson, R. P. (2005). Trends and perspectives in animal-attached remote sensing. *Front Ecol Environ.* **3(8)**, 437–444.
2. Chimienti, M., Cornulier, T., Owen, E., Bolton, M., Davies, I. M., Travis, J. M., Scott, B. E. (2016). The use of an unsupervised learning approach for characterizing latent behaviors in accelerometer data. *Ecol Evol*, **6**, 727-741.
3. van der Maaten, L., Hinton, G. (2008). Visualizing High-Dimensional Data Using t-SNE. *J. Mach. Learn. Res*. **9**, 2579-2605.
4. Berman, G. J., Choi, D. M., Bialek, W., Shaevitz, J. W. (2014). Mapping the stereotyped behaviour of freely moving fruit flies. *J. R. Soc. Interface* **11**, 20140672.
5. Bartumeus, F., Campos, D., Ryu, W. S., Lloret-Cabot, R., Méndez, V., Catalan, J. (2016). Foraging success under uncertainty: Search tradeoffs and optimal space use. *Ecol Lett.* **11**, 1299-1313.
6. Marques, J. C., Lackner, S., Félix, R., Orger, M. B. (2018). Structure of the Zebrafish Locomotor Repertoire Revealed with Unsupervised Behavioral Clustering. *Current Biology* **28**, 181-195.
7. Guilford, T. C., Meade, J., Freeman, R., Biro, D., Evans, T., Bonadonna, F., Boyle, D., Roberts, S., Perrins, C. M. (2008). GPS tracking of the foraging movements of Manx Shearwaters Puffinus puffinus breeding on Skomer Island, Wales. *Ibis*, **150**: 462–473.
8. Phillips, R. A., Xavier, J. C., Croxall, J.P. (2003). Effects of satellite transmitters on albatrosses and petrels *The Auk*, **120**, 1082-1090.
9. Igual, J. M., Forero, M.G., Tavecchia, G., Gonzalez-Solis, J., Martinez-Abrain, A., Hobson, K., Ruiz, W., Oro, D. (2005). Short-term effects of data-loggers on Cory's shearwater (*Calonectris diomedea*). *Mar. Biol*., **146**, 619-624.
10. Passos, C., Navarro, J., Giudici, A., González-Solís, J. (2010). Effects of extra mass on the pelagic behavior of a seabird. *The Auk*, **127(1)**, 100-107.

Posters

All abstracts are arranged for the posters below alphabetically, by the first author's surname

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Breed differences and heritability of cat behavior in a large questionnaire sample

M. K. Ahola[1234], K. Vapalahti[123], K. Tiira[123] and H. Lohi[123]

1 Research Programs Unit, Molecular Neurology, University of Helsinki, Helsinki, Finland; 2 Department of Veterinary Biosciences, University of Helsinki, Helsinki, Finland; 3 The Folkhälsan Institute of Genetics, Helsinki, Finland; 4 Section of Ecology, Department of Biology, University of Turku, FI-20014 Turku, Finland

The large variance in appearance in domestic cats (*Felis catus*) seen today is the result of thousands of years of domestication and dozens of years of intensive breeding [1]. This breeding has led to nearly a hundred recognized breeds, each having different characteristics distinguishing them from the other breeds [2]. Besides morphological differences, cat owners and breeders also report differences in behavior between the breeds. However, breed differences of behavior have been little studied [but see 3-6] and the heritability of feline behavioral traits remains unknown.

To examine the behavioral differences between cat breeds, we used an online feline health and behavior questionnaire directed to Finnish cat owners and collected a large (N=5726) survey data of home-living cats. The questionnaire was advertized on our web page (www.kissangeenit.fi) and on our Facebook page, and the questionnaire was also advertized by breed clubs and breeders. Our questionnaire measured nine different behavioral traits: shyness towards strangers; shyness towards novel objects; activity level; contact with people; aggressiveness towards strangers; aggressiveness towards family members; aggressiveness towards other cats; stereotypic wool-sucking; and stereotypic excessive grooming. Owners could evaluate their cat's behavior with a 5-point Likert-type scale, ranging from 'not at all' to 'very much', except for wool-sucking in which the answering options were 'never', '1-3 times in the cat's lifetime', '1-12 times per year', '1-4 times per month', '1-3 times per week', 'daily', 'many times per day', and 'most of the day'. Furthermore, we asked whether the owners think that their cat has a behavior problem. The owners could answer 'I don't know', 'no', 'yes, self-diagnosed', or 'yes, diagnosed by a veterinarian'. These behavioral traits have been studied before [3-7] and thus were of interest to this present study. Furthermore, wool-sucking has been suggested to be more common in Siamese cats [8] but this has been little studied. The questionnaire also included several background questions, such as breed, sex, age, and weaning age of the cat.

We analyzed the behavioral differences between cat breeds with multivariate logistic regression models. The best models that would include all the important environmental explanatory variables were selected using forward stepwise AIC model selection approach. By using this, we could control the effect of environment (e.g. weaning age) on behavior. We adjusted all P-values using false discovery rate. In each trait, we grouped cat breeds into clusters in which breeds are very similar to each other (P-value > 0.5) but differ from breeds in other clusters (P-value < 0.05). We calculated the heritability of behavior for two common cat breeds with an animal model based on a Markov chain Monte Carlo for generalized mixed models (MCMCglmm in R). We also included sex and owner of the cat in the analyses when they were significant predictors of behavior.

Our results showed that cat breeds differ in all studied behavioral traits. The magnitude of difference between breeds differed between the studied behavioral traits, with the highest differences seen in activity level and lowest in stereotypic behavior traits. As we controlled for the environmental effects in the analyses, these results indicated that behavior is partially genetically inherited. The heritability analyses agreed with this suggestion, as behavioral traits showed low to moderate heritabilities. However, due to the small sample size in heritability analyses, the confidence intervals are large and thus this requires further research. Our study indicates that cat behavior is heritable and genetic inheritance is the cause of breed differences of behavior. These results could hence be used as a tool for cat breeding.

References

1. Driscoll, C. a et al. (2007). The Near Eastern origin of cat domestication. *Science* **317**, 519–523.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

2. The International Cat Association. (n.d.). Recognized cat breeds. <https://www.tica.org/en/cat-breeds>. (Accessed: 12th April 2018)
3. Turner, D. C. (2000). Human–cat interactions: relationships with, and breed differences between, non-pedigree, Persian and Siamese cats. in *Companion Animals and Us: Exploring the relationships between people and pets* (eds. Podberscek, A. L., Paul, E. S. & Serpell, J. A.) 257–271. Cambridge University Press.
4. Hart, B. L. & Hart, L. A. (2013). *Your ideal cat: insights into breed and gender differences in cat behavior*. Purdue University Press.
5. Takeuchi, Y. & Mori, Y. (2009). Behavioral profiles of feline breeds in Japan. *J. Vet. Med. Sci.* **71**, 1053–1057.
6. Duffy, D. L., de Moura, R. T. D., & Serpell, J. A. (2017). Development and evaluation of the Fe-BARQ: A new survey instrument for measuring behavior in domestic cats (*Felis s. catus*). *Behavioural processes* **141**, 329-341.
7. Feaver, J., Mendl, M., & Bateson, P. (1986). A method for rating the individual distinctiveness of domestic cats. *Animal Behaviour* **34**, 1016-1025.
8. Bradshaw, J. W., Neville, P. F., & Sawyer, D. (1997). Factors affecting pica in the domestic cat. *Applied Animal Behaviour Science* **52**, 373-379.

Spatiotemporal navigation of swimming zebrafish larvae under water flow

Amir Asgharsharghi Bonab and Hernán López-Schier

Sensory Biology&Organogenesis, Helmholtz Zentrum München and Technische Universität München, Freising-Weihenstephan; asgharsharghi@gmail.com

Rheotaxis is an innate behavior in which animal orients to the direction of water current. Although visual signals are the most salient sensory cues, Zebrafish larvae is able to perform this behavior in the absence of visual information. We used behavioral data from controlled environment to reveal mechanosensory control of rheotaxis in larval zebrafish. Our data suggests that larval zebrafish performs rheotaxis by exploiting inhomogeneities in the flow, the most salient of which is radial differences in flow velocity. However, we note that radial velocity gradients are insufficient to identify flow direction because the lateral line in the zebrafish larva is a linear sensor-array that is bilateral-symmetric and whose axis of sensitivity is bidirectional. Our data reveal that during rheotaxis larval zebrafish use iterative patterns of yaw movements and discrete swim bouts, during which they capture low-dimensional snapshots of mechanical cues across the horizontal plane.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Eco-HAB - an automated and ecologically relevant system for assessment of social behaviors in mice.

Pawel M. Boguszewski[1], Alicja Puścian[2], Maciej Winiarski[1], Joanna Borowska[1], Joanna Jedrzejewska-Szmek[1] and Ewelina Knapska[1]

1. Nencki Institute of Experimental Biology of Polish Academy of Sciences; 2. Yale University School of Medicine; pmbogusz@gmail.com

Eco-HAB is an open source, RFID-based (radio-frequency identification) system for automated measurement and analysis of social preference and in-cohort sociability in mice. It was designed to overcome two major problems with social tests in rodents - irreproducibility (poor within- and cross-laboratory standardization and reproducibility) and high manpower costs of manual testing. Irreproducibility of conventional behavioral tests has been identified as one of the most important threats to science and its public understanding. I case of social behavior, even small changes in experimental conditions can produce significant modifications of the behavioral outcome. In Eco-HAB, animals live in ethologically relevant mouse group sizes in a spacious apparatus with shadowed areas and narrow tunnels, resembling natural burrows.

The Eco-HAB consists of four housing compartments, occupying four corners of a larger square, bridged by tube-shaped corridors. These corridors enable mice (cohorts of up to 12 subjects) to travel freely and select preferred areas within the available territory. Two chambers on opposite corners of the square offer access to food and water (ad libitum), and provide shelter and secluded places where mice can sleep and rest. The two other compartments have similar designs, except that they contain no food or water and one of the corners is equipped with an impassable, transparent, and perforated partition behind which an olfactory stimulus may be presented. Animals are tracked individually by subcutaneously injected microtransponders (125kHz MILFARE glass-covered transponders 9.5 mm length, 2.2 mm diameter, RFIP Ltd) that emit a unique 14 digits identification code when mice pass under RFID antennas placed on both ends of each corridor. Eco-HAB.rfid software is able to simultaneously receive and process transponder codes from up to eight RFID antennas. It is written in Delphi programming language and compiled with Borland Delphi 7.0 (Embarcadero Technologies) using ComPort Library ver. 4.11. Along with the behavioral assessment system, we designed and created a software package for analyzing data collected by Eco-HAB.rfid software written in the Python programming language (Python 2.7 with NumPy and SciPy libraries). It consists of functions for loading and merging raw data previously stored on the hard drive as well as for converting this data into a series of visits (referred to here as sessions) of each mouse to each of the four Eco-HAB compartments.

As the Eco-HAB is a computer-controlled system, it eliminates human handling and allows for continuous data collection lasting for days or even weeks with minimal presence of human observers. Since the position of each mouse in an Eco-HAB system can be tracked, a novel in-cohort measure of sociability, based on the tendency of mice for spending time together, can be assessed. Compared to manual tests of sociability, our system provides more reliable data, faster, and with less manpower for several key behavioral measures [1].

Key benefits:

- The system closely follows murine ethology and natural social behavior.
- Experimental apparatus closely resembles natural burrows.
- Requires no contact between a human experimenter and tested animals.
- Results are obtained faster, with less manpower, and without confounding factors.
- The developed protocols and standardized behavioral measures demonstrate high reproducibility.
- The Eco-HAB design is easily scalable, flexible and adaptable for species other than mice.
- It is a low-cost, open system where design and software is freely available www.eco-hab.com

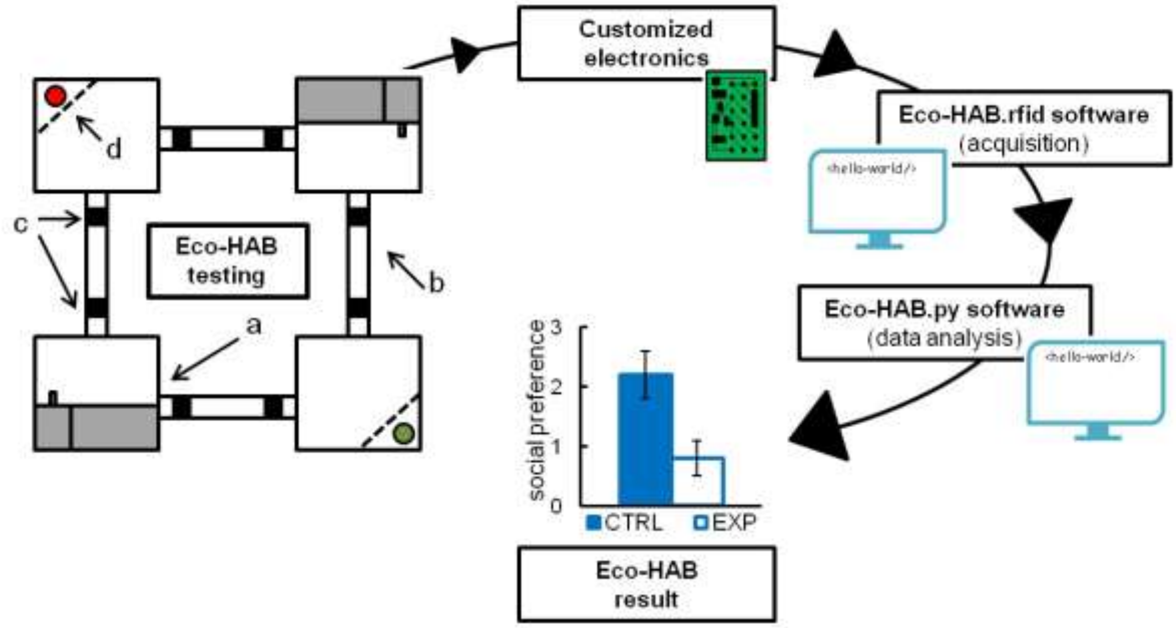


Figure 1. A schematic representation of Eco-HAB system and data processing. Eco-HAB consists of four housing compartments (a), tube-shaped inter-territorial passages (b), radio-frequency identification antennas (c), and impassable, perforated partitions behind which social and non-social (control) provenance stimuli may be presented (d, red/green dots). Food and water is available in housing compartments adjacent to those containing partitions. Eco-HAB is equipped with customized electronics and two software packages: Eco-HAB.rfid (for data acquisition and collection) and Eco-HAB.py (for filtering corrupted data segments and performing tailored analysis).

References

2. Puścian, A, Łęski, S, Kasprowicz, G, Winiarski, M, Borowska, J, Nikolaev, T, Boguszewski, PM, Lipp, H-P, Knapska, E (2016). Eco-HAB as a fully automated and ecologically relevant assessment of social impairments in mouse models of autism. *eLife* **5**.

CatWalk ™ XT Gait Analysis for Early Detection of Motor Symptoms in the 6-OHDA Rat Model of Parkinson's Disease

Jordi Boix-I-Coll[1,2], Daniela von Hieber[1] and Bronwen Connor[1]

1. Dept. of Pharmacology & Clinical Pharma; NeuroDiscovery Behavioural Unit, University of Auckland, New Zealand; j.boix-i-coll@auckland.ac.nz

Introduction

In recent years, computer-supported gait analysis has proven to be effective for the comprehensive assessment of gait changes in rodent models for a range of neurodegenerative and neurological disorders [1], [2]. However, proper validation and characterization of individual gait parameters is required for each specific neurological or neurodegenerative disorder. Parkinson's disease (PD) is a neurodegenerative disorder with progressive loss of dopaminergic (DA) neurons in the substantia nigra pars compacts (SNc) [3]. The degeneration of nigrostriatal innervation of the striatum and the subsequent reduced levels of DA in this area are responsible for the manifestation of the characteristic motor symptoms of PD [4]. The 6-hydroxydopamine (6-OHDA) PD model, is produced by a unilateral intracerebral injection of the neurotoxin 6-OHDA and therefore dopaminergic cell death occurs in the midbrain on one side of the brain only and asymmetric motor symptoms can be elicited [5]. The severity of symptoms in the 6-OHDA model depends on the site of injection and the dose of 6-OHDA used. In the 'full lesion' model or medial forebrain bundle (MFB) model, injections of the toxin into the MFB cause an acute and near complete loss of DA neurons in the SNc with significant loss of striatal DA content [6]. However, if the toxin is injected into the striatum, it reaches the SNc via retrograde transport along dopaminergic fibres and progressive degeneration of dopaminergic neurons can be observed with an overall milder pathology, known as the 'partial lesion' or 'striatal lesion' model [7]. With these established methods, it is often necessary to wait until the full extent of DA loss has occurred to observe a functional impairment. It is therefore advantageous to be able to identify reliable signs of functional impairment as early after lesioning as possible to allow for the assessment of potential therapeutic agents. Notably, gait disturbances have been reported to be the most common motor problems in PD patients and present themselves as the most constraining set of symptoms experienced by patients, finally depriving patients from most activities of normal daily living [8]. The aim of the present study therefore is to describe a proper methodological system to clearly identify which of the numerous gait analysis parameters are applicable to the characterization of gait disturbances in early stages of the 6-OHDA lesion rat models of PD.

Material and methods

Subjects

Forty-seven adult male Wistar rats (starting weight 250-300 g, 8-11 weeks old; University of Auckland Vernon Jansen Animal Resources Unit, Auckland, New Zealand) were used for this experiment. All the experimental procedures were in compliance with the NZ Animal Welfare Act (1999) and in accordance with the University of Auckland animal ethics approval. Ethical number R001513 and SOP 001344/6.

Surgical procedure

Animals were randomly allocated to one of four study groups either receiving a 6-OHDA injection into the striatum (Striatal cohort; n = 17), a 6-OHDA injection into the MFB (MFB cohort; n = 15) or a sham treatment (Sham striatum cohort; n = 6 and Sham MFB cohort; n= 9).

CatWalk™ XT gait analysis

The gait of unforced moving rats was analysed with the CatWalk™ XT system (Noldus Information Technology, Wageningen, The Netherlands) as described previously [1]. In an initial training period prior to 6-OHDA lesioning, rats were allowed to walk across the 70 cm long glass walkway for as many times as needed, but for a minimum of nine runs, in order to learn to walk in a straight line to the goal box positioned at the end of the walkway. Once trained, baseline testing followed prior to 6-OHDA lesioning and subsequent, weekly testing started 8 days after 6-OHDA lesioning and consisted of a minimum of three uninterrupted crossings (three runs). The CatWalk™ XT software was used to analyse runs that had at least three complete step cycles.

Gait parameters

All gait parameters examined in this study with the CatWalk™ XT system and showing significant differences are described below. Step cycle, swing phase, stand phase, stride length, print position and cadence have all been reported previously to correlate with speed and have thus been normalized before being statistically analysed. These parameters were normalized against the average speed of all animals by multiplication of each individual value with a normalization factor. The normalization factor F was calculated as follows:

$$\mathrm{F}_{\mathrm{n}_1,\mathrm{n}_2,\ldots m_1\ldots} = \frac{\mathrm{V}_{\mathrm{n}_1,\mathrm{n}_2,\ldots m_1\ldots}}{average\ \mathrm{V}_{m,n,\ldots}}$$

$F_{n_1,n_2,\ldots m_{1,\ldots}}$ representing the normalization factor for each test animal at a certain time point n, m,... etc.

$V_{n_1,n_2,\ldots m_{1,\ldots}}$ representing the individual speed at which each animal crossed the walkway at a certain time point n, m,... etc.

$average\ V_{m,n,\ldots}$ representing the average speed calculated for each time point

By normalizing the above mentioned parameters, we can confirm that reported significant differences are independent of the speed of each individual animal.

General parameters

For all the analysed parameters, faulty or incomplete step cycles at the start or end of runs were deleted and the classification of runs was performed on at least three uninterrupted step cycles.

Average speed: speed displayed during a recorded run from entering until leaving the walkway

Proportion of slow moving rats: the percentage of slow moving animals were plotted for each time point and treatment group. As discussed by Vlamings and co-workers [1] velocities between 20 and 50 cm/s are considered as slow walking. The speed at which an animal walks is very likely an indication of impaired motor abilities. Velocities below 20 cm/s are more attributable to explorative behavior, characterized by frequent pauses and accompanied by sniffing, thus all animals walking at less than 20cm/sec were excluded from the analysis of 'average speed' and all other speed associated variables [1].

Dynamic paw parameters

Step cycle: describes the time in seconds between two consecutive initial contacts of the same paw with the glass floor of the walkway and consists of the stand phase and swing phase:

1. Stand phase: reflects the duration in seconds of contact of a paw with the glass floor.
2. Swing phase: expresses the duration in seconds of no contact of a paw with the glass plate.

Cadence: describes how many steps are taken within one minute.

Coordination

Stand index: this is a measure for the speed at which a paw loses contact with the glass plate at the initiation of the swing phase.

Print position: this is the distance in cm between the position of the hind paw and the position of the previously placed front paw on the same side of the body (ipsilateral) and in the same step cycle. A positive value of the print position indicates that the hind paw is placed behind the front paw. A negative value of the print position indicates that the hind paw is placed in front of the front paw.

Initial dual stance: this is the first time in a step cycle of a front or hind paw that the contralateral front or hind paw also makes contact with the glass plate.

Phase Dispersion: this is a measure of inter-limb coordination and describes the temporal relationship between placement of two paws within a step cycle as described by the CatWalk™ XT software. Phase Dispersion is the moment of initial contact of a target paw expressed as a percentage of the step cycle time of an anchor paw, and is calculated as follows:

$$\text{Phase Dispersion [\%]} = \frac{\mathrm{IC}_{\mathrm{target}}[\mathrm{s}] - \mathrm{IC}_{\mathrm{anchor}}[\mathrm{s}]}{\text{step cycle}_{\mathrm{anchor}}[\mathrm{s}]} \times 100\%$$

IC describes the initial contact in seconds for either the target or the anchor paw. The anchor paw is always the front paw for diagonal phase dispersion (RF-LH or LF-RH).

Data was analysed with a paired samples t-test within treatment groups for each time point. Before statistical analysis was conducted, a Pearson's correlation analysis was conducted to investigate a possible association of the various phase dispersion variables with the average body speed.

Static parameters

Stride length: this is the distance between successive placements of the same paw.
Print length: this is the length in cm (horizontal direction) of the complete print. The complete print is the sum of all contacts with the glass plate.
Print area: this describes the surface area in cm^2 of the complete print.

Spontaneous forelimb asymmetry test (cylinder test) and drug induced rotations

The cylinder test and the rotational testing were used to confirm DA cell loss in the SNc.

Tissue processing and immunohistochemistry

Once behaviour testing have finished, animals were culled and brain tissue was prepared for immunohistochemistry.

Statistical analysis

All gait parameters were analysed using CatWalk™ XT software and expressed as mean ± s.e.m. The proportion of slow moving rats is expressed as a percentage of the total number of rats in a cohort. Two-way repeated measure ANOVA was used for the statistical evaluation of gait changes in the CatWalk™ XT. Influence of time, treatment (sham MFB, sham striatal, MFB lesion, and striatal lesion) or time × treatment on the different CatWalk™ XT parameters or on the proportional use of the left forelimb was evaluated. For multiple comparisons test, depending on whether the Levene's test was significant (assumption of homogeneity violated) or not significant, the Dunnett's post hoc or Tukey's post hoc tests were used, respectively for cases where the interaction time × treatment effect was significant. When the interaction was not significant, influence of the treatment factor was independently analysed. For the parameter Phase Dispersion a paired-samples t-test was used to analyse pairwise differences of diagonal or ipsilateral phase dispersion within each treatment group over time. Statistical significance for rotation testing was analysed using one-way ANOVA and group differences were determined by the Tukey's post-hoc test or Dunnett's post hoc.

The correlations of the number of neurons with positive immuno-staining for TH with gait parameters in the CatWalk™ XT test were evaluated by the Pearson's product correlation coefficient. Differences in the number of TH positive neurons in the substantia nigra for the different treatment groups were analysed using one-way ANOVA and Tukey's pots-hoc test, respectively.

All statistical tests were performed with SPSS 23.0 statistical software (IBM, Armonk, NY, USA). A *p* value of 0.05 or lower was considered statistically significant. No difference was observed for any of the gait parameter or behavioral test measures when the MFB and striatal sham lesion animals were compared.

Results

The gait of rats was evaluated with the CatWalk™ XT system prior to 6-OHDA lesioning (baseline testing) and then weekly after administration of either 6-OHDA or saline. Twenty-six variables showed significant differences at one or more-time points for the MFB lesion cohort compared to the sham cohort. Independency of results from interplay with velocity was achieved through normalization of previously reported speed-associated parameters step cycle, swing phase, stand phase, stride length, print position and cadence, against average speed. This ensured that any significant change observed was not due to the difference in walking speed, as reduced motivation in the MFB lesion cohort was observed frequently, but rather due to the impact of the manipulation on an animal's motor ability.

Our data demonstrated a significant reduction on the average speed in the MFB lesion group when compared with the sham MFB group starting at week 1 post lesion until the end of the experiment, as well as a significant difference when compared to striatal lesion cohort starting from week 1.

Cadence is a parameter that describe the number of steps per minute. The MFB lesion cohort showed a significant decrease in cadence in the MFB lesion group when compared with the sham MFB group starting 1 week post-lesion until the end of the experiment as well as a significant difference when compared to the striatal lesion group starting from week 1 [$F_{(18, 284)} = 2.136$, $p = 0.02$]. We also observed that the percentage of slow moving rats declined steadily in the sham and striatal lesion groups over the study period with both sham lesion groups reaching zero by 6 weeks post lesion, the MFB lesion cohort, however, showed an increasing number of slow-moving individuals during the first two weeks.

In order to further analyse the changes on the dynamic paw parameters observed in the MFB lesion animals, we studied the stride length parameter that describes the distance between successive placements of the same paw. Our data show that the normalized parameter stride length was significantly reduced for all paws (the left front [F

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

$(_{18, 284}) = 2.194$, p =0.05], the right front [F $(_{18, 284}) = 2.126$, p =0.05], the left hind [F $(_{18, 284}) = 2.114$, p =0.07], and the right hind [F $(_{18, 284}) = 2.161$, p =0.005]) in the MFB lesioned rats starting 1 week post-lesion until the end of the experiment.

The initial dual stance refers to the period of time during a step cycle when both the contralateral front and hind paws are in contact with the walkway and support the animals body weight. It is a regularity measure and variability gives the appearance of limping and postural instability. A significant increase in the length of contact of the left front [F $(_{18, 284}) = 2.483$, p =0.001] starting 1 week post-lesion until the end of the experiment. For the other paws no significant interaction of time x treatment was detected, but a significant effect of treatment was independently observed for the right front [F $(_{3, 40}) = 20.321$, p<0.001]; the left hind [F $(_{3, 40}) = 4.481$, p=0.011] and the right hind [F $(_{3, 40}) = 4.843$, p=0.008].

The normalized parameter print position, a further regularity measure, showed an increase in paw position in the MFB lesion cohort when compared to other treatments. We observed a significant effect of treatment for both the left paws [F $(_{3, 40}) = 4.644$, p =0.008] and the right paw [F $(_{3, 40}) = 2.894$, p =0.049].

Step cycle describes the time between two consecutive initial contacts of the same paw with the glass floor of the walkway. Our data showed no significant effect of time × treatment for any of the paw for the step cycle parameter but a significant effect was observed when treatment was studied independently for the left front [F $(_{3, 40}) = 4.843$, p=0.006], the right front [F $(_{3, 40}) = 2.801$, p=0.044] the left hind paw [F $(_{3, 40}) = 3.801$, p=0.017] and the right hind paws [F $(_{3, 40}) = 7.239$, p=0.01].

The step cycle is made up of two stages, the stand phase, which describes the time of contact of a limb with the floor of the walkway, and the swing phase, which describes the time of no contact of a limb with the floor. We found no effect of lesion over time for the stand parameter for any paw. When we examined the swing phases we found, however, a significant effect of treatment for the left front [F $(_{3, 40}) = 3.906$, p=0.017], the right front [F $(_{3, 40}) = 2.632$, p=0.046] and the left hind.

The speed at which a paw loses contact with the walkway is represented in the parameter stand index. Our data showed a significant difference for the left front paw in the MFB lesion group compared to the MFB sham group from 1 week post-lesion [F $(_{18, 284})= 2.647$, p <0.001]. The right front stand index parameter also showed a significant difference in the MFB lesion group compared with the sham MFB group from week 1 post-lesion interaction of time × treatment [F $(_{18, 284}) = 2.780$), p <0.001]. A similar observation was made for the hind paws Left hind stand index parameter showed a significant difference in the MFB lesion group compared with the sham MFB group from 2 to 5 weeks post lesion [F $(_{18, 284})= 5.108$, p =0.035]. The right hind stand index parameter also showed a significant difference in the MFB lesion group compared with the sham MFB group from 3 to 5 weeks post lesion [F $(_{18, 284}) = 3.257$, p =0.001). For all paws, the striatal lesion cohort did not show any significant difference when compared to the striatal sham group.

In order to identify the effect of 6-OHDA lesioning on inter-limb coordination, diagonal phase dispersion as well as ipsilateral phase dispersion were compared within treatment groups over time. In a perfectly balanced walking pattern it is not expected to find any differences when one side of the body or body axis is compared with the opposite side or axis. Any deviation from a coordinated walking pattern should be reflected by changes in one or both of these measures. Our data show a significant reduction in the , Left Front-Right Hind axis when compared to the Right Front- Left Hind axis starting at week 1 post lesion in both the striatal and the MFB lesion groups indicating a deviation on the coordinated walking pattern.

Finally, the print length and print area parameters were calculated on the base of the sum of all contacts of a paw with the glass floor. The more weight an animal puts onto a limb, the greater the contact area or print length value. Our data showed a significant increase in the right hind print length of the MFB lesion group when compared with the MFB sham group for weeks 2, 3, 4 and 6 post lesion as well as a significant difference when compared to the striatal lesion group starting from 2 weeks post lesion of animals in the MFB cohort [F $(_{18, 284}) = 1{,}824$), p =0.024]. No significant differences were observed for the other paws.

Spontaneous forelimb asymmetry test (cylinder test) and drug induced rotations

The cylinder test is a measure of sensorimotor forelimb function and was used as a reference to quantify the asymmetry of forelimb use in 6-OHDA lesioned animals compared to the sham-operated animals. Our results showed a significant reduction in contralateral forelimb use in the MFB and the striatal group. Drug-induced rotational behavior is a robust indicator of unilateral dopaminergic 6-OHDA lesioning. As expected, four weeks after 6-OHDA, highly significant rotational behavior was observed in the MFB lesion cohort when compared to sham MFB group (7.99 rotations/min and 0.10 rotations/min, respectively; $P < 0.001$). The striatal cohort also showed a significant increase in the rotation response after been challenged with Apomorphine when compared to its control (3.10 rotations/min, and 0.03 rotations/min; $P < 0.04$).

Quantification of dopaminergic neuronal degeneration following 6-OHDA lesioning

Stereological quantification revealed a significant reduction in the total number of tyrosine hydroxylase (TH) -positive neurons in the SNc of either the MFB or the striatal lesion cohorts (MFB: 2,199 ± 457 TH-positive cells, $P < 0.001$; Striatal: 11,803 ± 1,438 TH-positive cells, $P < 0.001$) when compared to sham groups (19,654 ± 2,246 and 20,015 ± 1,785, respectively) 6 weeks post lesion. No significant differences in the number of TH-positive neurons was observed between the two sham groups.

We also observed a significant reduction in the level of TH immunoreactivity in the striatum of animals in both the MFB lesion (1.41 ± 1.55 mm^3, $P < 0.001$) and the striatal lesion cohorts (9.66 ± 1.19 mm^3, $P < 0.001$) when compared to MFB sham (18.27 ± 0.2 mm^3) and striatal sham (18.52 ± 0.63 mm^3) groups 6 weeks after 6-OHDA lesioning. No difference in the level of TH immunoreactivity was observed in the striatum between the two sham groups.

Correlation of gait parameters with TH-positive neuronal cell loss

As we described on the results, the MFB lesion cohort present alterations in different gait parameters analyzed. In order to further analyzed this point, a correlation study was conducted to confirm the relationship between TH neuronal cell loss of the SNc and gait dysfunction in the MFB cohort and its control. A Pearson's correlation coefficient (R) was calculated for the average value of 27 significant gait parameters and the residual number of TH-positive neurons in the SNc of the MFB lesion cohort 6 weeks after lesioning.

Discussion

In this study, we identified twenty-one different and time-independent gait parameters that were significantly altered in the MFB 6-0HDA lesion rat model some as early as one week after the lesion. This suggests that gait analysis can provide a clinically relevant, low impact method of testing functional impairment.

Our data indicated that the overall walking speed and cadence of the MFB lesioned animals steadily declined when compared to the other groups. This data is reinforced with the percentage of slow moving rats' greatest in the MFB lesion cohort. Slow movement was defined as an average speed of 20-50cm/s [1]. Diminution of walking speed in different PD rat models have been previously reported [1], [9]. Interestingly, the alteration on average speed, cadence and the percentage of slow moving animals in the MFB lesion group was observed from the first week after the 6-OHDA lesion until the end of the experiment.

