Modelling of animal welfare

The development of a decision support system to assess
the welfare status of pregnant sows

Marc B. M. Bracke

Promotoren: Prof.dr. B.M. Spruijt
Hoogleraar in de Ethologie en Welzijn Dieren

Prof.dr.ir. J.H.M. Metz
Hoogleraar Technisch Ontwerp van Bedrijfssystemen in de Dierhouderij met bijzondere aandacht voor Milieu- en Welzijnsaspecten

Co-promotor: dr. W.G.P. Schouten
Universitair Docent, leerstoelgroep ethologie

Samenstelling promotiecommissie: Prof. dr. D.M. Broom (University of Cambridgy, UK), Prof.dr. J.M. Koolhaas (Rijksuniversiteit Groningen), Prof.dr.ir. A.J. van der Zijpp (Wageningen Universiteit) en Prof.dr.ir. R.B.M. Huirne (Wageningen Universiteit).

Modelling of Animal Welfare

The development of a decision support system
to assess the welfare status of pregnant sows

Marc B. M. Bracke

Proefschrift

Ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit
Prof.dr.ir. L. Speelman
in het openbaar te verdedigen
op dinsdag 17 april 2001
'S middags om vier uur in de Aula.

1613265

For ISBN number, etc., see p 150.

Abstract

Bracke, M.B.M., 2001. Modelling of animal welfare: The development of a decision support system to assess the welfare status of pregnant sows (Modelleren van dierenwelzijn: De ontwikkeling van een beslissingsondersteunend systeem om de welzijnsstatus van dragende zeugen in te schatten).

Keywords: farm animal welfare assessment, pigs, applied ethology, housing systems, model, computer, knowledge base, expert system.

A computer-based decision support system for welfare assessment in pregnant sows was constructed. This system uses a description of a husbandry system as input and produces a welfare score on a scale from 0 to 10 as output. Pregnant sows were chosen as a case in search for a formalised, i.e. structured, transparent, yet flexible procedure to 'objectively' assess the overall welfare status of farm animals in relation to the housing and management system based on available (and undisputed) scientific knowledge. The procedure to construct the welfare model and to calculate welfare scores is described. Decision making is based on the needs and distress of the animal thus from the perspective of the animal.

The sow welfare (SOWEL) model was validated using expert opinion in that there is a substantial agreement between pig welfare scientists and the model about the ranking of housing systems and, to a lesser degree, about the weighting of attributes of housing systems. The most important welfare-relevant attributes concern aspects of social contact, space, and substrate. The housing systems were roughly divided into low-, mid-, and high-welfare systems. Low-welfare systems were conventional housing in individual stalls and tethers, while high-welfare systems in our data set all provided substrate and outdoor access.

For practical applications further development of the decision support system is recommended, as well as ongoing validation, upgrading and extending of the model, e.g. to other species. The results show that integrated welfare assessment based on available scientific knowledge is possible.

PhD thesis, Institute of Agricultural and Environmental Engineering (IMAG). Livestock Production and Environmental Engineering Department, P.O. Box 43, 6700 AA Wageningen, The Netherlands.

Voorwoord (Preface)

Eind 1995 had ik besloten om, na een lange en afwisselende studie en een paar jaar in de praktijk als dierenarts, een baan te zoeken in het onderzoek. Een oude studiegenoot van de middelbare school, Eddie de Mayer, vertelde me dat er op het IMAG nog een vacature was, iets met dierenwelzijn en computers. 'Dat kan toch helemaal niet', dacht ik eerst, maar dat duurde niet lang. Al gauw had ik me enthousiast op m'n nieuwe uitdaging gestort.

Onder leiding van Alexander Udink ten Cate heb ik misschien wel duizend welzijnsbomen geplant. Daarvan is er uiteindelijk één overgebleven. En, omdat hoge bomen veel wind vangen, heb ik er maar een struik van gemaakt, met een welzijnsstam, behoeftetakken en attribuutsbladeren.

Toen begon ik deskundigen te vragen welzijnscijfers te geven. Dat lijkt misschien kunstmatig, maar het sluit vrijwel naadloos aan bij wat mensen van nature gewend zijn om te doen. Wij taxeren de welzijnsstatus van andere mensen en dieren, vaak via non-verbaal gedrag, zo automatisch, dat het soms bijna lijkt alsof we allemaal deskundigen zijn op het welzijnsgebied, en misschien is dat ook wel zo. Als je eenmaal een paar punten van de schaal hebt benoemd, is het ook verbazingwekkend gemakkelijk om bijvoorbeeld jezelf een welzijnscijfer te geven. Het proefschrift gaat over het zoeken naar een manier om cijfers voor dierenwelzijn toe te kennen op basis van beschikbare wetenschappelijke kennis. Wie daarin geïnteresseerd is, kan ik naar de diverse hoofdstukken van dit proefschrift verwijzen.

Ik wil iedereen bedanken die aan dit proefschrift heeft bijgedragen. In het bijzonder, mijn ouders, omdat ze me alle gelegenheid gaven om te studeren. Marjolein, omdat ze het heeft kunnen uithouden met haar promovendus die geregeld vroeg om een welzijnscijfer te geven op een schaal van 0 tot 10. Thomas en Marleen, omdat ze me, ook als welzijnsonderzoeker, een verrijkte omgeving bieden. Jos Metz vanwege zijn tactvolle management. Berry's brein, voor het scherp houden van het mijne. Alexander Udink ten Cate, Aalt Dijkhuizen, Herman Wierenga, en Willem Schouten voor hun support in de verschillende stadia van het project. De experts voor hun stimulerende gesprekken, de collega's op het IMAG, de studenten, de financiers, de leden van de Gebruikercommissie en alle overigen met wie ik onderdelen van het proefschrift heb besproken. Derhalve dank ik Adroaldo, André, Arianne, Beat, Beate, Bernard, Berry, Bert, Bo, Carolien, Chris, Claudia, Dan, David, Dinand, Dolf, Don, Doris, Dries, Durk, Ebby, Erica, Frank, Frans, Gerdien, Gert-Jan, Gerrit, Hans, Harold, Harry, Hay, Hein, Helmut, Herman, Huub, Ietje, Ilan, Ingrid, Ingvar, Iris, Jan, Jouke, Jeff, Jeremy, Jeroen, Joanna, Johan, Joop, Jos, Josef, Judith, Karel, Karin, Katelijn, Kees, Lenny, Marco, Marek, Marie-christine, Margriet, Marina, Martin, Menno, Michel, Mike, Mirjam, Moira, Nico, Nicoline, Paul, Per, Pernille, Peter, Pierre, Piet, Pieter, René, Rob, Ron, Rudi, Ruud, Sabine, Sandra, Silke, Teun, Ute, Victor, Wil, Willem, Wim, Xavier, en iedereen die ik verder nog vergeten ben.

NN08201, 2964

Stellingen

1. De welzijnsstatus van dieren kan in een getal worden uitgedrukt *(dit proefschrift).*

2. Welzijn laat zich niet adequaat voorschrijven met een lijst van minimumeisen *(dit proefschrift).*

3. Wanneer een objectieve inschatting van dierenwelzijn inhoudt dat wetenschappers dat zonder tussenkomst van hun menselijk gezichtspunt zou moeten doen, dan is een objectieve inschatting van welzijn niet mogelijk. In dat geval kunnen wetenschappers echter ook niet objectief vaststellen wat de kleur is van een appel *(ontleend aan Bekoff, M., L. Gruen, S.E. Townsend & B.E. Rollin, 1992. Animals in science: Some areas revisited. Animal Behaviour 44: 473-484).*

4. Een wetenschappelijke inschatting van de welzijnsstatus van dieren is een noodzakelijke, maar geen voldoende voorwaarde om te bepalen welk veehouderijsysteem moreel acceptabel is *(dit proefschrift; stelling deels ontleend aan David Hume, 1711-1776).*

5. Bij verdere ontwikkeling zal het computeralgoritme van het beslissingsondersteunend systeem uiteindelijk de hersenprocessen van dieren representeren waarmee zij hun eigen welzijn inschatten. Misschien dat we dan, tot onze schrik, zelfs moeten concluderen dat zo'n computersysteem zelf 'bewustzijn' heeft gekregen *(dit proefschrift, Berry Spruijt).*

6. Gezamelijke onderwijselementen voor studenten zootechniek en diergeneeskunde zullen een positieve bijdrage hebben op hun latere functioneren in de dierlijke productieketen *(ontleend aan ideeën van Hans de Vries en Pim Brascamp).*

7. Dierenartsen zouden een belangrijke rol kunnen spelen bij het monitoren van de dierenwelzijn op bedrijven, mits ze erin slagen om hun (patho-)fysiologische denkkader aanzienlijk te verbreden, en mits ze belangenverstrengeling weten te voorkomen.

8. Verbetering van het welzijn van landbouwhuisdieren is ook in het belang van de boer.

9. Oorlogen en rampen dragen bij aan het versneld verbeteren van de wereld *(vrij naar Heraklitus, ongeveer 530-470 voor Christus).*

Stellingen behorende bij het proefschrift **'Modelling of animal welfare: The development of a decision support system to assess the welfare status of pregnant sows.'**

Wageningen, 17 april 2001

Marc Bracke

Voor pa, ma, Marjolein, Thomas en Marleen

Contents

Chapter 1. General introduction

'Every British citizen in a lifetime of 70 years consumes 550 poultry, 36 pigs, 36 sheep and 8 oxen plus 10.000 eggs and dairy products equivalent to 18 tonnes of milk ' (Webster, 1995, p. 127). From these data each citizen can be estimated to be responsible for a total farm-animal lifespan equivalent to 2.5 times his/her own, most of which concerns life under intensive farming conditions, which has been criticised on welfare grounds. Society has called upon science to assess the animal's welfare status in different housing systems, and to find solutions for welfare problems. In fact, IMAG in Wageningen has a history of conducting research that involves the evaluation and design of housing systems for farm animals. Unfortunately, welfare is a complex problem (Dawkins, 1997). When one problem is solved, often other problems appear. For example, when individually-housed sows are kept in groups, problems with aggression between sows arise, which, in turn, may be remedied with a whole range of measures, from the provision of early social experiences (van Putten & Buré, 1997) to integrated group housing (Houwers *et al.*, 1993, 1996). What is required from science is not only to solve welfare problems, but also to provide a reasonably objective tool for integrated welfare assessment, which can support decision making by consumers, policy makers, farmers, extension and designers to evaluate existing and new (or improved) housing and management systems (Verkaik, 1975; Wierenga & Jongebreur, 1987; Ganzevoort, 1994; den Ouden, 1996).

Scientists seem to agree that animals have feelings that have a biological basis in the animal's various life functions or needs (Broom, 1998; Spruijt, in press). The welfare status can vary over a continuum from very good to very poor (Broom, 1996), and involves mechanisms to weigh the various positive and negative aspects of a given situation (Cabanac, 1971; McFarland, 1989; Spruijt *et al.*, in press).

A problem is that discussions about welfare, especially in society but also among scientists often concern many different examples, while a comprehensive framework seems to be lacking. Scientists dispute whether welfare should be defined in terms of feelings, or more directly in terms of biological functioning (Broom, 1996; Duncan, 1996; Anon., in press). Furthermore, they disagree which paradigm is best to measure welfare, e.g. whether this is the study of natural behaviour, consumer demand, measures of the HPA (hypothalamic-pituitary-adrenocortical) axis or (pre)pathological states. Although there are some causal models with relevance for welfare (e.g. Lorenz, 1978; Wiepkema, 1987; Hughes & Duncan, 1988; McFarland, 1989; Toates & Jensen, 1991), there is at present no comprehensive welfare theory with which verifiable predictions about welfare can be made (Haynes, 2000), nor is a single empirical measure for overall welfare presently available.

Nevertheless, over more than 30 years scientists have measured many aspects of biological functioning in a wide range of animals and housing conditions. They have measured aspects of behaviour, physiology, production and health. This has generated many findings that are evidently relevant for welfare. It appears that all housing systems have both positive and negative aspects (Fraser, 1995), and that a combination of measures is needed to assess welfare (Broom & Johnson, 1993). However, scientists presently do not know how to add these measures into an overall welfare judgement (Fraser, 1995).

Some models for overall welfare assessment have been described (e.g. Sundrum *et al.*, 1994) or are being developed (Johannessen *et al.*, 2000; Capdeville & Veissier, in press), but to date these don't have an explicit scientific basis, i.e. most of the considerations underlying these models remain implicit. Scientists have also tried formulating cut-off points (e.g. Barnett & Hemsworth, 1990; Wiepkema & Koolhaas, 1993). Cut-off points, however, have been severely criticised (Mendl, 1991). In particular, cut-off points cannot allow compensation between positive and negative aspects, which is a characteristic of welfare. Some authors have even concluded that scientists should not try to assess welfare overall (Fraser, 1995; Rushen & de Passillé, 1992).

This is to turn down an important request from society. As a result scientists write increasingly long welfare reports containing accumulations of scientific findings (e.g. Scientific Veterinary Committee, 1997), from which it is difficult to draw practical conclusions.

While everybody has his own personal opinions about welfare, the problem remains how to assess the overall welfare status of animals on a scientific basis. That is, if we had read all scientific papers, and if we furthermore knew all measurable facts about a housing system, we still would not know how to integrate that knowledge into one statement about the overall welfare status of the animals in that system.

This illustrates the topic of this thesis. More precisely, the aim is to find a formalised, i.e. structured, transparent, yet flexible way to 'objectively' assess the overall welfare status of farm animals based on available scientific knowledge.

In particular, we want to show how this can actually be done for one group of farm animals, for which we chose the case of pregnant sows, and for which we chose to assess welfare in relation to the housing system. The primary focus, therefore, is not to distinguish between individual farms that have only detailed differences in housing and management. The primary focus is to distinguish between different housing and management systems under stabilised conditions, i.e. types of farm that are used or are intended to be used for agricultural production over longer periods of time.

Furthermore, the aim is to assess welfare in a way that allows a quantitative expression (e.g. a score between 0 and 10) of what matters to the animals themselves, from their point of view. For this we presume that animals have a welfare status and consciousness, and we presume that scientific knowledge about the different biological functions is relevant to assess welfare. Thus, the question is not *whether* animals have welfare, but *how* to assess it (cf. Crook, 1983). This point is nicely illustrated in Gary Larson's cartoon 'Wildlife Management Finals' where a student is asked 'How much wood would a wood-chuck chuck, if a wood-chuck could chuck wood?' Similarly, our question is 'What is the welfare of pregnant sows in different housing systems assuming they have welfare?'

The welfare status is to be assessed overall, i.e. taking into account all positive and negative aspects (of a life) that matter to the sow. This includes all the animal's biological needs with all the states of need satisfaction and need frustration, rather than only a subset such as only the behavioural or only the (negative) stress-physiological components of welfare.

Even though welfare concerns what matters to the animals, we aim to answer a factual question, i.e. what is the welfare status in a given environment de facto, as a matter of fact. We are not trying to answer subsequent moral or political questions, such as what a given welfare status means for the political or moral acceptability of the housing system.

In search for a scientific basis we have been consensus-oriented, i.e. using known scientific facts, preferably published in peer-reviewed papers, that are undisputed (but not indisputable), and using 'all' available knowledge (Dawkins, 1997; Rushen, 1991), i.e. all apparently relevant findings from empirical research. Working with 'available knowledge' implies that our primary concern is not how to actually measure welfare empirically, nor how to interpret raw scientific data into a conclusion of the kind 'we found that...'. Our research concerns the next phase '... and these findings mean for welfare that ...'. This interpretation phase takes several reasoning steps from scientific findings and premises that derive from some scientific conceptual framework for welfare assessment, all the way to a single overall welfare score for a certain housing system. Therefore, it does not concern finding scientific proof, neither empirical, nor deductive, but it concerns dealing with uncertain information and finding the best possible assessment based on what is presently known.

As far as we know we are the first to try to formalise the reasoning process involved in overall welfare assessment. Formalisation requires that all steps are made explicit, and in a structured way so as to increase transparency, and to make hidden assumptions explicit (Tannenbaum, 1991; Sandøe & Simonsen, 1992; Fraser *et al.*, 1997), without getting stuck in deep philosophical issues about morality or about animal consciousness.

Since critical responses are to be expected, our further aim is to formalise overall welfare assessment in a flexible way that can adapt to new insights about welfare, to newly generated scientific findings, and to specific requirements resulting from different usages.

The methods used to reach these aims are to use techniques from information technology in order to be able to handle larger amounts of data, and to use interviews with experts to fill in gaps in knowledge.

The main aim of this thesis is to propose a formalised procedure to assess the overall welfare of farm animals based on available scientific knowledge. A decision support system for welfare assessment of pregnant sows in relation to their housing and management system is described and validated by expert opinion.

Overview of the thesis

This thesis contains nine chapters[1]. Chapter 2 introduces the strategy to develop a computer-based decision support system for welfare assessment in the case of pregnant sows. The strategy is the so-called Evolutionary Prototyping Method, which involves repeated upgrading of an initial prototype so as to end up with a 'final' version that has 'evolved' to be capable of adapting further when required. The initial prototype is described illustrating how overall welfare assessment can be performed explicitly on a scientific basis using general assumptions, such as to assess welfare based on needs.

The chapters 3 to 5 review those theoretical aspects of welfare that were found to be most relevant as indicated by the development of the prototype.

The 3rd chapter deals with a major objection, which is the view that overall welfare assessment is not possible, because it involves value judgements. We explain why we are more optimistic about its feasibility.

The 4th chapter reviews literature containing assessment tables and schemes (welfare models) in order to formulate recommendations for formalised overall welfare assessment.

1 The text of the chapters is formatted according to the *Netherlands Journal of Agricultural Science*, except for the chapters 2, 7 and 8, which are formatted for the journal to which they have been submitted.