Our results also showed an alteration in the stride length parameter in the MFB lesion cohort. These animals demonstrated a significant reduction in stride length when compared to the MFB sham group and the striatal lesion cohort for all four limbs. This parameter indicates successive placement of the same paw and the reduction observed in the MFB lesion cohort is an indication of gait abnormalities similar to those observed in humans. Both initial dual stance and print positions, other measurement related with the initial step cycle, were significantly altered in the MFB lesion cohort further indicating gait abnormalities in this group. Interestingly, we observed a significant increase in the stand index parameter for all four limbs of the MFB lesion cohort when compared to the MFB sham groups and the striatal cohort. This parameter is a measure for the speed at which a paw loses contact with the glass plate during one-step cycle. As previously mentioned, the MFB lesion cohort showed a significant reduction of average speed but a higher value of paw speed thus indicating an alteration in the footfall pattern and abnormalities in gait. Our data also showed that MFB lesion animals had a significantly higher value of right hind print length and print area starting from 2 weeks post lesion when compared to the MFB sham group and the striatal cohort. This value indicates a compensatory effect on the gait patter of those animals.

The effect of an MFB 6-OHDA lesion on inter-limb coordination was investigated by the parameter diagonal phase dispersion as well as ipsilateral phase within treatment groups over time. We observed a clear impact on diagonal phase dispersion. The most common footfall pattern for animals in this study was to alternate 'Ab' with the following sequence: left front – right hind – right front – left hind. Thus, any coordination deficit would naturally show up in the diagonal phase dispersion measure, which reflects exactly this footfall pattern, rather than in a parameter that represents an alternating pattern 'Aa' with a paw sequence of right front - right hind – left front – left hind, such as ipsilateral phase dispersion. Interestingly, the impairment of coordination of the diagonal axis seemed to be more prominent in the striatal lesion cohort compared to the MFB lesion cohort. This observation is the more surprising when looking at the results from the step-cycle analysis. The Swing phase for the right front/left hind axis of the MFB lesion cohort was significantly altered compared to its control, which supports the findings of the parameter diagonal phase dispersion. Yet, there were no measurable changes to any aspect of the step cycle for the striatal lesion cohort despite the significant changes to the diagonal phase dispersion in this group. Similarly, other parameters of coordination such as initial dual stance, stand index and print position show no or only minor changes in the striatal lesion group when compared to sham.

In the MFB lesion cohort mostly of the observed gait parameter disturbance paralleled symptoms of advanced human PD. Slower average speed in combination with reduced cadence and significantly altered swing phase for the right front/left hind body axis may reflect bradykinesia [2], [10], [11]. The altered stand index, stride length and print position also reflects the typical small, shuffling steps observed in patients. At the same time, the results reported are in accordance with those previously reported by Hsieh and co-workers [9]. Increased initial dual stance times reflect postural instability and prolonged double support duration in humans. While a significantly larger print area and print length was observed in the rat model, representing flat foot strikes in human patients, both of these parameters was increased only in the right-hind limb in our model, which was actually the non-impaired body side. Thus, this finding might be a compensatory measure rather than a true symptom.

In contrast to the MFB lesion cohort, we observed limited alteration of gait function in the striatal lesion cohort. Although we observed a minimal gait disturbance in this group with increased initial dual stance on the left front paw at 4 weeks post lesion and phase dispersion alterations from week 1 to week 6, these observations did not correlate with the functional alterations observed in the cylinder test and with rotational analysis. The limited gait dysfunction observed in the striatal lesion cohort may be due to the lower level of DA cell loss (39.83%) observed in the striatal lesion cohort compared to the MFB lesion cohort (88.79%) at 6 weeks post lesion.

Further supporting this, we observed a significant correlation between TH neuronal cell loss in the SNc and gait dysfunction measured over 27 parameters. The MFB lesioned cohort showed an average cell loss of >90% when compared to MFB sham with these animals showing a higher disturbance on gait parameters. Overall, we observed 4 of 27 gait analysis parameters with a strong correlation, and 18 parameters showing a moderate correlation with the number of TH positive neurons in animals with MFB lesioned animals at 6 weeks post lesion. As expected, our correlation data indicates a clear interdependency between cell loss and gait disturbances, with higher cell loss resulting in greater disturbance. This indicates that gait analysis is a reliable tool to identify the severity of the DA depletion in a 6-OHDA rat model. While a correlation between TH cell loss and gait disturbance has already been described rats [1], [9] to the best of our knowledge, our data provides a more extensive and definitive correlation between DA depletion and alterations in the different stages of gait.

In conclusion, we have demonstrated that gait analysis is a reliable method for early detection of motor deficiencies in a MFB lesion model of PD potentially allowing the detection of motor function impairment earlier than with rotational analysis. The gait deficiencies observed in the MFB 6-OHDA lesion resemble those experience by PD patients and the use of gait may provide a valuable tool for the evaluation of future therapeutic strategies for PD.

Ethical statement

All the experimental procedures were in compliance with the NZ Animal Welfare Act (1999) and in accordance with the University of Auckland animal ethics approval. Ethical number R001513 and SOP 001344/6.

References

[1] R. Vlamings *et al.*, "High frequency stimulation of the subthalamic nucleus improves speed of locomotion but impairs forelimb movement in Parkinsonian rats," *Neuroscience*, vol. 148, no. 3, pp. 815–823, 2007.

[2] K.-Y. Kwon, H. M. Lee, S. H. Kang, S. J. Pyo, H. J. Kim, and S.-B. Koh, "Recuperation of slow walking in de novo Parkinson's disease is more closely associated with increased cadence, rather than with expanded stride lengthe," *Gait Posture*, vol. 58, pp. 1–6, 2017.

[3] P. Damier, E. C. Hirsch, Y. Agid, and A. M. Graybiel, "The substantia nigra of the human brain: II. Patterns of loss of dopamine-containing neurons in Parkinson's disease," *Brain*, vol. 122, no. 8, pp. 1437–1448, 1999.

[4] H. Ehringer and O. Hornykiewicz, "Verteilung Von Noradrenalin Und Dopamin (3-Hydroxytyramin) Im Gehirn Des Menschen Und Ihr Verhalten Bei Erkrankungen Des Extrapyramidalen Systems," *Klin. Wochenschr.*, vol. 38,

no. 24, pp. 1236–1239, 1960.
[5] U. Ungerstedt, "6-Hydroxy-dopamine induced degeneration of central monoamine neurons.," *Eur. J. Pharmacol.*, vol. 5, pp. 107–110, 1968.
[6] R. L. M. Faull and R. Laverty, "Changes in dopamine levels in the corpus striatum following lesions in the substantia nigra," *Exp. Neurol.*, vol. 23, no. 3, pp. 332–340, 1969.
[7] S. Przedbroski *et al.*, "Dose-dependent lesions of the dopaminergic nigrostriatal pathway induced by instrastriatal injection of 6-hydroxydopamine," *Neuroscience*, vol. 67, no. 3, pp. 631–647, 1995.
[8] C.-S. Chuang, H.-L. Su, F.-C. Cheng, S. Hsu, C.-F. Chuang, and C.-S. Liu, "Quantitative evaluation of motor function before and after engraftment of dopaminergic neurons in a rat model of Parkinson's disease.," *J. Biomed. Sci.*, vol. 17, p. 9, 2010.
[9] T. H. Hsieh, J. J. J. Chen, L. H. Chen, P. T. Chiang, and H. Y. Lee, "Time-course gait analysis of hemiparkinsonian rats following 6-hydroxydopamine lesion," *Behav. Brain Res.*, vol. 222, no. 1, pp. 1–9, 2011.
[10] J. Jankovic, "Parkinson's disease: clinical features and diagnosis," *J. Neurol. Neurosurg. Psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.
[11] C. J. Turnbull and P. Fitzsimmons, "Advances in the treatment of the motor symptoms of Parkinson's disease," *J. R. Coll. Physicians Edinb.*, vol. 39, no. 1, pp. 29–31, 2009.

IL-13 signaling influences long-term potentiation of spatial learning

Tiroyaone M. Brombacher[1, 2,3], Martyna Scibiorek[1,2,3], Nokuthula Makena[4], Jacqueline Womersley[4], Ousman Tamgue[1,2,3] and Frank Brombacher[1,2,3]

[1] Cape Town Component, International Centre for Genetic Engineering and Biotechnology, Cape Town 7925, South Africa; [2] Division of Immunology, Institute of Infectious Disease and Molecular Medicine, Health Science Faculty, University of Cape Town, Cape Town 7925, South Africa; [3] South African Medical Research Council, Cape Town 7501, South Africa; 4 Department of Human Biology, University of Cape Town, Cape Town 7925, South Africa; [4] Department of Human Biology, University of Cape Town, Cape Town 7925, South Africa.

Background

Pro-inflammatory cytokines are known to play a role in the bidirectional communication between brain and immune system[1]. Recently, anti-inflammatory cytokines have been shown to also be expressed in the central nervous system (CNS), and to be essential for cognitive function [2, 3]. We recently showed that IL-13—possibly from immune cells—plays a role in the regulation of cognitive functions by stimulating astrocytes to produce brain derived neurotrophic factor that is known to play a significant role in memory formation [3].
In this study we aim to address the role of type 2 cytokine signaling via IL-4Rα on cognitive functions, in order to define IL-13 target cells involved, as well as the underlying molecular mechanism. We further address the consequence of inhibited signaling of IL-13 via IL-4Rα, leading to impairment of reference memory, but not learning.

Materials and Methods

To investigate immunological-relevant genes influencing cognitive function, mouse models deficient of IL-4Rα, IL-4/IL-13, and IL-4Rα/IL-13 were analyzed for their influence on learning and memory using the Morris water maze (MWM) as described by Brombacher et al.[3]. Briefly, during the acquisition phase, mice were given four 5-min trials a day for 4 consecutive days to locate a submerged circular platform placed ∼0.5 cm below water level in an open circular pool. A probe trial was performed on day 5 with the platform removed to test reference memory. Data were recorded using the EthoVision XT 8 automated tracking system (Noldus Information Technology, Leesburg, VA). Statistical analysis was performed using ANOVA and Bonferroni post hoc test or Student t test. Groups were run in alternating order on successive training days, and animal protocols were approved by the independent Animal Ethics Research Committee at the University of Cape Town (approval no. 050/015). To gain insight into cellular infiltrates various techniques were used.

Results

We show that IL4Rα or IL4Rα/13 deficiency does not impair learning in comparison to wild type control mice—whether simple or complex—however, reference memory is impaired. From loss of function approach, using the MWM, we demonstrate that IL-13 signaling via IL-13RαII is not the cause for observed impairment of memory formation.

Conclusions

While both IL-4 and IL-13 are known to be required for learning and memory, failure to signal via the common receptor IL-4Rα, impairs only reference memory, but not learning. This suggests that possible compensatory mechanisms and pathways employed in the absence of IL-4Rα are effective to uphold the processes of learning, but not sufficient to support reference memory, indirectly suggesting that IL-13 is critical for the process of memory formation.

References

1. Charlton, R.A., et al., *Associations between pro-inflammatory cytokines, learning, and memory in late-life depression and healthy aging.* Int J Geriatr Psychiatry, 2018. **33**(1): p. 104-112.
2. Derecki, N.C., et al., *Regulation of learning and memory by meningeal immunity: a key role for IL-4.* J Exp Med, 2010. **207**(5): p. 1067-80.
3. Brombacher, T.M., et al., *IL-13-Mediated Regulation of Learning and Memory.* J Immunol, 2017. **198**(7): p. 2681-2688.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Behavioral paradigm for testing pro sociality in rats

R. Bruls, M. Carrillo, V. Gazzola and C. Keysers

Netherlands Institute for Neuroscience, Social Brain Lab, Amsterdam, the Netherlands. r.bruls@nin.knaw.nl

Introduction

Prosocial behaviors are found throughout different animal species and are thought to be promoted by empathy, i.e. the ability to share another individual's feelings. Although human studies are revealing the neural correlates of empathy [1,2], what is needed are findings showing causality between these neural networks and behavioral responses to empathy. Animal models of social behaviors are a promising avenue for this endeavor. For example, it has been shown that a rat's reaction to another rat's distress is quantifiable throughout vicarious freezing [3]. Other studies have used distressed demonstrators to show that rats helped trapped [4] and soaked conspecifics [5] proving pro social behavior in rats. However, in order to link neural activity to prosocial behaviors, it is crucial to work with stressors that can be precisely timed, and linked to neural activity. Here, we propose a prosocial paradigm that uses shocks as a stressor, where observer rats choose between liberating either a cagemate receiving shocks or a cagemate receiving no shocks.

Methods

The goal of this experiment is to measure prosocial behaviour in Long Evans rats. Rats are housed in groups of three and are assigned either the role of observer (Obs), demonstrator 1 (Dem1) or demonstrator 2 (Dem2). Experiments are carried out in a custom-made experimental setup, which consists of three platforms equipped with an electric grid floor and placed adjacent to each other. The middle platform is separated by the two adjacent platforms by doors made of Plexiglas with perforated small holes. These doors can either be closed – no crossing between platforms possible – or open – animals can cross between platforms. The middle platform is equipped with two levers: L1 and L2. Pressing L1 or L2 will cause Door1 or Door2 to open, respectively. Observer animals are first trained to press only one lever. Once door opening is consistent, the second lever is added to the design. As a proof of principle that rats can distinguish between each lever granting access to a different platform, observers first perform a sucrose test. Here, sucrose is placed on one of the adjacent platform together with a light cue, while the other platform remains empty. Observer rats can use the light cue to choose which lever to press to access the sucrose platform. Once a given observer has reached the 70% preference cutoff measure for the sucrose platform, each demonstrator is placed on the adjacent platforms for a second test, which aims at testing for any side or cagemate preference. Here, the observer has to choose to enter the platform with Dem1 or Dem2. Once baseline preferences for either cagemate has been assessed, the observer is promoted to the prosocial testing, where one of the demonstrators receives footshocks (0.2 mA, 1 second, every 5 seconds for a total of 72 shocks). The observer is now faced with the decision to save the animal receiving footshocks or to open the door towards the non-shocked demonstrator. All experimental procedures were preapproved by The Institutional Animal Care and Use Committee of the Netherlands Institute for Neuroscience (IACUC-NIN-181102).

References

1. de Vignemont, F., Singer, T. (2006). The empathic brain: how, when and why? *Trends Cogn Sci* **10**:435–41.

2. Keysers, C., Kaas, J.H., Gazzola, V. (2010). Somatosensation in social perception. *Nat Rev Neurosci* **11**:417–28.

3. Carrillo, M., Migliorati, F., Bruls, R. et al. (2015). Repeated Witnessing of Conspecifics in Pain: Effects on Emotional Contagion. PLoS One **10**:e0136979.

4. Ben-Ami Bartal, I., Decety, J., Mason, P. et al. (2011). Empathy and Pro-Social Behavior in Rats. *Science* **334**:1427–30.

5. Sato, N., Tan, L., Tate, K. et al. (2015). Rats demonstrate helping behavior toward a soaked conspecific. *Anim Cogn* **18**:1039–47.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

NAT-1: Wireless EEG Technology for Preclinical and Behavioural Research.

Barry Crouch, Riedel Gernot and Bettina Platt

Institute of Medical Science, University of Aberdeen; Barry.crouch@abdn.ac.uk

The Neural Activity Tracker (NAT-1) is a newly developed wireless electroencephalograpy (EEG) data-logger, specifically engineered for rodent experimental and preclinical research applications. Four individually programmable recording channels, amplification, digitization and on-board data storage are combined with a 3-axis accelerometer in a single, lightweight, small form factor device (<4cm^2, 2.2g with battery). Powered by a single PR48 battery the NAT-1 allows for up to 48 hours of un-interrupted EEG recording and motion detection at sampling rates of up to 2 kHz. Wireless design also enables recording under freely moving conditions without the animal welfare concerns associated with tethered systems. EEG recording can be performed in a range of pre-existing settings, such as the animal's home cage and in various mazes and running wheels without modification. The NAT-1 is therefore ideally suited for sustained phenotypic characterisation of disease models, pharmaco-EEG applications and comprehensive studies of sleep and circadian activity.

Here we showcase the NAT-1's applications in pharmaco-EEG research through an investigation of the effects of pharmacological agents on NREM sleep EEG recorded while inside a Phenotyper (Noldus IT) home cage observation system as recently reported in [1]

The NAT-1's functions can also be optionally augmented with additional sensors through an on-board expansion socket. An infra-red (IR) receiver add-on enables intelligent epoch selection and analysis through precise time-stamping and synchronization of EEG with behavioural tracking data. EEG corresponding to specific behavioural events such as a lever press, proximity to an object / interaction partner, decision making, etc can therefore be identified and extracted with confidence.

Here we shall present a demonstration of temporal synchronization of EEG with positional tracking data obtained from the AnyMaze (Ugo Basile) video tracking package, using the IR add-on module. First, in a basic research application in which intelligent epoch selection has been utilized to study interactions among large scale brain networks during spatial decision making in wild type mice. Finally, we shall present a demonstration of how synchronization of EEG with tracking data enables high throughput automated EEG phenotyping of transgenic models of Alzheimer's disease (PLB2$_{APP}$). Briefly, behavioural data based characterization and sorting of EEG epochs prior to spectral analysis is used to create a detailed profile of how EEG band power is regulated by locomotor, navigational and cognitive factors during a single 10 minute Y-maze spontaneous alternation task. Data may then be analysed at the level of individual subjects or at the group level using similar means.

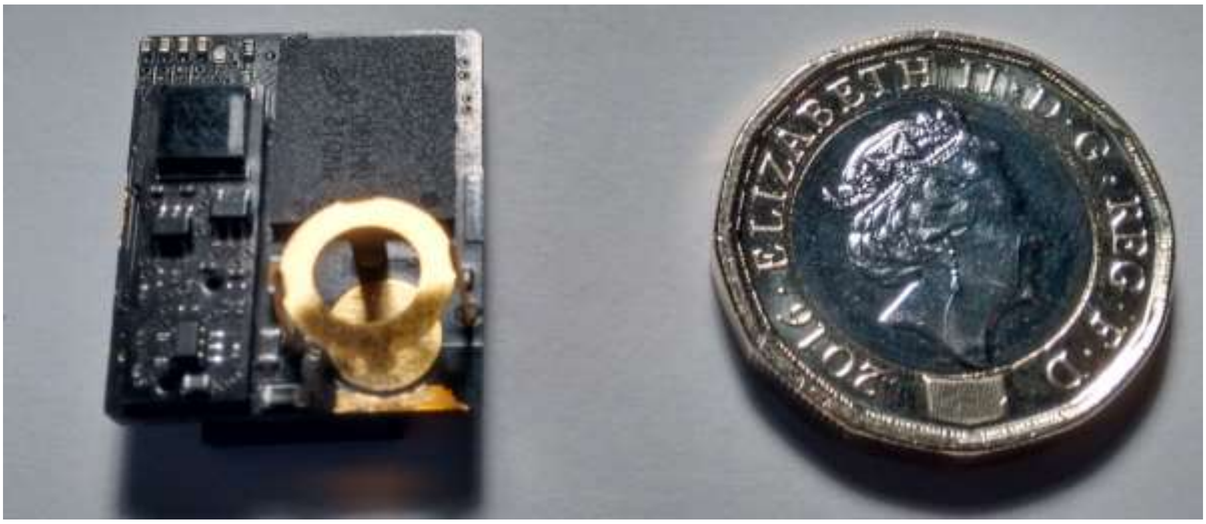


Figure 1. NAT-1 device + IR module (without battery), A UK £1 coin is presented for scale.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Ethical statement

EEG recordings were obtained using surgically implanted gold skull electrodes. The NAT-1 was evaluated through incorporation into ongoing medical research studies. All procedures to be presented were approved by the University of Aberdeen's Ethics Board and carried out with strict adherence to the Animal (Scientific Procedures) Act 1986, European (FELASA) legislation, UK Home Office regulations and best practice regarding post-operative care. Full details of procedures are described in [1] & [2].

References

1. Crouch, B., Sommerlade, L., Veselcic, P., Riedel, G., Schelter, B & Platt, B. (2018). Time-, Frequency- and direction-resolved communication within brain networks. *Scientific Reports* **8**, 1825 DOI:10.1038/s41598-018-19707-1.
2. Jyoti, A., Plano, A., Riedel, G. & Platt, B. (2010). EEG, activity, and sleep architecture in a transgenic AβPPswe/PSEN1A246E Alzheimer's disease mouse. *Journal of Alzheimer's Disease* **22(3),** 873–887.

Beam vs. Rod walking: comparison in sensitivity and reliability in transgenic Engrailed-2 mice

N. Czechowska, S.L. Baader

Institute of Anatomy, University of Bonn, Bonn, Germany. nico.czechowska@uni-bonn.de

Rotarod is one of the most often used techniques in assessing motor behavior in mice. Mice are placed on a rotating rod and the time to fall is registered. Having an appropriate apparatus at hand, this assay is easy to perform and is considered to have a reasonable high sensitivity [1]. However, when applying certain drugs to mice Rotarod revealed a yes or no answer rather than a dose response curve. The balanced beam test, in contrary, allowed for a graded measurement of motor behavior [2]. The balanced beam test in our lab consists of a 1 m long and 12 mm wide bar on which mice run from a starting point to a familiar box at the far end of the bar. Mice are videotaped from their back to follow their hind legs and their positioning on the stick. This allows one to measure the speed of walking but also the amount of slips and the capability of setting back slipped legs onto the bar. Other parameters to be measured are number of jumps or falls from the bar. The aim of this study was to compare both techniques in order to reveal differences in motor performance in transgenic mice which otherwise didn't show obvious ataxic behavior or muscle degeneration. In addition, we asked whether both techniques are suitable for analyzing motor learning, or forgetting previously learnt motor activities.

We compared two transgenic mice which showed either a cell type specific overexpression of a transcription factor called Engrailed-2 (En2). This mouse only shows an overexpression of En2 in Purkinje cells within the cerebellum (L7En2, [3]). In addition, we analyzed an En2 deficient mouse line ($En2^{-/-}$) which display En2 loss in all cells [4]. These mice were used as model mice to investigate motor performance since both mice didn't show ataxia or obvious motor disabilities.

Interestingly, both mice performed differently on the accelerating Rotarod test. At the first day of testing, we could not monitor significant differences in the time to fall off the rod between En2 overexpressing, En2 deficient and their corresponding wildtype mice. All mice fell off after about 100 seconds. While male L7En2 mice showed reduced motor learning during an initial five-day training period and a comparable memory consolidation as compared to wildtype mice, $En2^{-/-}$ mice learnt less than their wildtype littermates during the initial five-day period, but they forgot about what they learned within 30 days. This slight difference in motor performance was significantly more emphasized in the balanced beam test. When setting mice on the beam, slipping happened 20 times more often in L7En2 mice and ten times more often in En2 deficient mice as compared to wildtype mice. Correspondingly, walking speed was much increased in wildtype as compared to transgenic mice. In addition, resetting rear legs onto the bar happened significantly more often in wildtype as compared to transgenic mice. Variances in between experiments and mice were less in balanced beam tests as compared to Rotarod assays. It is however important to note, that the reliability of the experiments indeed depends on the age of the animals. Younger mice (postnatal day 15) tend to leave the bar although they are walking at a height of about 50 cm. Older mice, in contrary, tend to slide on their belly along the bar. This sliding, however, can be avoided by using thicker bars. In summary, the sensitivity and the reliability of the beam walking assay as well as the possibility to adapt the system makes the balanced beam test superior to the Rotarod analysis. Using both assays over a longer time period allows one to not only follow a temporary motor pattern, but enables to look at memory consolidation over time. Together with other motor behavior assays such as the single frame motion analysis, grip tests, or foot print analysis it is thus possible to dissect out differences in motor performance, in learning abilities and in pure muscle coordination problems.

The study was approved by Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen under the reference number: AZ-84-02.04.2016.A359.

References

1.Brooks S.P., Dunnett S.B. (2009). Tests to assess motor phenotype in mice. A user's guide. *Nature reviews Neuroscience* **10 (7)**, 519-529.

2. Joanna L.S., Rachael J.T., Terry A.B., Louise M.M., Gerard R.D. (2005). The mouse beam walking assay offers improved sensitivity over the mouse rotarod in determining motor coordination deficits induced by benzodiazepines. *Journal of Psychopharmacology* **19**, 221-227.

3. Baader S.L., Sanglioglu S., Berrebi A.S., Parker-Thomburg J., Oberdick J. (1998). Ectopic overexpression of Engrailed-2 in cerebellar Purkinje cells causes restricted cell loss and retarded extzernal germinal layer development at lobule junctions. *Journal of Neuroscience* **18 (5)**, 1763-1773.

4. Millen K.J., Wurst W., Herrup K., Joyner A.L. (1994). Abnormal embryonic cerebellar development and patterning of postnatal foliation in two mouse Engrailed-2 mutants. *Development* **120**, 695-706.

An agent-based model for the emergence of social styles in macaque societies.

R. Dolado[1], E. Gimeno[1], V. Quera[1] and F.S. Beltran[1]

[1]Institute of Neuroscience, University of Barcelona, Barcelona, Spain. ruth.dolado@ub.edu

The covariation hypothesis states that, in species of the genus *Macaca*, the differential handling of intra-group conflicts allows to classify these species in four social styles: grade 1, or "despotic"; grade 2, or "tolerant"; grade 3, or "relaxed"; and grade 4, or "egalitarian" [1]. Grade-1 and grade-4 species show different social style patterns. Grade-1 species display fierce agonistic behaviours resulting in steep hierarchies and asymmetric agonistic interactions, kinship determining the direction and outcome of both agonistic and affiliative interactions. Conversely, grade-4 species display a mild intensity of aggression, subordinates often responding to attacks with counter-attacks and, consequently, the amount of agonistic interactions increases resulting in weak and unstable hierarchies. Grades 2 and 3 are defined as intermediate patterns [1], combining behaviours associated with grade-1 and/or grade-4 species. However, the lack of consistent behavioural patterns exhibited by grade-2 and grade-3 species make it difficult to assign them to a particular grade [2], and could explain why some authors assign different groups of the same species to different grades [3,4]. In these cases, quantitative criteria based on social structure measurements could facilitate the process of assigning groups to a defined grade in macaque societies.

According to results from simulation studies using agent-based models [5], Hemelrijk [6] suggested an explanation about how different social styles in macaque societies could emerge as a result of the intensity of aggression, which ranges from high to mild, influences the gradient of hierarchy, and characterizes the different agonistic patterns referred to in the covariation hypothesis [1].

We have developed an agent-based model implemented in NetLogo to explore the four social styles in macaque societies and determine quantitative criteria to differentiate intermediate grades. Social interactions between virtual primates include agonistic and affiliative behaviours modulated by kin relationships. Agents interact according to the aggression rules proposed by Puga-González et al. [7], depending on the intensity of aggression (*stepdom* parameter). When the risk of losing an agonistic interaction is high, the virtual primates can perform affiliative interactions depending on an *anxiety* parameter (which ranges from calm to tense), a *kinfactor* parameter (the coefficient of relatedness between two given virtual primates) and a *stepsoc* parameter (the social style in macaque societies) according to the covariation hypothesis [1].

We ran a simulation study to obtain a set of 20 measurements of social structure defining aggression, affiliative and kinship patterns per each grade or social style. Following Hemelrijk [6], we progressively increased the intensity of aggression which in turn drove the emergence of the four patterns of aggression and provoked a set of correlated changes in the rest of social structure patterns (affiliative and kinship patterns). Preliminary results showed significant differences in 18 out of 20 social structure measurements between the four social styles. Results of a multiple pairwise comparison test allowed us to group the four grades into two clusters: grades 1 and 2 vs. grades 3 and 4. In accordance with the phylogenetic lineages of the *Macaca* genus [2], only high values of the intensity of aggression allowed us to define significant differences between the values of the social measurements obtained in intermediate grades (i.e., grades 2 and 3). These preliminary results provide quantitative criteria suitable to compare with naturalistic observations of *Macaca* species both in captivity and in the wild, and help defining their social styles. Although this research is focused on *Macaca* societies, agent-based models could be used to explore the social behaviour of others primate species with complex social systems.

References:

1. Thierry, B. (2004) Social epigenesis; In: B. Thierry, M. Singh, & W. Kaumanns (Eds.), *Macaque Societies: a model for the study of social organizations* (pp 267-290). Cambridge UK: Cambridge University Press.
2. Thierry, B., Iwaniuk, A.N. & Epllis, S.M (2000a). The influence of phylogeny on the social behaviour of macaques (primates: Cercopithecidae, genus *Macaca*). *Ethology* **106,** 713-728.
3. Berman, C.M., Ogawa, H., Ionica, C., Yin, H. & Li, J. (2008). Variation in kin bias over time in a group of

Tibetan macaques at Huangshan, China: contest competition, time constraints or risk response? *Behaviour* ***145***, 863-896.
4. Cooper, M.A. & Bernstein, I.S. (2008). Evaluating dominance styles in Assamese and Rhesus macaques. *International Journal of Primatology* ***29***, 225-243.
5. Hemelrijk, C. K. (1998). Risk Sensitive and Ambiguity Reducing Dominance Interactions in a Virtual Laboratory. In: Pfeifer, R., Blumberg, B., Meyer, J. A., & Wilson, S. W. (Eds.), *From Animals to Animats V* (pp. 255–262). Cambridge, MA: MIT Press.
6. Hemelrijk, C. K. (2002). Self-organization and natural selection in the evolution of complex despotic societies. *Biological Bulletin* ***202***, 283–288.
7. Puga-Gonzalez, I., Hildenbrandt, H., & Hemelrijk, C. K. (2009). Emergent patterns of social affiliation in primates, a model. *PLoS Computational Biology* ***5***(12).

Comparison of mouse motor performance on different RotaRod systems

E. Dreesen and G. Riedel

Institute of Medical Sciences, University of Aberdeen, Aberdeen, UK, r01esd17@abdn.ac.uk

Introduction

Drugs, brain damage and diseases as well as genetic modifications can have an effect on motor coordination and motor learning. It is important that this behaviour is measured during the development of new therapeutics or transgenic animals. Measurement in rodents is most commonly performed on the RotaRod because of its automated features [1]. Animals need to balance on a rotating rod and latency to fall is the analysed parameter. A longer latency is taken to index better coordination while a shorter latency can imply a possible impairment. The most frequently used protocols are the fixed speed paradigm for studying motor coordination and fatigue, and the accelerating paradigm for motor coordination and learning [2]. One problem that can have a significant impact on results is the wide variety of RotaRods currently on the market. Different systems can vary with respect to rod diameters, surface texture, minimum/maximum speed, and even the possibility to administer shocks when animals fall onto a grid floor. This will influence the reproducibility of results and thus both within as well as between laboratory reliability.

In the present study the motor performance of mice was tested on two RotaRod systems. The main aim was to determine the effect of small differences on the behavioural readout.

Methods

Female NMRI mice from our own breeding programme (Charles River, UK) aged 3 (n=35) and 5-6 (n=24) months (mo) were used. Each age cohort was randomly divided into two groups, one for each RotaRod system (see details below). The running order for each system was randomised and each day the animals were tested on a different lane based on a Latin square design. All animals were group housed with food and water available ad libitum under a 12/12 hour dark-light cycle (light on at 07:00 h). Both temperature (20°C± 1°C) and humidity (60−65%) of the holding room were regulated. Experiments were carried out in accordance with the European Communities Council Directive (63/2010/EU) and a project license with local ethical approval under the UK Animals (Scientific Procedures) Act (1986).

The animals were tested for motor coordination and motor learning on either a custom-made automated 4-lane accelerating Ugo Basile SRL (Italy) rat RotaRod NG ($n_{3\ mo}$ = 20, $n_{5\text{-}6\ mo}$ =12) as well as on a 4-lane accelerating TSE Systems (Germany) Rat RotaRod ($n_{3\ mo}$ = 15, $n_{5\text{-}6\ mo}$ = 12), shown in Figure 1. The only differences between both systems consisted of slightly wider grooves (2 mm compared to 1 mm) on the surface of the Ugo Basile SRL rod, more noise accompanying the rotating rod on the TSE system and rotating (Ugo Basile SRL) compared to fixed (TSE Systems) divisions between the lanes. The same accelerating protocol was used for both RotaRods and all tests were performed by the same experimenter, at the same time of day (morning) and in the same experimental room.

Mice were placed on the slowly rotating rod [5 rotations per minute (rpm)] against the direction of the rotation. The rod then progressively accelerated up to a speed of 45 rpm over a duration of 5 minutes. Four trials a day were performed for three consecutive days with a trial ending when either an animal fell off or the maximum trial time of five minutes had elapsed. Inter-trial intervals of 3 to 5 minutes were employed between trials. Latency to fall was automatically recorded and group differences between RotaRods for this parameter were analysed using two-way ANOVA (alpha = 0.05) and Bonferroni post hoc test in GraphPad Prism version 5.04.

Results

While testing on either system, certain animals' behaviour influenced their performance. Instead of focussing on staying on the rod they tried to stand up, touched the sides, turned around or moved move backwards, which caused them to fall off considerably faster than expected, skewing the results. Because of this behaviour these animals were excluded from analysis. They were identified on both RotaRod apparatuses and there was no manufacturer bias.