The 5th chapter discusses the biological basis for welfare assessment. Special attention is given to the idea to assess welfare based on needs as was used in the prototype. A review of the literature is supplemented with interviews with experts from various welfare-related disciplines to define welfare and to formulate a list of needs that could be used for actual welfare assessment.

The 6th chapter presents data from interviews with experts on pig welfare. They were asked to identify the main housing systems of pregnant sows, to give an overall welfare score to each system, and to explain the scores in relation to the attributes of the housing systems.

This information was needed for further development of the decision support system, because several versions that followed the construction of the prototype produced seemingly counterintuitive results (see Bracke *et al.*, 2000; Bracke *et al.*, in press). This indicated the need for further exploration of the feasibility of modelling the experts' reasoning process as well as the need for a frame of reference, in the form of welfare scores of the main housing systems in the model's domain, that could be used for preliminary validation while developing the model.

The 7th chapter describes the latest (i.e. 'final') version of the decision support system, including the welfare model and the formalised procedure that was used to construct the model on the basis of scientific statements and a list of needs.

The 8th chapter deals with the validation of the model using expert opinion. We compared the model's predictions of overall welfare scores for 15 housing systems and weighting factors for 20 welfare-relevant attributes with expert opinion solicited from welfare scientists using a written questionnaire.

The final chapter discusses some major points of criticism and practical implications of this work.

References

Anonymous, 2001. Scientists' assessment of the impact of housing and management on animal welfare. *Journal of Applied Animal Welfare Science*: In press.

Barnett, J.L. & P.H. Hemsworth, 1990. The validity of physiological and behavioural measures of animal welfare. *Applied Animal Behaviour Science* 25: 177-187.

Bracke, M.B.M., J.H.M. Metz & B.M. Spruijt, 2000. Welfare model to assess housing systems for pregnant sows. In: M.J.M Tielen & M.Th. Voets (Eds), Proceedings of the 10th International Congress on Animal Hygiene, Volume 1. Animal Health Service, Boxtel, pp. 377-380.

Bracke, M.B.M., J.H.M Metz & B.M. Spruijt, 2001. Development of a decision support system to assess farm animal welfare. Proceedings of the International Workshop on Assessment of Animal Welfare at Farm or Group Level, 27-28 August in Copenhagen. *Acta Agriculturae Scandinavica Section A, Animal Science, Supplement*: In press.

Broom, D.M., 1996. Animal welfare defined in terms of attempts to cope with the environment. *Acta Agriculturae Scandinavica Section A, Animal Science, Supplement* 27: 22-28.

Broom, D.M., 1998. Welfare, stress, and the evolution of feelings. *Advances in the Study of Behavior* 27: 371-403.

Broom, D.M. & K.G. Johnson, 1993. Stress and Animal Welfare. Chapman & Hall, London, 211 pp.

Cabanac, M., 1971. Physiological role of pleasure. *Science* 173: 1103-1107.

Capdeville, J. & I. Veissier, 2001. A method for assessing dairy cows welfare in a loose housing herd focussing on animal observations. Proceedings of the International Workshop on Assessment of Animal Welfare at Farm or Group Level, 27-28 August in Copenhagen. *Acta Agriculturae Scandinavica Section A, Animal Science, Supplement*: In press.

Crook, J.H., 1983. On attributing consciousness to animals. *Nature* 305: 11-14.

Dawkins, M.S., 1997. D. M. G. Wood-Gush Memorial Lecture: Why there has not been more progress in animal welfare research? *Applied Animal Behaviour Science* 53: 59-73.

Duncan, I.J.H., 1996. Animal welfare defined in terms of feelings. *Acta Agricultura Scandinavica Section A, Animal Science, Supplement* 27: 29-35.

Fraser, D., 1995. Science, values and animal welfare: Exploring the 'inextricable connection'. *Animal Welfare* 4: 103-117.

Fraser, D., D.M. Weary, E.A. Pajor & B.N. Milligan, 1997. A scientific conception of animal welfare that reflects ethical concerns. *Animal Welfare* 6: 187-205.

Ganzevoort, E., 1994. Dierenwelzijn moet meetbaar worden. *Platform* 8: 3-5.

Haynes, R.P., 2000. Do regulators of animal welfare need to develop a theory of psychological well-being? In: P. Robinson (Ed), Two Systems – One world. EurSafe 2000 Preprints. 2nd Congress of the European Society for Agricultural and Food Ethics, 24-26 August 2000. Centre for Bioethics and Risk Assessment, Copenhagen, pp. 135-138.

Houwers, H.W.J., R.G. Buré & J. Walvoort, 1993. Production aspects of integrated housing of sows with confined litters. *Animal Production* 56: 477.

Houwers, H.W.J., R.G. Buré & C. Lokhorst, 1996. Feeding method, management, housing equipment and sow behaviour for integrated group housing of sows. Report 96-04. IMAG-DLO, Wageningen, 102 pp.

Hughes, B.O. & I.J.H. Duncan, 1988. The notion of ethological 'need', models of motivation and animal welfare. *Animal Behaviour* 36: 1696-1707.

Johannesson, T., L. Alban & P.F. Johnsen, 2000. Weighting together measurements of farm animal welfare: A challenge. In: J.A.M. van Arendonck (Ed), Book of Abstracts of the 51st Annual Meeting of the European Association of Animal Production, Wageningen pers, Wageningen, p. 177.

Lorenz, K.Z., 1978. Vergleichende Verhaltensforschung: Grundlagen der Ethologie. Springer, Wien, 307 pp.

McFarland, D., 1989. Problems in Animal Behaviour. Longman Scientific & Technical, Harlow, 158 pp.

Mendl, M., 1991. Some problems with the concept of a cut-off point for determining when an animal's welfare is at risk. *Applied Animal Behaviour Science* 31: 139-146.

Ouden, M. den, 1996. Economic modelling of pork production-marketing chains, PhD. Wageningen Agricultural University, Wageningen, 168 pp.

Putten, G. van & R.G. Buré, 1997. Preparing gilts for group housing by increasing their social skills. *Applied Animal Behaviour Science* 54: 173-183.

Rushen, J., 1991. Problems associated with the interpretation of physiological data in the assessment of animal welfare. *Applied Animal Behaviour Science* 28: 381-386.

Rushen, J. & A.M.B. de Passillé, 1992. The scientific assessment of the impact of housing on animal welfare: A critical review. *Canadian Journal of Animal Science* 72: 721-743.

Sandøe, P. & H.B. Simonsen, 1992. Assessing animal welfare: Where does science end and philosophy begin? *Animal Welfare* 1: 257-267.

Scientific Veterinary Committee, 1997. The welfare of intensively kept pigs. Report to the Directorate General XXIV of the European Commission. Adopted 30. September 1997. Doc. XXIV/ScVc/0005/97, Scientific Veterinary Committee, Animal Welfare Section, Brussels, 199 pp.

Spruijt, B.M, 2001. How the hierarchical organization of the brain and increasing cognitive abilities may result in consciousness. *Animal Welfare*: In press.

Spruijt, B.M., R. van den Bos & F.T.A. Pijlman, 2001. A concept of welfare based on reward evaluating mechanisms in the brain: Anticipatory behaviour as an indicator for the state of reward systems. *Applied Animal Behaviour Science*: In press.

Sundrum, A., R. Andersson & G. Postler (Eds), 1994. Der Tiergerechtheitsindex-200/1994 - Ein Leitfaden zur Beurteilung von Haltungssystemen für Rinder, Kälber, Legehennen und Schweine. Verlag Köllen, Bonn, 211 pp.

Tannenbaum, J., 1991. Ethics and animal welfare: The inextricable connection. *Journal of the American Veterinary Medical Association* 198: 1360-1376.

Toates, F.M. & P. Jensen, 1991. Ethological and psychological models of motivation - towards a synthesis. In: J.A. Meyer & S. Wilson (Eds), From Animals to Animats. MIT Press, Cambridge, MA, pp. 194-205.

Verkaik, A.P., 1975. Rapport van de Commissie Veehouderij-Welzijn Dieren. Nationale Raad voor Landbouwkundig Onderzoek TNO, Den Haag, 220 pp.

Webster J., 1995. Animal Welfare: A Cool Eye Towards Eden. Blackwell, Oxford, 273 pp.

Wiepkema, P.R., 1987. Behavioural aspects of stress. In: P.R. Wiepkema & P.W.M. van Adrichem (Eds), Biology of Stress in Farm Animals: An Integrative Approach. Martinus Nijhoff, Dordrecht, pp. 113-133.

Wiepkema, P.R. & J.M. Koolhaas, 1993. Stress and animal welfare. *Animal Welfare* 2: 195-218.

Wierenga, H.K. & A.A. Jongebreur, 1987. Onderzoek welzijn landbouwhuisdieren. Pudoc, Wageningen, 116 pp.

Chapter 2. Strategy and prototype of a decision support system for welfare assessment in pregnant sows[1]

Abstract

Due to increasing empirical information on farm animal welfare since the 1960s, the prospects for sound decision making concerning welfare have improved. This chapter describes a strategy to develop a decision-making aid, a decision support system, for assessment of farm-animal welfare based on available scientific knowledge. Such a decision support system allows many factors to be taken into account. It is to be developed according to the Evolutionary Prototyping Method, in which an initial prototype is improved in reiterative updating cycles. This initial prototype has been constructed. It uses hierarchical representations to analyse scientific statements and statements describing the housing system. Welfare is assessed from what is known about the biological needs of the animals, using a welfare model in the form of a tree that contains these needs as welfare components. Each state of need is assessed using welfare-relevant attributes of the housing system and weighting factors. Attributes are measurable properties of the housing system. Weighting factors are assigned according to heuristic rules based on the principle of weighting all components (attributes and needs) equally, unless there are strong reasons to do otherwise. Preliminary tests of the prototype indicate that it may be possible to perform assessment of farm-animal welfare in an explicit way and based on empirical findings. The procedure needs to be refined, but its prospects are promising.

Keywords: animal welfare assessment, decision support system, pigs.

Introduction

Since the 1960s farm animal welfare has been the topic of moral and political debate (e.g. Harrison, 1964; Singer, 1975). It is a multifaceted concept, with both prescriptive and descriptive aspects (Fraser et al., 1997). The prescriptive aspect concerns questions of moral and political acceptability (Rushen and de Passillé, 1992). The descriptive aspect concerns welfare assessment per se, i.e. it is concerned with the question of What is the welfare status?, rather than What ought the welfare status to be?

A considerable amount of welfare research has been done; welfare reports (e.g. Scientific Veterinary Committee, 1997) and legislation, and codes of practice have been formulated, mainly attempting to establish minimum requirements for animal welfare. However, welfare is a quantitative variable and a standardised method (a welfare index) to assess the overall welfare status based on available scientific knowledge would have great political and moral (i.e. prescriptive) utility. It could be used to make legislation and product-quality control-

1 Paper by Bracke, M.B.M., J.H.M. Metz, A.A. Dijkhuizen & B.M. Spruijt, 2001. Development of a decision support system for assessing farm animal welfare in relation to husbandry systems: Strategy and prototype. *Journal of Agricultural and Environmental Ethics*: Accepted.

programmes less complex by prescribing one minimum overall-welfare level rather than minimum standards for many attributes of housing systems. This could give farmers the freedom to reach this minimum overall level in different, farm-specific ways. In addition, a welfare index could support decision makers to perform an integrated assessment of housing systems. Finally, it could be used to develop ethically defensible husbandry systems, when it shows that the animals' overall welfare status has indeed been improved substantially compared to conventional systems.

One major problem is that welfare is an ill-defined problem. At present it is still not clear how assessment of overall welfare should actually be performed. However, there has been a steady increase of empirical data on animals and their welfare (cf. Fraser and Broom, 1990). There have also been interesting attempts to construct practical welfare-models (e.g. Bartussek, 1990; Baxter and Baxter, 1984; Mellor and Reid, 1994; Taylor et al., 1995). It is generally agreed that many factors have to be taken into account, but the problem of weighting different parameters has not yet been solved. Our working thesis is that if welfare is indeed an appropriate, quantitative predicate of animals (e.g. Vorstenbosch, 1993; Broom, 1996), then it may be beneficial to use techniques from information technology. Welfare models must probably contain many parameters and employ complex calculation rules. Modern information technology is increasingly suited to collect and manage data (e.g. Date, 1995) and to perform calculations with various alternative models for welfare assessment. It also forces the assessment procedure to be performed in an explicit and formalised way. With this in mind, we set out to develop a decision support system that can be updated with new knowledge and that can help the end-user (e.g. politicians, farmers or animal welfare organisations) to make decisions about animal welfare. This chapter describes the strategy to develop such a decision support system and a first prototype that outlines an approach to identifying an explicit overall welfare-assessment procedure.

Strategy for decision support system development

According to Turban (1995) a 'DSS [decision support system] is an interactive, flexible, and adaptable CBIS [computer based information system], specially developed for supporting the solution of a non-structured management problem for improved decision making' (p. 84). A decision support system for welfare assessment, then, will assist in decision making as regards the 'non-structured' problem of how to assess welfare. An unstructured task is a task that is so poorly understood that the information to be used, the method of using the information, and the criteria for deciding whether the task is being done well cannot be specified (Alter, 1996, p. 133). For structured tasks such as standard diagnostic procedures in medicine expert systems may be developed, but for an unstructured task flexibility and adaptability are more important requiring the development of a decision support system. Welfare is also an unstructured task as it is not known exactly how to assess the welfare status in a systematic and objective way. Therefore, when employing information technology to the problem of welfare assessment, the development of a decision support system would seem to be appropriate.

A decision support system contains a knowledge base and a model base, which contain declarative knowledge and one or more models for problem solving respectively.

To develop a decision support system the Evolutionary Prototyping Method is commonly used (Turban, 1995), which starts with the development of a prototype that deals with a sub-problem or simplified version of the entire problem, and involves making a series of improved versions of the software based on immediate feedback from users. The steps taken in each development cycle include conceptual analysis, design and construction, testing, evaluation, and (making suggestions for further) upgrading. This process is likely to increase the ability of the DSS to be adaptable to changes in information requirements. In the course of time new insights about welfare assessment may be generated and the decision support system must have the flexibility to accommodate them.

In the first development cycle we produced a preliminary prototype to assess the welfare status of pregnant sows in a specified housing system, based on scientific statements and based on explicit calculation rules. We selected pregnant sows as a first 'case' to develop the methodology for welfare assessment in farm animals generally. Pregnant sows were chosen because relatively much is known about this group of animals. Below we will discuss the phases of conceptual analysis, design and construction, testing, evaluation, and upgrading for the prototype decision support system for pregnant sows.

Prototype decision support system

Conceptual analysis

The decision support system must support welfare assessment in relation to housing on a scientific basis. This specifies the problem space: the decision support system must deal with housing systems, scientific knowledge and the concept of welfare. These are its three kinds of input (Figure 1).

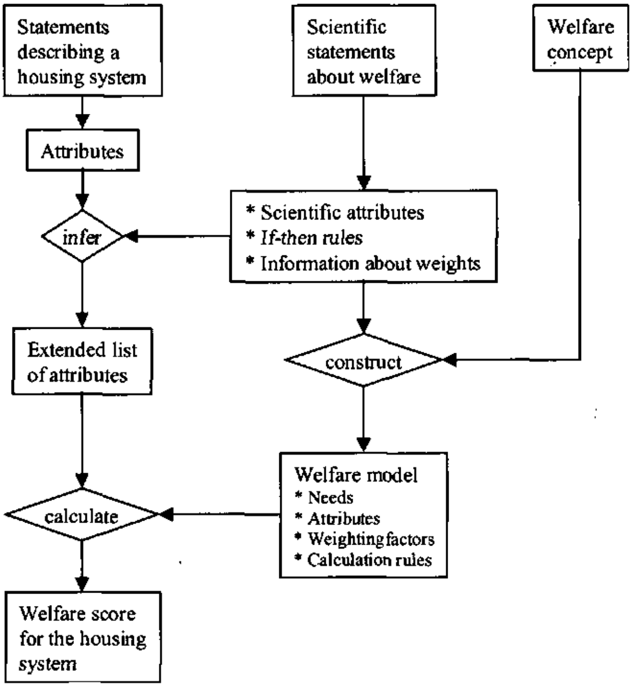


Figure 1. Animal welfare assessment diagram, which shows how the welfare status of animals in a housing system can be assessed on the basis of scientific statements and a concept of welfare.

The first kind of input is a description of a housing system. The term 'housing system' is used in a wide sense to include aspects of housing, management as well as performance criteria of the animals. We develop a decision support system for assessment at the housing-system level, but in such a way that individual farms can also be assessed. The description of the housing system must specify the living conditions of the animals. It must also allow making scientific inferences (see below).

Scientific statements derive from empirical research. They may be collected from the literature and from interviews with experts, and are stored in the knowledge base. Scientific statements are used to construct the welfare model (see below) but also to infer new truth-values about the attributes of the housing system. Inferring new knowledge is possible, because scientific statements often specify if-then relationships between the aspects of the environment and the welfare performance criteria of the animals, especially in terms of their behaviour, physiology, health and production. For example, in individual housing and when fed concentrates only, pregnant sows tend to develop stereotypic oral-behaviour patterns (Appleby and Lawrence, 1987). This allows inferring a truth-value about stereotypies based on information about housing and feeding. Ideally, a decision support system should allow this inferring function using scientific knowledge (Figure 1). The scientific statements and the statements describing the housing system together provide the objective basis for welfare assessment.

The third kind of input is the concept of welfare. Welfare can be defined in many different ways. Each definition of welfare may give rise to its own welfare model, which may be collected in the model base so as to serve the end-user. However, for actual decision making in the public domain, at least two criteria can be formulated. First, the welfare model should be in accordance with how welfare is understood in this public domain (Rushen and de Passillé, 1992). Scientists cannot just postulate some technical definition, e.g. that welfare is directly measurable as a function of cortisol levels (cf. Barnett and Hemsworth, 1990). For application in the public domain the concept of welfare must acknowledge the role of emotional states or feelings (Dawkins, 1980, 1990; Duncan and Petherick, 1991; Sandøe and Simonsen, 1992; Sandøe, 1996; Fraser et al., 1997) and theories of the animal mind that correspond with folk-psychology (e.g. Dennett, 1989; DeGrazia, 1996). Secondly, a welfare model must be based on knowledge of the facts, i.e. it must meet higher standards of objectivity than those required for personal purposes. Therefore, only objectively measurable parameters, including animal-based performance criteria and environmental parameters, should be used as parameters in the welfare model.

Even though at present there is no commonly accepted welfare theory available (Haynes, 2000), welfare scientists agree that animals have evolved cognitive and emotional systems ('needs') that help them to cope with changes and problems that threaten reproduction and survival in their natural environment (Wiepkema, 1987; Anon., in press). These cognitive and emotional systems are still operational, even in domesticated animals that are housed in environments that differ considerably from the natural environment. A pragmatic solution, therefore, is to assess animal welfare in a way that incorporates the animal's emotional states using the concept of needs. Needs are 'requirements, which are a consequence of the biology of the animal to obtain a particular resource or respond to a particular environmental or bodily stimulus' (Broom and Johnson, 1993). When scientists measure aspects of biological functioning (i.e. behaviour, physiology, pathology and production), they collect information that can be used to assess the animal's need states. Needs such as the need for food, water, thermal comfort, and social contact are the 'components' of welfare in that the overall welfare status of the animals is a function of their need states. Using needs for welfare assessment constitutes a form of functional decomposition, which entails a 'divide and conquer' strategy to solve complex problems by breaking them down into smaller sub-problems. The technique of functional decomposition is commonly used in systems science and in information technology (e.g. Cooper, 1992; Alter, 1996, p. 425). It can be used for welfare assessment when the complex problem of welfare assessment is broken down into a number of relatively more easily solvable assessments of need states. In this way the welfare model is based on biological processes and (neural) structures mediating the satisfaction of needs. A needs-based approach also supports completeness, because the various needs together must cover all aspects of welfare.

Needs may themselves be composed of sub-needs and must ultimately be determined from measurable attributes of the housing system. At the welfare side the needs in the welfare model will be relatively general for all species. At the attribute side, however, they will be more species specific (Baxter and Baxter, 1984). In assessing welfare from needs, and needs from attributes, weighting factors may be used. They should also be based as much as possible on the biology of the animals (Broom and Johnson, 1993). Since objective evidence for weighting factors is largely lacking (e.g. Taylor et al., 1995) we resorted to a (temporary) procedural solution, which will be explained below.

These deliberations about the modelling of housing systems, scientific knowledge and welfare underlay the construction of the prototype decision support system.

Design and construction

Since our main goal is to develop an explicit and formalised procedure for welfare assessment, the scope of the prototype was limited. It was restricted to only one housing system (individual housing) and contained only a small number of scientific statements. For knowledge representation we used a hierarchical representation-formalism, the 'tree'.

The prototype contains two tables with statements. One table contains 14 statements describing the housing system. The other table contains 36 scientific statements about pig behaviour, physiology, health and performance. Both tables were analysed in one tree, which has the term 'housing system' as root and attribute levels, i.e. properties of housing systems, as leaves (Figure 2). The tree contains a list of 56 attributes. Each of these attributes has two or more levels. For any one housing system only one level is true per attribute, thereby specifying one of its properties.

In total 34 out of the 56 attributes are relevant for welfare. The other, not directly welfare-relevant attributes in the tree help to determine what is true in the housing system about the welfare-relevant attributes. This is done by inference using if-then rules derived from the scientific statements (as described in the previous section).

To qualify as relevant for welfare an attribute must be a scientifically measurable parameter and it must be relevant in relation to at least one of the needs in the welfare tree. For example, attributes relevant to assess the need for food include regularity of feeding (predictability), meal size (controllability) and phasicity (pigs are biphasic animals, which means that they have two activity periods during the day, one in the morning and one in the evening, e.g. Hörning, 1992). The welfare-relevant attributes were incorporated into the welfare model.

Figure 2. Hierarchical representation of the entity 'housing system' in which the animals and their living environment can be described fully. The tree is read as 'a (housing) system has animals and environments, and the environment has objects, and the objects can be food, a floor, rooting substrate, conspecifics, the climate' etc.

```
Welfare has needs can be food  has  attributes can be   meal size can be ad lib
                                                                         much smaller than ad lib
                         water has attributes...        phasicity can be biphasic
                         thermocomfort ...                               monophasic
                         health...                      day activity can be day active
                         resting...                                         not day active
                         social contact...              regularity can be regular
                         moving...                                        not regular
                         rooting...                     stereotypies can be absent
                         body care can be grooming...                       present
                                          wallowing...  condition can be OK
                         ...                                             thin
                                                        eating tendency can be low
                                                                               high
```

Figure 3. Illustration of the prototype welfare tree. The decomposition of the need for food into its attributes is also included.

The main welfare model also has the form of a tree (Figure 3). It has 12 component needs, one for every behaviour system. The welfare tree is used to compute a welfare score by calculating weighted averages for every tree node from its components. Welfare is calculated from need scores, and need scores are calculated from attribute scores. The general formula is:

$$Sc = \frac{\sum_{1-n}(WF_i \times Sc_i)}{\sum_{1-n} WF_i}$$

where Sc = node score; WF_i = weighting factor of component$_i$; Sc_i = score of component$_i$; n: number of components connected to the node in the welfare tree; i: component number, with $1 < i < n$.

The attribute scores are derived proportionally from the welfare rank of the attribute levels within one attribute, and expressed on a scale from 0 to 10. For example, the attribute 'meal size' has two levels in the present prototype: 'ad lib' (i.e. 'all you can eat') and 'much smaller than ad lib'. The first level ranks higher for welfare than the second. (In the present stage of decision support system development we keep the rules simple and accordingly we ignore exceptions such as adverse consequences of inappropriate diets for ad lib feeding that should certainly be taken into account in a later stage.) The welfare ranks, rather than the absolute quantities of substrate, are transformed linearly into a numerical scale ranging from 0 (minimal, worst) to 10 (maximal, best). Accordingly, the attribute levels 'ad lib' and 'much smaller than ad lib' get scores of 10 and 0 respectively. An intermediate level would be assigned a score of 5, and so on.

Weighting factors are assigned to each attribute according to heuristic rules that are based on the principle of weighting attributes equally unless there are reasons to do otherwise. Such reasons are specified. For example, abnormal behaviour and pathological states receive a higher weight that is specified as equal to the number (n in the formula above) of attributes in the same list. Need scores for animals in a particular housing system are calculated for each need from its attribute scores and their weighting factors (cf. Table 1).

Table 1. Calculation of the state of the need for food in the prototype. The attributes of this need are assigned an equal weighting factor (WF) with the exception of 'stereotypies', which gets a high WF. The levels within every attribute get an attribute score on a scale between 0 (worst) and 10 (best). Since exactly one level per attribute is true for a housing system, we can determine its weighted attribute scores (shown in column 'housing system'). The score for the need for food is calculated as the weighted average attribute score (40 divided by 13 = 1.3, on a scale from 0 to 10).

Attribute	WF	Attribute level	Attribute score	Housing system
Meal size	1	Ad lib	10	
		much smaller than ad lib	0	0
Phasicity	1	Biphasic (two activity peaks per day)	10	10
		Monophasic (one activity peak)	0	
Day activity	1	Day active	10	10
		Not day active	0	
Regularity	1	Regular	10	10
		Not regular	0	
Stereotypies	7	Absent	10	
		Present	0	0
Condition	1	OK	10	10
		Thin	0	
Eat tendency	1	Low	10	
		High	0	0
Total	13			40

The overall welfare score (also scale 0 to 10) is calculated from the need scores in a similar way. The equality rule is also applied at the level of needs: equal weighting factors are attributed in principle to all needs on the same branch in the welfare tree. For example, the two components of the 'body care' need are 'grooming' (i.e. scratching in pigs) and 'wallowing'. These two needs are weighted equally to derive the score for 'body care. In turn, this need is weighted relative to the other needs at its own level (Figure 3). By continuing to calculate node scores as a weighted average from component scores, welfare is calculated in our prototype as a function of the (weighted) number of positive and negative attributes of the housing system.

Testing

The prototype was tested in four ways: a. by calculating the net weights of attributes, b. by constructing several versions of the welfare model, c. by applying the assessment method to compare housing systems and d. by comparing calculated scores with expert opinion.

a. The net weight of an attribute, i.e. its overall effect on welfare, can be calculated from its weighting factor and the place of that attribute in the welfare tree. The net weights range from 0.0038 to 0.083. This means that even the most important attribute contributes less than 10% to overall welfare.

b. Four versions of the welfare model were constructed and compared. In the main model the needs are ordered according to the behaviour systems. In another model the welfare tree was rearranged creating three main branches that represent physiological, behavioural and pathological needs. The third model skips the calculation of intermediate need-scores and calculates welfare directly from attribute scores. The fourth model is the TGI-200 (Tiergerechtheitsindex) model (Sundrum et al., 1994). This is a welfare assessment scheme that can be used for on-farm application, but it lacks a systematic account for the attributes in the model. The four models were (re)scaled to produce output on a scale between 0 and 10. For the (individual) housing system as described within the prototype, the TGI model generates the lowest score (2.3), the main model generates an intermediate score (3.4), and the other two models generate the highest scores (4.4 and 4.5). These model variations illustrate the principle of how, using information technology, different versions of the model can be generated and tested.

c. The calculation procedure was applied (outside the prototype) to four housing systems for pregnant sows: one individual and three group-housing systems, all without straw. In accordance with expectations, the system with individual housing received a lower score than the three group housing systems (Figure 4).

d. Finally, the calculated scores were compared with subjective scores obtained from an experienced pig-welfare ethologist. He was presented with a description of the housing system as analysed in the prototype and asked to rate both overall welfare and the different needs (Figure 5). There was a reasonable fit for some scores, e.g. for overall welfare, food, water, and grooming, but a difference for other scores such as health.

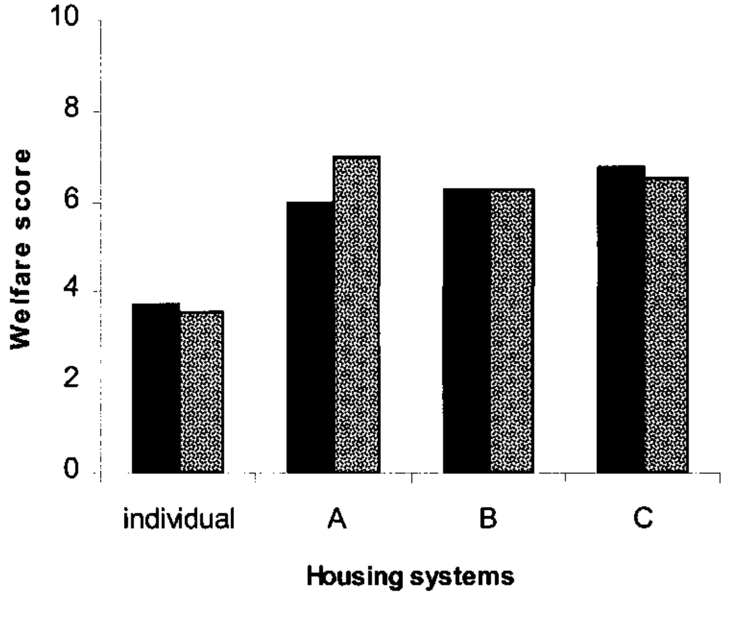


Figure 4. Welfare assessment of four housing systems for pregnant sows (one individual housing system and three strawless group-housing systems (A, B and C). Welfare scores (on a scale from 0 to 10) were subjectively assessed by a pig expert (dark columns) and calculated with the model (light columns).

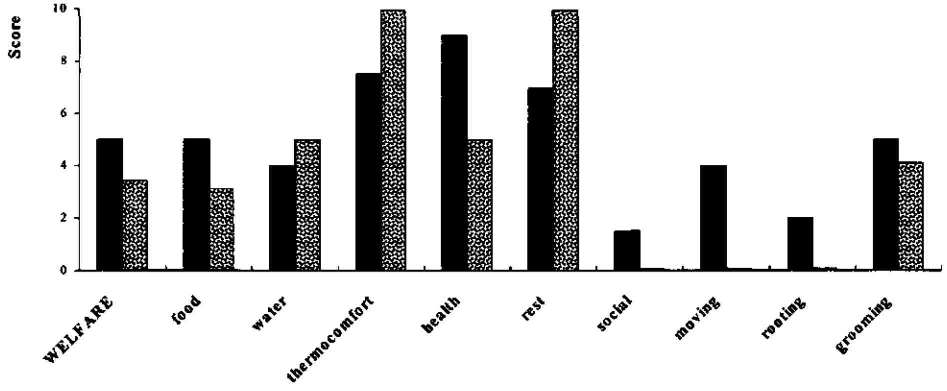


Figure 5. Comparing the welfare model with expert opinion. Scores for overall welfare and several needs for the housing system described in the prototype. Dark columns represent the subjective scores given by one pig expert. Light columns represent scores calculated with the welfare model in the prototype.

Evaluation and upgrading

The scope of the prototype is limited (only about pregnant sows, only one housing system and a limited number of scientific statements). The calculation procedure is simple. It uses a generic principle of calculating weighted averages, which is ultimately based on the ranking of levels within attributes. The more difficult task of comparing weightings between items (attributes or needs) is solved in a procedural way, by formulating heuristic rules based on the equality principle. Using procedures is one of the ways to improve decision making (Alter, 1996, p. 210). According to the equality principle, two items are given equal weight unless there are rational or scientific reasons for doing otherwise. This principle allows ignoring those instances where there may be some intuitive temptation to insert differential weightings, but where rational or scientific arguments for doings so are lacking. The equality principle provides a rational starting point for assigning weighting factors.

To further improve weighting rules, we are currently examining non-linear weighting (possibly in a hierarchical way, e.g. as suggested by Maslow, 1970). This may solve a potential problem for further development of the prototype procedure, which is that the weights of some attributes may become unacceptably small when the total number of attributes in the model increases.

Some differences were found between the calculated scores and expert opinion (Figure 5). The strength of the calculation procedure in the prototype is that all steps are made explicitly. This enables pinpointing the origin of such differences. Once the sources have been found, either the welfare model or the subjective assessment may require adjustment. The differences we found may have arisen because the number of attributes in the prototype was limited: each need score is derived from an average of only 2.4 attributes (range 1-7). As a result, the calculated need-scores tend to be either 0, 5 or 10. An expert, by contrast, is likely to take more aspects into account and to differentiate in greater detail. This problem may resolve when the prototype is developed further.

In the prototype we used a hierarchical representation-formalism (cf. Bracke et al., 1997). In principle, such a formalism can handle every possible statement concerning scientific knowledge or describing housing systems in a systematic way. In addition, trees seem suitable for constructing variations of the welfare model. The hierarchical organisation also facilitates evaluation of the welfare model, e.g. whether the model is complete, or whether important attributes are missing, and whether the model is balanced. In addition, the problem of multifunctionality of concepts can be handled. For example, the attribute 'day activity' (pigs are active during the light period) is relevant both for the need for food and for the need to rest, in that it is better for their welfare to have light during feeding and dark periods for resting. Such multifunctionality can easily be handled in our welfare tree by linking the attribute 'dayactivity' to both needs with reversed welfare-ranking of its levels. In general, trees were found to be valuable for welfare assessment, but other representation formalisms, such as procedural rules, fuzzy logic, (relational database) tables or object-oriented frames, may also be explored. In fact, we also used (if-then) procedural rules and had cast the tree in tables (cf. Table 1), but our main representation formalism was the tree.

The four welfare models in the prototype all had outcomes below 5. Roughly, a score of 5 means that only half of the relevant (weighted) attributes were fulfilled according to the sows' needs. This indicates that the housing system in the prototype, keeping sows individually without straw, is far from ideal for animal welfare. This conclusion is also supported by the results of applying the assessment procedure to the four existing housing systems (Figure 5). However, the scores generated by the models must be interpreted with care. From the scale that was used (between 0 and 10) it cannot be inferred where the threshold of acceptability should be drawn; this certainly does not have to be at the 5 level. The question what level is acceptable is logically distinct from the assessment of the welfare status, and is affected by factors other than animal welfare such as effects on human welfare and available alternatives. Furthermore, for proper comparison of scores a wider range of housing systems, including positive and negative controls (Wechsler et al., 1997) will have to be included in the analysis. It is important that a welfare model allows making proper distinctions between housing systems. In this respect the initial results of the prototype procedure, i.e. when applied to the four housing systems, showed promising results. The various outcomes also accord reasonably well with our own intuitions and with the opinion of the external expert. However, methodological concerns may be justified. It has been argued that scientists should not attempt to assess overall welfare objectively, because this is not possible (Fraser, 1995). However, in our opinion the practical utility of overall welfare assessment justifies making a serious attempt. Welfare assessment will always be relative to our present state of knowledge and our concept of welfare, but this should not refrain us from trying to reach consensus on these important issues.

Conclusions

In the prototype decision support system we developed a procedure to perform assessment of overall welfare in an explicit way that may, in the end, benefit actual decision making in politics and society. The prototype uses explicit statements describing a housing system and statements derived from empirical research, and it uses explicit procedural rules to assign scores and weighting factors. To our knowledge, such a procedure is unique in the field of animal welfare. However, the prototype is only the first cycle in the process to develop a decision support system according to the Evolutionary Prototyping Method. Further cycles will again include the stages of conceptual analysis, design, construction, testing, evaluation and upgrading before a final validation can be performed.