Figure 2 displays the results of the RotaRod system comparison for each age group. On the left the motor performance for 3 month old mice shows a significant effect of group (p=0.0013). Animals tested on the Ugo Basile apparatus show a longer latency to fall compared to mice tested on the TSE system, meaning the former stayed on the rod for a longer period of time and thus show seemingly better motor coordination. There is also a significant effect of trial (p<0.0001), but no interaction effect (p=0.3761). Both groups thus exhibit motor learning. The same conclusions can be made for the older animals (5-6 mo) whose results are shown on the right. A significant difference is found between the groups (p=0.0002) with animals tested on the Ugo Basile system again performing better. Both groups demonstrate motor learning (p <0.0001) without interaction effects (p=0.8352).

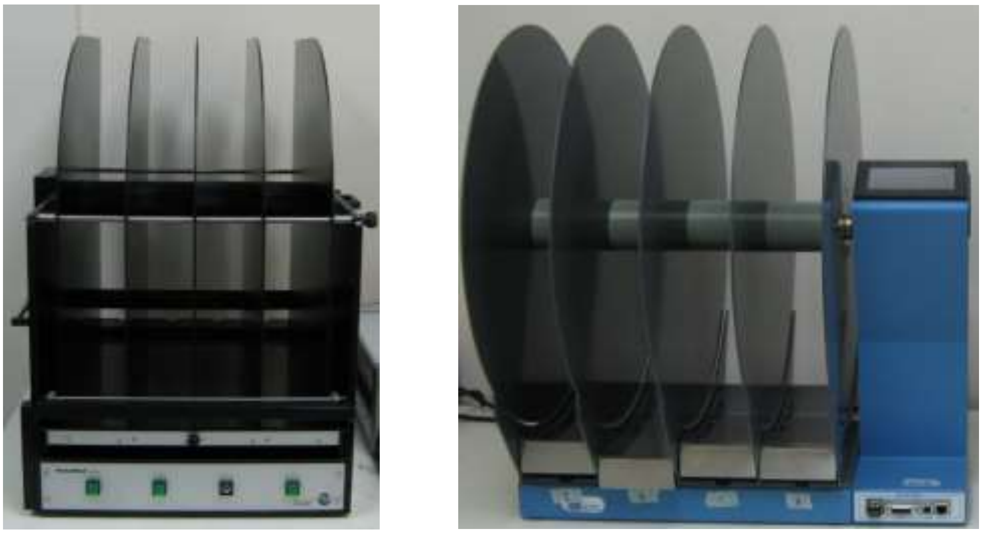


Figure 1: Pictures of the TSE RotaRod system (left) and the Ugo Basile (right) used in this study.

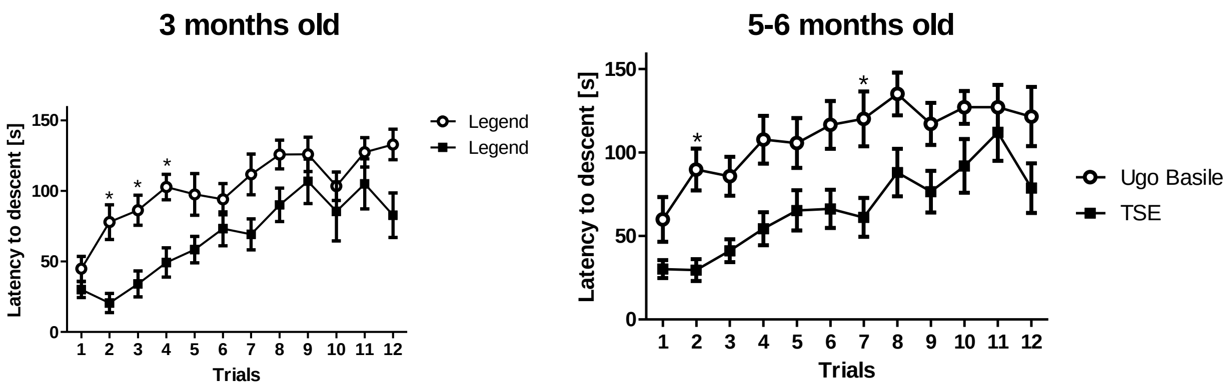


Figure 2. Comparison of motor performance on the Ugo Basile and TSE RotaRod in 3 month old (left) and 5 to 6 month old (right) NMRI mice. There is a clear significant main effect of group at both ages, with the animals performing the test on the Ugo Basile RotaRod having a longer latency to fall than the animals on the TSE RotaRod.

Discussion

Based on these results it seems that animals who have no apparent impairment perform worse on the TSE Systems RotaRod than on the Ugo Basile SRL RotaRod at 3 as well as 5-6 months old. As mentioned above, every effort was made to keep as many variables as possible consistent between tests. It appears that independent of age, NMRI mice find the TSE RotaRod more challenging. At present, our results can be linked to the three differences, namely the width of the grooves, the difference in noise and the divisions.

In conclusion, motor coordination may be strongly affected by small differences in RotaRod hardware. This will have considerable effects on studies looking into possible motor impairments of new transgenic models as phenotypes may be overlooked when tasks are too simple. The opposite would apply to the development of new therapeutic treatments, which may not show efficacy if the task is too difficult and impairments cannot be overcome. Since manufacturers are under considerable pressure to improve handling and hardware components by developing new versions of the RotaRod apparatus all the time, the effect of different hardware use on the reproducibility of earlier findings should not be overlooked.

References

1. Mann, A., & Chesselet, M.-F. (2015) Chapter 8 - Techniques for Motor Assessment in Rodents in LeDoux, M. S. *Movement Disorders (Second Edition)* (pp. 139–157). Boston: Academic Press.

2. Rustay, N.R., Wahlsten, D., Crabbe, J.C. (2003). Influence of task parameters on rotarod performance and sensitivity to ethanol in mice. Behavioural Brain Research, 141: 237-249.

Behavioural Probability Matrix: A Computational Interface for Human Behaviour Prediction in Smart Environments

R.E. Dunne

School of Computer Science, University of Manchester, Manchester. robert.dunne@postgrad.manchester.ac.uk

There is a rapidly growing interest in smart home and assistive living technology, and predicting human behaviour in these smart environments is a key component of moving from purely reactive to pre-emptive systems.

This project proposes a real time behavioural probability matrix (BPM) that can be used as an interface for applications to pre-empt user needs in smart home, ambient intelligence, or assistive living environments. The BPM provides a data structure for applications to access that will be updated in real time, as occupant sensor data is analysed, using the latest research developments in human behaviour prediction [3][2], particularly systems using deep learning [4][1] which currently produce the best results for human behaviour prediction.

The BPM contains for its first row a finite set of possible behavioural actions. Subsequent rows contain corresponding sets of probability distributions (that sum to one) that are stacked sequentially, in time series - giving an application the ability to take action based on the probability of an occupant's predicted next action, or several actions ahead.

Figure 1. Behavioural Probability Matrix (BPM)

$$BPM = \begin{bmatrix} b_1 & b_2 & b_3 & ... & b_n \\ P(b_1)_1 & P(b_2)_1 & P(b_3)_1 & ... & P(b_n)_n \\ P(b_1)_2 & P(b_2)_2 & P(b_3)_2 & ... & P(b_n)_n \\ P(b_1)_3 & P(b_2)_3 & P(b_3)_3 & ... & P(b_n)_n \\ P(b_1)_4 & P(b_2)_4 & P(b_3)_4 & ... & P(b_n)_n \\ ... & ... & ... & ... & ... \\ P(b_n)_n & P(b_n)_n & P(b_n)_n & ... & P(b_n)_n \end{bmatrix}$$

In the BPM the set of behaviours B = {b1, b2, b3, ..., bn} can be derived from existing behavioural ontologies [5]. Or created bespoke to suit the needs of the system.

Prediction probability values can be looked up in the BPM by accessing the corresponding element in the matrix. So for behaviour b1, the probability value for it occurring next would be at BPM1,1 or P(b1)1. For the probability value of it occurring as the second step of the occupant's currently observed activity, it would be BPM2,1 or P(b1)2, and so on.

Further work will need to be done on calculating the efficiency of looking up data to find the highest probability for each step, for different size matrices, where the BPM is of size |A|x|B|, where A = {b1, P (b1)1, ..., P (b1)n}, and B is the set of behaviours. A binary search algorithm for each matrix row could give a run time of O(log n) which may impact on real time lookups - this is identified as an area for continued research.

In summary, the BPM outlined here provides a bridge between the current work in deep learning human behaviour prediction, and applications in smart environments that will be able to pre-empt user needs.

References

[1] Sung Joon Choi, Eun Woo Kim, and Song Hwai Oh. Human Behavior Prediction for Smart Homes Using Deep Learning. 22nd International Sym- posium on Robot and Human Interactive Communication, 1(dataset 2):173– 179, 2013.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

[2] Abubaker Elbayoudi, Ahmad Lotfi, Caroline Langensiepen, and Kofi Ap- piah. Determining Behavioural Trends in an Ambient Intelligence Environ- ment. In Proceedings of the 9th ACM International Conference on PErva- sive Technologies Related to Assistive Environments - PETRA ’16, pages 1–8, New York, New York, USA, 2016. ACM Press.
[3] Sawsan M Mahmoud, Ahmad Lotfi, and Caroline Langensiepen. Behavioural pattern identication and prediction in intelligent environments.pdf. Applied Soft Computing, 13(4):1813–1822, 2012.
[4] Nhathai Phan, Dejing Dou, Hao Wang, David Kil, and Brigitte Piniewski. Ontology-based deep learning for human behavior prediction with explana- tions in health social networks. Information Sciences, 384:298–313, 2017.
[5] Natalia D´ıaz Rodr´ıguez, M. P. Cu´ellar, Johan Lilius, and Miguel Delgado Calvo-Flores. A survey on ontologies for human behavior recognition. ACM Computing Surveys, 46(4):1–33, 2014.

Using Changes in Behavior to Assess Pain and Pain-Relief in Mice Models

M. Durst[1], M. Sauer[1], T. Reding Graf[2], P. Jirkof[1]

1 Anesthesia and peri-operative Pain Research, Division of Surgical Research, University Hospital Zurich, Zurich, Switzerland. mattea.durst@usz.ch; 2 Visceral & Transplant Surgery Research, Division of Surgical Research, University Hospital Zurich, Zurich, Switzerland

Introduction

The work presented is part of a German Research Foundation (DFG) funded project on severity assessment in laboratory animals. Several methods to detect and grade pain, suffering and distress are tested for their feasibility, robustness and standardization. The involved working groups evaluate these methods in five animal species, different animal models e.g. pancreatitis, hepatectomy and laparotomy, and several laboratories. Our group at the University of Zurich aims to identify robust indicators to assess burden in laboratory mice and to analyze factors modulating pain and other negative states. Our goal is also to change from subjective parameters to more objective and automatic measurements. The gained expertise should later be used to refine experimental conditions.

Methods

On the one hand, we used a manually assessed Pain Score to investigate pain or reduced wellbeing. The Pain Score is a composite score including clinical signs and behavioral pain indicators [1]. On the other hand, we assessed changes in an animal's motivation that can occur with a reduced wellbeing resulting from pain. For this purpose, we tested the mice in tasks like nest building and burrowing behavior. Nest building was evaluated with a naturalistic nest complexity score [2]. A reduced nest complexity is associated with states of reduced wellbeing like pain. With the burrowing test we assessed the natural behavior of tunnel digging, another parameter that reflects the motivational state of an animal [3]. Here we measured the latency to start the motivational task of burrowing. An increased burrowing latency is seen in different models linked to pain and reduced wellbeing. Additionally, we automatically assessed the activity of the mice in their home cage over a period of 24h. The activity was measured via the moved distance and velocity of each mouse (Noldus EthoVision).

We applied the previous mentioned methods to grade severity in a mouse model of acute pancreatitis. The inflammation was evoked by the injection of Cerulein, a peptide commonly used to induce pancreatitis in rodent models. Male C57BL/6 mice were injected with 12 injections of Cerulein over a period of two days. We used a control group of mice that were injected with NaCl at the same frequency.

Results

The Pain Score was slightly increased in Cerulein injected animals on day 1 and the evening of day 2. We did not see any differences in both groups in the nest building. Nest complexity scores increased over time. Burrowing latency was higher in NaCl than in Cerulein treated animals, especially on the second day. Mice injected with Cerulein showed a decreased activity at day 1 and 2 of injection compared to NaCl treated mice. On day 3 Cerulein treated mice had a higher activity than NaCl. NaCl injected mice had a reduced activity on day 2 and 3 compared to the first day.

Conclusion

With the used methods, we are able to detect an impact of Cerulein on the animal's behavior in the pancreatitis model. The Pain Score and activity measurements indicate a negative influence, while the stable, equivalent nest building behavior does not confirm this. Cerulein injected animals perform better in the burrowing behavior test, which indicates that the NaCl treated mice, may have a more reduced wellbeing. Whether the behavioral changes in the current study are due to pain from the pancreatitis has to be clarified. Therefore, in the future we want to implement analgesia in this animal model to decide on whether the results indicate pain or are due to a different reason like the influence of Cerulein itself. The model's severity has to be further investigated by methods that are more sensitive to assess negative affective states like pain in the animal. These are for example a cognitive bias test, assessing self-administration of analgesia or a sugar consumption test to assess anhedonia. Assessing and managing pain in laboratory animals is crucial for animal welfare and experimental quality. With this work, we are contributing to an evidence based severity assessment and do our part to promote the 3Rs in animal experiments.

Ethical Statement

The animal experiments were approved by the Cantonal Veterinary Office, Zurich, Switzerland, under license no. 069/2014, and were in accordance with Swiss Animal Protection Law and conform to European Directive 2010/63/EU of the European Parliament.

References

1. Jirkof, P., et al., *Buprenorphine for pain relief in mice: repeated injections vs sustained-release depot formulation.* Lab Anim, 2015. **49**(3): p. 177-87.
2. Hess, S.E., et al., *Home improvement: C57BL/6J mice given more naturalistic nesting materials build better nests.* J Am Assoc Lab Anim Sci, 2008. **47**(6): p. 25-31.
3. Jirkof, P., et al., *Burrowing is a sensitive behavioural assay for monitoring general wellbeing during dextran sulfate sodium colitis in laboratory mice.* Lab Anim, 2013. **47**(4): p. 274-83.

NRX1-KO rats as a model to study cognitive impairment?

Sofie Embrechts and Luc Ver Donck

Janssen, Belgium; sembrech@its.jnj.com

Introduction

Neurexins and neuroligins are synaptic cell-adhesion molecules that connect pre- and postsynaptic neurons at the synapses. They mediate signaling across the synapses. In humans, alterations in genes encoding neurexins or neuroligins have been implicated in autism and other cognitive diseases.

In these experiments, we investigate the effects of deletion of the NRX1-α gene on a Sprague Dawley background compared to wild type littermates across a range of tasks to investigate the possible use of these animals for research in neurodegenerative diseases.

Methods

The animal testing facility is AAALAC accredited. Studies were performed in compliance with European Directive 86/609/EEC on Welfare and Treatment of Animals and Belgium law (Royal Decree dd. April 6, 2010). Test protocols were approved by the Institutional Ethical Committee on Animal Experimentation.

We obtained homozygous NRX1 knock-out (KO) rats and wild type (WT) littermates from internal breeding. The colony was originally sourced from Sage Labs (Cambridge, UK). They were group housed under standard housing conditions with 12/12 h light/dark cycle. Food and water were available *ad libitum*. All studies were performed on 10 to 13 weeks old animals of either sex.

A first cohort of animals (n= 20 WT, 16 KO) was evaluated in a battery of behavioural tests. The order of experiments was such that a test would have a minimal impact on the next and is as follows; first a V-maze experiment was performed to assess short term spatial memory. Next a pre-pulse inhibition (PPI) test was performed to evaluate whether neurexin may be involved in sensory motor gating. This was followed by assessment of spontaneous locomotor activity (LMA) in a novel environment. The fourth test was the Morris Water Maze (MWM) experiment, to investigate spatial learning and memory. Finally, the contextual fear conditioning test (FCT) was performed to evaluate formation of fear memory. In all tests (except PPI) video tracking using Ethovision XT (Noldus, NL) was applied to quantify animal behaviour. A naïve cohort was used to assess the moter function capacity in the rotarod test. Additionally, two naïve cohorts were tested in further experiments to investigate effects on spatial learning and memory in more detail in the water maze using an extended learning protocol and in a touch screen based spatial reversal test.

Results

In LMA and V-maze, NRX1-KO rats showed a significantly higher total distance travelled compared to WT rats ($p<0.005$), and this was more pronounced in female compared to male rats. In the rotarod NRX1-KO rats performed slightly better than WT's, and females slightly better than males. PPI only revealed effect of sex, but not of genotype. In the contextual FCT, we found significant effect of genotype during the conditioning phase: the NRX1-KO rats did not respond with a freezing response to the electrical foot shock. However, when confronted again with the same context 24h later, both WT and KO rats showed the same degree of freezing. In the V-maze test NRX1-KO rats show a decrease in the discrimination index, a measure of short term memory function, vs. WT rats. This deficit in short term spatial memory was confirmed in the MWM test, were NRX1-KO rats showed severely impaired spatial learning: after 3 sessions per day for 4 days, NRX1-KO rats were still not able to find the hidden platform, whereas WT rats did very well in this time frame.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

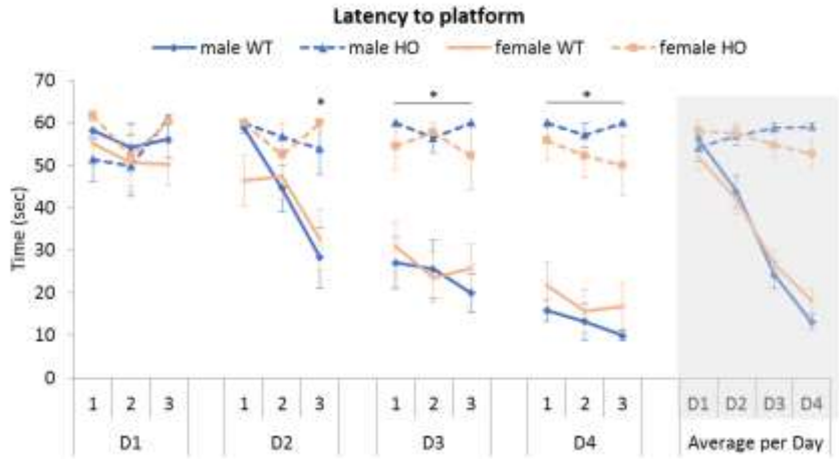


Figure 1: Water maze latency to reach the hidden platform shows normal learning by WT rats while NRX1-KO's (HO) fail to learn the platform location (meanSEM, n=7-10/group). ANOVA shows no effect of sex (p=0.3091), significant overall effects of genotype are seen (* p<0.0001)

We decided to further investigate the impairment in spatial learning and memory more in depth. Therefore, a new cohort of animals was evaluated in the water maze test using a more intense learning protocol to facilitate learning and memory formation: we increased the number of learning days to 8 and the number of trials per day to 4. This was followed by a reversal learning phase of 4 days where the platform was relocated to the opposite quadrant of the maze. Using this more intense learning protocol, NRX1-KO rats showed a slight improvement in learning, but still underperform vs. WT. In the subsequent reversal learning phase both groups show a learning curve, but NRX1-KO rats again perform worse compared to WT.

A third cohort of rats was evaluated in a reversal learning spatial separation paradigm in a touch screen set-up. When trained up to maximal performance at medium separation, NRX1-KO rats make less reversals and conversely need more trials per sessions, while response and collection latencies are slightly but significantly impaired vs WT rats. This difference is also found with minimal and maximal pattern separation.

Conclusion

The data from this study show that NRX1-KO rats show a modestly increased spontaneous explorative behavior compared to WT, and that this associated with slightly improved motor skills as observed in the rotarod. A surprising finding was that NRX1-KO rats did not show freezing behavior after having received a foot shock, but did so when exposed again to the same context 1 day later. This suggests that NRX1-KO rats perceived the environment as aversive and thus appear to have consolidated this memory, but it remains unclear why they did not display innate freezing behavior when exposed to the foot shock. The observation that there was no effect of genotype in PPI suggests that neurexin may not be involved in sensorimotor gating response.

The main finding in this study is that NRX1-KO rats display impaired memory function, both in terms of working memory as memory consolidation as revealed by the V-maze and water maze tests respectively. The lower performance in the spatial reversal task suggests that NRX1-KO's have impaired flexibility to adapt the responses to a changing rule. It is likely that deficient synapse formation and function because of NRX1-deficiency underlies these memory deficits. This provides an interesting model to evaluate pharmacological interventions to enhance cognitive function via modulation of synaptic plasticity.

References

Esclassan, F, Francois, J, Phillips, K G, Loomis, S, Gilmour, G. (2015). Phenotypic characterization of nonsocial behavioral impairment in neurexin 1α knockout rats. Behavioral Neuroscience 129:74-85.

The role of language in Neuropsychological rehabilitation of patients with developmental disorders

Pedro Ferreira Alves, Tâmara Rodrigues and Claudia Tirone

Portuguese Association of Relational-Historic Psychology; Portugal; psicologia.relacionalhistorica@gmail.com

Inspired by Luria's research, this paper aims to underline the basic function of language in clinical neuropsychological rehabilitation when inserting this novel form of behavior and mental structure. According to Luria (1959, 1966) through the word you can change the natural strength of stimuli, i.e. modify the well-known force of "Pavlov's rule". Our experimental research aims to show that spontaneous progress in development may be modified when initiating this organizer which affects impulses and motor reactions. To illustrate the decisive influence of language in shaping and structuring perception, attention, memory, planning and action, a sample composed of three Asperger children were given tasks using the "Piano-Card" as instrument (Quintino Aires, 2016), and employing different levels of difficulty with or without the support of language.

We used the Observer Software Program to collect and analyze observational records of performance understood as data to be classified according to formerly specified codes. These codes were built in order to describe the states the children and adolescent observed were in (either planning their activity or verifying it after carrying it out, or evading the task), the mistakes made and their nature, the need for help by examiner, and the total duration for a specific task. In this way it is possible to assess the point of socio-emotional development, this is to say, if speech has already acquired a regulating function for the children's and adolescent's behavior. This, in turn, allows psychologists and other mental health professionals to establish personalized intervention plans, attending to the capacities, difficulties and possibilities of each individual.

References:

1. Luria, A. R. (1959). The directive function of speech in development and dissolution. *Word: Journal of the International Linguistic Association*, **15**(3), 341-352. https://doi.org/10.1080/00437956.1959.11659703
2. Luria, A. R. (1966). *El papel del lenguaje en el desarrollo de la conducta*. Buenos Aires, Argentina: Ediciones Asociadas S.C.A.
3. Quintino Aires, J. M. (2016). Vigotsky-Luria na psicologia clínica no século XXI. In *De las neurociencias a la neuropsicología - El estudio del cerebro humano* (1st ed.). (Vol.1, pp. 301-385). Barranquilla, Colombia: Ediciones Corporación Universitaria Reformada

Measuring perseverative behaviors in lambs (Ovis aries): position habit acquisition and reversal learning task in the V-detour and T-maze tests

W. Florek, K. Barłowska, Ł. Wojciechowska, K. Horbańczuk, G.R. Juszczak, J.A. Modliński and S. Sampino.

Institute of Genetics and Animal Breeding of the Polish Academy of Science, Jastrżebiec (Magdalenka), Poland. s.sampino@ighz.pl

Introduction

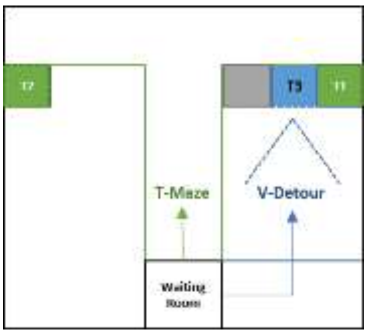
Insistence on sameness, inflexible adherence to routines, as well as restricted and fixated interest on habits, are key diagnostic symptoms of some neurodevelopmental diseases, including autism spectrum disorders (ASD) [1]. Perseveration is characteristic for children with ASD and is more evident when engaging in complex reversal learning tasks and in the presence of social disabilities [2-4]. Modeling these symptoms in animals has become essential to understand etiology and to develop therapeutic strategies. A variety of high-throughput assays to investigate social deficits and lower order forms of repetitive behaviors have been developed in rodents during the last decades [5-8]. However, mice have scarce face validity in case of higher order forms of repetitive behaviors, including perseveration and inflexibility. Particularly challenging is modeling high order repetitive behaviors in the context of accompanying social and communicative disabilities.

Aim

In the present study, we propose the lamb as a valuable translational model for children perseverative behaviors and inflexibility. We examined the behavior of 40-60 days old lambs (Ovis aries) facing a left/right position habit acquisition training followed by a reversal learning task in the V-detour and the T-maze tests, using the mother as social motivation. The aims of the study were: (i) to understand how lambs are able to solve the mazes and to remember the accomplished tasks; (ii) to examine the ability of lambs to change the acquired learned habit in response
to new environmental conditions; and (iii) to detect the occurrence of perseverative behaviors under a defined social separation paradigm.

Methods

Twenty-six dam-lamb couples were housed in groups of 4-5 pairs, and then transferred to the testing arena at the moment of the V-detour or T-maze tests (Figure 1). In both tests, each dam-lamb pair was first isolated from the flock mates into a $1m^2$ "waiting room", for 5 min. Subsequently, the dam was transferred into one of the "target" zones, T1, T2 or T3, depending on the test/trial, and the lamb was left isolated for further 5 min, in order to increase motivation to reunite with the dam. Lambs were then introduced to the V- or T-maze arenas and let to freely reach the dam using a left or right path for a defined amount of trials. Subsequently, lambs were challenged with
a reversal learning task in which the path previously chosen was not accessible (in the V-detour test), or the target position of the dam was switched to the opposite side (in the T-maze test). The subjects had to inhibit the initial learned response and readapt their spatial navigation pathways in order to reach the dam and obtain the social reward.



R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Figure 1. Schematic representation of the testing arena. In both tests lambs were isolated into the waiting room and the dams were transferred into one of the target zones, depending on the test/trial (T1-T2 in T-maze, T3 in V-detour tests). Subsequently, the lamb was transferred to either the T-maze (in green) or the V-detour (in blue) testing arenas. Dotted lines represent semitransparent fences.

V-Detour Test

At the age of 40 days, lambs were subjected to the V-detour test. After isolation, the lamb was transferred into the center of a V-shaped semitransparent maze, which was located just in front of the target zone (T3), adjacent to the Proximity Zone (Figure 2a). In each trial of the day 1, the lamb was let to reach the dam by detouring from a freely chosen side of the V-shaped fence. After every trial, animals were rewarded with 1 min dam-reunion period. The test was considered successfully completed when the lamb has reached the dam moving along the same side for three consecutive trials. On day 2 (3 days later), for the first three trials lambs were exposed to the same situation as in day 1 in order to reinforce the learned habit. In three subsequent trials the side previously chosen was closed with a semitransparent gate and the lamb had to detour the maze using the opposite path.

T-Maze Test

At the age of 60 days, lambs were subjected to the T-maze test. After isolation, the dam was transferred to either the left or right target zone, on the extremity of one of the lateral arms (Figure 2b). On day 1, the lamb was transferred into the "start" arm, being free to reach the dam for five consecutive trials. Each trial ended with 1 min reunion with the dam. On day 2 of the test (next day), for the first five trials the position of the respective dam was maintained as in day 1, to reinforce the learned habit. In the subsequent five trials, the target position of the dam was switched to the opposite arm of the maze, and the lamb had to reunite with the dam using the opposite path.

Outcome Measurements

All trials were video-recorded and subsequently analyzed using the automatic video-tracking software Noldus EthoVision® XT 11.5 (Figure 2). The total time spent by the lambs to complete the tasks, as well as the frequencies of entries into the different zones of the mazes were analyzed using Kruskal-Wallis test.

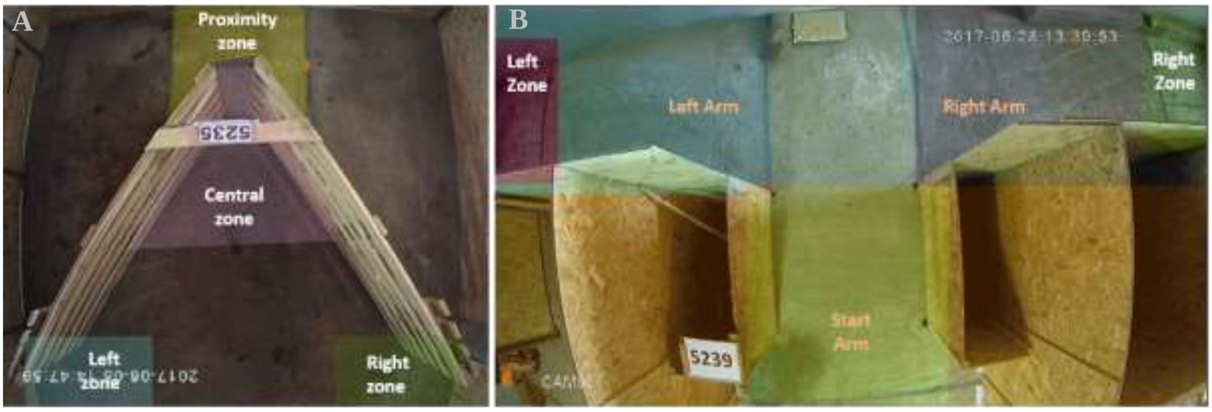


Figure 2. Photographic representation of the V-detour and T-maze arenas and the zones setting for Ethovision® XT analysis. (**A**) For the V-detour test, four different zones were defined. In each trial, the software calculated the total time spent by the lamb to reach the Proximity zone, as well as the frequency of entries into either the Central, Left and Right zones. (**B**) For the T-maze test, five different zones were defined. In each trial, the software calculated the total time spent by the lamb to reach the Left/Right zones adjacent to the dam, as well as the frequency of entries into either the Start, Left or Right Arm.

Preliminary Results

Position habit acquisition

All lambs were able to quickly learn how to reach the target zones in order to reunite with their dams. There was an effect of the trial on the time spent to reach the aim ($P<0.001$), with the maximum time spent on trial 1 of the day 1 in both tests, and a gradual tendency to solve the mazes within few seconds as further trials were performed. Subsequently, all lambs easily memorized how to solve the mazes within few seconds while subjected to similar trials on day 2, in both the V-detour and T-maze tests.

Reversal learning

All lambs showed worse performances when challenged in the reversal learning trials – i.e. when the previously chosen path was not accessible (in the V-detour test); or the target position of the dam was switched to the opposite side (in the T-maze test). The time spent to reach the aim in those trials was significantly higher compared to previous trials in all lambs analyzed ($P<0.01$). Moreover, we observed a variable degree of perseveration in the first reversal learning trial, measured as the frequency of entries into the previously learned paths in both the V-detour and T-maze tests ($P<0.001$).

Discussion and Conclusions

Perseveration is a key problematic behavior found in autistic persons and those with other developmental disorders. We designed a high-throughput battery of behavioral tests to examine this particular form of repetitive behavior in lambs. Our assays can be useful to study learning abilities, working memory, social attachment, and behavioral inflexibility in sheep models of human neurodevelopmental disorders. The development of new animal models is an important issue for the study of ASD etiology, as well as the establishment of new therapeutic and preventive strategies. The lamb has a big potential to become an efficient translational model, being a good alternative to rodents and primates.

Ethical Statement

All procedures were performed in accordance with the European Community regulation 86/609 and Polish laws for the use of animals in research. Experimental procedures were approved by the 2nd Local Ethical Committee of Warsaw (approval no.: WAW2/35/2017).

Funding

This research was funded by the Leading National Research Centre Scientific Consortium "Healthy Animal-Safe Food" (KNOW Ph.D. grant no. KNOW2015/CB/PRO1/47).

References

1. American Psychiatric Association. (2013). The *Diagnostic and Statistical Manual of Mental Disorders,* 5th ed. Washington, DC.

2. Coldren, J.T. and Halloran C. (2003). Spatial reversal as a measure of executive functioning in children with autism. *J Genet Psychol* **164**, 29-41.
3. Lionello-Denolf K.M., Mcllvane W.J., Canovas D.S., de Souza D.G., Barros R.S. (2008). Reversal learning set and functional equivalence in children with and without autism. *Psychol Rec* **58**, 15-36.
4. Yerys B.E., Wallace G.L., Harrison B., Celano M.J., Giedd J.N., Kenworthy L.E. (2009). Set-shifting in children with autism spectrum disorders: reversal shifting deficits on the Intradimensional/Extradimensional Shift Test correlate with repetitive behaviors. *Autism* **13**, 523-538.
5. McFarlane H.G., Kusek G.K., Yang M., Phoenix J.L., Bolivar V.J., Crawley J.N. (2008). Autism-like behavioral phenotypes in BTBR T+tf/J mice. *Genes Brain Behav* **7**, 152-163.
6. Pearson B.L., Pobbe R.L., Defensor E.B., Oasay L., Bolivar V.J., Blanchard D.C., Blanchard R.J. (2011). Motor and cognitive stereotypies in the BTBR T+tf/J mouse model of autism. *Genes Brain Behav* **10**, 228-235.
7. Pasciuto E., Borrie S.C., Kanellopoulos A.K., Santos A.R., Cappuyns E., D.Andrea L., Pacini L., Bagni C. (2015). Autism Spectrum Disorders: Translating human deficits into mouse behavior. *Neurobiol Learn Mem* **124**, 71-87.
8. Kas M.J., Glennon J.C., Buitelaar J., Ey E., Biemans B., Crawley J., Ring R.H., Lajonchere C., Esclassan F., Talpos J., Noldus L.P., Burbach J.P., Steckler T. (2014). Assessing behavioural and cognitive domains of autism spectrum disorders in rodents: current status and future perspectives. *Psychopharmacology (Berl)* **231**, 1125-1146.