In the prototype a hierarchical representation formalism was used. Functional decomposition was used to dismantle the complex problem of welfare assessment into smaller sub-problems, namely the assessment of needs. The weighting problem was solved in a procedural way, using the equality principle. The test results of the prototype augur well for further development cycles.

The next step is to work at a more theoretical level on the welfare model and the weighting rules, using both further literature research and interviews with experts to find useful solutions. In addition, housing systems will be inventoried and the knowledge base will be extended. In this way, the present prototype will be the starting point for developing an adaptable decision support system in which the welfare status of farm animals can be assessed explicitly on the basis of available scientific knowledge. Such a system can be updated with new scientific data, even after the final validation has been completed.

Acknowledgements

This work was funded by the Technology Foundation (STW) (through a joint grant from the Netherlands Ministry of Agriculture, Nature Management and Fisheries and the Netherlands Organisation for Scientific Research (NWO)), and the Dutch Society for the Protection of Animals (NVBD). We thank Dr. Jeff Rushen for his comments on an earlier draft.

References

Anonymous. 2001. Scientists' assessment of the impact of housing and management on animal welfare. J. Appl. Anim. Welfare Sci. In press.

Alter, S. 1996. Information Systems: A Management Perspective, 2nd edition. Benjamin/Cummings, Menlo Park, CA, 728 pp.

Appleby, M.C., Lawrence, A. B., 1987. Food restriction as a cause of stereotyped behaviour in tethered gilts. Anim. Prod. 45, 103-110.

Baxter, M.R., Baxter, S.H., 1984. How to delineate the sow's welfare needs? Ann. Rech. Vét. 15, 281-286.

Bartussek, H., 1990. Naturnähe in der Veredelungswirtschaft. Der Tiergerechtheitsindex. In: Bartussek, H., Eisenhut, M., von Haiger, A., Storhas, R., (Eds.), Naturnähe in der Veredelungswirtschaft. Ein Definitionskonzept. IGN, Gumpenstein, pp. 34-46.

Barnett, J.L., Hemsworth, P.H., 1990. The validity of physiological and behavioural measures of animal welfare. Appl. Anim. Behav. Sci. 25, 177-187.

Bracke, M.B.M., Metz, J.H.M., Udink ten Cate, A.J., 1997. Assessment of animal welfare in husbandry systems. In: Sørensen, J.T. (Ed.), Proceedings of the fourth International Symposium on Livestock Farming Systems. Wageningen Pers, Wageningen, pp. 231-237.

Broom, D.M., 1996. Animal welfare defined in terms of attempts to cope with the environment. Acta Agric. Scand. Sect. A. Animal Sci. Suppl. 27, 22-28.

Broom, D.M., Johnson, K.G., 1993. Stress and Animal Welfare. Chapman & Hall, London.

Cooper D., 1992. Oh! Pascal! Turbo Pascal 6.0. W.W. Norton & Company, New York.

Date, C.J., 1995. An Introduction to Database Systems, 6th edition. Addison-Wesley Publishing Company, New York.

Dawkins, M.S., 1980. Animal Suffering: The Science of Animal Welfare. Chapman and Hall, London.

Dawkins, M.S., 1990. From an animal's point of view: Motivation, fitness, and animal welfare. Behav. Brain Sci. 13, 1-61.

Degrazia, D., 1996. Taking Animal Seriously: Mental Life and Moral Status. Cambridge University Press, Cambridge, MA.

Dennett, D.C., 1989. The Intentional Stance. MIT Press, Cambridge, MA.

Duncan, I., Petherick, J.C., 1991. The implications of cognitive processes for animal welfare. J. Anim. Sci. 69, 5017-5022.

Fraser, A.F., Broom, D.M., 1990. Farm Animal Behaviour and Welfare, 3rd edition. Baillière Tindall, London.

Fraser, D., 1995. Science, values and animal welfare: Exploring the 'inextricable connection'. Animal Welfare 4, 103-117.

Fraser, D., Weary, D.M., Pajor, E.A., Milligan, B.N., 1997. A scientific conception of animal welfare that reflects ethical concerns. Animal Welfare 6, 187-205.

Harrison, R., 1964. Animal Machines. Vincent Stuart Ltd., London.

Haynes, R.P., 2000. Do regulators of animal welfare need to develop a theory of psychological well-being? In: Robinson, P. (Ed.), Two Systems – One world. EurSafe 2000 Preprints, 2nd Congress of the European Society for Agricultural and Food Ethics, 24-26 August 2000. Centre for Bioethics and Risk Assessment, Copenhagen, pp. 135-138.

Hörning, B., 1992. Artgemässe Schweinehaltung, Grundlagen und Beispiele aus der Praxis. Verlag C.F. Müller, Karlsruhe.

Maslow, A.H., 1970. Motivation and Personality, 2nd edition. Harper & Brothers, New York.

Mellor, D.J., Reid, C.S.W., 1994. Concepts of animal well-being and predicting the impact of procedures on experimental animals. Proc. of the Conference 'Improving the Well-being of Animals in the Research Environment'. ANZCCART, Sydney, pp. 3-18.

Rushen, J., de Passillé, A.M.B., 1992. The scientific assessment of the impact of housing on animal welfare - a critical review. Can. J. of Anim. Sci. 72, 721-743.

Sandøe, P., Simonsen, H.B., 1992. Assessing animal welfare: Where does science end and philosophy begin? Animal Welfare 1, 257-267.

Sandøe, P., 1996. Animal welfare and human welfare - are they the same kind of thing? Acta Agric. Scand. Sect. A. Animal Sci. Suppl. 27, 11-1.

Scientific Veterinary Committee, 1997. The welfare of intensively kept pigs. Report to the Directorate General XXIV of the European Commission. Adopted 30. September 1997. Doc. XXIV/ScVc/0005/97, Scientific Veterinary Committee, Animal Welfare Section, Brussels.

Singer, P., 1975. Animal Liberation: A New Ethics for Our Treatment of Animals. New York Review of Books, New York.

Sundrum, A., Andersson R., Postler, G., Schneider, H., Striezel, A., Walter, J., Winckler, C., 1994. Tiergerechtheitsindex - 200. Ein Leitfaden zur Beurteilung von Haltungssystemen. Kollen Druck, Bonn.

Taylor, A.A., Hurnik, J.F., Lehman, H., 1995. The application of cost-benefit dominance analysis to the assessment of farm animal quality of life. Soc. Indic. Res. 35, 313-329.

Turban, E., 1995. Decision Support Systems and Expert Systems, 4th edition. Prentice-Hall, London.

Vorstenbosch, J., 1993. The concept of integrity: Its significance for the ethical discussion on biotechnology and animals. Livest. Prod. Sci. 36, 109-112.

Wechsler, B., Fröhlich, E., Oester, H., Oswald, T., Troxler, J., Weber, R., Schmid, H., 1997. The contribution of applied ethology in judging animal welfare in farm animal housing systems. Appl. Anim. Behav. Sci. 53, 33-43.

Wiepkema, P.R., 1987. Behavioural aspects of stress. In: Wiepkema, P.R., van Adrichem, P.W.M., (Eds), Biology of Stress in Farm Animals: An Integrative Approach. Martinus Nijhoff, Dordrecht, pp. 113-133.

Chapter 3. Review part 1: Is overall welfare assessment possible?[1]

Abstract

Several authors have concluded that scientists should not attempt to perform overall animal welfare assessment (OWA). They argue that scientists have continued to fail to make progress in this area and that value judgements are inherently involved in OWA for which science cannot provide answers. We take a more positive attitude toward OWA and argue that scientists should avoid creating a self-fulfilling prophecy. OWA is necessary for making actual moral and political decisions. Science has already accumulated much relevant information about welfare and this information should be applied in decision making.

The task of OWA is to assess welfare based on knowledge of the biological needs of animals. Weighting of welfare-relevant factors constitutes a problem. However, when scientists cannot provide empirical data to solve weighting issues, this does not mean that rational answers cannot be found, e.g. in the form of procedural rules. OWA is conceived as a problem of multi-criteria decision making with fuzzy information. It focuses on the descriptive aspect of welfare, i.e. on what the welfare status of the animals really is without taking an ethical stance. The welfare status of animals depends on their biology and on the way animals assess their own welfare. It does not depend on how it happens to be perceived by us. Even though OWA necessarily remains a human activity, it is not arbitrary, nor does it allow of multiple 'correct' answers. OWA is a descriptive activity that can achieve more and more accuracy as science proceeds.

Keywords: welfare assessment, housing, decision support, weighting, ethics.

Introduction

Concern for animal welfare is an issue for many people. Expressing concern about welfare often presupposes making an assessment of the overall welfare-status. Many people are convinced about the validity of their personal assessment of the welfare status of animals. However, differences in opinion appear hard to resolve.

To help resolve these differences we are presently working on a model to assess the welfare status of farm animals on a scientific basis. Our goal is to develop a tool to perform overall welfare assessment (OWA), which can be used to support moral and political decision-making. For this purpose we are developing a kind of expert system, a decision support system (Bracke *et al.*, 1999). Such a system requires a method to assess welfare in an explicit way and on a scientific basis.

In this and the two following chapters we will discuss various considerations for performing OWA. The chapters 4 and 5 deal with available assessment tools and the biological basis for OWA respectively. In the present chapter we will discuss the methodological question whether it is theoretically possible to perform OWA on a scientific

1 Paper by Bracke, M.B.M., B.M. Spruijt & J.H.M. Metz, 1999. Overall animal welfare assessment reviewed. Part 1: Is it possible? *Netherlands Journal of Agricultural Science* 47: 279-291.

basis. We are optimistic and believe OWA is possible.

However, scepticism about OWA is widespread, even among applied ethologists. Several prominent authors in the field of applied ethology have concluded that science is limited in its ability to determine 'overall' welfare or to compare welfare in disparate environments (Fraser, 1995; Rushen & de Passillé, 1992; Dawkins, 1998). Some authors (e.g. Fraser, 1995; Tannenbaum, 1991) have provided a forceful methodological argument that may be used to justify this scepticism about OWA. They argued that welfare cannot be defined and studied as a purely technical, scientific concept, because underlying value-related assumptions are inherently involved in OWA (Sandøe & Simonsen, 1992; Fraser, 1995; Fraser *et al.*, 1997). In this chapter we will argue against a sceptical interpretation of this methodological argument and explain why we believe it is possible to perform OWA on a scientific basis.

Welfare definitions

Welfare definitions have been reviewed by several authors (Rushen & de Passillé, 1992; Anon., 1992; Fraser *et al.*, 1997; Stafleu *et al.*, 1996), and were discussed in two relatively recent international conferences (see Journal of Agricultural and Environmental Ethics 6, Supplement 2: 1993; Acta Agriculturae Scandinavica Section A, Animal Science, Supplement 27: 1996). Below we will discuss points of disagreement and consensus to arrive at a subjective definition of welfare. Subjective definitions of welfare tend to strengthen the sceptic view that OWA is not possible, but the discussion of welfare definitions will also provide a clue to reject scepticism.

There appears to be general consensus regarding a number of properties relating to welfare (partly extracted from Broom & Johnson, 1993). (1) Welfare is a characteristic of animals, i.e. it is a descriptive property of animals; it is not a property of the environment. Since it is generally agreed that animals have a welfare status, we will not be dealing with the Cartesian argument that animals may be automata lacking consciousness (Bermond, 1997), nor with the application of Ockham's razor so as to deny radically that welfare states can be known (Kennedy, 1992). (2) Welfare can range on a continuum from very poor to very good. (3) The results of scientific measurements are relevant for welfare assessment. (4) The scientific assessment of animal welfare requires that a variety of measures must be employed. There is, at present, no standard welfare 'thermometer'.

Welfare has been defined in various ways. Examples of much cited definitions of welfare are 'living in harmony with the environment and with itself, both physically and psychologically' (Lorz, 1973), and 'the welfare of an individual is its state as regards its attempt to cope with its environment' (Broom, 1986). Widely different concepts have been used to define welfare. They include (listed tentatively from the more objective to the more subjective): pre-pathological states (Moberg, 1985), stress, coping, fitness and adaptation (e.g. Broom, 1986), predictability and controllability (Wiepkema, 1982), harmony (e.g. Lorz, 1973; Hughes, 1976), emotional states, wants, subjective feelings, suffering (Dawkins, 1988; Duncan & Petherick, 1991, Sandøe, 1996).

A distinction has been made between subjective and objective definitions of welfare (Mendl, 1991; Rushen & de Passillé, 1992; Sandøe *et al.*, 1996; Barnard & Hurst, 1996; Sumner, 1996). Objective definitions relate welfare directly to measurable parameters (e.g. Broom, 1986). They tend to emphasise the importance of biological functioning and seem 'inspired' by the question how welfare can be measured in a scientific, objective way. By contrast, subjective definitions define welfare in terms of subjective emotional states of animals, i.e. as what matters to the animals from their point of view (e.g. Dawkins, 1988, 1990; Duncan, 1996). Subjective definitions are more prevalent in the philosophical literature (e.g. Singer, 1990; Regan, 1983) and seem 'inspired' by the question how animals ought to be treated ethically (Rushen & de Passillé, 1992).

Subjective definitions have the problem how emotional states can be studied scientifically (Mason & Mendl, 1993). On the other hand, objective definitions seem to have counterintuitive implications. Examples include cases of adaptive pain (Dawkins, 1980) and an unfelt tumour (example from Mason & Mendl, 1993). Objective definitions of welfare have these problems precisely because they are objective and accordingly fail to satisfactorily accommodate the perspectival (point of view) nature of welfare (Sumner, 1996).

Stafleu *et al.* (1996) suggested an alternative classification that may help resolve these points of dispute. They distinguish three levels to define welfare: conceptual, explanatory and operational. Operational definitions concern the question how welfare can be assessed or measured in practice; explanatory definitions concern how welfare may be conceived within a scientific framework; conceptual definitions identify the meaning of the concept of welfare at a philosophical level. This classification made by Stafleu *et al.* may relate to the subjective-objective dichotomy in two ways. If subjective definitions apply at the conceptual level and objective definitions apply at the explanatory or operational level, then apparent disagreement would be resolved by pointing out the difference in abstraction levels. On the other hand, if the objective-subjective dichotomy applies within the conceptual level, then the dispute is more fundamental and important theoretically, but may have relatively little practical (operational) implications. Different approaches to welfare often lead to similar conclusions (Duncan & Fraser, 1997). The impact of cases like adaptive pain, unfelt tumours and injured unconscious animals are probably of little importance for actual OWA in relation to housing.

Furthermore, true proponents of an objective welfare definition are becoming rare. Broom has generally been cited as a scientist defining welfare as an objective state. Recently, however, Broom (1998) re-emphasised the importance of feelings as part of his concept of welfare. Maybe there is more consensus on the issue whether subjective feelings of animals are a central part of the concept of welfare than would appear from the subjective-objective dichotomy. Sentience is generally accepted as a necessary condition for welfare. 'When people express concern about animal welfare, it is precisely the conscious experience of suffering that worries them most' (Dawkins, 1998, p. 306). Non-sentient objects like machines, computers or plants do not have a welfare status, at least not in a sense that is relevant in a socio-political context (Stafleu *et al.*, 1996).

Therefore, when all agree that emotional states are an important part of welfare and some believe emotional states are sufficient to define welfare, it seems rational to opt for a subjective definition of welfare as a starting point for OWA. In effect, we suggest to restrict the use of the term welfare to the representation of what matters to animals from their point of view. We use the term welfare to denote the animal's quality of life as it is experienced and valued by the animal itself, i.e. its prudential value (Sumner, 1996). A central role in welfare is taken by the emotional states that can be ascribed to animals. The welfare state of an animal is determined by all the emotional states and only the emotional states in so far as they are experienced subjectively by that animal.

Restricting the use of the term welfare to subjective states does not correspond with some intuitive notions which people may have about welfare. Per definition, a drugged animal that is kept in a permanently euphoric state has a high welfare status, even though it may be questioned whether this is morally acceptable. Alternatively, it would be a category mistake to believe that an unconscious animal that is injured has poor welfare (contra Broom, 1998). We would prefer to give such infringements another name, e.g. harm to integrity, but would not say that the animal's welfare was affected. We believe that to restrict the use of the term welfare to subjective emotional states will benefit discussions about welfare, because it can provide some conceptual clarity, which is much needed in this area. At the explanatory level, however, it is necessary to explain the relationship between subjective emotional states on the one hand and theoretical concepts that can be measured, at least in principle, on the other hand. At the operational level, finally, welfare assessment must necessarily depend fully on observable attributes, for, in the end, all information about animal subjectivity must necessarily be derived from what is or can be observed about the animals.

Because subjective definitions of welfare are the least 'measurable' of all definitions, they make OWA even more vulnerable to sceptical criticism than objective definitions do. Before we will discuss the main methodological argument we will briefly discuss a factual argument against OWA.

A factual argument against OWA

Rushen & de Passillé (1992) stated that research has not been successful to perform OWA despite substantial political pressure. This observation may justify a recommendation not to attempt OWA, but it fails to support the sceptic conclusion that scientists should not attempt OWA at all. Furthermore, even a well-intended recommendation comes with the risk of creating a self-fulfilling prophecy, warranting the question that scientists really tried hard enough to perform OWA.

The clause 'despite substantial political pressure' indicates that previous efforts should be taken seriously. However, to our knowledge a complete review of 'failed' attempts to perform OWA has never been published. This makes it difficult to evaluate whether previous attempts have been exhaustive.

Finally, even if this claim were true at present, the factual argument will loose force over time. Developments in other fields, such as neurobiology or information technology, may bring new perspectives and possibilities for performing OWA.

Some minor aspects of methodological scepticism

Methodological scepticism about OWA concerns the limits of human cognition to know the welfare status of animals.

One argument is that there seems to be a general methodological problem with OWA because similar attempts in other fields have also failed. Well known are methodological problems with IQ tests and the apparent lack of human health and welfare indices. Opposing this criticism we recognise the development and use of index scores in several related fields. Examples include the field of medical technology assessment (e.g. McDowell & Newell, 1987; Streiner & Norman, 1995), psychiatric tests (the DSM IV, American Psychiatric Association, 1994) as well as tools to quantify environmental impacts (e.g. van Lenthe *et al.*, 1996). Who hasn't been subjected to tests in school or to a psychological test for a job interview? Several of these indices have proven their use. When developed carefully and when used with care, indices may be extremely valuable for certain purposes. This also applies to an index for animal welfare.

A related argument is that welfare is not a constant feature. Welfare is a transient state which has multiple attributes and which is different under various circumstances (Dawkins, 1998). Similarly Fraser *et al.* (1975) argued that stress (read 'welfare'), like disease, cannot be put on a unidimensional scale. Welfare varies in nature; it varies over time; and it varies between individuals in a group. From this argument it follows that it is important to specify the exact circumstances. When this is done, we may, for example, state that the welfare status of a certain group of animals over a specified period of time is '7' on a scale from 0 to 10. This is a very short statement that conveys much information in a very efficient way. It specifies that welfare is reasonably good, but could still be better. Ideally, the scoring system should be transparent, in that a full explanation for this score can be given, including a specification of all underlying normative assumptions that were used to reach the welfare score (Sandøe & Simonsen, 1992; Fraser *et al.,* 1997). Such a methodology for OWA could be of substantial benefit for decision making concerning animal welfare.

The argument that values are inherently involved

Fraser (1995) argues that science has a limited ability to compare welfare in systems that differ in a large number of features. Welfare is not a single attribute, like the height of a building, which can easily be measured in meters. On the contrary, welfare is a complex attribute, more like the safety of a building. The safety of a building will be different for different types of users of the building (p. 105). Similarly, different people will judge the welfare status of animals as very different, because they emphasise different attributes of housing systems as most important for welfare.

Since welfare is a complex attribute, a large number of variables (behaviour, physiology, productivity, and health) must be taken into account. An overall welfare-judgement is possible when one system outperforms the other in every respect (cf. also Taylor *et al.,* 1995). However, a variety of measures are likely to yield a complex picture, with certain important advantages favouring each system. Fraser (1995) argues that science cannot provide the data that are necessary to decide which is the better system in such cases. In the absence of scientific facts value judgements are inherently involved in deciding which advantages of each system are most important.

This argument does not deny that animals have a welfare status about which science can provide much relevant information to help reducing value-based differences in people's ideas about animal welfare (e.g. Fraser, 1995, p. 106). Rather, the argument suggests that because science cannot solve value-based differences that are inherently involved in OWA, scientists should not attempt to perform OWA.

We have attempted to cast the reasoning steps of this sceptical argument into the following syllogism:

a. OWA requires combining many different attributes of housing systems that point in different directions.
b. Combining attributes that point in different directions requires weighting them against each other.

If a and b then c.

c. OWA requires weighting attributes.
d. Weighting attributes inherently involves value judgements.

If c and d then e.

e. OWA involves inherent value judgements.
f. Descriptive statements cannot logically refute or confirm value judgements.

If e and f then g.
g. Descriptive statements cannot logically refute or confirm OWA.
h. Science only makes descriptive statements.

If g and h then i.
i. Science cannot logically refute or confirm OWA.
j. What cannot be done, should not be attempted.

If i and j then k.
k. Science should not attempt OWA.

Reply to the syllogism

The syllogism above contains 11 statements (a-k) involving five inference steps. The final conclusion (k), which is a value judgement (that scientists ought not to attempt OWA), is drawn from (i), which is a description (that science cannot prove OWA). This inference requires statement (j) (that ought implies can, cf. Griffin, 1992) to avoid a naturalistic fallacy (which is statement (f) that descriptions cannot logically refute or confirm value judgements). This poses the question what science can and should do.

Bekoff *et al.* (1992) correctly pointed out that an assessment of animal welfare is always an assessment from a human's point of view. Scientists tend to be very careful when making statements about welfare, especially when it concerns the subjective states of animals, because as scientists they have a tradition of preferring parsimonious explanations that generally have been considered to exclude animal subjectivity. This sceptical attitude with respect to welfare stands in sharp contrast with the apparent ease with which 'ordinary' persons seem to be able to assess the welfare state of animals with which they have at least some familiarity. Often such private claims about welfare are made with a strong conviction that they are accurate and valid.

Recently, there has been a revival of putting the concept of welfare into an evolutionary perspective (e.g. Barnard & Hurst, 1996; Dawkins, 1998). For a moment we would like to put the human cognitive capacities regarding animal welfare in an evolutionary perspective. Humans have evolved as hunter-gatherers and farmers. Being able to recognise states of animal welfare, such as sick or aggressive animals, is likely to have increased fitness in humans. Wemelsfelder *et al.* (1998) found high inter- and intra-observer consistency of spontaneous subjective assessments of pig behaviour. This finding supports the suggestion that human cognitive abilities may be quite accurate in representing welfare. In line with this view the role of science could be to provide accurate factual data as well as to expose possible defects in our cognitive abilities to assess welfare.

The biological sciences that underlie welfare research seem to be unable to provide proof about animal consciousness. Opposing this demand, it has been argued that it may be unreasonable to demand proof in the case of welfare, because the underlying sciences themselves generally concern issues of statistical significance (the 5% probability level) rather than proof (Sambraus, 1981). Accordingly, instead of demanding proof, the task of OWA should be to provide the best possible solution or prediction of what the welfare status is, based on the scientific data that are available at the time the assessment is made. It follows that OWA is relative to the level of factual knowledge and this level may change over time depending on scientific progress.

Furthermore, the idea that science only makes descriptive statements (statement (h) in the syllogism above) may be questioned. Science is not value-free (cf. Rollin, 1990). Values are involved not only at the periphery of science, such as in deciding whether to allocate money to a project, but also more internal to science. Many descriptive activities require interpretations that involve value judgements. Since the quality of science, both internally and at its periphery, can be better or worse, *all* sciences inherently involve values. In these senses, a scientific approach to OWA can be expected to be value laden as well. It follows that the (normative) claim that underlying value notions should be made explicit (e.g. Fraser, 1995; Fraser *et al.*, 1997) applies to all scientific activities and is not specific for OWA. Therefore, the inherent involvement of values in OWA cannot support a sceptical attitude unless values are involved in OWA in additional ways. We will discuss two such ways. One concerns the weighting problem and the other is about ethical concerns.

The weighting problem

The weighting problem concerns the statements (a) to (e) in the syllogism. Although the statement (a) that many different attributes must be weighted, increases the complexity of welfare assessment, and therefore also the risk of making errors, the methodological issue concerns the unitary case of weighting two attributes against each other. An example, given by Fraser (1995, p. 106), is to weigh freedom from coccidiosis and freedom of movement. It would be impossible to weigh such attributes against each other, especially when people differ about what is more important and when scientific data cannot decide between them. Another example concerns tethered sows: "Observer A ... may conclude that the welfare of ... sows tethered in stalls is high because the animals are well fed, reproducing efficiently and free from disease and injury. Observer B ... concludes that the welfare of the same animals is poor because they give vocalisations that are thought to indicate frustration, and they escape from the stalls whenever the chance arises. Observer C ... agrees that the sows' welfare is poor because stalls are unnatural environments which prevent the animals' natural behaviour" (Fraser *et al.*, 1997, p. 202). The sceptic may be tempted to conclude from such examples that everybody can be right about welfare.

However, the way humans perceive welfare is not constitutive of the welfare status of animals. If observer A disagrees with observer B, then at least one of them must be mistaken. In the absence of proof, there is a second-best alternative, namely that procedure in which all available evidence has been taken into account in the most rational way. Rational OWA requires that reasons be provided for why certain aspects are believed to be more important than others are. Rational OWA implies that the assessment is based on observable and measurable data and that the point of view of the animals is taken into account as much as possible.

Animals are able to compare widely different aspects of their environment (McFarland, 1989). Where animals can compare, they themselves are the norm because welfare was defined as what matters from their point of view. Where animals cannot compare, e.g. when assessments concern effects that apply over longer periods of time or between individuals in a group, a method for OWA must resort to other rational procedures (Sandøe *et al.*, 1996). The starting principle may be 'all to count for one, and nobody for more than one'. Adjustments to this principle are necessary, but must comply with the constraint of rationality, which includes, for example, the requirements of consistency and reflective equilibrium. The method of finding a reflective equilibrium uses our most basic intuitions to develop principles which, in turn, are used to evaluate our more peripheral intuitions (e.g. DeGracia, 1996). In this respect OWA may be more akin to the social sciences and to philosophy than to the natural sciences.

In the end, animals must provide the norm of what is more important freedom from coccidiosis or freedom of movement. The extremes are easy to weigh. Undoubtedly, the mildest case of coccidiosis is better than the severest-possible restriction of space, while severe coccidiosis is undoubtedly worse than a mild restriction of space. Therefore, weighting of rather incompatible attributes, such as coccidiosis and space, is possible at least in extreme cases.

The relative self-evidence involved in extreme cases may have led several authors to argue that instead of trying to assess overall welfare scientists should take a problem-oriented approach, which entails identifying, rectifying and preventing welfare problems (Fraser, 1995; Rushen & de Passillé, 1992). However, the problem-oriented approach fails to be a proper alternative for OWA. In order to establish that a problem really is a problem and to establish that it is important, one must have some grasp of the overall context. For example, severe lameness is obviously a real welfare problem. If this is a 'fact', then, by the same token, we have sufficient knowledge to use this kind of knowledge as a basis for OWA: the problem-oriented approach constitutes a way of doing OWA (cf. Taylor *et al.,* 1995). This would support our optimism about OWA. However, the problem-oriented approach is a rather narrow approach to OWA. It only concerns extreme cases and takes into account only the negative aspects of welfare. A more comprehensive view must also attempt to take into account less extreme cases as well as positive aspects of welfare.

When weighting involves less extreme cases, a standard is needed to verify whether the outcomes of OWA are valid. However, no golden standard is available, because the private minds of animals will probably never be directly assessable (Nagel, 1974; Dawkins, 1993; Mason & Mendl, 1993). This increases the uncertainty involved in OWA, but it may not be an insurmountable problem. OWA is not the only field lacking a golden standard. The same problem occurs in related fields, which we referred to above such as medical technology assessment and psychiatric tests. Validation of OWA may be performed in different ways. These may be based on a combination of optimism about our human cognitive ability to assess welfare and the ability of science to expose errors in this cognitive ability.

Further progress in OWA beyond simple cases of weighting attributes requires finding a rational way to perform OWA in a systematic and explicit way. It is not a simple task to take into account all available data (Rushen, 1991; Dawkins, 1997) and it is likely to require a multidisciplinary approach (Sandøe & Simonsen, 1992) involving multi-criteria decision making with fuzzy information. It would be naïve to believe that politicians can make decisions based on an accumulation of facts provided by scientists. The interpretation of facts involves weighting evidence, which is as much part of science as the gathering of data. In this respect OWA is itself part of the natural sciences.

Several authors have also shown to be optimistic about solving the weighting problem. For example, Morton & Griffiths (1985) propose that the various bodily and behavioural signs can usefully be given scores so that an accumulated score of pain can be obtained. Taylor *et al.* (1995) and Appleby (1997) suggest that cost-benefit analysis may provide a solution to the weighting problem. Fraser (!, personal communication) is also optimistic that schemes for combining different measures into some kind of 'welfare score' can be developed if we really want to or need to, and that we could achieve some reasonable consensus among experts about which items are more or less important. This point illustrates that the authors we have quoted above may not be themselves as sceptical about OWA as would appear from the text.

In our next chapter we will review many other authors who have attempted OWA in a systematic and explicit way. This has made their work vulnerable for criticism, but, we will show, also open for further improvement.

Ethical concern

A second candidate why values may be thought to be additionally involved in OWA constitutes ethical concern. Fraser *et al.* (1997) interpret the task of making values involved in OWA explicit as identifying values underlying ethical concern. Like Tannenbaum (1991) Fraser maintains that there is an 'inextricable connection' between animal welfare and ethical values. The term 'freedom' that was used to illustrate the weighting problem above, also has an ethical connotation (Fraser & Broom, 1990). However, welfare assessment per se is logically distinct from making an ethical assessment. Ethical statements are ought-statements, i.e. they are prescriptive. Welfare statements are is-statements, i.e. they are descriptive. It is a naturalistic fallacy to derive an ought from an is (cf. statement (f) in the syllogism above). A situation with poor welfare for animals may very well be ethically acceptable, for example when the interests of the animals are outweighed by the interests of human beings, as may be the case in certain types of animal experimentation. Therefore, OWA and ethical assessment are logically distinct from one another (cf. Fraser & Broom, 1990, p. 256; Rushen & de Passillé, 1992; Broom, 1996; Barnard & Hurst, 1996; Dawkins, 1998).

OWA is not sufficient for ethical decision making. However, OWA is important and maybe even necessary for ethical decision making. There is widespread agreement that welfare is a morally relevant property of animals (e.g. Vorstenbosch, 1993). Because the relationships between humans and animals in all human societies are so intimate that many of our decisions affect their welfare, we cannot ignore our ethical responsibility. Because of this responsibility we cannot do without OWA. Making factual claims, either explicit or implicit, about the welfare status of animals is unavoidable in human society. Therefore, OWA is necessary for practical reasons. In addition, OWA may be performed for purely scientific reasons, namely when viewed as the attempt to find out what the welfare status of animals really is, as objectively as we possibly can and on the basis of the best available knowledge.

Because OWA is a descriptive activity, it must be possible to assess the welfare status of animals (provided they really have such a welfare status). However, welfare also has a prescriptive element in its meaning. Since welfare concerns what matters to animals from their point of view, it does inherently involve 'prescriptions' made by animals. By saying that welfare is poor, we identify, as it were, prescriptions issued by animals, which are of the kind 'let this not happen to me'. However, all prescriptions also have a descriptive element in their meaning (Hare, 1981, p. 22). Even matters of taste, which are highly subjective, have a factual aspect that can be reported objectively. Mutatis mutandis, OWA involves the attempt to assess welfare descriptively, even though welfare is itself a prescriptive property of animals. Also, OWA does not violate the principle of parsimony, because the goal is not to explain observable phenomena based on mental states, but to assess mental states and welfare based on observable phenomena.

Conclusions

During the last few years we have been working on a project to find a method to assess welfare (OWA) on a scientific basis. This work has made us optimistic about its feasibility. However, we also perceive a sceptical attitude within the scientific community that may jeopardise progress in this field of research, because such scepticism may become a self-fulfilling prophecy. Our optimism has its origin in the belief, which is shared by many people, that OWA based on knowledge of the facts is possible. This is because OWA can be conceived as multi-criteria decision making with fuzzy information that concerns the descriptive activity to determine what matters to animals from their point of view. Values are inherent in OWA in the way values are inherent in all scientific activities (and especially in many fields of biology). However, as a descriptive activity OWA is logically distinct from making an ethical assessment. When performing OWA many uncertainties, especially concerning the weighting of attributes, must be dealt with. For this a multidisciplinary approach is needed. Knowledge about the animal's behaviour and physiology should be used in a systematic way to assess their welfare, because the animals themselves should provide the norm of what is important for their welfare. Where science fails to provide final answers, OWA may benefit from more fundamental research. However, we also stressed the importance of finding rational procedures to integrate the large amount of knowledge that is already available in a systematic and explicit way. We are optimistic that such a method for OWA can be developed.

Acknowledgements
We wish to thank Dr. David Fraser for comments on an earlier version of this paper.

References

Anonymous, 1992. Report of the scientific veterinary committee (animal welfare section) on the welfare of laying hens kept in different production systems. Report: VI/5291/92-EN. CEC, Brussels, 147 pp.

American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorders (DSM-IV), 4th edition. American Psychiatric Press, Washington, USA.

Appleby, M.C., 1997. Life in a variable world: Behaviour, welfare and environmental design. *Applied Animal Behaviour Science* 54: 1-9.

Barnard, C.J. & J.L. Hurst, 1996. Welfare by design: The natural selection of welfare criteria. *Animal Welfare* 5: 405-433.

Bekoff, M., L. Gruen, S.E. Townsend & B.E. Rollin, 1992. Animals in science: Some areas revisited. *Animal Behaviour* 44: 473-484.

Bermond, B., 1997. The myth of animal suffering. In: M. Dol, S. Kasanmoentalib, S. Lijmbach, E. Rivas & R. van den Bos (Eds), Animal Consciousness and Animal Ethics: Perspectives from the Netherlands. Van Gorcum, Assen, pp. 123-143.

Bracke, M.B.M., J.H.M. Metz, A.A. Dijkhuizen & B.M. Spruijt, 1999. The development of a decision support system to assess the welfare status of pregnant sows. In: Th. Wensing (Ed), Production Diseases in Farm Animals: 10th International Conference 1998. Wageningen Pers, Wageningen, p. 235.

Broom, D.M., 1986. Indicators of poor welfare. *British Veterinary Journal* 142: 524-526.

Broom, D.M., 1996. Animal welfare defined in terms of attempts to cope with the environment. *Acta Agriculturae Scandinavica Section A, Animal Science, Supplement* 27: 22-28.

Broom, D.M., 1998. Welfare, stress, and the evolution of feelings. *Advances in the Study of Behavior* 27: 371-403.

Broom, D.M. & K.G. Johnson, 1993. Stress and Animal Welfare. Chapman & Hall, London, 211 pp.