Momentary Leadership and Collective Motion in Black Neon Tetra (*Hyphessobrycon herbertaxelrodi*) Schools

E. Gimeno[1], F.S. Beltran[1], R. Dolado[1] and V. Quera[1]

[1]Institute of Neuroscience, University of Barcelona, Barcelona, Spain. ruth.dolado@ub.edu

Fish schools need to reach a consensus about their travel directions, such as when searching for food or migrating. Decision-making in such systems can be studied by analyzing the effects of leadership on collective motion, where a leader is defined as an individual who initiates a movement towards a certain direction and is followed by the other group members [1]. Empirical studies have found that leadership in fish schools can emerge from individual differences in body size [2], hunger or food motivation [3], exploratory tendency or boldness [4] and knowledge or training experience [5]; all these factors can also interact with one another [3, 6]. However, some authors state that leadership in fish schools is temporary and can occur as a result of a self-organized process based on the interactions among the individuals [7-9].

We conditioned a fish school of black neon tetra (*Hyphessobrycon herbertaxelrodi*) to find food in a container in an experimental tank, while another group was allowed to swim freely with no particular goal. The individuals were similar in body size, task knowledge and hunger level in order to minimize sources of individual variability that could determine leadership. In the conditioned group, fish were given 10 minutes to find the food container in the tank, and the trial ended when all the individuals were inside the container. This process was repeated six times a day for eight days. The unconditioned group also received habituation trials to ensure that its exposure to the experimental tank was comparable to that of the conditioned group. The length of the habituation trials for the unconditioned group was determined by the mean daily latency the conditioned group took to reach the food container. Fish were not marked or tagged, since this invasive procedure was not necessary to determine leadership within each trial. This study was conducted according to the institutional guidelines for the care and use of laboratory animals established by the Ethics Committee for Animal Experimentation of the Universitat de Barcelona.

We used the video recordings made on the last two days of training (days 7 and 8) for both groups and obtained individual trajectories. Then, we compared their collective motion and used a method developed to measure leadership in pigeons to determine whether their motion was leaderless [10]. Nagy's method provides momentary correlational delays (i.e., time intervals that maximize correlations between movement vectors) at every time step of a trial [10], but we observed that at some time steps the differences between individual leadership scores were very small. In those cases, there was no clear leadership. We therefore computed the standard deviation of individual leadership scores for every time step ($SD\tau$) and the median of the standard deviations for the entire trial (Mdn_{SD}). We excluded the time steps that had a $SD\tau < Mdn_{SD}$ with the understanding that no clear momentary leader existed at those time steps (SD-filtered results). On the basis of momentany leadership scores, we computed the mean proportion of leadership duration (the average length of momentary leadership episodes divided by the total number of time steps taken into account) and the leadership persistence (the proportion of time steps in which a momentary leader is again the leader in the next time step). We also tested the relationship between the mean momentary leadership of each fish and its average distance from the school center and individual speed.

Our findings suggest that black neon tetras schools are leaderless societies, since they displayed an unstable leadership consisting of brief episodes of individual fish leading the group motion, at least in groups made up of individuals of similar size, knowledge and hunger level. Momentary fish leaders tended to be at the front; however, the tendency to lead did not necessarily stem from faster swimming speeds, and further research is needed to clarify whether individuals have a positional preference for being at the front or the rear of the school. Training affected only some aspects of the collective motion of black neon tetra, goal-driven schools being faster and more cohesive than spontaneous, goalless ones, but directional information transmission flowed from front to rear in the group, regardless of the training they had received.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

References:

1. Krause, J., Hoare, D., Krause, S., Hemelrijk, C.K., Rubenstein, D.I. (2000). Leadership in fish shoals. *Fish and Fisheries* **1**, 82–89.
2. Reebs, S.G. (2000). Can a minority of informed leaders determine the foraging movements of a fish shoal? *Animal Behaviour* **59**, 403–409.
3. Webster, M.M. (2016). Experience and motivation shape leader-follower interactions in fish shoals. *Behavioral Ecology* **12**, 77–84.
4. Harcourt, J.L., Ang, T.Z., Sweetman, G., Johnstone, R.A., Manica, A. (2009). Social feedback and the emergence of leaders and followers. *Current Biology* **19**, 248–252.
5. Ioannou, C.C., Singh, M., Couzin, I.D. (2015). Potential leaders trade off goal-oriented and socially oriented behavior in mobile animal groups. *The American Naturalist* **186**, 284–293.
6. Reebs, S. (2001). Influence of body size on leadership in shoals of golden shiners, *Notemigonus crysoleucas*. *Behaviour* **138**, 797–809.
7. Couzin, I.D., Krause, J., Franks, N.R., Levin, S.A. (2005). Effective leadership and decision-making in animal groups on the move. *Nature* **433**, 513–516.
8. Pavlov, D.S., Kasumyan, A.O. (2000). Patterns and mechanisms of schooling behavior in fish: A review. *Journal of Ichthyology* **40**, S163–S231.
9. Radakov, D.V. (1973). *Schooling in the Ecology of Fish*. New York: Halsted Press.
10. Nagy, M., Vasarhelyi, G., Pettit, B., Roberts-Mariani, I., Vicsek, T., Biro, D. (2013). Context-dependent hierarchies in pigeons. *Proceedings of the National Academy of Sciences* **110**, 13049–13054.

Validation of an automated system for monitoring nocturnal activity in group-housed monkeys

J. K. Green[1,2] and C. L. Witham[1,2]

1 Institute of Neuroscience, Newcastle University, Newcastle-upon-Tyne, UK. j.k.green1@ncl.ac.uk; 2 Centre for Macaques, MRC Harwell, Salisbury, UK.

Background

Sleep and nocturnal activity are rarely monitored in captive primates, despite accounting for half of their daily lives. Studies in humans and other species show that sleep disruption can be related to stress. Recent improvements in camera technology and reductions in price have made it easier to record nocturnal behavior, however it is very time-consuming to manually analyze video footage. It is essential to have automated methods of analyzing this footage for the monitoring of nocturnal activity to become standard. For this project we have developed and validated a simple method for measuring movement in group-housed Rhesus macaques, with a particular focus on monitoring welfare in laboratory macaques.

Developing an Automated System

We used high definition cameras and infra-red lighting at a breeding colony of Rhesus macaques to film the monkeys between 19:00 and 07:00 hours. The monkeys were housed in twenty different mixed sex groups in the colony (7-22 animal per group; age range 0-22 years). A total of 450 nights of footage was filmed from 20 different groups. For the automated analysis we took the absolute difference in pixel intensity between two frames (inter-frame interval of 0.25s) and calculated the mean, variance, skewness and kurtosis across the image [1]. We then applied a baseline correction to the data to allow comparison between different cameras and smoothed the corrected data (with a moving average) to improve the signal to noise ratio.

Validating the Automated System

From the 450 nights of footage 31 one-hour periods of footage of eight different groups was selected to develop and validate the automated scoring method. We manually scored the footage for the start and end of each bout of movement. A bout was considered finished when there had been no movement for three seconds. Both manual and automated data was converted to 1s intervals. For the manual data each interval was scored either 0 (no movement) or 1 (movement). A K nearest neighbor classification model (K=10) was fitted to 5000 data points taken from eight of the one-hour periods. We used the other 23 one-hour periods to validate the model, achieving a classification accuracy of 94%. Figure 1A shows the correlation between the manual and automated scores; each score is presented as the % of the intervals with movement across the individual one-hour periods (the r-square value for this data is 0.989). We used a Bland-Altman plot to explore this correlation in more detail (Fig. 1B).

Application

To show a potential application of the model we investigated the seasonality of group sleep onset time. This was done by applying the model to footage from the 450 nights and calculating the group sleep onset as the first 3 minute period with no movement. The mean group onset time varied from 20:30 in winter to 22:45 in summer, showing a very clear peak in summer corresponding to the later sunset at that time of year (Figure 1C).

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

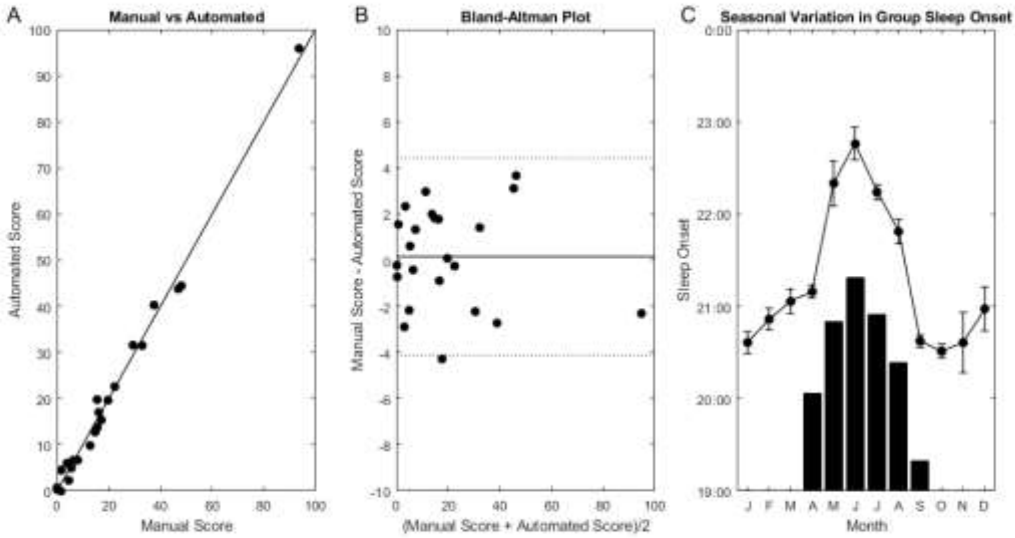


Figure 1: (A) Correlation between manual movement score and automated movement score. Solid line shows line of unity. (B) Bland-Altman plot for results in A. Solid line shows mean difference and dotted lines show mean ± 1.96 standard deviations. (C) Seasonal variation in group sleep onset. Error bars show mean sleep onset ± standard error of mean, black bars show average time of sunset for each month.

Conclusion

Our proposed method is very simple to apply and gives a good approximation of the level of movement at group level. It could in principle be used to monitor any captive species where it is possible to film under infra-red light and where it is possible to get good coverage of the enclosure (it would be difficult in enclosures with a lot of solid barriers or large amount of foliage).

Ethical statement

The study was conducted at a Home Office licensed UK Rhesus macaque breeding center and involved non-invasive techniques only. The project proposal was approved by the local Animal Welfare and Ethical Review Board. The housing and husbandry of the animals in the colony complies with both the Home Office Code of Practice and the ARRIVE guidelines [2].

References

1. Dawkins, M. S. (2003) ‘Behaviour as a tool in the assessment of animal welfare’ Zoology **106**: 383-387
2. Kilkenny, C. Browne, W. J. Cuthill, I. C. Emerson, M. Altman, D. G. (2010) ‘Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research’ PLoS Biol **8**: (6)

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Voluntary wheel running behavior as a measure for severity assessment

C. Häger[1], L.M. Keubler[1], S.R. Talbot[1], S. Biernot[1], N. Weegh[1], M. Buettner[1], S. Glage[1] and A. Bleich[1]

Institute for Laboratory Animal Science and Central Animal Facility, Hannover Medical School, Hannover, Germany. Bleich.andre@mh-hannover.de

An evidence-based graded severity assessment is pivotal in laboratory animals during scientific procedures, not only regarding the legal obligations and the demand for standardized high-quality data, but also with regard to ethical justification. However, scales grading severity in laboratory animals that incorporate specific parameters of the respective experimental model and that are furthermore scientifically sound and routinely applicable are still scarce. Particularly, there is a lack of objective standardized severity assessment parameters that are non- or minimally-invasive and easily applicable.

By inducing a graded intestinal inflammation via application of increasing doses of DSS (dextran sodium sulfate; 1% and 1.5%) or water only (0% DSS) differentiated manifestations of colitis were analysed in terms of clinical signs (general appearence, general activity, body weight), histologies and by monitoring of voluntary wheel running (VWR) behaviour using a fully automated home cage based system. 10-13 week old, female C57BL/6J (B6) mice were single housed in home cages with free access to a running wheel (Revolyzer® 3TS system, software DASY Lab 11.0 preclinics GmbH, Germany) that allowed automatic and undisturbed 20 hour monitoring of wheel rotations (WR_{20}) from 12 pm to 8 am daily, leaving a 4 hour interval for general maintenance and experimental manipulations. For determination of the steady state running performance an adaption phase of 14 days was chosen before experiments. For subsequent analysis the mean of the last three days of the respective adaption phases was set as baseline to calculate the relative change in %.

Clinical scoring indicated marginal signs of compromised welfare, whereas reduced running performance was directly related to the induced colitis grade. Additionally, VWR-behaviour was influenced by serial blood sampling. Moreover, in a mouse model of restraint stress (immobilization for 1 hour on 10 consecutive days) a reduction of VWR was observed in restraint-stressed mice and in corresponding control mice undergoing repeated faecal sample collection (sampling for 2 hours in a seperately cage). An unsupervised k-means algorithm based cluster analysis of body weight and wheel running data from the colitis model enabled the discrimination of cluster borders to define distinct levels of severity. Further testing of the cluster model with data from the restraint stress model enabled an unbiased individual severity assessment. The integration of VWR-behaviour in a k-means based cluster model provides a novel approach suitable for individual severity grading in laboratory mice. Particularily with regard to the 3R´s principle, this approach have to be evaluated under group housing conditions.

Ethical statement

This study was conducted in accordance with the German law for animal protection and the European Directive, 2010/63/EU. All experiments were approved and permitted by the Lower Saxony State Office for Consumer Protection and Food Safety (LAVES, license 15/1905).

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Measuring Prosocial Behavior in Rats

J. Hernandez-Lallement, V. Gazzola, C. Keysers

Social Brain Lab, Netherlands Institute for Neuroscience, Amsterdam, The Netherlands; j.hernandez-lallement@nin.knaw.nl

Introduction

Showing what is commonly referred to as prosocial behaviors is appreciated and rewarded in human society, whereas anti-social actions often lead to isolation and reclusion. How can we scrutinize the brain processes associated with behaviors resulting from the perception of other's emotions? While neuroimaging can explore vicarious activations using correlational approaches, it cannot influence neuronal activity (hence limiting insights into causality), and have a limited spatial resolution. Here, we present a rodent model of empathy-driven prosocial behavior, suited for cutting-edge neuroscientific manipulations, which scrutinizes the response to other's distress in the form of prosocial (help) or anti-social behaviors (aggression).

Methods

We train rats to initiate a trial by performing a nosepoke which triggers the extension of two levers. In the training phase, pressing either lever delivers n = 1 sucrose pellets. Once behavior is acquired, one lever is modified so that it requires twice as much strength to be operated. Once preference for one of the lever (expected to be the easy lever) is developed, rats are promoted to the testing phase where pressing the preferred lever leads to an electric shock delivered to an adjacent partner. Accordingly, if other's distress carries a negative value exceeding the cost of higher effort for some individuals (prosocial) but not for others (aggressors), the first group should switch preference while the second should stick with the lower effort option. Importantly, the second group of individuals provide means to study instrumental aggression, i.e., hurting others to obtain food. We tested several types of costs (magnitude, delays, effort), different strains of rats (Long Evans, Sprague Dawley) as well as different gender in the propensity to behave pro-socially to avoid the pain of others.

Results

We found that a subset of animals switched their previously acquired preference upon association of the preferred lever with a shock to a conspecific. Since most animals preferred the easy lever in no shock sessions, switching to the hard lever can be interpreted as a cost, and hence as pro-social behavior. Interestingly, a subset of animals kept their initial preferences, although these animals were able to switch levers in a subsequent control where the previously non preferred lever was associated with higher reward magnitude.

Conclusion

This behavioral paradigm provides means to scrutinize both pro- and anti-social behaviors. Indeed, while the ***first group*** allows examining how the perception of other's distress **promotes pro-social behavior**, the ***second group*** enables the exploration what predisposes some individuals to **disregard the distress of others.**

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Applying Home Cage Behaviour Observations as a Pain Assessment Assay in a Collagen Induced Arthritis Rat Model.

S.F. Izzard[1] and J. Cruden[2]

In-vivo In-vitro Technology, GlaxoSmithKline, Stevenage, United Kingdom. sam.f.izzard@gsk.com

Background

Rheumatoid Arthritis causes inflammation of the joints. It is an autoimmune disease where your own immune system attacks your own body. Symptoms consist of joint pain, swelling, stiffness, tiredness, depression, irritability and flu like symptoms. Around 2 million people in the UK suffer from rheumatoid arthritis, and about 75% of these are women. Onset commonly begins between the ages of 40-50, but can start at any age (Hertsmstherapy.org.uk, 2018). Experimental models of rheumatoid arthritis have contributed immensely to understanding the pathogenesis as well as the treatment of this debilitating disease.

The rat Collagen Induced Arthritis (CIA) model can be used to unravel mechanisms involved in the development of arthritis and are an effective model to study the effect of new therapeutics. There are various treatment strategies from the use of conventional disease modifying anti-rheumatic drugs to biologicals. Unfortunately, not all patients respond to existing treatment and so there is still a strong need for new treatments. All animal studies were ethically reviewed and carried out in accordance with Animals (Scientific Procedures) Act 1986 and our internal Policy on the Care, Welfare and Treatment of Animals. The aim of this work is to optimize the animal model for CIA, and to minimise animal suffering.

Method

Sixteen female Lewis rats, from Charles River Laboratory, housed in groups of four per cage, were allocated to the behavior observation study. On arrival to the animal room the rats were checked carefully and acclimatised for seven days prior to the start of the study. CIA was induced in three groups and the fourth group acted as a control. (no CIA). Rats were housed in clear Polysulphone plastic cage (see Figure 1) with a raised stainless steel bar lid measuring 58cm long, 38cm wide, 22cm high. The cage had 2cm depth of wood shavings, and an aspen wood chew block 2cm x 2cm x 10cm long (Datesand) with a handful of soft shredded tissue for nesting and suspended Maxi Fun Tunnels 15cm long, 8cm circumference (LBS), and a red shelf (Rosper Engineering).

Figure 1; cage with enrichment devices.

Rats were fed ad libitum with 5LF2 PMI Eurodent diet from PMI International Certified LabDiet®, and laboratory grade animal drinking water was available via water bottles attached to the cages. Cage bases and trays were replaced weekly, disposable enrichment replaced as required and the racks were swapped for new racks fortnightly. Throughout the facility there was piped radio tuned into a local commercial station during the light phase. The light cycle was on a 12:12 light/dark with light phase from 06.00h – 18.00h and a gradual increase or decrease of lighting over a ten-minute dawn; dusk period. Video recordings were made in conjunction with the welfare and study observations made during the working day.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Low level light CCTV cameras and web video CCTV video manager software equipment were supplied by Tracksys; the video software was from Noldus. The night vision camera had an infra-red light source; this allowed recording of rats' movements during active period hours (lights off) without undue disturbance. The PC screen was only switched on when the camera views needed to be checked to ensure they had not been moved during any room duties (normally at 16.00h on test days for no more than 10 minutes); at all other times the screen was turned off. Each cage was videoed daily for six hours for the duration of the study for each group of rats (up to 21 days). The first hour of footage from the start of the dark phase, was sampled, as an initial scan throughout the footage indicated this was the period the rats appeared to be the most active which has been observed in other studies of laboratory rats (e.g. [1]Hurst et al, 1997).

Behavioural data was collated from video records (see Figure 2) using fixed interval instantaneous focal sampling, with a note of activity, posture and location in the cage made at one minute intervals from 18.05h – 19.05h which equated to the first three hours of dark phase. This equates to a total of 36 scans per rat per recording session for location, position and activity. The behavioural observations included, rearing, social interactions, grooming and locomotion.

Figure 2; Screen shot of pre-arthritis and onset of arthritis. Top left cage are the control animals.

The behavioural ethogram used was developed in part from the following published ethograms: [2]Hurst et al (1997), [3]Abou Ismail et al (2010), and cage side observations to define rat activity during scan sampling. For the purpose of instantaneous sampling, and to keep some consistency in the project, an activity was defined as what an animal was doing whilst it was mobile or stationary, posture was defined as the exact deportment of the body and, location of the rat in the cage, which included whether they were under the shelf or in the tunnel.

In this study we evaluate the preliminary results comparing the difference in behaviour prior to and during the onset of arthritis.

References

1. Hertsmstherapy.org.uk. (2018). *Osteoarthritis and Rheumatoid Arthritis | HMSTC Letchworth | HMSTC*. [online] Available at: http://www.hertsmstherapy.org.uk/arthritis/ [Accessed 6 February 2018].
2. Hurst, J. L., Barnard, C. J., Nevison, C. M., and West, C. D. (1997) Housing and welfare in laboratory rats: Welfare implications of isolation and social contact among males. *Animal Welfare,* 6 329-347.
3. Abou-Ismail, U. A., Burman, O. H. P., Nicol, C. J., and Mendl, M. (2010) The effects of enhancing cage complexity on the behaviour and welfare of laboratory rats. *Behavioural Processes*, 85 172-180.

Use of Cortisol Measurement and Novel Tank Testing to Evaluate Stress and Anxiety Responses to Exercise Training in Zebrafish

A.B. Jones

Department of Health Professions, Northeastern State University, Broken Arrow, OK, USA. Jones70@nsuok.edu

While it is well known that exercise is generally beneficial, it is becoming increasingly important to understand the psychological and physiological consequences (both positive and negative) to individuals who engage in exercise. Exercise regimens tailored to the individual serve to provide optimal benefits while reducing the potentially stressful effects of exercise (ex: overtraining, burnout, or injury). Personalized, or precision, medicine tailors interventions, including exercise prescriptions, to the individual based on their predicted response or risk of disease. This idea of precision exercise training to alleviate stress and anxiety is especially important because the dose (duration, intensity) of exercise that provides benefits is very individualized.

Zebrafish (*Danio rerio*) are becoming increasingly involved in experiments involving stress research as well as those studies investigating the benefits of exercise. This may partly be due to the zebrafish stress-integrating system: it is organized as a hypothalamic-pituitary-interrenal (HPI) axis compared to the mammalian hypothalamic-pituitary-adrenal (HPA) axis, and there are many homologies. Like mammals, zebrafish respond to a variety of stressors, such as handling, heavy metals, organic pollutants, rapid temperature changes, predator confrontations, confinement/restraint, overcrowding, and novel environments, through increased cortisol levels released through the gills into the water. Cortisol is also the main corticosteroid released in humans during times of stress.

The protocol described here uses behavioral testing and physiological measurements to determine stressfulness of a given exercise regimen in an individual zebrafish. Adult, wild-type male and female zebrafish were individually housed and exercised at a moderate intensity (approximately 300 L/hr current) using a swim tunnel (see Figure 1). Prior to exercise training, each fish was evaluated for baseline anxiety-like behaviors using the novel tank test (NTT) and stress hormone levels by cortisol excretion. Fish were exercised at 10 minutes, then 20 minutes, and finally 30 minutes, 3 times per week. NTT and cortisol levels were measured after 2-4 weeks at each exercise duration before it was increased. Control fish were placed in the swim tunnel for the corresponding amount of time but it was not running. After completion of all exercise training, fish were again measured for anxiety-like behaviors and cortisol concentrations to determine the stressfulness of the training regimen. Anxiety-like behaviors were reported based on the percentage of time spent in the upper part of the novel tank during a 5-minute protocol. Increased time spent in the upper half of the novel tank compared to baseline indicated reduced anxiety-like behaviors while decreased time spent in the upper section indicated increased anxiety-like behaviors. Free cortisol concentrations were collected from small holding water samples and analyzed using EIA (Cayman Chemical, USA).

This combination of methodologies allows for evaluation of both behavioral and physiological responses to a given bout of exercise training. These behavioral and physiological measurements were analyzed to determine whether an individual fish responded positively or negatively to a given exercise protocol and training could then be adjusted to reach a beneficial level.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Figure 1. Individualized exercise training in a swim tunnel.

The study described here was approved by the Northeastern State University Institutional Animal Care and Use Committee and was conducted in conformity with ethical and humane principles of research.

Action Detection from Egocentric Videos in Daily Living Scenarios

G. Kapidis[1, 2], E. van Dam[1], R. Poppe[2], L. P. J. J. Noldus[1], R. C. Veltkamp[2]

1 Noldus Information Technology, Wageningen, The Netherlands; 2 Department of Informatics and Computer Science, University of Utrecht, Utrecht, The Netherlands
{georgios.kapidis, elsbeth.vandam, lucas.noldus}@noldus.nl, {r.w.poppe, r.c.veltkamp}@uu.nl

Introduction

Human Action Recognition (HAR) as a computer vision subfield has seen impressive progress due to the growth of deep learning based methods. The field of egocentric vision has received significant attention from the activity recognition community as the amount of relevant work indicates ([1]–[8]). Recent detection methods rely on neural network structures to describe the content of video frames as well as the sequential relations in terms of objects [2], [7], activities [1], [3], [4], [6], locations [5], [7], interactions [8] or combinations of different scene aspects. In the domain of activities of daily living and specifically in that of older adult healthcare, egocentric vision is becoming acknowledged as a means of activity analysis and understanding [9]. A nice review for egocentric activity recognition can be found in [10].

In this work, we focus on HAR from an egocentric video perspective for the analysis of indoor scenes and activities. We are using Long Short-term Memory (LSTM)[11] a deep learning method designed for sequential data that allows us to analyze videos in the temporal domain, efficiently making use of the inherent structure that exists in a sequence of video frames. First experiments for distinguishing between standing/sitting/no-action classes in the Activities of Daily Living (ADL) dataset [12] reveal promising results towards classification of the performed actions of the camera wearer.

Methodology

We are following the example of [5] to produce a dataset of objects that can be classified into activity groups for each video frame. We do this by applying a state-of-the-art object detector on the videos comprising the ADL dataset, saving the detections and using them to produce train and test splits for the LSTM training scheme.
More specifically, we use the YOLO (You Only Look Once) object detector [13] trained to detect the ADL20 set of objects from [5] which consists of typical objects found in an indoor scene e.g. fridge, oven, television, door. To train the LSTM we process the object detections per frame, into a binary vector with length the number of available objects (20 in this case) with ones in the indices of detected objects for the current frame and zeros otherwise. Since the LSTM is able to model the temporal dynamics when having a sequence as input, we take advantage by using sets of five frames. Furthermore, we change the frame rate to a single frame per second (1 fps) with the effect of enlarging the period covered from one training example from 33 milliseconds to 1000, without complicating the input by having a very long sequence. For our training scheme, a single LSTM layer is sufficient to model the three proposed action groups. The action groups are the result of mapping the action classes proposed in [12]. The activity mapping is shown in Table 3.

Table 3: ADL action classes mapped to standing and sitting action groups with the addition of the background class for the frames that have been annotated as having no occuring activity. In the 3rd and 4th row the frames for each class in the train and test sets, respectively.

Group	No-action	Standing	Sitting
Activity	Background	Combing hair, make up, brushing teeth, dental floss, washing hands/face, drying hands/face, enter/leave room, adjusting thermostat, laundry, washing dishes, moving dishes, making tea, making coffee, drinking water/bottle, drinking water/tap, making hot food, making cold food/snack, mopping in kitchen, vacuuming, taking pills, making bed, cleaning house, using mouth wash, grabbing water from tap	Eating in kitchen, watching tv, using computer, using cell, reading book, writing, putting on shoes/socks, drinking coffee/tea

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Train frames	1208	6701	3734
Test frames	2890	10338	7791

Experiments, Results and Discussion

After experimenting with the training hyperparameters of the LSTM, we discovered the optimal values for the number of training iterations, learning rate, sequence size, mini-batch size and dropout as shown in Table 4. The network consists of one LSTM layer, followed by a fully connected layer without a non-linear activation so it can be used for inference. The loss function is weighted softmax cross entropy, with Adam [14] for optimization. The weighting of each class is an additional hyperparameter to account for the imbalance of the action groups in the train set.

Table 4: LSTM hyperparameters and architecture

Training steps (epochs)	1500 (~62)
Learning rate	10^{-4} to 10^{-6} in 100 steps with polynomial decay
Sequence size	5
Mini-batch size	96
Dropout	0.85
LSTM layers	1
Hidden units	20

Our single best execution in terms of overall accuracy has 73.08% of the frames in the test set classified correctly. It is important to note the imbalance of the test set as shown at the 4th row of Table 3. Per class results are shown in Table 5. The low recall in the no-action case does not affect the overall accuracy significantly, because the number of test samples is low compared to the other two classes.

Table 5: Per class recall and precision results, max is 1.

	No-action	Standing	Sitting
Recall	0.01	0.92	0.75
Precision	0.11	0.71	0.77

The preliminary results show that it is possible to classify the frames when a person is 'standing' versus when 'sitting' successfully based on the underlying objects and activities; however, the 'no-action' frames are easily mistaken as one of the other groups, due to the aforementioned class imbalance. Another reason is the objects that signify each action group and the background. The background does not depend on specific objects to be detected, rather on random occurrences of objects, which are characteristic for other actions. A 'fridge' and an 'oven' found in a 'no-action' frame, would suggest a food related action at test time, which belongs in the 'standing' group because the number of training samples with this label is greater and as such, distinctive of the 'standing' action group.

In future work we intend to study the objects that individually describe each activity and action group in order to identify which object classes (and lack thereof) represent specific actions.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676157.

References

[1] M. Ma, H. Fan, and K. M. Kitani, "Going deeper into first-person activity recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1894–1903.

[2] G. Bertasius, H. S. Park, S. X. Yu, and J. Shi, “First Person Action-Object Detection with EgoNet,” in *Proceedings of Robotics: Science and Systems*, 2017.
[3] K. Nakamura, S. Yeung, A. Alahi, and L. Fei-Fei, “Jointly Learning Energy Expenditures and Activities Using Egocentric Multimodal Signals,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6817–6826.
[4] Y. Poleg, A. Ephrat, S. Peleg, and C. Arora, “Compact CNN for indexing egocentric videos,” in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016, pp. 1–9.
[5] G. Kapidis, R. Poppe, E. van Dam, L. Noldus, and R. Veltkamp, “Where Am I? Comparing CNN and LSTM for Location Classification in Egocentric Videos,” in *SmarterAAL’18 Workshop, IEEE International Conference on Pervasive Computing and Communication (PerCom)*, Athens, Greece, 2018.
[6] G. Vaca-Castano, S. Das, and J. P. Sousa, “Improving egocentric vision of daily activities,” in *Image Processing (ICIP), 2015 IEEE International Conference on*, 2015, pp. 2562–2566.
[7] G. Vaca-Castano, S. Das, J. P. Sousa, N. D. Lobo, and M. Shah, “Improved scene identification and object detection on egocentric vision of daily activities,” *Comput. Vis. Image Underst.*, vol. 156, pp. 92–103, Mar. 2017.
[8] R. Yonetani, K. M. Kitani, and Y. Sato, “Recognizing micro-actions and reactions from paired egocentric videos,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2629–2638.
[9] S. Karaman, J. Benois-Pineau, R. Mégret, V. Dovgalecs, J.-F. Dartigues, and Y. Gaëstel, “Human Daily Activities Indexing in Videos from Wearable Cameras for Monitoring of Patients with Dementia Diseases,” in *2010 20th International Conference on Pattern Recognition*, 2010, pp. 4113–4116.
[10] T.-H.-C. Nguyen, J.-C. Nebel, and F. Florez-Revuelta, “Recognition of Activities of Daily Living with Egocentric Vision: A Review,” *Sensors*, vol. 16, no. 1, p. 72, Jan. 2016.
[11] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[12] H. Pirsiavash and D. Ramanan, “Detecting activities of daily living in first-person camera views,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2847–2854.
[13] J. Redmon and A. Farhadi, “YOLO9000: Better, Faster, Stronger,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6517–6525.
[14] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *ArXiv14126980 Cs*, Dec. 2014.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Uppsala University Behavioral Facility (UUBF) Has The Capability And Competence To Conduct Behavioral Studies In mouse, Rat and Fish

Asa Konradsson-Geuken, Klas Kullander, Erika Roman and Svante Winberg

Uppsala University, Sweden; asa.konradsson-geuken@farmbio.uu.se

Uppsala University Behavioral Facility (UUBF) is a non-profit core facility supported by the Faculty of Medicine and Pharmacy, Uppsala University, Sweden. UUBF's main aims are to provide administration- and organization services of behavioral tests for internal and external research groups. We offer a large array of behavioral tasks for mouse, rat and fish, e.g. tasks for exploration, motor behaviors, sensorimotor processing, learning, memory and behavioral profiling. UUBF provides equipment together with protocols for behavioral experiments, and assists with data analysis, interpretation, and advanced statistical analyses. We also provide assistance with writing of ethical applications and training and guidance in experimental design. Further, UUBF offers biannual graduate courses on animal behavior e.g. "How to study behavior in vertebrates with focus on fish and rodents". You are welcome to contact UUBF for discussing your future behavioral experiments.