Dawkins, M.S., 1980. Animal Suffering: The Science of Animal Welfare. Chapman and Hall, London, 149 pp.

Dawkins, M.S., 1988. Behavioural deprivation: A central problem in animal welfare. *Applied Animal Behaviour Science* 20: 209-225.

Dawkins, M.S., 1990. From an animal's point of view: Motivation fitness and animal welfare. *Behavioral and Brain Sciences* 13: 1-61.

Dawkins, M.S., 1993. Through our Eyes only. The Search for Animal Consciousness. W.H Freeman and Company, Oxford, 192 pp.

Dawkins, M.S., 1997. D.M.G. Wood-Gush Memorial Lecture: Why has there not been more progress in animal welfare research? *Applied Animal Behaviour Science* 53: 59-73.

Dawkins, M.S., 1998. Evolution and animal welfare. *The Quarterly Review of Biology* 73: 305-328.

DeGrazia, D., 1996. Taking Animals Seriously. Mental Life and Moral Status. Cambridge University Press, New York, 302 pp.

Duncan, I.J.H., 1996. Animal welfare defined in terms of feelings. *Acta Agricultura Scandinavica Section A, Animal Science, Supplement* 27: 29-35.

Duncan, I.J.H. & D. Fraser, 1997. Understanding animal welfare. In: M.C. Appleby & B.O. Hughes (Eds), Animal Welfare. CAB International, Oxon, pp. 19-31.

Duncan, I.J.H. & J.C. Petherick, 1991. The implications of cognitive processes for animal welfare. *Journal of Animal Science* 69: 5017-5022.

Fraser, A.F. & D.M. Broom, 1990. Farm Animal Behaviour and Welfare, 3rd edition. Baillière Tindall, London, 437 pp.

Fraser, D., 1995. Science, values and animal welfare: Exploring the 'inextricable connection'. *Animal Welfare* 4: 103-117.

Fraser, D., J.S.D. Ritchie & A.F. Fraser, 1975. The term 'stress' in a veterinary context. *British Veterinary Journal* 131: 653-662.

Fraser, D., D.M. Weary, E.A. Pajor & B.N. Milligan, 1997. A scientific conception of animal welfare that reflects ethical concerns. *Animal Welfare* 6: 187-205.

Griffin, J., 1992. The human good and the ambitions of consequentialism In: E.F. Paul & F.D. Miller Jr. (Eds), The Good Life and the Human Good. Cambridge University Press, Cambridge, pp. 118-132.

Hare, R.M., 1981. Moral Thinking. Its Levels, Method and Point. Clarendon Press, Oxford, UK.

Hughes, B.O., 1976. Behaviour as an index of welfare. In: Proceedings of the Fifth European Poultry Conference, Malta, pp. 1005-1081.

Kennedy, J.S., 1992. The New Anthropomorphism. Cambridge University Press, Cambridge, 194 pp.

Lenthe, J. van, L. Hendrickx, C.A.J. Vlek & W. Biesiot, 1996. Using Decision Analysis for the Integrated Assessment of the Climate Change Problem. IVEM, Groningen, 111 pp.

Lorz, A., 1973. Tierschutzgesetz. C.H. Beck'sche Verlagsbuchhandlung, München, 272 pp.

Mason, G. & M. Mendl, 1993. Why is there no simple way of measuring animal welfare? *Animal Welfare* 2: 301-319.

McDowell I. & C. Newell, 1987. Measuring Health: A Guide to Rating Scales and Questionnaires. Oxford University Press, Oxford, 342 pp.

McFarland, D., 1989. Problems in Animal Behaviour. Longman Scientific & Technical, Harlow, 158 pp.

Mendl, M., 1991. Some problems with the concept of a cut-off point for determining when an animal's welfare is at risk. *Applied Animal Behaviour Science* 31: 139-146.

Moberg, G.P., 1985. Biological response to stress: Key to assessment of animal well-being? In: G.P. Moberg (Ed), Animal Stress. American Physiological Association, Bethseda, pp. 27-49.

Morton, D.B. & P.H.M. Griffiths, 1985. Guidelines on the recognition of pain, distress and discomfort in experimental animals and an hypothesis for assessment. *Veterinary Record* 116: 431-436.

Nagel, T., 1974. What is it like to be a bat? *Philosophical Review* 83: 435-450.

Regan, T., 1983. The Case for Animal Rights. University of California Press, Berkeley, 425 pp.

Rollin, B.E., 1990. The Unheeded Cry: Animal Consciousness, Animal Pain and Science. Oxford University Press, Oxford, 308 pp.

Rushen, J., 1991. Problems associated with the interpretation of physiological data in the assessment of animal welfare. *Applied Animal Behaviour Science* 28: 381-386.

Rushen, J. & A.M.B. de Passillé, 1992. The scientific assessment of the impact of housing on animal welfare: A critical review. *Canadian Journal of Animal Science* 72: 721-743.

Sambraus, H.H., 1981. Der Nachweis von Leiden bei Tieren. EU MT, Zürich, Switserland.

Sandøe, P., 1996. Animal and human welfare: Are they the same kind of thing? *Acta Agriculturae Scandinavica Section A, Animal Science, Supplement* 27: 11-15.

Sandøe, P. & H.B. Simonsen, 1992. Assessing animal welfare. *Animal Welfare* 1: 257-267.

Sandøe, P., M.H. Giersing & L.L. Jeppesen, 1996. Concluding remarks and perspectives. *Acta Agriculturae Scandinavica Section A, Animal Science, Supplement* 27: 109-115.

Singer, P., 1990. Animal Liberation, 2nd edition. Avon Books, New York, 301 pp.

Stafleu, F.R., F.J. Grommers & J. Vorstenbosch, 1996. Animal welfare: Evolution and erosion of a moral concept. *Animal Welfare* 5: 77-84.

Streiner, D.L. & G.R. Norman, 1995. Health Measurement Scales: A Practical Guide to their Development and Use, 2nd edition. Oxford University Press, Oxford, 231 pp.

Sumner, L.W., 1996. Welfare, Happiness and Ethics. Clarendon Press, Oxford, 239 pp.

Tannenbaum, J., 1991. Ethics and animal welfare: The inextricable connection. *Journal of the American Veterinary Medical Association* 198: 1360-1376.

Taylor, A.A., J.F. Hurnik & H. Lehman, 1995. The application of Cost-Benefit Dominance analysis to the assessment of farm animal quality of life. *Social Indicators Research* 35: 313-329.

Vorstenbosch, J., 1993. The concept of integrity. Its significance for the ethical discussion on biotechnology and animals. *Livestock Production Science* 36: 109-112.

Wemelsfelder, F., E.A. Hunter, A.B. Lawrence & T. Mendl, 1998. The inter- and intra-observer reliability of spontaneous qualitative assessment of pig behaviour. In: I. Veissier & A. Boissy (Eds), Proceedings of the 32nd Congress of the International Society for Applied Ethology. INRA, Clermont-Ferrant, p. 63.

Wiepkema, P.R., 1982. On the identity and significance of disturbed behaviour in vertebrates. In: W. Bessei (Ed), Disturbed Behaviour in Farm Animals. Hohenheimer Arbeiten 121. Verlag Eugen Ulmer, Stuttgart, pp. 7-17.

Chapter 4. Review part 2: Assessment tables and schemes[1]

Abstract

In the scientific literature several attempts have been made to systematically assess the overall welfare-status of animals in relation to housing and management. This chapter reviews assessment tables and schemes that have been constructed to this end. These tables and schemes have a tabular format that allows an assessment of housing systems using a list of welfare-relevant attributes (properties of the housing system). Rather than identifying deficits, the focus of this review is on finding positive recommendations for the purpose of developing a method for overall welfare assessment (OWA) on a scientific basis. The main recommendation is to use the tabular format as representation formalism for OWA. The concept of linked tables provides the key to performing OWA on a scientific basis in an explicit and systematic way.

Keywords: assessment tables, model.

Introduction

Much has been written about assessment of animal welfare. The main focus has been either on showing that animals have a welfare status (e.g. Rollin, 1990) or on how scientific measurement(s) may say something about welfare (e.g. Broom & Johnson, 1993). These studies have generally stopped short of putting theory to practice. Relatively few publications exist where authors have actually attempted to perform overall welfare assessment (OWA), i.e. to give a quantified, 'objective' judgement about the overall welfare-status of animals in relation to housing conditions. What has been published hasn't received much attention, maybe because quantifying welfare is considered to be subjective and impossible. We are presently developing a method to perform OWA in an explicit and systematic way on the basis of scientific knowledge. In the previous chapter (Bracke *et al.*, 1999a) we explained why we are optimistic and believe that scientific OWA is possible. In this chapter we will review papers in which an attempt to perform OWA is made. We will focus on papers that contain assessment tables and schemes. The aim is to extract recommendations for further development of a procedure to perform OWA in a more objective way.

Overview of assessment tables and schemes

The literature on OWA is diverse. Some authors provide a practical assessment tool that can be applied on farms (Bartussek, 1986; Bock, 1990; Sundrum *et al.*, 1994; Beyer, 1998). Bartussek's TGI (Tiergerechtheitsindex) is even used in actual political decision-making in Austria (Bartussek, 1999). Other papers are more theoretical, containing only an attempt to provide a methodological framework for more objective assessment of animal welfare (Mellor

1 Paper by Bracke, M.B.M., J.H.M. Metz & B.M. Spruijt, 1999. Overall animal welfare reviewed. Part 2: Assessment tables and schemes. *Netherlands Journal of Agricultural Science* 47: 293-305.

& Reid, 1994; Taylor *et al.*, 1995; Baxter & Baxter, 1984). Most authors focus on one species, mainly poultry (Duncan, 1978; Brantas, 1981; Hurnik & Lehman, 1988; Hughes, 1990; Appleby & Hughes, 1991; Nilsson, 1997) and pigs (Baxter & Baxter, 1994; Anon., 1985, 1989; Schlichting & Smidt, 1989; Konerman & van den Weghe, 1989; Svendsen & Svendsen, 1997). We also found papers on elephants (Kiley-Worthington, 1989) and pension horses (Beyer, 1998). Some publications cover more than one species (Sundrum *et al.*, 1994; Bartussek, 1986; Fraser, 1983). Others provide an assessment scheme for one type of housing system only (e.g. Nilsson, 1997, for hens in battery cages), or compare a number of specified housing systems for one category of animals (e.g. Brantas, 1981; Hurnik & Lehman, 1988, for laying hens; Svendsen & Svendsen, 1997, for pregnant sows). Some focus only on behaviour (Brantas, 1981; Fraser, 1983; Schlichting & Smidt, 1989), while others include a wider range of attributes such as health, physiology and management (Duncan, 1978; Anon., 1985, 1989; Nilsson, 1997). Not all authors actually present overall scores; those which do present overall scores include Duncan (1978), Brantas (1981), Fraser (1983), Kiley-Worthington (1989), Bartussek (1986), Sundrum *et al.* (1994) and Beyer (1998). If possible, it would seem preferable to calculate overall scores, but it requires additional assumptions as to how overall scores may be obtained from component scores.

While all authors attempt to provide a biological basis for OWA, their approach differs. Some have a designer's background (Baxter & Baxter, 1984; Bartussek, 1986). Others use economical theories, especially cost-benefit analysis (Mellor & Reid, 1994; Taylor *et al.*, 1995) or derive principles from the social sciences that strongly emphasise the use of statistical evaluation (Beyer, 1998).

In the various papers on OWA two main formats can be distinguished: assessment schemes and assessment tables. First, we will explain the tabular format of tables and schemes, then we will give a few examples.

The formats of assessment tables and schemes are rather similar: they both specify welfare-relevant attributes (Tables 1 and 2). They differ in that an assessment scheme is a generic tool that can be used for OWA, whereas an assessment table only illustrates OWA for a limited number of specified housing systems.

An assessment table (Table 1) is a matrix with welfare-relevant attributes specifying the rows of the matrix and housing systems specifying the columns. The cells of the matrix contain scores per attribute for every housing system. An attribute score is often expressed in terms of plusses and minuses, but may also be a numerical value. The overall-welfare score of a housing system is derived from its attribute scores with the help of some calculation rule, e.g. summation or calculation of the (weighted) average score.

Table 1. Format of an assessment table. The first column contains some examples of welfare-relevant attributes. The other columns represent housing systems (HS). In this example each attribute score ranges over a scale from – to +. The overall-welfare score is some function of the attribute scores (e.g. sum or average; average in this example).

Attributes	HS1	HS2	HS3
Space	-	+/-	+
Climate	-	-	+
Abnormal behaviour	-	+	+
Etc.	...	...	...
Overall welfare:	-	+/-	+

Table 2. Format of an assessment scheme. The first two columns contain some examples of welfare-relevant attributes and their levels. The last column contains attribute scores for every attribute level. In this table the attribute 'space' has 3 levels; the other attributes have 2 levels. Per attribute the score ranges from – to +. Per attribute exactly one level is true for any given housing system. The overall-welfare score is some function of the attribute scores (e.g. average or sum). When an average calculation rule is chosen a housing system with 2 m^2/animal, natural ventilation and high levels of abnormal behaviour would receive an overall-welfare score of '-'.

Attribute	Levels	Attribute score
Space	2 m^2/animal	-
	3 m^2/animal	+/-
	4 m^2/animal	+
Climate	natural ventilation	-
	thermocontrolled	+
Abnormal behaviour	high levels	-
	no abnormal behaviour	+
Etc.	...	...

Assessment schemes have a rather similar format, but they can be applied to all housing systems for which the scheme was designed. As a result the matrix of an assessment scheme lacks columns with specified housing systems. An assessment scheme (i.e. welfare model) is made up of a list of attributes, their levels (i.e. the possible values each attribute can take), their welfare-scores and a rule to calculate overall welfare from the attribute scores (Table 2).

In an assessment table Duncan (1978) compared the welfare status of laying hens in cages and pens with access to litter. He evaluated both systems with 9 items including health, physiology, behaviour, abnormal behaviour (feather pecking), management and production. Each item generates between 0 and 3 ticks, i.e. advantages, per housing system. He expressed overall welfare as the total number of advantages per system: 6 ticks for hens in cages and 8 for hens with access to litter.

During one year Brantas (1981) observed the behaviour of laying hens in three different housing systems: battery cages, litter with slats and get-away cages. Of the 37 behaviours that he considered relevant for welfare, Brantas ranked the means over the housing systems when they differed more than 10%. After explicitly excluding the use of additional weighting factors, he calculated rank totals, which showed that battery cages performed worse than either get-away cages or deep litter.

In an assessment table Schlichting & Smidt (1989) evaluated five different housing systems for fattening pigs using 10 items, which were assessed on a 5-point scale. The items were the main behaviour systems as well as the level of abnormal behaviour. Konerman & van den Weghe (1989) quantified health and hygiene for the same housing systems. These two papers together will be referred to as the KTBL assessment in the remainder of this article.

Assessment schemes were produced by Sundrum *et al.* (1994) for various farm animals, which they call the TGI-200 (Tiergerechtheitsindex, also called Animals Needs Index, Bartussek 1997) after Bartussek (1986). In these TGI schemes between 0 and 7 points are assigned to each of a number of specified attributes of a housing system. The TGI score is the overall sum of points, which has a maximum of 200 points. For example, for sows a pen size of more than 2.8 m^2/sow receives 7 points, but a pen size of 1.6 m^2/sow receives only 1 point. Other attributes include the number of floor types in the pen, access to pasture, trough width, nose rings and group size. In total 48 attributes are relevant for pregnant sows and these are organised into eight 'influence areas' ('Einflussbereiche'), which are movement, ingestion, social behaviour, rest, comfort/exploration, eliminative behaviour, hygiene and stockmanship.

The goal for OWA is to develop an assessment scheme. Application of this scheme to actual cases (housing systems) results in construction of an assessment table. Such a table is necessary for development and validation of an assessment scheme for OWA.

Terminology and other suggestions for standardisation

The various authors use different terms to identify and classify characteristics of housing systems that are relevant in OWA.

Generally used concepts to classify welfare-relevant characteristics of housing systems are either needs (e.g. Baxter & Baxter, 1984; Hurnik & Lehman, 1988) or freedoms (Appleby & Hughes, 1991; Mellor & Reid, 1994; Webster, 1995). Bartussek (1986) and Sundrum *et al.* (1994) use the term 'influence area' and also 'functional area' ('Functionskreis'). We prefer to use the term needs. In the next chapter we will explain why.

Concepts to identify characteristics of housing systems are more diverse. Several authors use the term 'indicators' (Fraser & Broom, 1990; Hughes, 1990). Taylor *et al.* (1995) use the term WRF (welfare-relevant factor), which they define as 'all housing and management variables relevant to or indicative of [animal welfare]' (Taylor *et al.*, 1995, p. 315). Beyer (1998) uses the term 'item' and Baxter & Baxter (1984) use the terms 'performance requirement', 'performance criteria', and 'performance specifications'. Also in use are the terms 'parameter' (Brantas, 1981), 'aspect' (Appleby & Hughes, 1991) and '(dis)advantages' (Duncan, 1978; Appleby & Hughes, 1991; Svendsen & Svendsen, 1997). The diversity of these terms is potentially confusing and further developments in the field of OWA would most certainly benefit from a unified terminology.

We propose to use the term 'attribute'. This term is adopted from the literature on conjoint analysis, which is a statistical tool for multivariate data analysis (Hair *et al.*, 1995) to evaluate the quality of a product. The term 'attribute' has been applied in the context of welfare assessment by den Ouden *et al.* (1997). An attribute is a housing characteristic, which may also be a characteristic of the animals, e.g. 'pen size' and 'production'. An attribute has two or more levels, which are properties of housing systems. For example, the level of the attribute 'pen size' may be '5 m^2'; the level of 'production' may be '24 piglets/sow/year'. For every housing system exactly one level is true per attribute. In addition, an attribute may, but need not be relevant for welfare. Whether or not this is the case is part of the welfare assessment procedure. When it is relevant for welfare it is assigned welfare value, which we will call 'attribute score'. From these attribute scores an overall-welfare score may be calculated.

Besides terminology, standardisation would also be welcomed in two further respects.

First, one scale should be used to present overall scores. Authors differ widely in how they present overall scores. For example, Fraser (1983) calculates the Behavioural Deprivation Index in percentages; Bartussek (1986) equates optimal welfare with 37 points; Sundrum *et al.* (1994) equate optimal welfare with 200 points. Since each scale can logically be transformed into a numerical scale, comparison between authors would improve when overall scores were expressed in a standardised way, e.g. on a scale from 0 to 10.

Secondly, constraints should be internal to an assessment scheme. Assessment schemes must set constraints that specify the class (domain) of housing systems for which they are designed. Constraints may concern the category of animals (e.g. Sundrum *et al.*, 1994) or the type of housing system; for example only battery cages (Nilsson, 1997). Constraints may be external or internal to the assessment procedure. An example of an external constraint is that legal constraints must be met before the evaluation scheme is used (e.g. Taylor *et al.*, 1995). The assessment schemes by Sundrum *et al.* (1994) contain the external constraints that housing systems with fully slatted floors and completely-outdoor systems are excluded, and that it is not allowed to calculate an overall score when the sum of points for one of the eight 'influence areas' is 0. By contrast, in Nilsson's scheme legal requirements have been incorporated in the list of attributes as constraints, i.e. as condemnation variables (i.e. minimum requirements). When applying an assessment scheme, external constraints may be overlooked and, therefore, we recommend using constraints that are internal, i.e. incorporated into the list of attributes, because this prevents unintended misapplication of the tool.

Further evaluation of assessment tables and schemes

Below we will evaluate some further aspects of the assessment procedure. These are the quality of the output, calculation, weighting and scaling, and the list of welfare-relevant attributes used to assess welfare.

Output of OWA

Constraints that restrict the application of the assessment tables and schemes to a specific category of animals or a specific set of housing systems, may help simplify the task of designing the assessment procedure. On the other hand, Fraser (1983) and Bartussek (1986) each present a scheme with application to a wider range of housing systems and species. This approach has intuitive appeal because it unifies the approach across species. However, this approach may also be limited in the ability to accommodate welfare requirements that are specific to certain categories of animals or to certain types of housing systems.

The value of the overall results depends, among other things, on the range of housing systems that has been assessed. Some authors only include a few housing systems (e.g. Duncan, 1978; Hurnik & Lehman, 1988). This reduces the value of the table, especially when the overall results also lie close together. With the exception of Duncan, authors who produced overall scores tended to generate clear distinctions between housing systems. Incorporation of a larger number of housing systems, especially when these differ widely with respect to both attributes and overall welfare, will improve the quality of the assessment. Beyer (1998) even statistically quantified the relationship between housing systems and welfare by calculating so-called z-values. A z-value of −1.5, for example, means that a score obtained for a housing system lies 1.5 standard deviations below the mean. Statistical analysis should be used in OWA, but it may be difficult to find a proper set of reference housing-systems to make this approach valid. Our point to include a wider range of housing systems in an assessment is not trivial. For, whereas the statement 'the welfare score is 7' is virtually meaningless, it becomes meaningful when it is situated in the context of a set of scores for different housing systems, which range, for example, from 3 to 10.

Several authors provide cut-off points or lines below which welfare is considered to be unacceptably low. Bartussek (1986) sets the level at 21 out of 37 points. Nilsson's scheme has politically-set % levels which increase over time. Fraser (1983) compared welfare across species and stated that a behavioural deprivation index of 25% results in aberrations in behaviour and that a 50% reduction is 'clearly stressful' (p. 16). Appleby & Hughes (1991) incorporated Duncan's idea (Duncan, 1978) of a welfare plateau into their assessment cube.

They set this plateau at 2/3 of the sum of welfare value contributed by each of three equally important attributes, namely enrichment, group size and density.

Cut-off points have been criticised for being arbitrary and subjective (Mendl, 1991). Furthermore, the concept of a cut-off point is ambiguous. It may indicate the level at which the welfare status is (very) low or it may indicate the level at which the welfare status is considered to be ethically unacceptable. Taken as an acceptability line the concept of a cut-off point falls outside the scope of OWA taken as a descriptive activity (Bracke *et al.*, 1999a). When taken in the first sense cut-off points establish distinct classes of welfare from what is actually a continuous variable that ranges from very good to very poor (e.g. Broom & Johnson, 1993). Using cut-off points in this way may be necessary for practical reasons. However, since such points are inherently arbitrary, we recommend that scientists abstain from drawing up such classes if possible.

Calculation, weighting and scaling

Whenever an overall score is calculated some calculation rule is employed to derive this score from the attribute scores. Constraints that specify, in the case of assessment schemes, when the scheme can be used to calculate welfare, have been discussed above. Two further aspects involved in calculation are weighting factors and interactions.

An interaction exists when the contribution of one attribute score to overall welfare depends on one or more other attribute-scores. In OWA interactions may be the rule, rather than the exception. For example, in pigs the value of a wallowing pool depends on the environmental temperature, and the value of rooting substrate depends on the feeding regime. In the papers reviewed here we have found no apparent suggestions as to how interactions may be handled in OWA. Dealing with interactions remains an issue that requires further attention.

Some suggestions have been made with respect to the use of weighting factors. The constraints (see above) may include considerations of weighting. They may specify minimum requirements before a more quantified approach to welfare is deemed acceptable. Weighting factors varying between 3 and 10 are explicitly used in Nilsson's scheme. In the other papers weighting factors are set at 1, either explicitly (e.g. Brantas, 1981) or implicitly. Beyer (1998) points out that weightings are also affected by the number of items, i.e. attributes or rows in the table. She uses this strategy to increase the importance of roughage in her scheme (p. 39). Another example can be found in Kiley-Worthington (1989), who incorporates 4 different attributes (out of 14) about space, but only one for the combined attribute 'food, water and shelter'.

All authors who presented overall scores used additive calculation-rules: (weighted) component scores were added to determine the overall score. Other ways to calculate overall scores include using a multiplicative rule and using interpretative skills rather than calculation. For example, Mellor & Reid (1994), who did not calculate overall scores, proposed an assessment scheme in which the overall score is to be interpreted from a set of five component scores, which are the five freedoms. They suggest setting the overall score at least as low as the lowest component score. This suggestion is in accordance with certain intuitive reasoning about welfare (cf. Maslow, 1970). However, it may have counterintuitive implications when taken too far, e.g. when it were taken to imply that the most negative feeling ever experienced by an animal would define its welfare status. Our main point here is that OWA may involve other than additive calculation rules.

Taylor *et al.* (1995) are sceptical about weighting and argue, instead, for an economic theory called Cost Benefit Dominance. This theory relies only on the ranking of housing systems within attributes. One housing system is better than another system when there is complete dominance of all attributes of that system over the alternative system. However, as Taylor *et al.* recognise, this theory runs into practical problems because systems will rarely be better in all aspects with respect to welfare. This is also confirmed by the other assessment schemes and tables reviewed here. Taylor *et al.* make several suggestions to solve this theoretical difficulty, but each of these reintroduces (aspects of) the weighting problem.

The ranking of levels within attributes is the basis of OWA (Brantas, 1981; Taylor *et al.*, 1995). It has substantially more objective validity than the consequent weighting and calculation of overall welfare. This ranking requires that attributes apply generally, i.e. across housing systems. It is because of this requirement, that the tabular format is especially suitable for OWA. All assessment tables and schemes reviewed here have a tabular format and employ only attributes that apply across housing systems. The attribute scale should range between the worst and best possible conditions within the constraints. This is made most concrete when the set of housing systems in the assessment table cover the full range for every attribute (as is done by Brantas, 1981; Hurnik & Lehman, 1988; Hughes, 1990). In any case it is helpful when the range is specified (e.g. that the scale ranges from 1, worst, to 5, best). Failure to do so (e.g. Duncan, 1978; Svendsen & Svendsen, 1997) complicates evaluation of the results. Furthermore, the scale should have a neutral mid-point, e.g. a 3-point scale or a 5-point scale. Beyer (1998) takes this one step too far. She requires that existing housing systems be distributed normally over the attribute scale. This makes welfare too relative. Although an assessment of welfare does depend on the range of levels the attributes can take (i.e. the domain of the assessment scheme), welfare does not depend on the number of housing systems that happen to have certain attributes (characteristics). A final remark about attribute scales is that they should only reflect local information about the attribute and its levels. Several authors also take into account considerations concerning the weighting of different attributes against each other and in relation to overall welfare. This is expressed in an assessment table or scheme in that for one attribute the scale ranges, for example, from - - to + and for another attribute it ranges from - to + + +. We suggest that such considerations should be specified explicitly in the calculation rule and in setting weighting factors, and not at the level of the attribute scores. The assessment of attributes should focus on the relationship between truth-values (what is true) and its scores related to welfare. For example, the assessment of the level of stereotypies should focus on how increases in stereotypies are evaluated on its own sub-scale. The reason why attribute scores should be set locally is that in an explicit assessment procedure every assumption should be open for criticism. The ranking of levels within attributes has a rather firm basis, whereas the weighting of attributes against one another is much more hypothetical. For this reason, considerations of weightings across attributes and concerning the calculation of overall scores should be formulated as separate assumptions and not be mixed with assignment of attribute scores.

In assigning attribute scores authors of assessment tables generally were not explicit about the kind of scales they used, but authors of assessment schemes tended to use linear scales for the relation between truth-values and attribute scores (e.g. Beyer, 1998; Bartussek, 1986; Sundrum *et al.*, 1994; Nilsson, 1997). However, whether linear scales are appropriate remains to be shown.

Welfare relevant attributes describing housing systems

Attributes that are relevant for welfare include aspects of the environment, behaviour, health and physiology (Duncan, 1978; Hughes, 1990). The first aspect concerns design criteria; the latter three aspects are animal-based attributes, also called performance criteria (Baxter & Baxter, 1984; Rushen & de Passillé, 1992).

Beyer (1998) identified lying and feeding facilities, surrounding-building, stockmanship and outdoor exercise as relevant attributes for welfare assessment of pension horses using factor analysis. Although factor analysis can be a useful tool to find relationships between attributes, a risk associated with the use of factor analysis is that too much emphasis may be placed on contingent correlations at the expense of biological relations. By contrast, Bartussek (1986) classified environmental attributes according to the contact points with the animal: space, conspecifics, floors, air and stockmanship. Although his scheme lacks important resources such as food and water, it is, nevertheless, an interesting idea that relates to the concept of skin-lesions as a measure of welfare (cf. Ekesbo, 1981). Similarly, in Nilsson's scheme (1997) welfare is assessed according to three extending circles: the animal, the pen, the building (L. Keeling, personal communication). Such a logical ordering according to contact points between environment and animals, provided it retains the biological meaning of resources to the animal, seems a more reasonable way to organise environmental attributes.

Animal-based attributes can be ordered according to the types of response, which the animal has available to interact with the environment, namely behavioural and physiological. Physiology includes 'normal' physiology, which includes all states where homeostasis is maintained, and patho-physiology or pathology. Pathology can be organised hierarchically according to the specific diseases that have been described, which can be organised according to the main physiological systems involved, e.g. respiratory, urogenital, digestive, nervous and metabolic disorders. These same systems also organise the responses of 'normal' physiology. Production parameters are a subclass of physiological parameters that concern mainly aspects of metabolism and reproduction. Stress-physiology is a class of physiology that is particularly relevant for welfare. It concerns situations that involve attempts to cope with situations of reduced predictability and controllability. Stress-physiology may be regarded as an intermediate between 'normal' physiology and pathology.

Having a complete list of welfare-relevant attributes is a necessary condition for overall assessment, i.e. the attributes in a given assessment procedure must adequately cover all main fields of welfare. For this reason the health status, physiological requirements such as respiration, osmoregulation, nutrition and thermoregulation, as well as behavioural opportunities are necessary components of welfare assessment. Several tables and schemes seem to meet this criterion, e.g. Mellor & Reid (1994), Hurnik & Lehman (1988), Nilsson (1997) and the KTBL assessment. For other authors this is less obvious. For example, Beyer (1998) and Sundrum *et al.* (1994) mainly focussed on design criteria for practical reasons and the extent such an approach allows adequate assessment of performance criteria is not easily answered.

In a full specification of an assessment table for OWA both truth-values and attribute scores should be given for every attribute. This requires a detailed description of every housing system. In some papers the systems were hardly described, but Svendsen & Svendsen (1997) included simple drawings of the pen-layout and the working group on pig housing (Anon., 1985, 1989) included pen-layouts together with a detailed table describing the systems. The tabular format has advantages for this purpose too, because it forces to be explicit and systematic.

A further advantage of using the tabular format may be realised when tables are linked as in a relational database (Date, 1995). Linked tables may allow making various sources of information available for welfare assessment. One table, in which housing systems are described, can be linked to the assessment table by the names of the housing systems. Similarly, the table for overall assessment can be linked to supporting tables by the attributes (e.g. Fraser, 1983; Sundrum *et al.*, 1994). Konerman & van den Weghe (1989), who assess the health status in relation to housing systems for fattening pigs, also provide an illustration how this is done. Their final table contains three attributes: infection pressure, claws and limbs and (health) control. Each of these is assessed in a separate table. For example, (health) control is assessed in a separate table from the attributes group size, accessibility/reachability, building/unit size and (age) uniformity, and the overall results of this table are used as a component in the final table. In this way linked tables can support defining operational definitions of compound concepts.

Finally, tables may be linked to support formalising the relationship between welfare and its scientific basis. When scientific knowledge is collected in a table, attributes can provide the link to the assessment table. This is a very important feature, as it may support providing an explicit scientific basis for OWA. No paper reviewed here does this, and, as a result, the scientific basis remains exemplary and mostly unspecified. We conclude that linking of tables may support the construction of a formalised, i.e. explicit and systematic, procedure for OWA in all aspects of its problem space, namely the description of housing systems, welfare values and the scientific basis. Doing so within the framework of a relational database allows securing data integrity (Date, 1995) and it allows dealing with large amounts of data. This seems required for OWA, because welfare is a complex problem, which depends on many factors (Dawkins, 1997). In the papers discussed here, all tables and schemes are limited in size; the number of rows is maximally 48 (Fraser, 1983). Overall welfare assessment that aspires approaching the ideal of taking into account all available data (Duncan, 1978; Rushen & de Passillé, 1992), therefore may benefit from database technology.

Recommendations

In OWA an attempt is made to descriptively assess the overall welfare status of animals from what is known about their biology and about their living conditions, i.e. about attributes of the housing system in which they are kept. OWA involves a problem of multi-criteria decision making with fuzzy information. The main task for OWA is to increase the degree of objectivity involved in making overall-welfare judgements. The best way to do this is to make all steps between the attributes of the housing system and welfare explicit and perform them according to some systematic procedure. Only after the entire procedure has been made explicit will it be possible to criticise assumptions and systematically search for improvements. The tabular format seems to be a suitable tool for making OWA explicit and systematic (Webster, 1995). In this format housing systems (in columns) can be assessed in a systematic and analytic way according to a list of welfare-relevant attributes (in rows). We have used this format successfully in representing the arguments pig experts use to explain their scores for welfare in relation to housing conditions (Bracke *et al.*, 1999b).

Some minor recommendations include standardisation of terminology and scales. We suggest using the term 'attributes' and a scale between 0, worst and 10, best. Constraints should be specified and preferably be stated within the table to avoid erroneous application.

On the other hand, cut-off points that specify what level of welfare is still acceptable, should not be given, because scientific OWA is a descriptive activity that is logically distinct from ethical assessment.

What is important for OWA is to specify the relations between welfare, housing and scientific knowledge.

The format of an assessment table focuses on the relation between welfare and designated housing systems. To develop a scheme for OWA an assessment table should be constructed which includes a number of housing systems that cover a wide range of animal environments. The assessment of these housing systems can provide useful reference scores to facilitate the interpretation of newly obtained scores for other housing systems.

Much work remains to be done concerning the weighting of attributes and calculation rules. So far, additive calculation rules that did not use additional weighting factors were used most often. The basis for calculation is the ranking of the levels within attributes. Therefore, in OWA the attributes should apply across housing systems and attribute (welfare) scores should be assigned from a local perspective, i.e. without taking into account considerations concerning the weighting between attributes. The weighting of attributes and the calculation rule, which is needed to calculate overall welfare from attribute scores, should be specified separately. Because weighting is more problematic than the ranking within attributes, sub-scores should be retained and presented in addition to the overall scores.

The relationship between attribute scores and descriptive properties of the housing systems should be made explicit. This may be done by providing a full description of the housing systems, including, for example, drawings of the pen-layout, as well as by linking tables. Linked tables, as in a relational database, may also support assessment of compound attributes and support assessment on an explicit scientific basis. Database technology also allows that many attributes are taken into account. However, some logical and/or biological ordering principle is needed to ensure that welfare is assessed overall, that welfare is assessed on an objective and scientific basis while retaining the idea that the point of view of the animals is definitive of its welfare status. How this can be done will be discussed in the next chapter.

The main conclusion of this chapter is that the tabular format and the concept of linked tables are necessary to solve the main difficulty in OWA, namely to evaluate overall welfare in a procedural and explicit way that is open for criticism and, most importantly, allows further improvement.

References

Anonymous, 1985. Barn Types for Fattening Pigs: An Evaluation (in Dutch). IMAG, Wageningen, 65 pp.

Anonymous, 1989. Barn Types for Sows and Piglets: An Evaluation (in Dutch). IMAG, Wageningen, 80 pp.