Ethical Statement: the researchers using our facility are holding their own approved ethical permits

Validation and psychometric properties of an adapted multitasking measuring tool

A. Kopacz, C. Biele and A. Zdrodowska

National Information Processing Institute, Warsaw, Poland. akopacz@opi.org.pl

Interruptions are often unexpected breaks in current activity, that can introduce a new task, thus forcing a person to move from one action to another in an unplanned manner [1]. This generates a multitasking situation in which more than one task in the same time period is performed. Multitasking refers to the ability to perform two or more tasks simultaneously and sequentially by switching from one task to another [2]. In our study, we validate Polish adaptation of the UNRAVEL task [3] as a tool to measure functioning during multitasking.

The purpose of the original UNRAVEL procedure is to investigate the effect of interruptions on task performance. It engages executive processes in which the task is to focus attention on the correct element and to navigate to the next task-relevant element. In this sequential task steps have to be performed in a particular sequence, and correct performance depends on placekeeping in the sequence. The current study aims to adapt the UNRAVEL task into Polish and to verify whether the Polish adaptation can yield valid and reliable data about multitasking performance. Secondly, the study investigates the psychometric properties of an adapted version of the UNRAVEL task. Because the research in the area of multitask functioning suggests the existence of the age-related differences [4,5], we decided to investigate the properties of the Polish version of the UNRAVEL task for groups of both younger and older adults (N=121).

The Polish adaptation differs from the original in that we reduced the number of steps in the procedure from seven to five and used an acronym WINDA (meaning "an elevator" in Polish). Shortening of the procedure is expected to lessen the difficulty of the task for older participants. Each step of the WINDA sequence requires a two-alternative forced choice applied to one feature of a presented stimulus (a letter and a digit). For example, the W step involves choosing whether the letter on the screen is a capital or lowercase, the N step requires a choice about whether the digit in the stimulus is odd or even. For each step, the letter in the sequence mnemonically relates to one of the two candidate responses, which are unique for each step (e.g., "W" for "Wielka" (capital), "N" for "Nieparzysta" (odd); their opposites are "M" for "Mała" (lowercase), and "P" for "Parzysta" (even), respectively. In the original version three of the seven rules refer to visual attributes of the stimuli: a font style (underline or italic), a color (red or yellow) and localization on the screen (above or below the box), two of them refer to digits: the digit is even or odd and the digit is less than or more than 5, and two refer to letters: the letter is near to or far from the start of the alphabet and the letter is vowel or consonant. The five rules in the polish version of the UNRAVEL task were made analogically. Two rules refer to digits: the digit is other than 1 or 8 or equal to 1 or 8 and the digit is odd or even, two refer to letters: the letter is capital or small and the letter is near to letter A or Z in alphabet, and one rule is perceptual: the character is below or above the box.

In order to test the properties of the WINDA task and to compare it to the original UNRAVEL task, we re-enacted a set of analyses completed by original creators of UNRAVEL [3]. Our analyses for WINDA task paralleled Altmann's results for UNRAVEL experiment in most instances but also revealed a moderating effect of age (younger vs. older adults). The main finding of the adapted WINDA study, mirroring Altmann's UNRAVEL results, is that interruptions lead to an increased chance of resuming at the wrong step in a sequential task during a cognitively engaging activity and that the interruption effect differs in age groups. We examine the potential of the adapted procedure to produce individual differences and verify the reliability of sequence errors as a measure of individual differences in placekeeping operations. Similarly to the original task, reliability for sequence errors is high, both in the group of younger and older adults, suggesting that the WINDA task is a reliable tool, suitable for measuring individual differences in multitasking performance. Obtained effects suggest that the adaptation process that we employed for the creation of the WINDA task may be utilized to generate other language adaptations of this tool and the particular rules used in WINDA, as well as probably the acronym itself (defined as the set of rules), could be exchanged for another without compromising the procedure as a tool to measure sequential control. This opens the possibilities for the creation of different versions of the task (characterized by different levels of difficulty) targeted at the specific subject groups.

Ethical statement

All procedures were conducted in accordance with the research ethics guidelines at SWPS University of Social Sciences and Humanities in Warsaw.

Acknowledgement

Publication was supported by a National Science Centre under grant number 2012/07/D/HS6/02533.

References

1. Miyata, Y., Norman, D.A. (1986). Psychological issues in support of multiple activities. In: D. A. Norman, S. W. Draper (Eds.), User centered system design: New perspectives on human-computer interaction (pp. 265-284). Hillsdale, NJ: Erlbaum.

2. Dzubak, C.M. (2008). Multitasking: The good, the bad, and the unknown. The Journal of the Association for the Tutoring Profession 1, 1-12.

3. Altmann, E.M., Trafton, J.G., Hambrick, D.Z. (2014). Momentary interruptions can derail the train of thought. Journal of Experimental Psychology: General 143, 215-226.

4. Craik, F.I., Salthouse, T.A. (2000). The handbook of aging and cognition. Mahwah, NJ: Erlbaum.

5. Craik, F.I.M, Bialystok, E. (2006). Planning and task management in older adults: Cooking breakfast. Memory & Cognition 34, 1236–1249.

Mouse behavior pipelines for various disease models

K. Kraitsy[1], A. Jelem[1] and S. Badurek[1]

Affiliation [1]Preclinical Phenotyping, Vienna Biocenter (VBC), Dr. Bohr-Gasse 3, 1030 Vienna, Austria; Sylvia.badurek@vbcf.ac.at

In our facility, we provide behavioral assays to answer questions on disease models in mice. Beside other subjects, we address the field Learning and Memory as well as Anxiety and Depression.

Symptoms of anxiety and depression or deficits in learning and memory can occur independently. However, anxiety and depression-like symptoms [1] can influence learning and memory function.

We established pipelines to phenotype mouse lines, addressing questions on Learning and Memory as found in Alzheimer's Disease [2] or Dementia, congenital mental deficits as well as several cognitive disorders.

In another line of research, we combine the TSE PhenoMaster System [3], which is a highly flexible metabolic cage assemble, with well-known but also cutting-edge motoric assays.

Thus, we established behavior pipelines addressing Parkinson's Disease models, but also models for stroke or traumatic brain injuries [4].

By using also, the TSE PhenoMaster System, we developed another pipeline as well, which can be used for mouse models addressing metabolic questions like thermoregulation, respiration and/or calorie consumption.

All the animal work was conducted according the Austrian and European laws and guidelines for animal research.

References:

1. Ueda, H., Sasaki, K., Halder, S.K., Deguchi, Y., Takao, K., Miyakawa, T., Tajima, A. (2017). Prothymosin alpha-deficiency enhances anxiety-like behaviors and impairs learning/memory functios and neurogenesis. *J Neurochem* 2017 Apr: 141 (1): 124-136. doi: 10.1111/inj.13963.Epub 2017 Feb 21.

2. Perusini, J.N., Cajigas, S.A., Cohensedgh, O., Lim S.C., Pavlova I.P., Donaldson Z.R., Denny C.A. (2017). Optogenetic stimulation of dentate gyrus engrams restores memory in Alzheimer's disease mice. *Hippocampus* 2017 Oct; 27 (10): 1110-1122. doi: 10.1002/hipo.22756. Epub 2017 Jul 20.

3. Robinson, L. Plano, A., Cobb, S., Riedel, G. (2014). Long-term cage activity scans reveal lowered exploratory behaviour in symptomatic female Rett mice. *Behav Brain Res* 2013 Aug 1;250:148-56. doi: 10.1016/j.physbeh.2014.01.018. Epub 2014 Feb 8.

4. Kraitsy, K., Uecal, M., Grossauer, S., Bruckmann, L., Pfleger, F., Ropele, S., Fazekas, F., Gruenbacher, G., Patz, S., Absenger, M., Porubsky, C., Smolle-Juettner, F., Tezer, I., Molcanyi, M., Fasching, U., Schaefer, U. 2014. Repetitive long-term hyperbaric oxygen treatment (HBOT) administered after experimental traumatic brain injury in rats induces significant remyelination and recovery of sensorimotor function. *PlosOne* 2014 May 21;9(5):e97750. doi: 10.1371/journal.pone.0097750.eColection 2014.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Cognitive ability characterization of rat model with alpha-synuclein overexpression

Charlotte Laloux, Cédrick Lachaud, David Devos and Regis Bordet

Lille Neurosciences Federation, INSERM U1171, Lille University; charlotte.laloux@univ-lille2.fr

Due to the encouraging success achieved with therapeutic interventions aimed at treating the motor symptoms in Parkinson's disease (PD), a greater appreciation of the non-motor aspects of the disease has emerged. One third of these patients suffer from cognitive impairment, sometimes preceding the classical motor symptoms, which has consequences for patient's quality of life and mortality. Mainly frontal dysexecutive function but also memory and recognition are affected; these disease induced cognitive impairments are less/not sensitive to dopamine replacement therapy. Progressive degeneration of nigrostriatal dopamine neurons and the pathological deposition of the protein alpha-synuclein in intraneuronal inclusions are the neurobiological basis for PD. A model of AAV-mediated overexpression of alpha-synuclein in midbrain dopamine neurons have been developed and described in rats, showing a progressive dopaminergic neurodegeneration associated to motor symptoms related to PD (Kirik et al, 2002; Decressac et al, 2012; Bourdenx et al, 2015). The present study aims to assess cognitive function of this model in order to determine whether behavior of AAV induced alpha-synuclein overexpression rats resembles those observed during early PD manifestations.

Twenty four male adult Sprague Dawley rats were enrolled in the cognitive function test (French ethical comity authorization, APAFIS#6281-2016072517437766v3). AAV2/9 vectors carrying the human mutant p.A53T alpha-synuclein were injected into both substantia nigra of fourteen rats and AAV-GFP marker in ten rats (as controls). A comprehensive behavioral battery of tests was performed every 4 weeks for motor skills, anxiety, working memory and object recognition ability until 16 weeks. In parallel, touchscreen based operant tasks were performed for visual discrimination and attention (5 choices serial reaction time).

Behavioural parameters analysis revealed that motor skills impairment occur within the first 2 months after the AAV-hα-syn injections, as well as anxiety behavior in the elevated plus maze and working memory in spontaneous alternation test. Analysis of touchscreen based tasks and object recognition are still ongoing. Interpretation of the results in this study has to take into account the sensorimotor impairment induced by AAV-hα-syn injections. Analysis of the different parameters of interest as measured in the adopted test battery have to be related to other subtle parameters such as activity of the animal or response latency in each test and correlation between tests is also required. Observing relevant cognitive impairment in a progressive synuclein - based PD model would offer relevant evidence to study underlying neurophysiological mechanisms and propose potential therapeutic target to improve patient's quality of life.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Motion Capture Validation in Immersive Data Visualization Environment

I. Rito Lima[1], D. Birch[2], A. Faisal[1]

1 Brain and Behaviour Lab, Department of Bioengineering, Imperial College London, London, United Kingdom. ines.rito-lima17@imperial.ac.uk, a.faisal@imperial.ac.uk; 2 Data Science Institute, Imperial College London, London, United Kingdom. david.birch@imperial.ac.uk

Introduction

In the emerging topic of Human-Data Interaction (HDI), immersive data visualization environments allow exploring the social interaction of groups of researchers with data, providing that their spatial movement may be captured and correlated with the data displayed. Imperial College's Global Data Observatory (GDO) is a high resolution circular environment covering 313º where 15 people can simultaneously explore and interact with data through visual analytics. The space is equipped with multiple Microsoft Kinects, allowing crowd motion tracking and approximated eye gaze detection through a system presented as Holokinect [1]. As the largest in Europe, the GDO offers a unique scenario to develop research on human mobility patterns and data interaction. In this context, the aim of this paper is to describe the technologies involved in the GDO environment, discuss the research possibilities that they offer and most importantly to validate the space as human mobility analysis' centre, promoting its use in multidisciplinary research projects while exploring potential collaboration synergies.

Infrastructure and Usability

The GDO, a Large High-Resolution Display (LHRD) assembles 64 monitors powered by 32 computers and five Microsoft Kinects v2 depth-cameras is able to track body movement and eye gaze of several subjects simultaneously. It allows the social interaction of 15 people with data, which has proven to decrease the difficulty of gaining insight into data [1]. The GDO has been used to explore dynamic Bitcoin transactions [2], fuzzy logic [3], and neuroimaging data [4]. However, it presents many possibilities, in particular to study human mobility patterns, both in small indoor and large outdoor scales (either using the Holokinect system or online satellite data, respectively). The former includes the study of human mobility in urban areas, which may contribute to the emerging subject of smart cities and how they operate. The Kinect devices have previously proven to be useful in game control [5], robot control [6,7], and to improve the tracking process and loss of sight in PET scan [8]. In the context of HDI, it allows the study of how groups interact and organize when given particular tasks, such as searching for anomalies in a large data display. Further the eye gaze detection system can provide detailed feedback on the efficacy of the data displays in attracting the attention of the participants. This feedback could be used to understand and detect particular anomalies in the data displayed which are apparent visually but not easily identified in the dataset itself. The use of eye gaze to control an adaptive visualization task is proven to increase cognitive abilities [9].

Validation Methodology

To evaluate and validate the GDO with Holokinect (see Figure 1.a) as an immersive environment to measure human mobility behaviour, a comparative experiment involving a distinct motion capture system was carried out. In this experiment a subject wearing a mocap suit performed specific representative movements inside the GDO space, in different positions of the grid presented on Figure 1.b. By operating simultaneously, the Holokinect system recorded the same set of movements. In this way, data collected from both systems can be compared, and the accuracy of both systems evaluated, according to the position inside the GDO. This allows analysis of the impact of distinct configuration settings, such as considering weighted joints to better approximate the biomechanical hierarchy and constraints. The comparative equipment used in the experiment was the igs-180 mocap suit, from Synertial, that incorporates 17 sensors each of which is a 9-axis inertial movement units (IMU), capturing the movement on a 60Hz rate. The data is saved as a BVH file, containing the rotation in XYZ axes for each joint in Euler angles (deg). The data recorded from the Holokinect system is stored on a different format in an online DataBase platform, containing the XYZ coordinates of the set of joints considered by the Holokinect system on a rate of 30 Hz. Thus, the first step of the comparing methodology involves converting BVH files to Kinect files and vise versa allowing assessment of the autoencoding error. The second stage assesses the difference between the measures of both systems, both in Euler angles, first considering all joint equals, and then considering the weighted joints. This data processing is performed for each recording in a different position on the grid,

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

averaging the error across joints. Ultimately, it is possible to present the detected errors according to the position on the grid, varying from the centre to different radius circumferences, building a heat map of the accuracy of the movement capture inside the GDO. This variability may be, for example, due to sight lines and occlusion of participants. This is ongoing work, representing a fundamental basis for studying accurately human mobility patterns inside the GDO immersive environment.

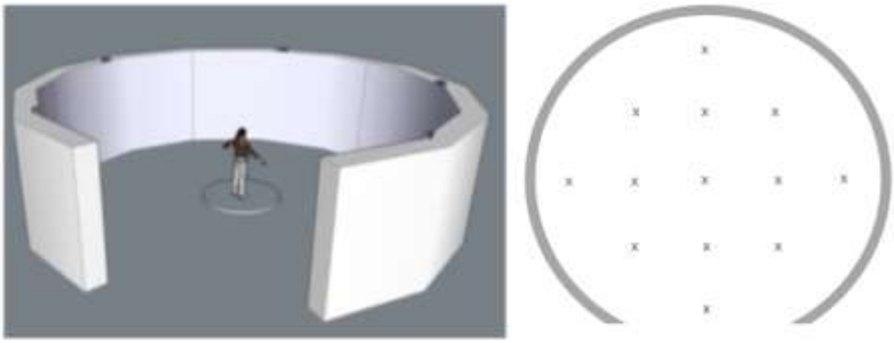


Figure 1. (a) GDO environment and Holokinect montage; (b) top view of the grid used for recordings, 1m distance between two consecutive crosses.

References

1. Chao, L., Birch,D., Xu, Z., Guo, Y. (2017) HoloKinect: A real-time scalable fusion framework for multiple target tracking. Internal Report, Data Science Institute, Imperial College London.

2. McGinn, D.; Birch, D.; Akroyd, D.; Molina-Solana, M.; Guo, Y.; Knottenbelt, W.J. (2016) Visual izing dynamic Bitcoin transaction patterns. *Big data*, 4, 109–119.

3. Molina-Solana, M.; Birch, D.; Guo, Y.k. (2017) Improving data exploration in graphs with fuzzy logic and large-scale visualisation. *Applied Soft Computing*, 53, 227–235.

4. Wall, M.B.; Birch, D.; Yong, M.Y. (2016) Opportunities and considerations for visualising neuroimaging data on very large displays. *F1000Research*, 5.

5. Zhang, Z. (2012) Microsoft kinect sensor and its effect. *IEEE multimedia*, 19, 4–10.

6.Du, G.; Zhang, P.; Mai, J.; Li, Z. (2012) Markerless kinect-based hand tracking for robot teleoperation. *International Journal of Advanced Robotic Systems*, 9, 36.

7. El-laithy, R.A.; Huang, J.; Yeh, M. (2012) Study on the use of Microsoft Kinect for robotics applications. *Position Location and Navigation Symposium (PLANS), 2012 IEEE/ION*, 1280–1288.

8. Noonan, P.J.; Howard, J.; Tout, D.; Armstrong, I.; Williams, H.A.; Cootes, T.F.; Hallett, W.A.; Hinz, R. (2012) Accurate markerless respiratory tracking for gated whole body PET using the Microsoft Kinect. *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012 IEEE*, 3973–3974.

9. Steichen, B.; Carenini, G.; Conati, C. (2013) User-adaptive Information Visualization: Using Eye Gaze Data to Infer Visualization Tasks and User Cognitive Abilities. *Proceedings of the 2013 International Conference on Intelligent User Interfaces; ACM: New York, NY, USA, IUI '13*, 317–328.

Ethomics and Brain Activity During Extreme Road Conditions: Race Car Driver on TopGear Racetrack

I. Rito Lima, S. Haar, P. Orlov, A. Faisal

Brain and Behaviour Lab, Department of Bioengineering, Imperial College London, London, United Kingdom.
ines.rito-lima17@imperial.ac.uk, s.haar@imperial.ac.uk, p.orlov@imperial.ac.uk, a.faisal@imperial.ac.uk

Introduction

Understanding the relation between brain signals and body movements while driving in extreme scenarios brings us closer to assess human behaviour and the related neural processes. ETHOMICS (ethology+genomics) is the approach to measure comprehensively natural perceptual input (vision), motor output (eyes and limbs movements) and brain activity. This will allow to foster the development of specific in-vehicle and in-body technologies able to prevent critical conditions and improve driving safety. With this purpose, the problem of drivers' fatigue is being widely explored, however, the neurological response under challenging driving conditions, where the skill of the driver comes in play, and his responses can save life, has not been carefully studied yet. To this end, a Formula E champion drove a sports car in the Silverstone circuit under extreme conditions (high speed, low visibility and road slipperiness), while we recorded his brain activity (using EEG), his eyes movements and his hand/foot movements. Understanding how the brain and consequently the body operates under these conditions provides new insights in human reaction mechanisms and, ultimately, can be used to develop autonomous cars and their safety procedures. The data collected was analysed in respect to the GPS position on the driving circuit. The initial results obtained from the real experimental scenario suggest correlations between body and eye movements and concrete brain waves, in relation to the curvature of the track and the unexpected events caused by the extreme driving conditions. The test drive was prescheduled by the racing team, who race in these conditions frequently, for promotional purposes. It enabled us this unique scientific observation of motor expertise in the wild. Although unlikely needed, emergency response units were present.

Related Work

Most studies on driving have been focusing on understanding, measuring and evaluating drivers' attention [1,2], fatigue [3], and drowsiness [4], which are major causes of road accidents. Some of those studies used these measurements to implement brain machine interfaces (BMI) to reverse these states [1,4]. The use of neurological data to determine and prevent these states is an emerging field. It was shown that drivers intention to perform an emergency braking task is detected in the brain and muscles previously to the behavioural response, in almost 130ms [5]. That corroborates that drivers' intentions can be detected in the brain before the initiation of the motor response. Here we try to characterize the responses of a top skill race driver under extreme driving conditions and unexpected events using his neural activity and eye movements.

Method

To explore the relation between brain, body and eye movement, in respect to the position on the track, an experienced professional race car driver (20+ years racing experience) drove an Audi R8+ (610 bhp) on the TopGear Race track, in wet weather conditions with 3-5ºC (contributing to the decrease of visibility and slipperiness of the road). Driver assistance systems on the car were turned off. The driver was equipped with: a 32-channels wireless EEG system with dry electrodes (LiveAmp - BrainVision); an eye tracking system (ETG 2W A - SMI); inertial measurement units (IMUs) on his hands and feets (MTw Awinda - Xsens); and an Apple watch to monitor his heart rate. The Audi R8 used on the experiment was equipped with a GPS system and a camera recording the inside of the car during the entire experience. The video recordings were used to annotate the data and detect events (drifts and loss of control). The data collected from the eye glasses allowed to determine eye gaze. As pupils' movements are recorded, the glasses could also be used to determine the percentage of eyelid closure (PERCLOS), in order to incorporate the measurement of drowsiness in our study, as in [4]. For EEG data handling, MATLAB toolbox EEGLAB was used to perform pre-processing tasks (removing bad channels). The collected data was analysed in MATLAB in order to relate data both in space and time. Figure 1 presents the system architecture used on the experiment, regarding the equipments involved, the conditions in which the experiment occurred, and the data analysis process.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

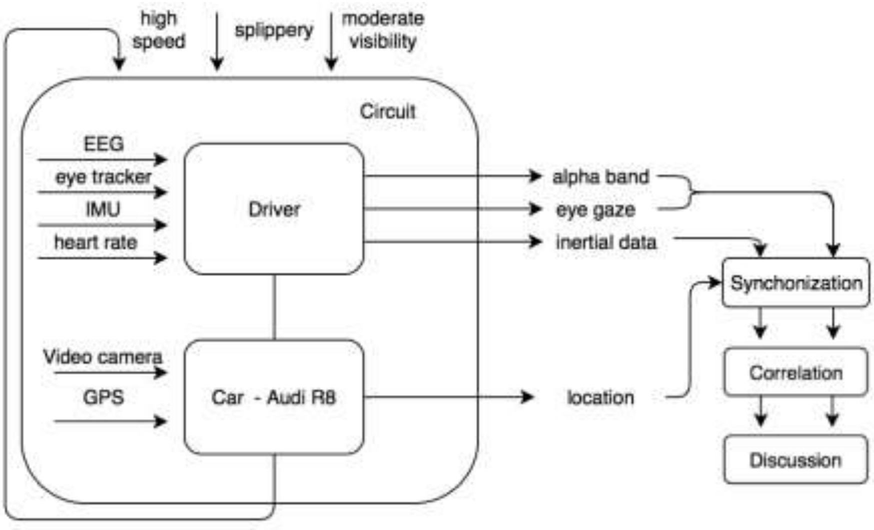


Figure 1. Architecture of the experiment and data processing analysis.

Results

Timestamp synchronization was performed across all data streams. The EEG alpha band and eye gaze were analysed together, as this band power is strongly connected to eyes' activity and state of alertness [2,6]. The inertial data from the IMUs enable analysis of the abruptness of both hand and foot movements in respect to the position on the circuit track (driving straight requires less movement unless unexpected events such as drifting occur). Figure 2 illustrates some preliminary results of the first driving lap. In Figure 2.a, the colour of data represents the power of the alpha band whereas the size of the circle represents the magnitude of the eye movements. In Figure 2.b, each colour corresponds to one of the sensors, and the size of the circle represent the abruptness of movements. S indicates the starting point and the arrow the direction of the car movement. As shown, it is possible to identify measurable changes in brain and motor activity beyond circuit curves, resulting from moments of loss of control of the car, for instance, during drifts caused by the slipperiness of the road. It is also possible to identify longer periods of eye fixation when the driver starts a circuit straight line, and an increase on the alpha band in the second half of the curves.

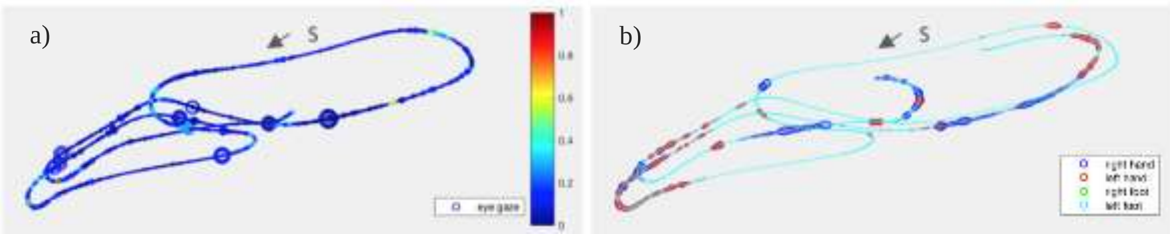


Figure 2. First round of data across the circuit track obtained by GPS data and (a) EEG data with a heat map for the alpha band and the eye gaze fixation times; (b) RMS of the gyroscopes data from the four sensors.

Additional analysis will include further assessment on the brain waves and eye movement patterns during moments of drifts, where there is an increase on body movement (especially the hands), as well as an increase on the neural response mechanism.

References

1. Busso, C., Jain, J. (2012) Advances in Multimodal Tracking of Driver Distraction, *Digital Signal Processing for In-Vehicle Systems and Safety*, *Springer*, 253-257.
2. Baldwin, C. L., Roberts, D. M., Barragan, D., Lee, J. D., Lerner, N., & Higgins, J. S. (2017). Detecting and Quantifying Mind Wandering during Simulated Driving. *Frontiers in Human Neuroscience*, *11*, 1–15.
3. Lal, S. K. L., Craig, A., Boord, P., Kirkup, L., & Nguyen, H. (2003). Development of an algorithm for an EEG-based driver fatigue countermeasure. *Journal of Safety Research*, *34*(3), 321–328.
4. Li, G., & Chung, W. Y. (2017). Combined EEG-Gyroscope-tDCS Brain Machine Interface System for Early Management of Driver Drowsiness. *IEEE Transactions on Human-Machine Systems*, *48*(1), 50–62.
5. Haufe, S., Treder, M. S., Gugler, M. F., Sagebaum, M., Curio, G., & Blankertz, B. (2011). EEG potentials predict upcoming emergency brakings during simulated driving. *Journal of Neural Engineering*, *8*(5).
6. Nicolas-Alonso, L. F., & Gomez-Gil, J. (2012). Brain computer interfaces, a review. *Sensors*.

Sequential analysis to quantify variability in Canine Abnormal Repetitive Behaviour

B. A. Loftus[1], L. Asher[2] and R. A. Casey[1]

1 Dogs Trust, London, UK and The University of Bristol, Bristol, UK. Bethany.loftus@dogstrust.org.uk, Rachel.casey@dogstrust.org.uk; 2 Newcastle University, Newcastle, UK, lucy.asher@ncl.ac.uk;

Background: Abnormal Repetitive Behaviours (ARBs) displayed by dogs include behaviours such as tail chasing, shadow chasing, and repetitive licking. These behaviours have been variably described as stereotypic[1,2] or compulsive[3] behaviours, however there is no consensus on how to classify ARBs into either category. According to common definitions, a stereotypy is an invariant and repetitive behaviour with no apparent function, whereas 'compulsive' behaviours relate to variable goal directed behaviours which are abnormally repeated, sustained or out of context. The expression of ARB can vary between individuals both in phenotype and in how variable the behaviour appears to be. Some individuals display highly repetitive and invariant ARB whereas others appear to be variable in presentation. In other species, the degree to which ARB is invariant is suggested to be important in indicating dysfunctional behaviour control mechanisms. Repeated 'goal directed' behaviours are suggested to reflect deficits in areas of the brain involved in planning and attaining goals (the prefrontal cortex) which may be classed as a 'compulsive' behaviour[4], whereas 'fixed' and invariant ARBs are hypothesized to reflect dysfunction in areas of the brain responsible for motor control (the basal ganglia)[5] and may be described as stereotypic.

Aim: The aim of this study was to develop and compare measures of variability within and between ARB sequences from a sample of kennel housed dogs, in order to identify a suitable method of quantifying variability. Understanding the degree of variability within and between ARB sequences may enable classification of ARB in dogs as compulsive *or* stereotypic.

Method: The subjects were 47 adult dogs which displayed any form of abnormal repetitive behaviour, housed in one of 13 Dogs Trust and RSPCA re-homing centres. All dogs were videoed in their kennels on a standard day for a period of five hours. Screening criteria for each case was based on whether the dog repeated the reported 'abnormal' behaviour more than three times consecutively, or for more than 10 seconds continuously, repeated over more than three bouts of the behaviour. Noldus Observer™ was used to analyse the sequences of behaviours within periods of ARB. Each type of ARB was analysed according to individual behaviour components which comprised the sequence (example sequence: spin right, spin right, spin right, spin left, spin left).

There is no gold standard for measuring repetitiveness in ARB sequences, therefore measures of variability were either developed or adapted for the purpose of this study based on an understanding of the nature of ARBs or were based on measures used in other studies[6–8]. The measures included both within bout variability (how dependent the current behaviour is on the preceding behaviour) and between bout measures of variability (how similar separate bouts of ARB were to one another). The measures of variability were applied to the 47 dog's sequences of ARB.

Analysis and Results: The structure between measures were explored with Principle Components Analysis. The seven variability measures grouped onto three factors which highlighted different aspects of variability. One factor measured 'between bout' variability, another measured 'within bout' variability comparing the observed sequence of behaviours to random, and a third measured sequential dependency. All three factors are aspects of repetition in abnormal repetitive behaviour and may be useful when considering how they relate to behavioural control. The degree of variability in ARB may indicate the degree of dysfunction in control of behaviour, which can be tested for using behavioural tasks such as a two-choice gambling task, set shifting tasks, or a reversal learning task[5].

No distinct categories into which dogs could be classified based on the variability of ARB were identified in this study, since the majority of ARB displayed was on a continuum from variable to relatively repetitive. Nine of forty-seven dogs displayed ARB that was entirely a repetition of the same behaviour component, the remaining subjects ranged from almost entirely repetitive to very variable. The definition of stereotypy suggested by Mason[9]

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

is '*a repetitive behaviour induced by frustration, repeated attempts to cope, and/or CNS dysfunction*'. Based on the results of our data, few subjects displayed entirely repetitive behaviour, and the rest displayed ARB on a continuum which suggests that the majority of dogs showing ARBs in the study population cannot be described as stereotypic. Whether the majority of dogs therefore fall into the 'compulsive' behaviour category requires further research.

References

1. Denham, H. D. C., Bradshaw, J. W. S. and Rooney, N. J. (2014) 'Repetitive behaviour in kennelled domestic dog: Stereotypical or not?', *Physiology and Behavior*. Elsevier Inc., 128, pp. 288–294.

2. Protopopova, A., Hall, N. J. and Wynne, C. D. L. (2014) 'Association between increased behavioral persistence and stereotypy in the pet dog', *Behavioural Processes*, 106, pp. 77–81.

3. Dodman, N. H. *et al.* (2016) 'Genomic risk for severe canine compulsive disorder, a dog model of human OCD', *International Journal of Applied Research in Veterinary Medicine*, 14(1), pp. 1–18.

4. Mills, D. and Luescher, A. (2006) 'Veterinary and pharmacological approaches to abnormal repetitive behaviour', in Mason, G. and Rushen, J. (eds) *Stereotypic Animal Behaviour. Fundamentals and Applications to Welfare*. Wallingford: CAB International, pp. 286–324.

5. Garner, J. P. (2006) 'Perseveration and stereotypy—systems-level insights from clinical psychology', in Mason, G. J. (ed.) *Stereotypic Animal Behaviour: Fundamentals and Applications to Welfare*. Wallingford: CABI International, pp. 121–152.

6. Brilot, B. O. *et al.* (2009) 'Quantification of abnormal repetitive behaviour in captive European starlings (Sturnus vulgaris)', *Behavioural Processes*, 82(3), pp. 256–264.

7. Vickery, S. S. and Mason, G. J. (2005) 'Stereotypy and perseverative responding in caged bears: Further data and analyses', *Applied Animal Behaviour Science*, 91(3–4), pp. 247–260.

8. Bakeman, R. and Quera, V. (2011) *Sequential analysis and observational methods for the behavioral sciences*, *Sequential Analysis and Observational Methods for the Behavioral Sciences*.

9. Mason, G. (2006) 'Stereotypic behaviour in captive animals: fundamentals, and implications for welfare and beyond.', *Stereotypic Animal Behaviour: Fundamentals and Applications to Welfare*, pp. 325–356.

Measuring human running to inform mechanical ageing of running shoes

Ryan Martin and Tom Allen

Sports Engineering TEAM, School of Engineering, Manchester Metropolitan University, UK; t.allen@mmu.ac.uk

Almost a third of adults in England are obese, with an annual cost to the National Health Service of ~6 million [1]. The obesity issue is partly due to the UK population being less active today than in the 1960s, and the Government want people to exercise more [2]. Running is an accessible and social activity well suited to getting more people exercising regularly. A recent survey estimated the UK running population at ~10 million [3], Sport England predict ~2 million adults regularly participate in athletics [4] and ParkRun UK claim >1.5 million people have participated in their runs [5]. Benefits of regular exercise are clear, but poorly conditioned recreational athletes are more prone to injury [6].

An estimated 37 to 56% of runners sustain injuries, with ~50 to 70% of these classified as 'overuse injuries' [7]. Foot pronation and ground reaction forces are injury factors amongst runners [8], and running shoes degrade with use, reducing their ability to limit any harmful impact forces [6]. Running shoes also affect muscle activity, influencing fatigue, comfort, work and performance [9]. Mechanical ageing can be used to impose, and measure, degradation, repeatedly compressing the shoe to mimic running [10]. The aim of this work was to assess the feasability of measuring ground reaction forces to individualise mechanical ageing of running shoes. Ethical Approval was obtained from the local ethics committee.

Mechanical ageing parameters came from a participant (22 year, 76 kg, 1.86 m & 21.8 BMI) running over a force plate (Z17068AA, Kistler, sampling at 2,000 Hz, with data collected and analysed using Bioware 5.1.1.0) while wearing shoes (Kalenji, EKIDEN, size 8.5). The shoes were naturally aged with the participant running outdoors (time and distance monitored, Strava GPS device), and an identical shoe was mechanical aged in a servo hydraulic device (Instron, fitted with 5 kN load cell), with compression applied via a rigid heel surrogate. The mechanically aged shoe was subjected to a dwell waveform [10], involving a loading cycle based on the participants foot-strike followed by a pause (0.5 s) to represent the time between strikes. Shoe degradation from natural and mechanical ageing was assessed at set running distances, by comparing the energy absorbed in a mechanical loading cycle. This research helps link mechanical ageing to specific runners, informing shoe design and guidance as to when they should be replaced.

References

[1] Public Health England (2017). Health matters: obesity and the food environment <https://www.gov.uk/government/publications/health-matters-obesity-and-the-food-environment/health-matters-obesity-and-the-food-environment--2>. Accessed 6 February 2018.

[2] Public Health England (2016). Health matters: getting every adult active every day <https://www.gov.uk/government/publications/health-matters-getting-every-adult-active-every-day/health-matters-getting-every-adult-active-every-day>. Accessed 6 February 2018.

[3] SMS INC (2014). UKs running population reaches remarkable 10.5 m says sports marketing surveys inc. < http://www.sportsmarketingsurveysinc.com/uks-running-population-reaches-remarkable-10-5m-says-sports-marketing-surveys-inc/>. Accessed 6 February 2018.

[4] Sport England (2016). Active People Survey 10. < https://www.sportengland.org/media/11746/1x30_sport_16plus-factsheet_aps10.pdf>. Accessed 6 February 2018.

[5] ParkRun UK <http://www.parkrun.org.uk/>. Accessed 6 February 2018.

[6] Cook, S. D., Kester, M. A. & Brunet, M. E. (1985). Shock absorption characteristics of running shoes. *The American Journal of Sports Medicine* **13**, 248-253.

[7] Mechelen, V. W. (1992). Running injuries. A review of the epidemiological literature. *Sports medicine* **14**, 320-335.

[8] Nigg, B. M., Baltich, J., Hoerzer, S. & Enders, H. (2015). Running shoes and running injuries: mythbusting and a proposal for two new paradigms: 'preferred movement path' and 'comfort filter'. *British Journal of Sports Medicine* **49**, 1290-1294.
[9] Nigg, B. M. (2001). The Role of Impact Forces and Foot Pronation: A New Paradigm. *Clinical Journal of Sport Medicine* **11**, 2-9.
[10] Nadine M. Lippa, Paul K. Collins, Jason Bonacci, Scott G. Piland, James W. Rawlins & Trenton E. Gould (2017). Mechanical ageing performance of minimalist and traditional footwear foams, *Footwear Science* **9**, 9-20.

Can fish-catching and sweeping tasks reproduce natural whisker movements in pinnipeds?

Alyx Milne, Catherine Smith, Robyn A Grant

Division of Biology and Conservation Ecology, Manchester Metroplitan University, UK, robyn.grant@mmu.ac.uk

Pinnipeds (seals, sea lions and walruses) have the most sensitive whiskers of all mammals. All pinnipeds are able to use their whiskers to discriminate between different textures, shapes and sizes of objects. Sea lions are thought to be able to move their whiskers actively, which can be demonstrated during a dynamic somatosensory task, such as ball-balancing [1]. However, the degree to which seals and walruses can move their whiskers has not yet been documented.

This presentation documents a new method to easily encourage natural whisker movements in pinnipeds. All experiments were carried out with approval by local ethics committees at Blackpool Zoo, Rhyl Seaquarium and Hardjevik Dolfinarium. Sea lions, walruses and seals were filmed during fish catching, where the fish were thrown to the animal, and fish sweeping, where the fish was swept across the mystacial pad. The video was then analysed using custom behavioural software, the Manual Whisker Annotator (MWA) [2], which provided measurements of head orientation, whisker angles, whisker amplitudes and whisker asymmetry. Both fish sweeping and fish catching tasks caused protraction of the whiskers towards the food item. Sea lions had the largest whisker movements, however, these were not as strong as during the ball-balancing task (in [1]). We conclude, that moving food items can encourage natural whisker movements in pinnipeds; however, an even more dynamic task, such as ball-balancing, will encourage larger movements. Focusing on encouraging natural whisker sensing movements might provide important behavioural and sensory enrichment. Future work might focus on designing a swimming fish delivery system to encourage more natural sensing behaviors in pinnipeds in captivity.

Figure 1: Example video stills showing whisker protractions and asymmetry during fish catching. Figure 6 in [1].

References

1. Milne, A. and Grant R.A. (2014). Characterisation of whisker control in the California sea lion (*Zalophus Californiaus*) during a complex, dynamic sensorimotor task. *Journal of Comparative Physiology A,*
2. Hewitt, B., Yap, M. and Grant, R. (2016). 'Manual Whisker Annotator (MWA): A Modular Open-Source Tool'. *Journal of Open Research Software,* 4 (16), p.p. 1-7.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Analysing facial expressions with the help of FaceReader – New methodological perspectives in moral developmental research

Anne M Möller and Brigitte Latzko

Leipzig University, Germany; anne_mareike.moeller@uni-leipzig.de

Recently, within moral developmental research moral emotions became increasingly centre of interest [9]. Moral emotions like shame and guilt play a key role in transforming social-moral norms into corresponding behaviour [1]. Moreover, these emotions show to what extent moral rules are assumed as personally binding [7].

A prominent approach to assess moral emotions and immoral behaviour is the so called happy victimizer paradigm [6]. Within this paradigm a half standardised interview is used in which moral emotion attributions can be indicated with the help of hypothetic moral conflict situations [8].

Although this approach led to impressive empirical evidence for the happy victimizer phenomenon [6] describing that young children attribute positive emotions to a moral wrongdoer even though they know the moral rule, there are still some methodological limitations:

Firstly, verbal abilites of children aged 3 to 4 are still limited [5]. Little is known whether they can not express different emotions or whether they do not experience them. Secondly, it remains unclear how far verbal attributions of moral emotions in the happy-vicitmizer-interview are influenced by cognitive control and terms of social desirability [3] – especially when working with older children with emerging (social) cognitive competences.

To overcome these methodological limitations the present study aims at exploring an extended approach for the measurement of moral emotions by assessing affective behavioral responses. Thus the study wants to examine whether the emotions reported in the interview correspond with spontaneous affective behavioral responses. It is assumed that these affective behavioral responses would be related to those emotions mentioned in the interview in terms of valence.

To test this hypothesis a pilot-study was conducted supplementing the happy victimizer interview with two different moral conflict situations by measuring facial emotional expressions (i.e. micro-expressions) [2]. Micro-expressions are cognitively controlled less because they are immediate and unconscious. Therefore they can neither be manipulated nor hidden [4].

Within the pilot-study 20 pupils (10 male/10 female), aged between 7 and 14 ($M = 9.8$; $SD = 2.9$), were interviewed and video-taped. In order to compare micro-expressions with reported emotions children were asked how the protagonist in the story would feel. The video-based data at the interval of 1/10 of a second after this stimulus was analysed afterwards with the help of FaceReader 7.0 [10] and ObserverXT 14.0.

Point-biserial correlation analysis was used to test the hypothesis. Analyses revealed that in one moral conflict situation there is a positive correlation between negative reported emotions and scared facial expressions ($r = .58$, $p < 0.05$). Additional descriptive statistical analyses indicated a strong relation between negative reported emotions and negative micro-expressions in general (47.1%) whereas no relation was found with regard to positive emotions.

Although findings, given the small sample size, should be interpreted with caution they emphasize the need of extended methodological approaches for measuring moral emotions and underline the potentials of FaceReader technology as innovative method in the field of moral developmental research. Furthermore FaceReader can be used to gain deeper insights in moral emotions, especially concerning mixed emotions. Advantages of the FaceReader and methodological implications for further research (e.g. selected intervals and testsetup) shall be discussed in detail.

References

1. Arsenio, W. F., Gold, J., & Adams, E. (2006). Children's conceptions and displays for moral emotions. In M. Killen & J. G. Smetana, (Eds.), *Handbook of moral development* (pp. 581-609). Mahwah, NJ: Erlbaum.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

2. Ekman, P., Friesen, W. V, & Hager, J. C. (2002). *The Facial action coding system*. Salt Lake City: Research Nexus eBook.

3. Dys, S. P., & Malti, T. (2016). It's a two-way street: Automatic and controlled processes in children's emotional responses to moral transgressions. *Journal of Experimental Child Psychology*, *152*, 31–40.

4. Haggard, E. A., & Isaacs, K. S. (1966). Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy. In L. A. Gottschalk & A. H. Auerbach (Eds.), *Methods of research in psychotherapy* (pp. 154–165). New York: Appleton-Century-Crofts.

5. Harris, P. L. (1989). *Children and emotion: The development of psychological understanding.* Oxford: Blackwell.

6. Krettenauer, T., Malti, T., & Sokol, B. W. (2008). The development of moral emotion expectancies and the happy victimizer phenomenon: A critical review of theory and application. *European Journal of Developmental Science*, *2*(3), 221–235.

7. Malti, T., & Latzko, B. (2017). Moral Emotions. In *Reference Module in Neuroscience and Biobehavioral Psychology* (Vol. 1). London: Elsevier.

8. Nunner-Winkler, G., & Sodian, B. (1988). Children' s Understanding of Moral Emotions. *Child Development*, *59*(5), 1323–1338.

9. Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). What's moral about the self-conscious emotions. In J. L. Tracy, R. W. Robins, & J. P. Tangney (Eds.), *The self-conscious emotions: Theory and research* (pp. 21–37). New York: Guilford Press.

10. Uyl, M. J. Den, & Kuilenburg, H. Van. (2005). The FaceReader : Online facial expression recognition. In *Proc. Measuring Behavior* (pp. 589–590).

Effect of conditioned pain modulation on somatosensory profile in surgical models of osteoarthritis pain in rats and dogs

C. Otis[1,2], E. Labelle[1,2], B. Monteiro[1,2], M. Moreau[1,2], K.A. Cristofanilli[1], D. Aoudj[1], M. Tardif[1], J.P. Pelletier[2], J. Del Castillo[1], B. Lussier[1,2], E. Troncy[1,2]

1. GREPAQ (Animal Pharmacology Research Group of Quebec), Faculty of Veterinary Medicine, Université de Montréal; 2. Osteoarthritis Research Unit, University of Montreal Hospital Centre, Canada

Introduction

The objective was to compare the effect of conditioning stimulus (CS) on somatosensory changes in animal models of osteoarthritis (OA) pain.

Materials and Methods

OA model was surgically-induced in right stifle at day (D) 0. Rats (*n*=48/Sprague-Dawley/female/ovariectomized (OVX)) were divided in 4 groups: OA-OVX negative control: placebo/ OA-OVX positive analgesic (PA): oral pregabalin 30mg/kg and subcutaneous carprofen 5mg/kg daily/ naïve-OVX/ naïve-normal. OA-dogs (*n*=6/ sterilized: female=4; male=2) received a placebo. Quantitative sensory testing (QST) included punctate tactile [rats; paw withdrawal threshold (PWT); BSL; D7; D14; D21; D35; D49: D56] and mechanical [dogs; BSL, week (W)4; W7; W9] allodynia evaluation, pre- and post-CS. The CS was mechanical pressure (ear) in rat, and ischemic (thoracic limb) in dog. Mixed model and Student tests were applied (α=0.05).

Results

Dynamic CS increased QST (at baseline, BSL) threshold in rats ($p < 0.023$), but not in dogs ($p > 0.922$).
Paw (secondary) allodynia (ipsilateral) was noted in OA rats: punctual (D7; $p < 0.004$) for OA-OVX-PA group, and during the whole follow-up (D7-D56) for OA-OVX-placebo ($p < 0.035$). An increased PWT post-CS was observed in the OA-OVX-placebo group, corresponding to an efficient conditioned pain modulation –CPM– (D21; D35; $p > 0.107$).
A central sensitization (contralateral PWT) was observed in OA-OVX-placebo rats compared to OA-OVX-PA (at D14; $p = 0.004$) and to all other groups (at D56; $p < 0.004$). The post-CS contralateral PWT increase counteracted the secondary allodynia present in pre-CS at both time-points (D14 and D56; $p > 0.062$) in OA-OVX-placebo. An efficient CPM effect was observed for all time-points in OA-OVX-PA and OA-OVX-placebo rats.
Primary allodynia was present in OA dogs at W4 with a nadir at W7; $p = 0.038$ without detecting any CPM effect, but a high level of stress was present in all dogs.

Discussion

Animals with OA developped allodynia, supporting peripheral and central sensitization. Post-CS response was variable in rats within time, supporting a possible fatigue in CPM (activation–alteration). In laboratory dogs, the stress-induced analgesia (manifested at BSL and for the whole follow-up) complexified the CPM manifestation.

Acknowledgements

The authors would like to thanks ArthroLab's personnel for their techinical assistance. There was not proprietary interest or funding provided for this projet. Dr Beatriz Monteiro is a recipient of a Vanier Canada Graduate Scholarship.

References

1. Fingleton C., Smart K.M., Doody C.M. (2017). Exercise-induced Hypoalgesia in People With Knee Osteoarthritis With Normal and Abnormal Conditioned Pain Modulation. The *Clinical journal of pain* **33**, 395-404.
2. Klinck M.P., Mogil J.S., Moreau M., Lascelles B.D.X., Flecknell P.A., Poitte T., Troncy E. (2017). Translational pain assessment: could natural animal models be the missing link? *Pain* **158**, 1633-1646.
3. Le Bars D., Dickenson A.H., Besson J.M. (1979). Diffuse noxious inhibitory controls (DNIC). I. Effects on dorsal horn convergent neurones in the rat. *Pain* **6**, 283-304.
4. Yarnitsky D., Arendt-Nielsen L., Bouhassira D., Edwards R.R., Fillingim R.B., Granot M., Hansson P., Lautenbacher S., Marchand S., Wilder-Smith O. (2010). Recommendations on terminology and practice of psychophysical DNIC testing. *European journal of pain* **14**, 339.
5. Youssef A.M., Macefield V.G., Henderson L.A. (2016). Pain inhibits pain; human brainstem mechanisms. *NeuroImage* **124**, 54-62

WhiskerMan: A Suite of Tools for Model-Based Rodent Whisker Tracking and Touch Detection

R.S. Petersen*, M.H. Evans*, M.S.E. Loft, A. Colins

Division of Neuroscience & Experimental Psychology, University of Manchester, Manchester, UK.
R.Petersen@manchester.ac.uk, Mathew.Evans@manchester.ac.uk. (*equal contributions)

Application of machine vision to behavioural research is making it possible to track the position and shape of objects or organisms across time with unprecedented precision. Nominally a simple task with markers and modern image processing libraries, tracking remains challenging when markers cannot be used. Further difficulties arise in the presence of occlusion, distractor objects, or if 3D information is required, necessitating the fusion of information from multiple cameras.

Measuring rodent whisker movements is one such challenging tracking problem. Mice are the most widely used model organism in mammalian biomedicine and whisker-mediated actions are a crucial part of their behavioural repertoire. There is therefore substantial interest in the problem of 'whisker tracking'. During natural active exploration, rodents rhythmically move their whiskers against objects in their surroundings ('whisking'). Accurate estimates of whisker shape are critical, as they can be used to estimate the mechanical forces and moments that drive mechanotransduction in the whisker follicle [1]. In the case of the whisker system, it is both difficult and undesirable to add markers to the whiskers as such a load may alter the animal's whisking behaviour.

Our approach was to image mouse whisking using a high speed camera (1000 frames per second; high-power infrared LED illumination) and to develop a parameterised model-based whisker tracker – WhiskerMan (see Figure 1 [2, 3]). Each whisker of interest is modeled as a quadratic Bezier curve. In a given video frame, the initial conditions of each Bezier curve were set by extrapolation from the solution of the previous frame. The Bezier curves were then fitted to the image using local gradient-based search with behaviourally appropriate regularisation.

We have recently generalized WhiskerMan to 3D. Whisking is commonly studied by imaging of the whiskers in the horizontal plane defined by the two eyes and the nose. However, whisker movement actually has a substantial 3D component and whisker-object contact can involve mechanical forces with a vertical component. To estimate 3D whisker shape, the whiskers were imaged using one high-speed camera for the horizontal plane and a second for the coronal ('face on') plane. A calibrated mapping from 3D coordinates to horizontal and vertical camera projections was calculated by moving a small object of known dimensions (accurate to <0.1mm) within the field of view. The WhiskerMan tracking algorithm was extended to fit 3D quadratic Bezier curves based on their projections on the two image planes.

Due to the combination of regularized fitting, use of temporal constraints and a parametric model with few (6-9) free parameters per whisker, WhiskerMan is robust to events such as whisker cross-overs that cause errors in previous trackers. WhiskerMan also permits multiple whiskers to be individually tracked at once. It was possible to image from a large proportion of the whisker array including Greek whiskers β-δ, and rows B-E within arcs 1-4. Figure 1 illustrates successful tracking of three whiskers simultaneously. Unlike with previous trackers, it was possible to obtain a 3D description of position and shape, including azimuthal and elevation angles and 3D curvature, for each whisker.

A second challenge in whisker tracking is to identify periods when whiskers touch an object. These are behaviourally important events but may only occur sparsely amongst millions of video frames. To detect touch onset and offset time with millisecond accuracy, in large volumes of video data (thousands of behavioural trials, millions of video frames in a typical experiment), we developed a semi-automatic touch detection GUI. Machine vision was used to identify object locations (a metal pole) in each video frame. This information was combined with whisker tracking outputs to determine the distance to the pole from the tracked whisker in each frame. A distance threshold identified candidate touches, which could then be quickly verified and curated to millisecond

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

precision through visual inspection in a Matlab based GUI. In turn, human-classified touches were used to refine detection thresholds for subsequent videos. The location and direction of touch (an important cue for sensing) is also measured.

In summary, WhiskerMan allows fast and accurate estimates of rodent whisker angle and curvature in 3D, and touch events with millisecond accuracy, with minimal human input. Whilst developed to describe mouse whisking behavior, the model-based approach outlined above may be useful to other tracking problems in behavioural research.

All experimental procedures were approved by the United Kingdom Home Office authorities and the University of Manchester ethical review.

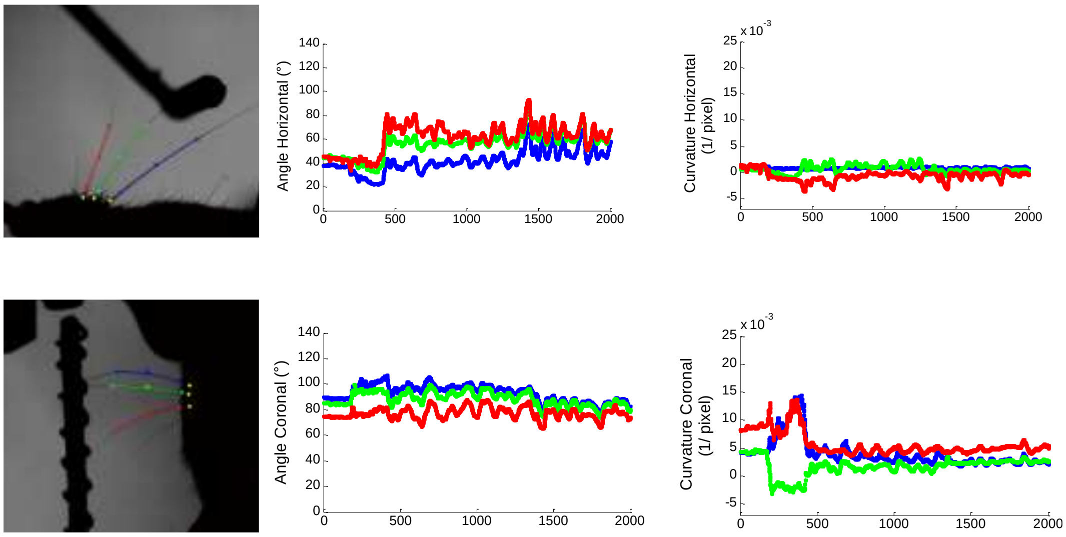


Figure 1. Output of WhiskerMan 3D.. The horizontal (top) and coronal (bottom) camera views are shown on the left. Overlaid red, green and blue traces are Bezier curve fits to each of three whiskers tracked simultaneously. Estimated whisker angle (middle) and curvature (right) of each whisker for a 2 second trial (sampled at 1kHz).

References:

1. Campagner, D., Evans, M. H., Loft, M. S., & Petersen, R. S. (2018). What the whiskers tell the brain. *Neuroscience*. *368*, 95 – 108.

2. Campagner, D., Evans, M. H., Bale, M. R., Erskine, A., & Petersen, R. S. (2016). Prediction of primary somatosensory neuron activity during active tactile exploration. *Elife*, *5*.

3. Bale, M. R., Campagner, D., Erskine, A., & Petersen, R. S. (2015). Microsecond-scale timing precision in rodent trigeminal primary afferents. *Journal of Neuroscience*, *35*(15), 5935-5940.

Visual Capture of Gait in Redirected Walking

Y. Rothacher[1], A. Nguyen[2], B. Lenggenhager[3], A. Kunz[2] and P. Brugger[1,4]

1 Department of Neurology, Neuropsychology Unit, University Hospital Zurich, Zurich, Switzerland. yannick.rothacher@usz.ch 2 Innovation Center Virtual Reality, ETH Zurich, Zurich, Switzerland. 3 Department of Psychology, Cognitive Neuropsychology, University of Zurich, Zurich, Switzerland. 4 Zurich Center for Integrative Human Physiology (ZIHP), University of Zurich, Switzerland.

Introduction

Virtual reality (VR) applications using head mounted displays (HMDs) allow users to deeply immerse in a virtual world. There is, however, one obstacle to a full-blown immersion, dubbed the “locomotion problem”. It concerns the user’s navigation through large virtual environments. While virtual worlds can easily be expanded infinitely, physical locomotion remains constrained by the size of the available room. Redirected walking is a technique that allows users of virtual reality applications to explore virtual environments larger than the available physical space [1]. This is made possible by manipulating the walking trajectory of users through visual rotation of the virtual surroundings, without the user noticing the interference. This causes the user to correct for the rotation by walking on a curved pathway (see Figure 1). At the extreme, a user could proceed infinitely far straight forward in the virtual environment while walking in full circle in reality. Apart from its value for designing VR-applications, the issue of redirection is of considerable interest for behavioral and psychological research. During walking, redirection causes a mismatch between visual and bodily feedback, the latter comprising vestibular, proprioceptive and somatosensory cues. Experimentally induced sensory mismatch situations have long been used in psychology to study various processes such as action monitoring [2] or feeling of agency [3]. In this study we focused on the important, yet unsolved question of how individual differences influence detection thresholds of redirected walking. Knowledge of these differences will not only help improve virtual reality applications, but also deepen our understanding of how humans process multisensory conflicts during locomotion.

Methods

In an explorative study with 60 healthy participants (age 18-35) we determined individual redirected walking detection thresholds (“redirection thresholds”), using a Bayesian adaptive method for threshold estimation. Participants also underwent comprehensive cognitive testing and an assessment of psycho-physical traits. Tests included visual-dependence measurements in various contexts (rod-and-frame task, assessment of the Romberg quotient, measurement of vection susceptibility), but also addressed non-visual body perception and control (blind veering, balance stability, interoception, somatosensory amplification). We used a linear mixed model procedure, accounting for participants as a random factor, to analyze the relation of test performances with individual redirection thresholds.

Results

When testing the assessed test performances univariately, a positive relation of individual redirection thresholds with the performance in the rod-and-frame test and with vection onset-time emerged. A negative relation with redirection thresholds was found for the Romberg quotient and blind sway. When combining the tested variables together in one model, only the effect of the rod-and-frame test performance remained significant.

Discussion

Our results allow to pinpoint the neuropsychological factors associated with an individual’s sensitivity to detect manipulations of gait while walking in a virtual reality environment. Of all tested variables, performance in the rod-and-frame test showed the most prominent association to redirection thresholds. Specifically, the more visual dependent a participant was in the rod-and-frame task, the worse he/she performed at detecting redirected walking manipulations. This result supports the view that visual dominance over body-related signals constitutes a “visual capture of gait”, which hampers the detection of any locomotor perturbation. The paradigm of redirected walking may be used for various further research questions regarding the processing of multisensory signals during active

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

walking. Such research-based implementations of redirected walking could also be applied to specific clinical populations suffering from impairments of action monitoring or sensory integration during locomotion (see [4]).

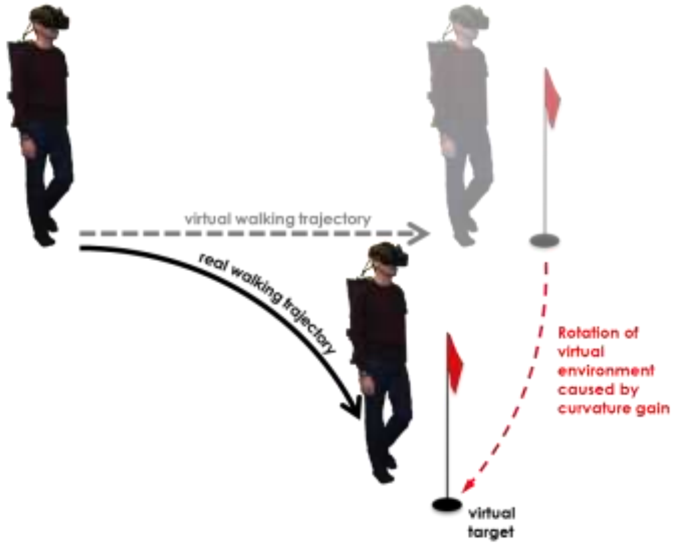


Figure 1. Illustration of a walking trajectory manipulation through redirected walking (note: the unit of curvature gain corresponds to 1/radius of trajectory curvature).

Ethical Statement

All experimental procedures were approved by the "Kantonale Ethikkomission Zurich" (BASEC Nr. 2016-01153).

References

1. Razzaque, S., Kohn, Z., Whitton, M. C. (2001). Redirected Walking. *Proc. Eurographics*, 289–294.
2. Bock, O. (1992). Adaptation of aimed arm movements to sensorimotor discordance: evidence for direction-independent gain control. *Behav. Brain Res.* **51**, 41–50.
3. Farrer, C., Franck, N., Georgieff, N., Frith, C.D., Decety, J., Jeannerod, M. (2003). Modulating the experience of agency: a positron emission tomography study. *Neuroimage* **18**, 324–333.
4. Kannape, O.A., Perrig, S., Rossetti, A.O., Blanke, O. (2017). Distinct locomotor control and awareness in awake sleepwalkers. *Curr. Biol.* **27**, R1102–R1104.

Agora: a complex social recognition paradigm

Sheila Sanchez Garcia, Gernot Riedel

IMS, UoA

Introduction: A variety of neuropsychiatric disorders are characterized by disruptions in social behaviour and social recognition, including depression, autism spectrum disorders, bipolar disorders, obsessive-compulsive disorders, and schizophrenia. Developing tools for the assessment of social skills in mouse models is essential to further advance in the understanding of these diseases. Social interaction in rodents and especially mice is highly dependent on smell (1). Our specific interest in smell related behaviours arose from our current work on understanding mechanisms of Parkinson's disease (PD). An early phenotype of PD is anosmia (2) and this may be translated to animals using a behavioural task that carries features of human social interactions.

The Agora was the central place in ancient Greek cities (see Fig. 1). In this new paradigm, we seek to imitate the agora as the central square for social activities within a community. This concept is still alive today in many cities, villages and communes in Europe, in which the central town square hosts restaurants, benches, shops, or museums and churches at the edges of the agora. Consequently, social encounters and active social exchanges are more likely at the perimeter and the central area is merely an open space.

Figure 1. Village squares in southern France (left) and in Valbonne (right).

Method: In the Agora for rodents (see Fig. 2), stranger mice are confined in cubicles around the perimeter of the square and thus can select between up to 5 partners for social interaction. They are free to choose and in the discrimination phase are conflicted with a novel and 4 familiar interaction partners.

Figure 2. AGORA maze with an inside decagon and peripherally attached cubicles (Ugo Basile).

Animals and Protocol: The animals used in the experiments were all wildtype male mice 15-17 weeks old at the beginning of the experiment. The main principle of this test is based on the free choice by a subject mouse to spend time in any part of an open circular arena (Agora) to which 5 cubicles are attached housing an animal inside each. It consists of two experimental phases, the second of which includes a stranger animal never encountered before replacing one of the familiar interaction partners. To quantify social tendencies of the experimental mouse, the main task is to measure the time spent with the stranger mouse in comparison with the familiar conspecifics. This is achieved using video observation (AnyMaze) and data extraction from tracking of movement within the

Agora. Thus, the experimental design of this test allows evaluation of preference for social novelty in contrast to the propensity of time spent with a previously encountered familiar mouse.

First Results and Conclusion: Using this experimental set up, we validated that there was no bias for any cubicle during habituation (not shown) or during the initial encounter, when 5 stranger mice were placed into the cubicles. Upon re-exposure and inclusion of a novel stranger, the test protocol produced robust social recognition in wildtype mice. Male mice can readily identify a stranger mouse and spent significantly more time in a zone adjacent to its cubicle compared to the 4 recently encountered conspecifics (see Fig. 3). This paradigm therefore provides a more realistic behavioural scenario and offers the possibility of determination of more complex interaction paradigms, for example comparison between recent and remote memory of social partners.

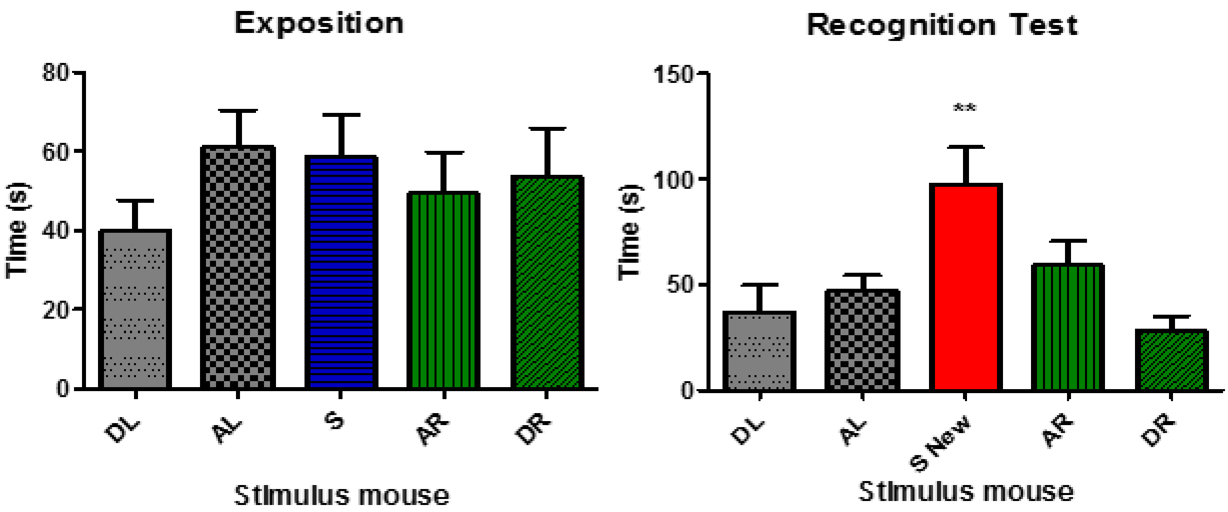


Figure 3. Male WT mice show robust social recognition for 5 stimulus mice in the Agora paradigm. Interaction time of test mouse with each stimulus mouse during Exposure (A) and Recognition Test (B). Bars represent the mean density ± SEM. **P=0.0016

Ethical statement: Mice were housed and tested in accordance with UK Home Office regulations. All experimental procedures were subject to the University of Aberdeen's Ethics Board and conducted in accordance with the European Directive on the Protection of Animals used for Scientific Purposes (2010/63/EU) and the Animal (Scientific Procedures) Act 1986.

References:

1. Zou J, Wang W, Pan YW, Lu S, Xia Z. Methods to measure olfactory behavior in mice. Curr Protoc Toxicol. 2015 Feb 2;63:11.18.1-21.

2. Rey NL, Wesson DW, Brundin P. The olfactory bulb as the entry site for prion-like propagation in neurodegenerative diseases. Neurobiol Dis. 2018 Jan;109(Pt B):226-48.

Using Whisker Movements And Locomotion To Measure Neurodegenerative Disease In Rodent Models

U. Simanaviciute[1,2], M.H. Yap[3], B.M. Hewitt[2,3] and R.A. Grant[2]

1. School of Biological Sciences, University of Manchester, Manchester, UK; ugne.simanaviciute@student.manchester.ac.uk; 2. School of Biology & Conservation Ecology, Manchester Metropolitan University, Manchester, UK; robyn.grant@mmu.ac.uk; 3. School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, Manchester, UK

As the lifespan of the world population is extending, the incidence of neurodegenerative disease is growing. Millions of people are affected every year by the loss of function and/or structure of neurons in the central nervous system. Many researchers have been looking at animal models to further understand neurodegenerative diseases and possible treatment types. Although rodents are not the best models for human neurodegenerative diseases, they still remain one of the cheapest and quickest mammals to show results [1]. Most studies on rodents involve either intrusive techniques or extensive training for traditional motor coordination tasks; however, developments in behavioural task designs have suggested that whisking and locomotion might be a highly quantitative way to measure neurodegeneration. It has been shown that tracking natural whisking behaviour can show significant differences between wild-type and transgenic models of Amyotrophic Lateral Sclerosis [3] and Huntington's Disease [2]; indeed, in some cases, whisking was found to be an even more sensitive test of motor dysfunction than any previous behavior tests [2]. This study aims to characterise how whisking, locomotion and measures of exploratory behaviour can help us to understand neurodegenerative disease in nine mouse models of Alzheimer's disease, Amyotrophic Lateral Sclerosis, Cerebellar Ataxia, Cortical Development Disorders, Parkinson's disease and Huntington's disease. All animals were filmed using the same methods and analysed using the same software, so that they could be reliably compared. Whisker movements of many of these mouse models have never been described before.

This study uses state-of-the-art behavioural set-ups, to collect high-speed video data quickly and non-invasively from rodents. The video was then analysed using previously developed custom behavioural software, including the Automated Rodent Tracker (ART) [4] and the Automated Rodent Whisker Tracker (ARWT) [5]. This gives highly quantitative and precise movement measurements of whisker frequency, amplitude and angular position, as well as locomotion speed. We present here data to show that mouse whisking, locomotion and exploratory behaviours are affected in mouse models of neurodegeneration. As tracking whisker and locomotion behaviour is non-invasive, highly quantitative and automated, it allows easy collection of large amounts of precise data. In addition, whisking and locomotion can be collected from freely moving animals and eliminates the need for animal training. We recommend that whisking and locomotion be incorporated in to behavioural test batteries of rodent models.

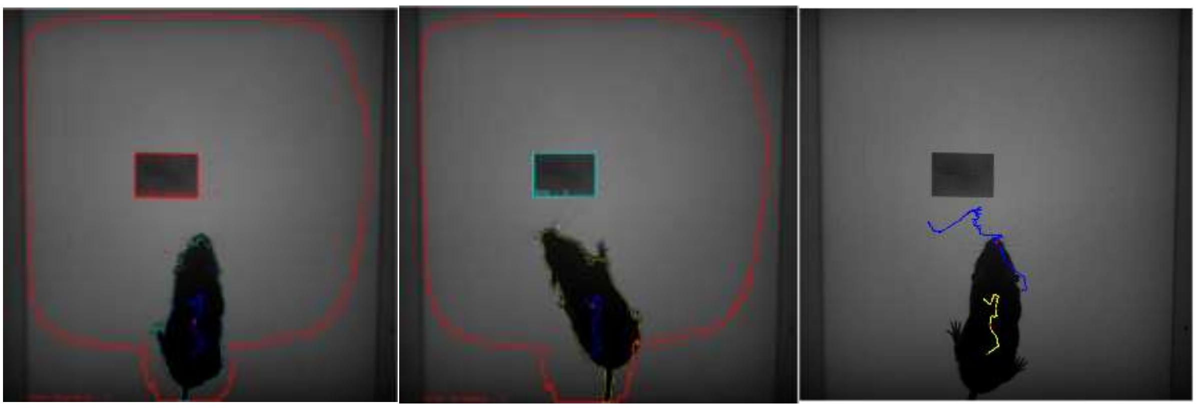


Figure 1. Left: a single video tracking from ART from an example clip of a transgenic mouse with Cerebellar Ataxia. Middle: a single video tracking from ARWT from the same example video. Right: overlaid nose tip

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

(blue) and body (yellow) tracking seen in the batch processing review window from ART from the same example clip.

References

1. Burns, Terry C., Matthew D. Li, Swapnil Mehta, Ahmed J. Awad, and Alexander A. Morgan (2015) 'Mouse Models Rarely Mimic the Transcriptome of Human Neurodegenerative Diseases: A Systematic Bioinformatics-Based Critique of Preclinical Models'. *European Journal of Pharmacology* 759: 101–17, **no. 10.1016**

2. Garland, H., N. I. Wood, E. A. Skillings, P. J. Detloff, A. J. Morton, and R. A. Grant. (2017) Characterisation of Progressive Motor Deficits in Whisker Movements in R6/2, Q175 and Hdh Knock-in Mouse Models of Huntington's Disease. *Journal of Neuroscience Methods*, **no. 10.1016**

3. Grant, R. A., P. S. Sharp, A. J. Kennerley, J. Berwick, A. Grierson, T. Ramesh, and T. J. Prescott. (2013) 'Abnormalities in Whisking Behaviour Are Associated with Lesions in Brain Stem Nuclei in a Mouse Model of Amyotrophic Lateral Sclerosis'. *Behavioural Brain Research* 259, **no. 10.1016**

4. Hewitt, Brett M., Moi Hoon Yap, Emma F. Hodson-Tole, Aneurin J. Kennerley, Paul S. Sharp, and Robyn A. Grant (2017) 'A Novel Automated Rodent Tracker (ART), Demonstrated in a Mouse Model of Amyotrophic Lateral Sclerosis'. *Journal of Neuroscience Methods*, **no. 10.1016**

5. Hewitt, B., Yap, M. H., Hodson-Tole, E., and Grant, R. A. (In submission). 'Fully Automated Whisker Detector'. *IET Image Processing*.

Measuring Anxiety-Like Behavior in Pet Dogs

S. Sulkama[1,2,3], M. Ahola[1,2,3], K. Tiira[1,2,3] and H. Lohi[1,2,3]

1 Department of Veterinary Biosciences, University of Helsinki, 00014 Helsinki, Finland. sini.sulkama@helsinki.fi; 2 Research Programs Unit, Molecular Neurology, University of Helsinki, 00014 Helsinki, Finland; 3 The Folkhälsan Institute of Genetics, 00290 Helsinki, Finland

Background:

Anxiety disorders are among the most common disorders in humans affecting almost one sixth of people at some point in life [1] and are part of behavioral problems observed in domestic dogs [2, 3]. Dogs suffer from various breed-specific phobias, compulsions and other traits, including fear towards strangers or new situations, noise phobia, separation anxiety, impulsivity and hyperactivity, and stereotypic behaviors, that respond to human anxiety disorders and can be measured. In this project, we utilize the presence of natural neuropsychiatric phenotypes in dogs. There is a large dog population potentially available for studying anxieties. Due to similarities in anxiety-like behaviors between dogs and humans, dogs appear as excellent natural animal models in human anxiety research. In this study, we defined the following specific aims:

- To collect the largest existing behavioral data from dogs using our online behavioral questionnaire targeted to dog owners and breeders.
- To analyze the behavioral questionnaire data across anxieties to identify the low- and high-anxiety dogs and the extend of anxiety comorbidity.
- To invite selected dogs from the low- and high-anxiety categories for behavioral tests to confirm the anxiety traits and the validity of the questionnaire data.

Methods:

Because our aim is to reach a great number of dogs from different breeds, we developed an owner-completed online anxiety questionnaire (www.koirangeenit.fi/questionnaires/behavior/) to study anxieties in dogs. The behavioral questionnaire includes background questions which cover various demographic, environmental and behavioral features related to each dog's life history, and questions about various anxiety-like behaviors such as fear, noise phobia, separation anxiety, ADHD-like behavior, compulsion, aggression and fear of high places or surfaces. We have already collected over 13 500 questionnaire responses from over 200 breeds of dogs. To analyze the massive data from questionnaires, we developed a script that automatically categorizes dogs according to the responses in different anxiety traits. We defined three groups for each anxiety trait: low, moderate and high. Dogs in the "low" group were reported no or rare history of the trait and its frequency. In the "high" group dogs showed the trait "often" or "always" with high frequency so that it disturbs dog's daily routines. Our preliminary results show that noise phobia and fear are the most common anxieties in dogs, and separation anxiety and aggressiveness the most uncommon anxiety traits. Furthermore, dog breeds show remarkable differences in prevalence of all anxiety traits.

Although the behavioral questionnaire has been shown to accurately phenotype canine behavior [4], as a secondary objective of our study is to develop lab-based measures for dogs, which parallel the neurocognitive tasks that have been used in humans. We will test a subsample of dogs from various breeds with behavioral tests in order to validate the questionnaire within our study population. Validity is assessed by comparing owner's ratings in the behavioral questionnaire against the behavioral test codings of the dogs in standardized situations. We will develop behavioral tests to measure as many anxiety traits as possible. We already developed a behavioral test to measure fearfulness and piloted a test for impulsive/hyperactive behavior. The behavioral test for fearfulness included the following three parts: (1) a stranger approaching and trying to pet the dog, (2) a free 2-minute exploration in a novel space while the owner remained passive and (3) a reaction to a novel human-like moving object. In the first part, the owner stayed passive with loose leash while the test leader approached the dog from the front. Test leader stopped in front of the dog, bent over the dog and tried to pet the dog on the head. In the second part, the owner stood passive in the corner inside a marked area (1 m × 1 m) while the dog was allowed to freely explore the novel place for 2 min. The floor of the room was marked with crossing lines to measure the activity and movement of the dogs using the number of crossed lines. The dog's activity (=lines crossed) and time spent within the close vicinity of the owner were measured. In the third part the novel object resembling a human figure was built on top of the remote controlled toy car, moving sideways in relation to the dog before stopping (distance to the dog was 15 m). The dog was unleashed when the novel object had stopped and the owner was advised to follow instructions of the test leader. With each dog, we waited as long as the dog felt safe to sniff the object, and some dogs were also fed with treats insuring that no fear towards the object was left. The delay of inspecting and the degree of independence of approaching (before, together or after the owner)

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

the object were measured. All test parts were video recorded and behavioral data was later analyzed from the videos. For hyperactive behavior, we intend to test a subsample of dogs using accelerometry-based activity monitors to measure their motor activity levels. We will monitor 30 hyperactive dogs and 30 dogs with low activity levels non-stop for one week period to figure out whether owner's questionnaire ratings and activity levels measured with activity monitors correlate with each other. To measure cognitive impulsion in dogs we piloted the touch screen-based Stop Signal Task (SST) which measures response inhibition and impulsive action, and which have earlier been used for cognition tasks in dogs and rhesus monkeys successfully [5, 6]. We assume that for the most impulsive and hyperactive dogs, inhibiting their responses is more difficult, they cannot wait and they make more mistakes. The final test for impulsion is still in the process of being planned because we would like to develop a test which requires less training of dogs. To measure fear of high places or surfaces we will introduce different surfaces, stairs and high places for a few dozens of dogs and video record the test for later analyses.

Measuring some anxiety-like behaviors such as noise phobia in dogs is ethically dubious because tests may cause discomfort for dogs. However, in Finland we have the Finnish mental test and one part of the test measures dog's reactions towards gunshots. We have the behavioral questionnaire responses of many dogs which have performed the Finnish mental test as well. Thus we can compare noise phobia scores from our questionnaire and the score from the gunshot part of the Finnish mental test without causing any additional discomfort for the dogs. For separation anxiety and compulsions we can use video data recorded by owners to confirm the severity of the anxiety trait and to compare behavior and our behavioral questionnaire data.

This study has potential to gain novel insights about anxiety-like behaviors in dogs with the largest existing validated behavioral data of pet dogs. Accelerometers provide us a new tool for measuring activity levels of dogs and catch the hyperactive behavior.

References:

1. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives* of *General Psychiatry* **62**: 593.

2. Overall KL, Dunham AE, Frank D (2001). Frequency of nonspecific clinical signs in dogs with separation anxiety, thunderstorm phobia, and noise phobia, alone or in combination. *Journal of the American Veterinary Medical Association* **219**: 467–473.

3. Blackwell EJ, Bradshaw JWS, Casey RA (2013). Fear responses to noises in domestic dogs: Prevalence, risk factors and co-occurrence with other fear related behavior. *Applied Animal Behavior Science* **145**:15–25.

4. Tiira K, Lohi H (2014). Reliability and validity of a questionnaire survey in canine anxiety research. *Applied Animal. Behavior Science* **155**: 82–92.

5. Range F, Aust U, Steurer M, Huber L (2008). Visual categorization of natural stimuli by domestic dogs. *Animal Cognition* **11**: 339–347.

6. Liu S, Heitz RP, Bradberry CW (2009). A touch screen based stop signal response task in rhesus monkeys for studying impulsivity associated with chronic cocaine self-administration. *Journal of Neuroscience methods* **177**: 67–72.

What Makes Fido an Elite Competitor?

Matthew Sullivan, Stuart Spence, Hannah Lockwood and Huw Lloyd

Division of Biology & Conservation Ecology, Manchester Metropolitan University, UK; m.sullivan@mmu.ac.uk

Tri-axial accelerometers have been used extensively with the aim of activity recognition in humans and, less so, in animals. At each sampling point the accelerometer data is labelled with the behaviour being displayed at that instant. The raw data is then windowed and new derived features are calculated for each window of data. These new features, such as mean and variance of the data, are then applied to a labelled training subset of the data by a machine learning or classification technique to build an algorithm to see how well the behavioural category is correctly assigned. The most successful algorithm is then applied to the unlabelled test dataset. Much of the literature is motivated by the potential for rehabilitation and therapy of movement disease and disorders such as in stroke and Parkinson's disease. This involves seeing where, for example, patterns of gait vary between patient and control subjects.

A few domesticated animal species have been selected for their abilities in competitive arenas, such as horses and working dog breeds. Canine agility competitions test speed and manoeuvrability. There is in the UK a national league and championship structure and it is not unreasonable to describe the most successful dogs as elite canine athletes. This paper introduces an ongoing project looking at accelerometer signatures of elite dogs and their ordinary domestic pet counterparts when performing particular behaviours. The aim of this work is to be able to identify components of actions such as jumps in order to better understand the strengths of particular dogs. The focus will be on methods development including data processing, videography and, uniquely, qualitative data from dog owners and breeders, and other stakeholders.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Measuring Mouse Vision Using Innate Behavioural Responses

R. Storchi*, F.P. Martial*, J. Wynne*, S. Ryan**, C.G. Twining*, T. F. Cootes*, R. Killick** and R.J. Lucas*

***Faculty of Biology, Medicine and Health | University of Manchester, Manchester, UK**
****Department of Mathematics and Statistics | Lancaster University, Lancaster, UK**
Corresponding Author: riccardo.storchi@manchester.ac.uk

Optogenetics and stem cell based treatments provide new opportunities for treating retinal dystrophies [1, 2]. The development of these treatments largely relies on preclinical studies performed in mice. An important step in evaluating treatment effectiveness is represented by behavioural assays of mouse vision. Many commonly used tests are based on sub-conscious, reflex responses [3, 4] whose activity is only indirectly related to perceptual vision. More relevant tests largely rely upon learned associations between visual stimuli and either positive or negative conditioned stimuli [5] and are inherently throughput because they require long training periods and only allow association with single visual stimuli. An alternative and more humane approach to assess vision relies on measurements of mouse spontaneous behaviour. These tests rely on the hypothesis that when mice detect a change in their visual environment they naturally change their behaviour. These tests are currently low throughput as they rely on very simplified measures of behaviour such as distance moved within an open arena [1] or the time required to move from a light to a dark area [6]. Here we show that, by combining a better experimental design with more sophisticated behavioural measures based on changepoint analysis, we can obtain reliable high throughput readouts of mouse vision. We designed an open field apparatus to capture mouse behaviour simultaneously with multiple cameras while stimulating from the top (Figure1A&B) and to perform reliable tracking by using animal silhouette and body markers (Figure 1C-F). We then performed three series of experiments designed to capture a large repertoire of innate behavioural responses to different visual stimuli (Figure 2) and we measured contrast sensitivity (Figure 2A) as well as visual acuity (Figure 2C). We then analysed behavioural time series by extracting the time at which changes in behaviour were detected and by classifying these changes (Figure 3A&B; changepoint detection was based on mean changes and performed with PELT algorithm [7]). Our method, fully automated, allows us to detect reliable behavioural responses on a single trial basis and to collect multiple trials within a single experimental session (Figure 3C&D).

Ethical Statement:

Experiments were conducted in accordance with the Animals, Scientific Procedures Act of 1986 (United Kingdom) and approved by the University of Manchester ethical review committee. The work is funded by a David Sainsbury Fellowship (NC/P001505/1) from National Centre for Replacement Refinement and Reduction of Animals in Research (NC3Rs).

References:

[1] Lagali PS et al (2008) Light-activated channels targeted to ON bipolar cells restore visual function in retinal degeneration. Nat Neurosci(6): 667-675. DOI:10.1038/nn.2117

[2] Barnea-Cramer et al (2016) Function of human pluripotent stem cell-derived photoreceptor progenitors in blind mice. Sci Rep(6):29784. doi:10.1038/srep29784

[3] Lucas RJ et al (2001) Characterization of an ocular photopigment capable of driving pupillary constriction in mice. Nat Neurosci 4(6): 621-626. DOI:10.1038/88443

[4] Kretschmer F et al (2015) A system to measure the Optokinetic and Optomotor response in mice. J Neurosci Meth 256:91-105. DOI:10.1016/j.jneumeth.2015.08.007

[5] Nishiguchi KM et al (2015) Gene therapy restores vision in rd1 mice after removal of a confounding mutation in Gpr179. Nat Commun 23(6):6006. doi: 10.1038/ncomms7006.

[6] Bourin M, Hascoët M (2003) The mouse light/dark box test. Eur J Pharmacol. 463(1-3):55-65. PMID: 12600702

[7] Killick R et al (2012) Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association* 107(500) 1590-1598.

Characterization of Behavior and Voluntary Alcohol Intake in Wistar and Lister Hooded Rats

N. Tjernström and E. Roman

Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden, nikita.tjernstrom@farmbio.uu.se

Introduction

Preclinical researchers are faced with the challenge of selecting the optimal model organism based on the research question addressed. Few comparisons between different rat strains are available [1]. Sometimes anecdotal reports and laboratory traditions guide the choice. Therefore, studies comparing different rat strains may refine preclinical research.

The multivariate concentric square field™ (MCSF) test is based on forced exploration, but the animal can freely choose between different environments aiming at provoking behaviors associated with activity, risk taking and shelter seeking [2, 3]. In the final interpretation, the descriptive parameters are brought together into the functional categories general activity, exploratory activity, risk assessment, risk taking and shelter seeking [2].

When comparing Wistar rats from different vendors, pronounced differences in voluntary alcohol intake have been demonstrated [4, 5]. However, with the aim of using Lister hooded rats in future experiments, little information about voluntary alcohol intake in this strain was available.

This study aimed to evaluate behavioral profiles and voluntary alcohol intake in Wistar rats after an interruption of breeding at the vendor, and to compare Wistar rats with Lister hooded rats.

Material and methods

Animals

Male rats of two different strains, Lister hooded (HsdOla:LH) and Wistar (RccHan:WI) rats (Envigo, Horst, the Netherlands) were delivered at 6 weeks of age (n=21 of each strain). The animals were housed 3 per cage in transparent type IV cages with raised lids, wood-chip bedding and two pieces of paper as enrichment. The cages were kept in an animal room on reversed light/dark cycle (lights off at 6:00 am) with a masking background noise. Once the alcohol intake studies began the rats were single housed in high type III cages with wood-chip bedding and two pieces of paper until the end of the study. The experimental protocol and use of animals in this study was approved by the Uppsala Animal Ethical Committee, and was consistent with the Swedish Legislation on Animal Experimentation (Animal Welfare Act SFS1998:56) and the European Union Directive on the Protection of Animals Used for Scientific Purposes (2010/63/EU).

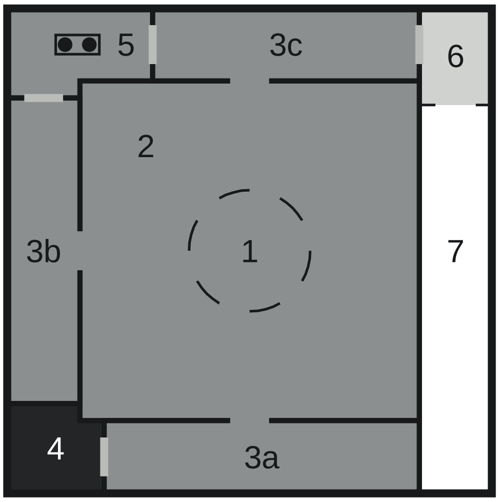


Figure 1. Schematic layout of the MCSF arena. The shading of the zones represents illumination levels in the arena. The arena is divided into zones: 1, the central circle; 2, the central field; 3a-c, corridors; 4, the dark corner room; 5, the hurdle with hole board; 6, the slope; and 7, the bridge [6].

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

The multivariate concentric square field™

Before the alcohol intake studies began the animals underwent testing in the multivariate concentric square field™ test (MCSF) (Figure 1). The MCSF is used for behavioral profiling of rodents [2, 3].The animal is placed in an arena and able to explore and freely choose between different environments during 20 minutes. The MCSF contains a dark corner room (DCR) which represents a safe area were the animal can seek shelter. It also has two risky areas, a brightly illuminated bridge and a central circle. Lastly the animal can explore an elevated hole board in the hurdle. The different areas are connected with corridors which also provide areas for the animal to explore. At the start of the trial the animal is placed in the central field facing the wall without any openings.

The room that the MCSF is placed in has the same background noise as the animal room and the experimenter leaves the room during the trial, to make sure not to influence the animal's behavior. The trials are recorded using a camera mounted in the ceiling above. Noldus Ethovision® XT 12.0 (Noldus Information Technology, Wageningen, The Netherlands) is used to track the animal's movements during the trial. Behaviors that can't be tracked automatically is manually scored by the experimenter. These behaviors are rearing, grooming, stretch attend posture (SAP), boli and urine left in the arena after the trial.

Voluntary alcohol intake

A modified intermittent two-bottle free-choice paradigm was used, with three consecutive 24-hour sessions per week for six weeks [7]. Each rat had access to two bottles, one with tap water and the other with a 20% (v/v) alcohol solution. Bottle positions were changed for each session to avoid position preference. After each 24-hour session the bottles were weighed to measure intake of water and alcohol. Alcohol preference was calculated as alcohol intake (g) divided by total fluid intake (g). Drinking sessions took place on Tuesdays, Wednesdays and Thursdays. Changing of cages was done on Fridays to make sure not to disturb the animals during alcohol intake sessions.

Statistical analyses

One Lister hooded animal was excluded due to health reasons. Statistical analyses were carried out using Statistica 13.2 (Dell Inc., Tulsa, OK, USA). Differences were considered statistically significant at $p<0.05$. Most of the data did not show normal distribution according to the Shapiro-Wilk's W test, hence nonparametric tests were used. Between group differences were analyzed with Mann-Whitney U-test. Intragroup differences over time was analyzed with Friedman ANOVA followed by Wilcoxon Matched Pairs Tests where appropriate. Behavioral profiles in the MCSF test were presented using the trend analysis, which is based on grouping behaviors within the same context and then ranking the individuals against each other. The rank values are then summed into a sum rank for the functional categories general activity, exploratory activity, risk assessment, risk taking and shelter seeking [2].

Results

MCSF

The result from the trend analysis is shown in Figure 2. The only functional behavior category that differed between the strains was risk taking, where Lister hooded rats exhibited higher level of risk taking behavior.

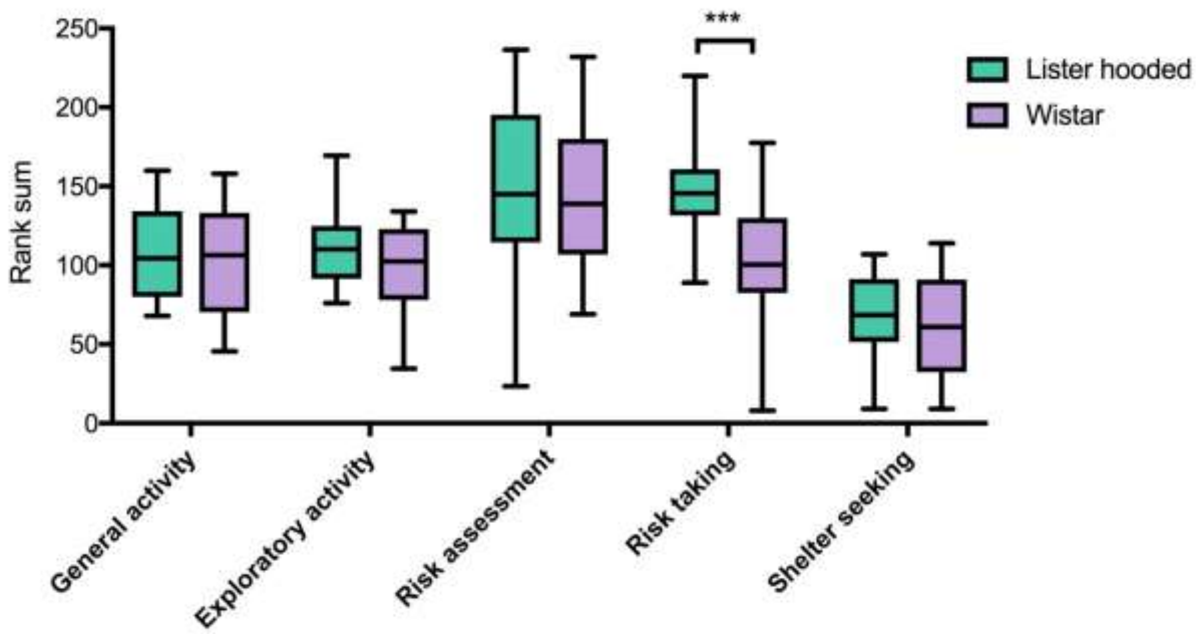


Figure 2. Rankings of the functional behavioral categories in the MCSF trend analysis. Data are shown as median with upper and lower quartiles, min and max. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ comparing Lister hooded and Wistar rats (Mann-Whitney U-test).

Voluntary alcohol intake

Alcohol intake (g/kg) and preference (%) for the two strains are shown in Figure 3. All 18 sessions are represented in the figure. No significant differences in voluntary alcohol intake was found during the six weeks of access. Lister hooded rats displayed a significantly higher preference during six of the sessions. The variance within the group was larger for Lister hooded rats, most obvious in alcohol preference on the first drinking sessions of each week.

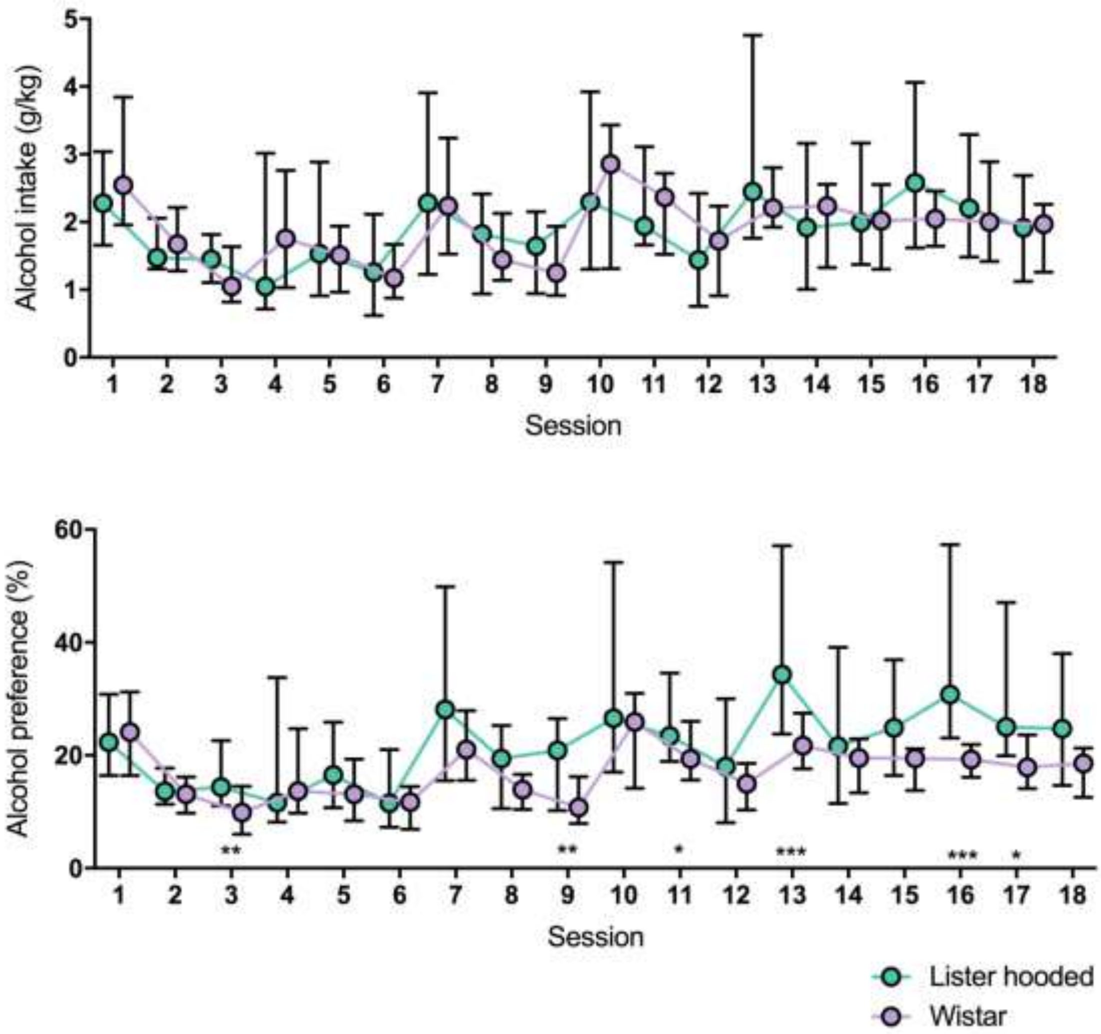


Figure 3. Alcohol intake (g/kg) and preference (%) in Lister hooded and Wistar rats during the 18 sessions of voluntary alcohol intake during a period of 6 weeks. Data are shown as median with interquartile range. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ comparing Lister hooded and Wistar rats (Mann-Whitney U-test).

Data representing the average intake and preference on the first, second and third day of access is shown in Figure 4. No significant difference between the strains was found when considering the alcohol intake. Lister hooded had significantly higher preference than Wistar rats on the second and third alcohol session of the week. When comparing the difference between alcohol sessions within the respective strains, significant differences were found between all alcohol sessions for both strains with significantly higher intake and preference on the first day of access.

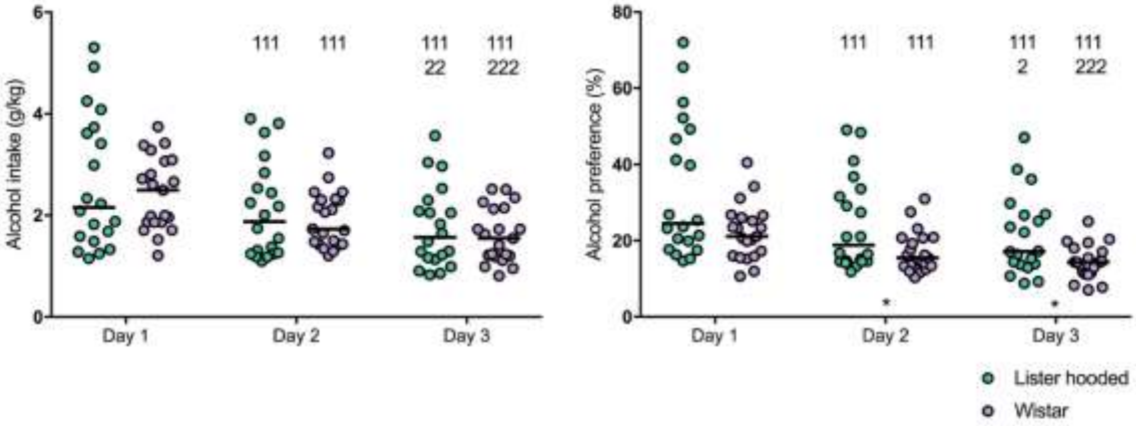


Figure 4. Average alcohol intake (g/kg) and preference (%) in Lister hooded and Wistar rats during the three weekly drinking sessions. Data are shown as scatter dot plots with the line at median. *p<0.05, **p<0.01, ***p<0.001 comparing Lister hooded and Wistar rats (Mann-Whitney U-test), [1]p<0.05, [11]p<0.01, [111]p<0,001 compared to the intake on the first day of access within the respective strain (Wilcoxon's matched pairs test), [2]p<0.05, [22]p<0.01, [222]p<0.001compared to the intake on the second day of access within the respective strain (Wilcoxon's matched pairs test).

Discussion

The results from this study revealed smaller differences between Wistar and Lister hooded rats, than previously have been demonstrated when comparing Wistar rats from different vendors [4, 5, 8]. The behavioral difference that could be seen in the trend analysis was that the Lister hooded rats displayed a higher level of risk taking behavior relative to Wistar rats. The parameters that make up risk taking behavior are related to the illuminated bridge and the central circle. The behavior of Lister hooded and Wistar rats in the MCSF has been evaluated previously [3]. The study found that Wistar rats spent more time in the DCR and took longer to enter the central circle. The trend analysis was not used at that time but these behaviors are related to shelter seeking and low risk taking behavior, which thus are in agreement with the findings in this study. In a previous comparison, Wistar rats had high thigmotaxis in the open field test [9]. On the other hand, with large enough sample size individual differences have been demonstrated, and both high and low risk taking individuals, respectively, among outbred Wistar rats [10]. The Lister hooded rats' behavior are not as well characterized as for example Wistar rats. They are commonly used in operant techniques that require high attentional function [11]. A more risk taking behavior is also beneficial when faced with such a task. The operant chamber is an unfamiliar environment and to learn the requirements of the task they need to be active and nose poke the illuminated holes to eventually succeed and get the reward.

When it comes to voluntary alcohol intake no difference between Wistar and Lister hooded rats were found during the 18 sessions. However, Lister hooded rats had a higher alcohol preference towards the end of the study. Moreover, pronounced individual differences within the Lister hooded rats were noted. The alcohol intake of Wistar rats in this study as well as in a parallel, unpublished study, indicates a lower median intake as well as a lower variance within the group compared to previous studies [7, 12]. Based on these results the Lister hooded rats would be more suitable for phenotypic division into high- and low-drinking subgroups, which previously has been useful in studies using the Wistar strain [7, 12]. The reason for this change in drinking behavior seen in the Wistar rats is not known, but it could be due to an interruption in breeding caused by an outbreak of parvo virus at the vendor, or a change in breeding material. However, despite several previous studies utilizing the large individual differences in voluntary alcohol intake among the Wistar rats [4, 7, 12], a batch-dependent effect cannot be excluded.

In conclusion, comparisons between laboratory animal strains, as here when comparing behavioral profiles and voluntary alcohol intake are important and may facilitate the choice of animal strain used in preclinical research.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Acknowledgements

Funding from the Alcohol Research Council of the Swedish Alcohol Retailing Monopoly, the Swedish Brain Foundation and the Facias Foundation supported this study. The behavioral characterization was conducted with support from the Uppsala University Behavioral Facility (UUBF).

References

1. Cavigelli SA, Michael KC, Ragan CM (2013) Behavioral, Physiological, and Health Biases in Laboratory Rodents - a basis for understanding mechanistic links between human personality and health. In: Anim. Personal. Behav. Physiol. Evol. University of Chicago Press, pp 441–498

2. Meyerson BJ, Jurek B, Roman E (2013) A Rank-Order Procedure Applied to an Ethoexperimental Behavior Model—The Multivariate Concentric Square Field^{TM} (MCSF) Test. *J Behav Brain Sci* **2013**:350–361

3. Meyerson BJ, Augustsson H, Berg M, Roman E (2006) The Concentric Square Field: A multivariate test arena for analysis of explorative strategies. *Behav Brain Res* **168**:100–113

4. Palm S, Roman E, Nylander I (2011) Differences in voluntary ethanol consumption in Wistar rats from five different suppliers. *Alcohol* **45**:607–614

5. Goepfrich AA, Gluch C, Friemel CM, Schneider M (2013) Behavioral differences in three Wistar Han rat lines for emotional reactivity, cognitive processing and ethanol intake. *Physiol Behav* **110**–111:102–108

6. Lundberg S, Abelson KSP, Nylander I, Roman E (2017) Few long-term consequences after prolonged maternal separation in female Wistar rats. PLoS ONE.

7. Momeni S, Roman E (2014) Subgroup-dependent effects of voluntary alcohol intake on behavioral profiles in outbred Wistar rats. *Behav Brain Res* **275**:288–296

8. Palm S, Hävermark Å, Meyerson BJ, Nylander I, Roman E (2011) When is a Wistar a Wistar? Behavioral profiling of outbred Wistar rats from five different suppliers using the MCSF test. *Appl Anim Behav Sci* **135**:128–137

9. Momeni S, Sharif M, Ågren G, Roman E (2014) Individual differences in risk-related behaviors and voluntary alcohol intake in outbred Wistar rats. *Behav Pharmacol* **25**:206–215

10. Ohmura Y, Yamaguchi T, Futami Y, Togashi H, Izumi T, Matsumoto M, Yoshida T, Yoshioka M (2009) Corticotropin releasing factor enhances attentional function as assessed by the five-choice serial reaction time task in rats. *Behav Brain Res* **198**:429–433

11. Momeni S, Segerström L, Roman E (2015) Supplier-dependent differences in intermittent voluntary alcohol intake and response to naltrexone in Wistar rats. Front Neurosci.

Strain differences in acoustic/vibrational startle response in zebrafish larvae

R. van den Bos1, G. Flik1 and M. Gorissen1

1Department of Animal Ecology and Physiology, Faculty of Science, Radboud University, Nijmegen, the Netherlands
ruudvdbos@science.ru.nl, ruudvandenbos1@gmail.com

Introduction

Zebrafish are still gaining momentum in many areas of biomedical research. Many strains of zebrafish exist, such as AB, TL and TU. As these strains are not genetically homogeneous, differ in genetic make-up and appearance due to mutations (TL *versus* AB/TU), and laboratories are rearing their own offspring in as yet not standardized conditions, results of studies may vary between institutes - and even within institutes.

For instance, we [1,2,3; in preparation] and others [e.g. 4,5] have shown that adults and larvae of the AB and TL strain differ in physiology (AB higher levels of cortisol than TL), mRNA expression levels (among others of those of the stress-axis) and behaviour (inhibitory avoidance, swimming patterns, shoaling behaviour, motor activity and startle habituation). TL has two mutations compared to AB giving it a different physical appearance: long fins *versus* short fins and spots *versus* stripes. These spots are caused by a mutation in *connexin 41.8* [6]. Yet, this connexin 41.8 gene is also present in other organs than skin, i.e. eye, heart and brain, and may thus potentially affect (many aspects of) behaviour.

In a previous study [1], we showed that larvae of AB strain habituated faster than larvae of the TL strain to a series of acoustic/vibrational stimuli. To assess whether the mutation in *connexion 41.8* has an effect on startle habituation, we compared startle behaviour of larvae of the AB and TL strain to larvae of the leopard strain, which also has spots due to a mutation in the connexin 41.8 gene, yet does not carry the long fin mutation. In addition, we compared startle habituation of larvae of the AB strain with those of the TU strain, which is also a wild-type (striped) zebrafish, and of which differences in mRNA expression profiles and behavioural have been reported [7,8].

Materials and Methods

Five day old larvae of the AB, TL, leopard (LEO) and TU strains (reared under 28.5 degrees Celsius and 14L/10D in Petridishes with E3 medium from day 0 onwards), were exposed to a protocol in Daniovision (Noldus Information Technology, Wageningen, the Netherlands) consisting of 10 min habituation, followed by 10 acoustic/vibrational stimuli with a 20 second inter-stimulus interval (ISI), a period of 10 minutes rest, followed by 30 stimuli with a 1 second ISI [1]. Habituation to repeated stimuli occurs normally more strongly in the second series than the first series. Experiments were done in 2-3 biological replicates to eliminate batch differences and statistical analysis was used to assess consistency of responses. The dependent variable was: maximum velocity following a stimulus (Vmax (mm/s)) in 1-second intervals. Subjects showing a response less than 15 mm/s on the first stimulus were discarded from further analysis [1; AB, TL and LEO: <12%; TU ~25%]. Statistical analysis was done using repeated measures Analysis of Variance (ANOVA) followed by post-hoc testing where appropriate using IBM SPSS version 23 for Windows (IBM, Armonk, NY, USA). P-values smaller than or equal to 0.05 (two-tailed) were considered significant.

Ethical statement: All experiments were carried out in accordance with the Dutch Animals Act (http://wetten.overheid.nl/BWBR0003081/2014-12-18), the European guidelines for animal experiments **(**Directive 2010/63/EU; http://eur-lex.europa.eu/legal-content/NL/TXT/HTML/?uri=CELEX:32010L0063 and institutional regulations. Larvae were euthanized by exposure to ice slurry for at least 20 minutes.

Results

Figure 1 (panels A-D) shows the maximum velocity (mm/s) across the time period when 10 stimuli with a 20s interval were presented. The profiles clearly show an increase in maximum velocity when a stimulus is presented, underlining its validity for measuring acoustic/vibrational startle responses. In addition the profiles show differences in baseline values as well as their variation between strains, with AB having a low baseline value with little variation as we have observed earlier [1].

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

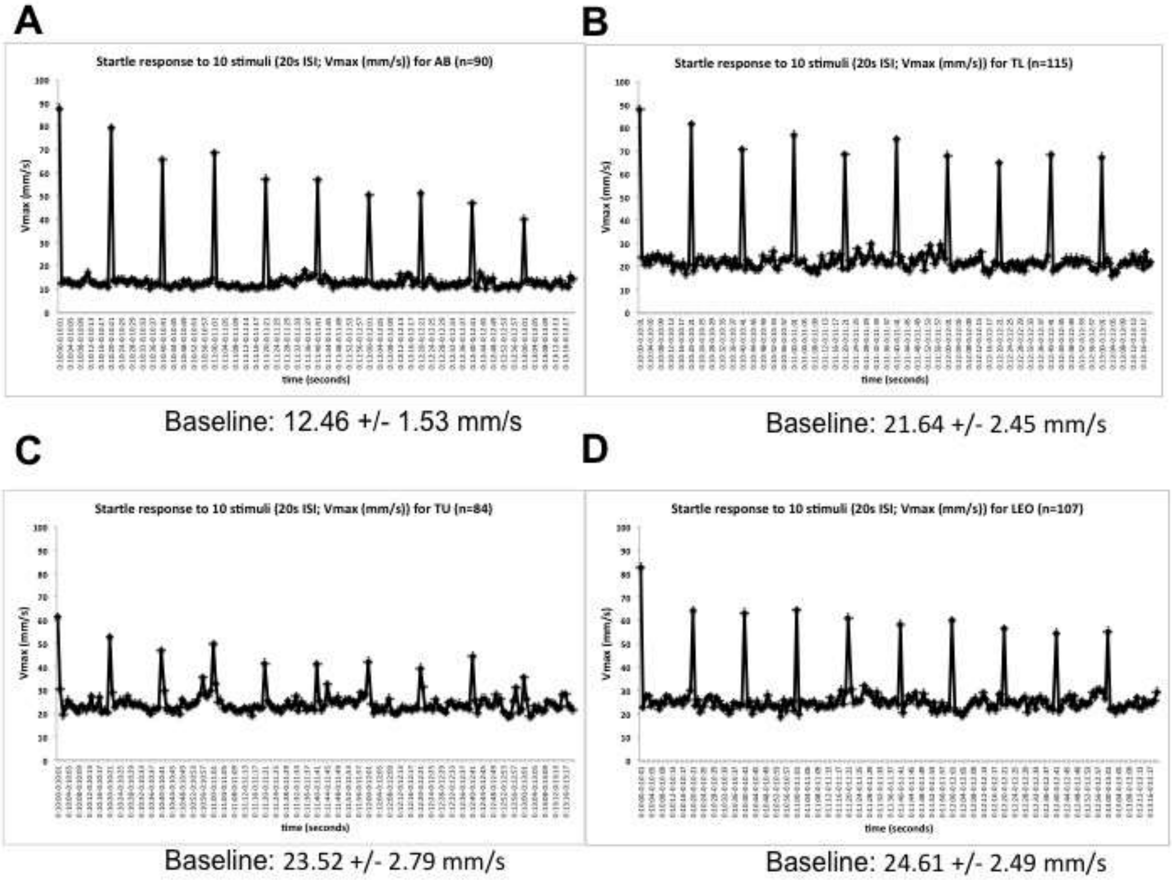


Figure1, Panels A-D: Startle response (means; maximum velocity (mm/s)) to 10 stimuli with 20s ISI of AB, TL, leopard (LEO) and TU 5dpf larvae. Values in panels indicate mean (+standard devation) of baseline (intervals between stimuli).

No differences were observed between larvae of the AB, TL and LEO strain in the startle response to the first stimulus (Figure 2, panels A and C). Like in our earlier experiments larvae of the AB strain showed a stronger habituation than larvae of the TL strain, both when exposed to 10 stimuli with 20s ISI (Figure 2, panel A) and 30 stimuli with 1s ISI (Figure 2, panel C). In both cases larvae of the leopard strain showed startle values in between those of AB and TL (Figure 2, panel A: strain * stimulus: $F(18,2781)=2.724$, $p<0.001$; panel C: $F(58,8642)=2.756$, $p<0.001$). Defining habituation more precisely as the ratio of the startle response to the 4th, 8th, 12th, 16th and 20th stimulus to the 1st stimulus revealed a highly significant strain effect ($F(2,299)=8.009$, $p<0.001$) with ratio’s from AB larvae differing from those of TL/LEO (Tukey HSD). Parcelling this out revealed that ratio’s of AB larvae differed from those of TL/LEO larvae at the 4th stimulus, while at the 8th stimulus and 12th stimulus ratio’s of AB/LEO larvae differed from those of LEO/TL larvae. No differences were found at the 16th and 20th stimulus.

Larvae of the TU strain showed lower startle values than larvae of the AB strain at the first stimulus (Figure 2, panels B and D). Also thereafter the startle response of larvae of the AB and TU strain differed, especially when considering the first series of stimuli (Figure 2, panel B: strain * stimulus: $F(9,1548)=2.893$, $p\leq0.002$; panel D: $F(29,4959)=2.399$, $p<0.001$). As to the habituation at the 4th, 8th, 12th, 16th and 20th stimulus, larvae of the TU larvae habituated slower than larvae of the AB strain (strain: $F(1,171)=8.585$, $p<0.004$; p-values <0.05 at 4th, 12th, 16th and 20th stimulus).

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

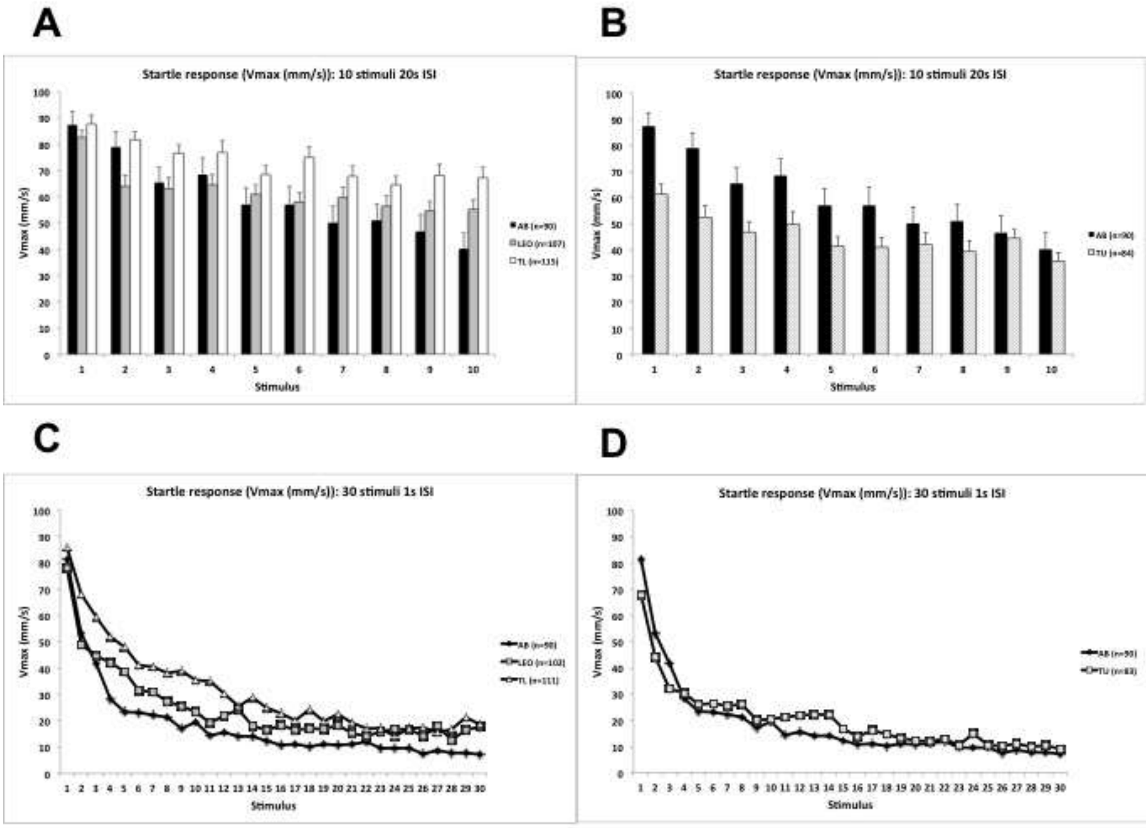


Figure2: Panel A: Startle response (mean+SEM) to 10 stimuli with 20s ISI of AB, TL and leopard (LEO) 5dpf larvae. Following a significant interaction term, one-way ANOVA's for different stimuli showed significant effects for stimulus 2 and stimuli 6-10; post hoc testing (Tukey HSD) showed that for stimulus 2: [AB=TL]=/=[LEO], for stimulus 6: [AB=LEO]=/=[TL], for stimulus 7-9: [AB=LEO]=/=[LEO=TL] and for stimulus 10: [AB]=/=[TL=LEO]. Panel B: Startle response (mean+SEM) to 10 stimuli with 20s ISI of AB and TU 5dpf larvae. Following a significant interaction term, Student t-tests showed significant differences for stimulus 1-6 and stimulus 8. Panel C: Startle response (means only for sake of clarity) to 30 stimuli with 1s ISI of AB, TL and leopard (LEO) 5dpf larvae. Following a significant interaction term, one-way ANOVA's for different stimuli showed significant effects for stimulus 2-21 and stimulus 23-30; post-hoc testing (Tukey HSD) showed that for stimulus 2,3,6-11,14,15: [AB=LEO]=/=[TL], for stimulus 4,5,13,16,17,19,23,26,27,29,30: [AB]=/=[LEO=TL], for stimulus 12,20,21,25,28: [AB=LEO]=/=[LEO=TL], for stimulus 18: [AB]=/=[LEO]=/=[TL] and for stimulus 24: [AB=TL]=/=[TL=LEO]. Panel D: Startle response (means only for sake of clarity) to 30 stimuli with 1s ISI of AB and TU 5dpf larvae. Following a significant interaction term, Student t-tests showed significant differences for stimulus 1,3,11,15,17 and 18.

Discussion

The data of this study show that differences in startle response habituation between AB and TL may be partly caused by the *connexin 41.8* mutation as larvae of the leopard strain showed values in between those of AB and TL. This suggests that this mutation is not 'an innocent and cosmetic' mutation but may affect many (behavioural) paradigms. Future studies should unravel its underlying mechanism and its consequences for using TL as strain in (behavioural) experiments. Currently we are testing whether differences in cortisol levels may also explain differences in startle behaviour. In addition, the data show that AB and TU show differences in startle behaviour, while both are wild-type fish. Recent studies [7,8] have also showed differences between AB and TU regarding behaviour and mRNA expression.

The data of this study emphasize that strain differences may affect (habitation to) startle responses and hence may have an effect on results (within and) between laboratories as well as on the choice for the optimal strain for (behavioural) research.

References

1. Bos, van den R., Mes, W., Galligani, P., Heil, A.H., Zethof, J., Flik, G., Gorissen, M. (2017). Further characterisation of differences between TL and AB zebrafish (Danio rerio): Gene expression, physiology and behaviour at day 5 of the larval stage. *PLoS ONE* **12(4)**, e0175420. https://doi.org/10.1371/journal.pone.0175420.

2. Bos van den R., Zethof, J., Flik, G., Gorissen, M. (2017). Light regimes differentially affect baseline transcript abundance of stress-axis and (neuro)development-related genes in zebrafish (Danio rerio, Hamilton 1822) AB and TL larvae. *Biology Open*, **6**, 1692-1697 doi:10.1242/bio.028969.

3. Gorissen, M., Manuel, R., Pelgrim, T. N. M., Mes, W., de Wolf, M. J. S., Zethof, J., Flik, G., van den Bos, R. (2015). Differences in inhibitory avoidance, cortisol and brain gene expression in TL and AB zebrafish. *Genes, Brain and Behavior*, **14**, 428-438.

4. Séguret A, Collignon, B., Halloy, J. (2016) Strains differences in the collective behaviour of zebrafish (Danio rerio) in heterogeneous environment. *Royal Society Open Science,* **3**, 160451. (doi: 10.1098/rsos.160451).

5. Gao, Y., Zhang, G., Jelfs, B., Carmer, R., Venkatraman, P., Ghadami, M., Brown. S.A., Pang, C.P., Leung, Y.F., Chan, R.H.M. *et al.* (2016) Computational classification of different wild-type zebrafish strains based on their variation in light-induced locomotor response. *Computers in Biology and Medicine,* **69,** 1–9. (doi:10.1016/j.compbiomed.2015.11.012).

6. Watanabe, M., Iwashita, M., Ishii, M., Kurachi, Y., Kawakami, A., Kondo, S., Okada, N. (2006). Spot pattern of leopard Danio is caused by mutation in the zebrafish connexin41.8 gene. *EMBO Reports*, **7**, 893-897.

7. Holden, L.A., Brown, K.H. (2018) Baseline mRNA expression differs widely between common laboratory strains of zebrafish. *Scientific Reports,* **8**, 4780 (DOI:10.1038/s41598-018-23129-4).

8. Vignet, C., Bégout, M-L., Péan, S., Lyphout, L., Leguay, D., Cousin, X. (2013) Systematic screening of behavioral responses in two zebrafish strains. *Zebrafish,* **10,** 365–375. (doi:10.1089/zeb.2013.0871).

Acknowledgements

We would like to thank Tom Spanings for excellent fish care and Kari Bosch for assistance with the experiments.

Fast Development of Customized Rodent Behavior Recognition with Semi-Automated Labeling

Elsbeth A. van Dam[1], Malte Lorbach[1], Ruud A.J. Tegelenbosch[1], Lucas P.J.J. Noldus[1]

[1]**Noldus Information Technology, Wageningen, The Netherlands**

Observation, measurement and analysis of rodent behavior is an essential ingredient of behavioral neuroscience studies and objective quantitative behavioral data are a prerequisite for reliable research results. Because manual recording of behavior by human observers is time-consuming, tedious and subjective, there has been an ongoing effort to develop techniques for automated measurement. Since the 1990s, we have seen automated behavioral observation tools evolve from simple video tracking systems able to record the movement of a laboratory rat in a test arena, with behavioral classification limited to "moving" vs. "not moving" [1][2], to software able to recognize ten different categories of behavior in a rodent from overhead video [3]. However, in experiments that deviate from the pre-defined research setups or in situations where behaviors are performed differently due to drug treatment, genetic variance or disease progress (such as excessive grooming or stereotypic behavior), researchers are left with the laborious and error-prone method of manually scoring their data.

One solution for this has been introduced with the Janelia Automatic Animal Behavior Annotator (JAABA) [4]. With this interactive labeling tool the user presents a small set of hand labeled video segments to the system, which uses Active Learning to incrementally train a classifier. Subsequently, the classification results are shown back to the user who can then correct the labels and retrain the system. The usefulness of this tool was shown by Van den Boom *et al*. [5], who reached an accuracy of 94% (F-score) for the labeling of fast excessive self-grooming in SAPAP3 knock-out mice, that was not recognized by the generic automated recognition system. The classifier needed ~12 minutes of annotated video to reach reliable performance and it took the authors less than a few days to set up a trained system.

To improve on this concept further, Lorbach *et al*. [6] recently proposed a Clustering-based Active Learning framework for labeling social rodent behavior. They make the interactive labeling more efficient by querying the user to label specific automatically selected segments of the video that the system can learn from most. In [6] the segments are selected based on a combination of a data-driven Active Learning approach [7] and the properties of the Dirichlet Process that facilitate the discovery of rare classes [8]. Another way of selecting segments is to use a discriminative approach described in [9]. Here the criteria for the usefulness are based on the uncertainty of the classifier. The participants of a user study had to label only 300 1-s video clips selected from 9 videos of 15 min each to achieve the same accuracy as with training with the fully labeled dataset. With this interactive approach the manual labeling time was reduced from ~7 hours for the entire dataset to 18.7 minutes, which illustrates the potential of active learning for reducing manual labeling effort.

In this research we are extending the querying framework from Lorbach by experimenting with more informative feature sets and targeted initialization of the underlying classifiers. The above-mentioned frameworks classify behavior only on tracked locations and derived features, such as relative body points per frame and sliding window statistics, and require those trajectory features to be calculated prior to the learning stage. The feature set used by Van Dam et al. (2013) for the generic recognition of specific rodent behaviors adds body shape and optical flow features to the trajectory features, yielding a richer dataset with better classification capabilities. Furthermore, we experiment with initializing the active learner with pre-trained, generic behavior classifiers. Finally, we present the results in a user-friendly prototype with integrated feature acquisition.

For this research we use the rodent behavior datasets described in [3] and [9]. These datasets were recorded with approval of authorized ethical committees. For detailed descriptions we refer to [3] and [9].

References

[1] Spruijt, BM, Hol, T, Rousseau, J (1992). Approach, avoidance, and contact behavior of individually recognized animals automatically quantified with an imaging technique. *Physiol Behav* 51(4):747-752.

[2] Noldus, LPJJ, Spink, AJ, Tegelenbosch, RAJ (2001). EthoVision: A versatile video tracking system for automation of behavioral experiments. *Behav Res Methods, Instruments, Comput*. 33(3):398-414.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

[3] Van Dam, EA, Van der Harst, JE, Ter Braak, CJF, Tegelenbosch, RAJ, Spruijt, BM, Noldus, LPJJ (2013). An automated system for the recognition of various specific rat behaviours. *J Neurosci Methods* 218(2):214-224.
[4] Kabra, M, Robie, AA, Rivera-Alba, M, Branson, S, Branson, K (2012). JAABA: interactive machine learning for automatic annotation of animal behavior, *Nature Methods* 10(1):64–67.
[5] Van den Boom, BJG, Pavlidi, P, Wolf, CJH, Mooij, AH, Willuhn, I (2017). Automated classification of self-grooming in mice using open-source software. *J Neurosci Methods* 289:48-56.
[6] Lorbach, MT, Poppe RW, Van Dam, EA, Noldus, LPJJ, Veltkamp, RC (2016). Clustering-based Active Learning in Unbalanced Rodent Behavior Data. *Proc of the Int Workshop on Visual observation and analysis of Vertebrate And Insect Behavior (VAIB)*.
[7] Dasgupta, S, Hsu, D (2008). Hierarchical sampling for active learning. *Proc. Conf. Machine Learning (ICML)*: 208–215.
[8] Haines, TSF, Xiang, T (2013), Active Rare Class Discovery and Classification Using Dirichlet Processes. *International Journal of Computer Vision*, 106(3):315–331
[9] Lorbach, M (2017). Automated Recognition of Rodent Social Behavior. *Ph.D. dissertation*, Utrecht University, Utrecht, The Netherlands.

Do Freestalls Restrict Dairy Cows In Their Lying Behavior? Comparing postures of dairy cows in cubicles and on pasture.

E. van Erp-van der Kooij[1], O. Almalik[1], D. Cavestany[2] and F.J.C.M. van Eerdenburg[3]

1Department of Applied Biology, HAS University of Applied Sciences, Den Bosch, the Netherlands. L.verp@has.nl; 2Faculty of Animal Science and Veterinary Medicine, University of the Republic, Montevideo, Uruguay. Daniel.cavestany@gmail.com; 3Department of Farm Animal Health, Faculty of Veterinary Medicine, Utrecht University, Utrecht, the Netherlands. f.j.c.m.vaneerdenburg@uu.nl.

Exhibiting natural lying behaviour is an important indicator for the health and welfare of dairy cows [1-5]. Cubicles for dairy cows are often too narrow and too short, causing difficulties lying down or standing up. In too short or narrow cubicles, cows cannot display all their natural lying postures. Bedding material and divider type also influence and possibly hinder lying behaviour [6,7]. Furthermore, in a cubicle barn cows cannot choose in which direction they lie, which might be important because cows seem to prefer a North-South position [8]. In this study we compared lying postures of 574 dairy cows on pasture in Uruguay and the Netherlands and 506 cows in cubicle barns in the Netherlands. Lying postures were recorded by taking photographs of cows lying down using ordinary handheld cameras and determining lying position using the illustrations, see Figure 1. Cows were photographed while lying down in the cubicles or in the pasture, approaching them carefully and without handling or disturbing them. Therefore, no ethical review of the study was necessary. Students were trained by the authors to compare the photographs with the pictures, and the first 100 photographs were double-checked by the authors. Indoors we also recorded cubicle bedding and divider type, while outdoors in Uruguay orientation of the lying cows was recorded as well. Our aim was to determine whether lying postures differed between cows in cubicles and on pasture. If cubicles restrict cows in their lying posture, this could suggest a decreased welfare.

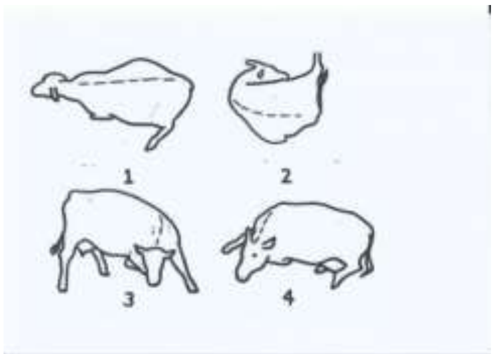


Figure 1. Lying postures of dairy cows: long (1), short (2), wide (3) and narrow (4).

Lying postures were determined by studying photographs of the cows and comparing results. Photographing each cow and comparing the pictures required much handiwork, automation of this process would be welcome. Using automatic cameras and vision techniques such as image recognition could reduce the amount of labour and standardize results. Our results showed significantly more long and wide postures for cows on pasture and more short and narrow postures in cubicles. 55,5% of cows on pasture prefer lying in a North-South direction. Cows in cubicles showed significantly more long and wide postures and less short and narrow postures on soft bedding compared to hard bedding ($P<0,01$). Cows in cubicles with English type dividers showed more long and wide postures than cows in cubicles with R-shaped or U-shaped dividers. We conclude that cubicles and dividers restrict cows in their lying behavior, since lying postures differed significantly from that on pasture, implying a decreased welfare. Certain types of bedding and dividers restricted cows less than others, so housing systems can be improved. For cattle welfare, we advise to stimulate housing systems where cows can lie unrestricted and choose their lying position as well as their orientation. In future research, automatic recording of lying behaviour and lying posture of cows, using vision techniques, would improve efficiency as well as accuracy of results in this field of research.

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

References:
1. Cook, N.B., T.B. Bennett, and K. V Nordlund (2004). Effect of Free Stall Surface on Daily Activity Patterns in Dairy Cows with Relevance to Lameness Prevalence. *Journal of Dairy Science* **87**: 2912–2922.
2. Ito, K., M.A.G. von Keyserlingk, S.J. Leblanc, and D.M. Weary (2010). Lying behavior as an indicator of lameness in dairy cows. *Journal of Dairy Science* **93**: 3553–60.
3. Kilgour, R.J. (2012). In pursuit of "normal": A review of the behaviour of cattle at pasture. *Applied Animal Behaviour Science* **138**: 1–11.
4. Munksgaard, L., and H.B. Simonsen (1996). Behavioral and Pituitary Adrenal-Axis Responses of Dairy Cows to Social Isolation and Deprivation of Lying Down. *Journal of Animal Science* **74**: 769–778.
5. Sepúlveda-Varas, P., D.M. Weary, and M. a G. von Keyserlingk (2014). Lying behavior and postpartum health status in grazing dairy cows. *Journal of Dairy Science* **97**: 1–10.
6. Fregonesi, J. a, D.M. Veira, M. a G. von Keyserlingk, and D.M. Weary (2007). Effects of bedding quality on lying behavior of dairy cows. *Journal of Dairy Science* **90**: 5468–72.
7. Van Gastelen, S., B. Westerlaan, D.J. Houwers, and F.J.C.M. van Eerdenburg (2011). A study on cow comfort and risk for lameness and mastitis in relation to different types of bedding materials. *Journal of Dairy Science* **94**: 4878–88.
8. Begall, S., J. Cerveny, J. Neef, O. Vojtech, and H. Burda (2008). Magnetic alignment in grazing and resting cattle and deer. *Proceedings of the National Academy of Science* **105**: 13451–13455.