Appleby, M.C. & B.O. Hughes, 1991. Welfare of laying hens in cages and in alternative systems: Environmental, physical and behavioural aspects. *World's Poultry Science Journal* 47: 109-121.

Bartussek, H., 1986. Ergänzung zum Artikel 'Vorschlag für eine Steiermärkische Intensivtierhaltungsverordnung'. Irdning (1985). *Der Österreichische Freiberufstierartzt* 97: 4-15.

Bartussek, H., 1997. Animals Needs Index ANI (Tiergerechtheitsindex TGI-35L) for assessment of animals well being in housing systems for Austrian proprietary products and legislation. In: J.A.M. Arendonk (Ed), Book of Abstracts of the 48th Annual Meeting of the European Association for Animal Production. Wageningen Pers, Wageningen, p. 190.

Bartussek, H., 1999. A rieview of the animal needs index (ANI) for the assessment of animals' well-being in the housing systems for Austrian proprietary products and legislation. *Livestock Production Science* 61: 179-192.

Baxter, M.R. & S.H. Baxter, 1984. How to delineate the sow's welfare needs? *Annales Recherches Véterinaire* 15: 281-286.

Beyer, S., 1998. Konstruktion und Überprüfung eines Bewertungskonzeptes für pferdehaltende Betriebe unter dem Aspekt der Tiergerechtheit Ph.D. Wissenschaflicher Fachverlag, Giessen, 135 pp.

Bock, Ch., 1990. Zur Beurteilung tiergerechter Laufställe für Milchvieh, KTBL-Schrift 339. KTBL, Darmstadt, 82 pp.

Bracke, M.B.M., J.H.M. Metz & B.M. Spruijt, 1999a. Overall animal welfare assessment reviewed. Part 1. Overall animal welfare assessment reviewed: Is it possible? *Netherlands Journal of Agricultural Science* 47: 279-291.

Bracke, M.B.M., J.H.M. Metz, B.M. Spruijt & A.A. Dijkhuizen, 1999b. Overall welfare assessment of pregnant sow housing systems based on interviews with experts. *Netherlands Journal of Agricultural Science* 47: 93-104.

Brantas, G.C., 1981. Interpretation of behaviour observations. In: L.Y. Sørensen (Ed), Proceedings of the First European Symposium on Poultry Welfare, Køge, Denmark, June 9-12, 1981. World Poultry Science Association, Copenhagen, pp. 45-49.

Broom, D.M. & K.G. Johnson, 1993. Stress and Animal Welfare. Chapman & Hall, London, 211 pp.

Date, C.J., 1995. An Introduction to Database Systems, 6th edition. Addison-Wesley Publishing Company, Reading, 839 pp.

Dawkins, M.S., 1997. D.M.G. Wood-Gush Memorial Lecture: Why has there not been more progress in animal welfare research? *Applied Animal Behaviour Science* 53: 59-73.

Duncan, I.J.H., 1978. An overall assessment of poultry welfare. In: L.Y. Sørensen (Ed), Proceedings of First Danish Seminar on Poultry Welfare in Egglaying Cages. National Committee for Poultry and Eggs, WPSA, Copenhagen, pp. 81-88.

Ekesbo, I., 1981. Some aspects of sow health and housing. In: W. Sybesema (Ed), Current Topics in Veterinary Medicine and Animal Science, Martinus Nijhoff, The Hague, pp. 250-266.

Fraser, A.F., 1983. The behaviour of maintenance and the intensive husbandry of cattle, sheep and pigs. *Agriculure, Ecosystems and Environonment* 9: 1-23.

Fraser, A.F. & D.M. Broom, 1990. Farm Animal Behaviour and Welfare, 3rd edition. Baillière Tindall, London, 437 pp.

Hair, J.F. Jr., R.E. Anderson, R. Tatham & W.C. Black, 1995. Multivariate Data Analysis, 4th edition. Prentice Hall, Englewood Cliffs, 745 pp.

Hughes, B.O., 1990. Welfare in alternative housing systems for laying hens. Eigth European Poultry Conference. Feria de Barcelona, Barcelona, pp. 199-210.

Hurnik, J.F. & H. Lehman, 1988. Ethics and farm animal welfare. *Journal of Agriculture and Environmental Ethics 6, Supplement* 1: 305-318.

Kiley-Worthington, M., 1989. Ecological, ethological, and ethically sound environments for animals: Toward symbiosis. *Journal of Agricultural Ethics* 2: 323-347.

Konermann, H. & H. van den Weghe, 1989. Subkriterium "Tiergesundheit und Hygiene" In: C. Fuchs, J. Gartung & J. Hoeges (Eds), Haltungssysteme für Mastschweine. Vergleich, Bewertung, Verbesserungsansätze. KTBL-Schrift 335. KTBL e.V., Darmstadt, pp. 82-88.

Maslow, A.H., 1970. Motivation and Personality, 2nd edition. Harper & Brothers, New York, 411 pp.

Mellor, D.J. & C.S.W. Reid, 1994. Concepts of animal well-being and predicting the impact of procedures on experimental animals. In: R.M. Baker, G. Jenkin & D.J. Mellor (Eds), Improving the Well-being of Animals in the Research Environment, Proceedings. ANZCCART, Sidney, pp. 3-18.

Mendl, M., 1991. Some problems with the concept of a cut-off point for determining when an animal's welfare is at risk. *Applied Animal Behaviour Science* 31: 139-146.

Nilsson, C., 1997. Dispensprogram för medgivande av undantag från kravet att höns för äggproduktion inte får inhysas I andra burar än sådana som uppfyller hönsens behov av rede, sittpinne och sandbad. Beslut 1997-08-28. Jordbruks Verket, Jönköping, Sweden.

Ouden, M. den, J.T. Nijsing, A.A. Dijkhuizen & R.B.M. Huirne, 1997. Economic optimization of pork production-marketing chains I. Model input on animal welfare and costs. *Livestock Production Science* 48: 23-37.

Rollin, B.E., 1990. The Unheeded Cry: Animal Consciousness, Animal Pain and Science. Oxford University Press, Oxford, 308 pp.

Rushen, J. & A.M.B. de Passillé, 1992. The scientific assessment of the impact of housing on animal welfare: A critical review. *Canadian Journal of Animal Science* 72: 721-743.

Schlichting, M. & D. Smidt, 1989. Subkriterium Tierverhalten. In: C. Fuchs, J. Gartung & J. Hoeges (Eds), Haltungssysteme für Mastschweine. Vergleich, Bewertung, Verbesserungsansätze. KTBL-Schrift 335. KTBL e.V., Darmstadt, pp. 71-88.

Sundrum, A., R. Andersson & G. Postler (Eds), 1994. Der Tiergerechtheitsindex-200/1994 - Ein Leitfaden zur Beurteilung von Haltungssystemen für Rinder, Kälber, Legehennen und Schweine. Verlag Köllen, Bonn, 211 pp.

Svendsen, J. & L.S. Svendsen, 1997. Intensive (commercial) systems for breeding sows and piglets to weaning. *Livestock Production Science* 49: 165-179.

Taylor, A.A., J.F. Hurnik & H. Lehman, 1995. The aplication of Cost-Benefit Dominance analysis to the assessment of farm animal quality of life. *Social Indicators Research* 35: 313-329.

Webster, J., 1995. Animal Welfare: A Cool Eye Towards Eden. Blackwell, Oxford, 273 pp.

Chapter 5. Review part 3: Welfare assessment based on needs and supported by expert opinion[1]

Abstract

Welfare concerns what matters to animals from their point of view. What matters to animals is their state of need. Satisfaction and frustration of needs are associated with emotional states, the subjective experience of which directly determines the welfare status of an animal. Because emotional states are difficult to assess, overall welfare assessment (OWA) is best approached as an assessment of needs.

For actual OWA a list of needs must be formulated. Different authors have formulated different lists. From these lists a concept need-list was constructed. For validation the needs-based approach for OWA was discussed in interviews with 21 experts in the field of ethology and other welfare-related sciences. The experts generally used mental terminology to define welfare, but when asked to classify their definition of welfare, many preferred a definition in terms of measurable parameters or a combination of both mental terms (feelings) and measurables. Most experts believed that welfare can be assessed objectively and that the problem of OWA is indeed best approached through an assessment of needs. Experts differ as to the exact composition of the list of needs. A list of needs is formulated which we intend to use for OWA in the case of sows.

Keywords: interviews, expert, animal welfare assessment model, pigs.

Introduction

This chapter is the third and last of a series of three reviews on the topic of overall farm-animal welfare assessment (OWA). These chapters highlight three areas of concern.

The first (Bracke *et al.*, 1999b) dealt with the methodological question whether OWA is possible in principle. We defined welfare as what matters to animals from their point of view. This means that the welfare status of an animal is fully determined by the quality of its emotional states, including their sign (positive or negative), intensity and duration. A major problem is how the private minds of animals can be assessed by scientific methods (Nagel, 1974; Dawkins, 1993, 1998; Mason & Mendl, 1993). We suggested to regard OWA as the attempt to make the best possible assessment based on what is known scientifically. Accordingly, OWA is conceived as the descriptive activity that involves multi-criteria decision making with fuzzy information. Based on these considerations we believe that OWA is possible, but the question remains how it should be done in a systematic and explicit way.

The second review chapter (Bracke *et al.*, 1999c) reviewed assessment tables and schemes that have been published to find useful recommendations as to how OWA may be performed. The basic format for OWA was identified as a table in which housing systems (in columns) are compared and evaluated using a list of criteria (welfare-relevant attributes, in rows). This assessment table must be linked with other (supporting) tables to make OWA fully explicit.

1 Paper by Bracke, M.B.M., B.M. Spruijt & J.H.M. Metz, 1999. Overall animal welfare reviewed. Part 3: Welfare assessment based on needs and supported by expert opinion. *Netherlands Journal of Agricultural Science* 47: 307-322.

The present chapter deals with the problem of how OWA can be performed on a scientific basis. The suggestion is that overall welfare can be assessed from an assessment of needs. We have applied this idea and constructed a prototype decision support system, which is a computer-based information system that was designed to examine the feasibility of performing OWA in a systematic way (Bracke *et al.*, 1999a). Because performing OWA on the basis of an assessment of needs is a very basic assumption, we conducted interviews to examine the degree of consensus for this assumption. The results of these interviews will be presented in this chapter. The aims of this chapter are to specify how overall welfare assessment (OWA) can be performed on the basis of biological needs, to examine the degree of consensus among experts for such a needs-based approach and to specify a list of needs that may be used for actual OWA in pigs.

A biological basis for welfare

Biological organisms are regulated by homeostatic control mechanisms, which support survival and reproduction in the environment of evolutionary adaptation (EEA). The higher vertebrates are goal-directed (Toates, 1986). They have a number of more or less distinct motivational systems. These systems, which we call needs, can be thought of as intervening variables which have functionally related sets of behaviours or physiological responses that can be activated by a certain class of stimuli and deactivated by a specific event or behaviour. Classical examples of motivational systems are hunger, thirst, sex and thermoregulation. Each motivational system serves a proximate goal (reference point, set point or Sollwert). These goals have been formed in the course of evolution and are, therefore, strongly similar for individuals of the same species (Wiepkema, 1987). In order for responses to be functional in achieving the goal the animal surveys its environment and compares Sollwert (the goal) and Istwert (the actual state of the world). Discrepancies between Istwert and Sollwert cause activation of behavioural and physiological responses with the aim of reducing the discrepancy and restoring homeostasis. The degree of the discrepancy may vary (i.e. the animal may be more or less hungry, thirsty etc.), and changes in this 'central motivational state' (Toates, 1986), which we call 'state of need', give rise to observable changes in behaviour and physiology (e.g. when more hungry the animal may run faster toward food).

Higher vertebrates show certain flexibility in the way they achieve their goals, i.e. they can take different courses of action to obtain a certain goal. This flexibility also requires that animals must monitor the effectiveness of their responses. Emotions, such as pleasure and fear, are functional elements in this monitoring process, in that they strengthen (when positive) or weaken (when negative) the use of a particular behavioural or physiological response (Fraser & Duncan, 1998). Emotions are causally related to behaviour (e.g. Dawkins, 1993; Broom, 1998). They function as signals in the brain to coordinate the responsiveness in a certain direction. These signals produce a coordinated state, which is generally appropriate for coping with categories of challenges (e.g. danger). Such a state is recognised by the animal. For example, the higher vertebrates are able to recognise these internal emotional states as is shown by their ability for so-called drug-discrimination learning, which is a much used technique to investigate the subjective effects of drugs on the control of behaviour (Overton, 1991).

Positive reward occurs in cases where discrepancies between Istwert and Sollwert are reduced or minimised. Animals are attracted to objects and events associated with positive reward (positive emotions). Events or stimuli that are rewarding act as reinforcers, in the sense that they tend to strengthen a response. Conversely, negative emotions (aversions) occur when discrepancies between Istwert and Sollwert increase in magnitude or continue to exist. Animals will attempt to avoid such situations, which they find aversive. Especially negative is stress, i.e. '(a) a protracted failure of the animal to maintain alignment between its reference values and the actual state of the world and (b) the absence of an assessment of near-future realignment' (Toates, 1995, p. 31).

Because animals have different needs and because they often cannot serve different goals at the same time, animals also have regulatory mechanisms to solve cases of conflict between them. For example, an animal may have to choose between food and escape from danger. Animals are generally very well able to make 'decisions' among different possible courses of action. Such decisions involve a cost-benefit evaluation, which requires a common currency (McFarland, 1989). Most likely, this common currency is the rewarding value that represents the expected benefits of each alternative course of action. In this model, animals, like humans, are supposed to act so as to maximise positive affective states and minimised negative ones. They maximise reward (Cabanac, 1971). This implies that animals themselves assess their different states of need and this overall assessment constitutes their welfare.

OWA based on needs

From the argument above it follows that for overall welfare assessment (OWA) the various states of need of an animal must be assessed and integrated as much as possible in the way the animals themselves perform the assessment. Below, we will specify the concept of 'needs' further.

An important distinction is between instrumental and intrinsic relevance. Instrumentally relevant are those aspects that are a means to the end of OWA; i.e. they are relevant because they give information about other aspects, which are more intrinsically relevant. Intrinsic aspects are ends in themselves. For example, straw is instrumental for pig welfare, because it provides substrate to root. Rooting is intrinsically relevant when rooting is itself rewarding. Rooting would be instrumentally relevant if it were only a means to an end, e.g. a means to obtain food. In the latter case straw could be regarded as intrinsically relevant if it had a dietary value for pigs. The primary task for OWA is to determine what is intrinsically relevant for welfare and how these are affected by other aspects in an instrumental way.

Intrinsically relevant for welfare are all and only the emotional states of animals (Bracke *et al.*, 1999b). However, it is difficult to assess the sign (positive or negative), intensity and duration of all emotional states separately. More suitable for an objective assessment of welfare is the assessment of needs. If emotions represent a state of the organism which has a biological function for a particular need, that state must be accessible for measurement and this indirectly reveals an aspect of the subjective state of the animal. For operationalisation of OWA we postulate a (positive) (cor)relation between biological functioning and subjective welfare. This postulate receives general support in the scientific literature (e.g. Broom, 1998; Duncan, 1993; see also Fraser *et al.*, 1997), but exceptions (which we will discuss below) have also been recognised. Under the assumption that the state of need is a direct reflection of how animals subjectively experience this state emotionally, we may regard needs as intrinsically relevant in the assessment of welfare. They can be assessed objectively and this provides the scientific basis for OWA.

It follows that only proximate needs are intrinsically relevant for welfare. Welfare concerns the proximate causation of behaviour, rather than its ultimate function. Ultimate goals such as survival and reproduction have shaped proximate needs in the course of evolution, but survival and reproduction per se do not matter to animals from their point of view (Duncan & Petherick, 1991). For example, what matters to a female animal in oestrus is a proximate need to mate, rather than the ultimate goal of fertilisation. Because our concept of needs for OWA is closely linked to emotional states, it is very similar to, but more general than, animal 'wants' (cf. Duncan & Petherick, 1991; Duncan, 1996). It is also similar to Rollin's concept of Telos (e.g. Rollin, 1990), when conceived as the genetically and environmentally constrained nature of animals, 'from which flow certain interests and needs, whose fulfilment *matter* to the animal' (p. 203). For the same reason, our concept of needs differs from the concept of needs as suggested by Hurnik and collaborators (Hurnik & Lehman, 1985, 1988; Hurnik, 1993). For example, longevity may be instrumental in OWA, but it is not intrinsically relevant, because animals do not have the concepts of life and death (Webster, 1995, p. 15).

Motivational systems are complex systems. For example, reaching satiety is regulated by various internal and external feedback signals, including the energetic value of the food as well as oropharyngeal signals associated with palatability, chewing and swallowing. Needs, such as the need for food, can be often be de-composed further into component set-points. As a result, OWA can be regarded at a conceptual level (cf. Stafleu *et al.*, 1996) as a hierarchical assessment: welfare can be de-composed into a set of needs, which can be de-composed further into component set-point states, from which welfare can be assessed.

In the literature general consensus exists that an aggregate of several different measures should be used to assess welfare (e.g. Broom & Johnson, 1993). In fact many attributes (characteristics, aspects) of housing and management affect the welfare status of farm animals. In order to specify how these attributes affect welfare, a needs-based approach seems the most appropriate (cf. also Dawkins, 1998). It offers an organising principle and provides a way to check whether the list of welfare-relevant attributes is complete. In addition, a risk in OWA is that one component is taken for the whole (Rushen, 1991). The single most important reason to use a needs-based approach for OWA is that it supports assessment of welfare *overall*: it helps to identify welfare problems (frustrated needs) and it helps to identify gaps in our scientific knowledge (to assess the state of need properly).

Scientific paradigms

An assessment of the state of need includes an assessment of the degree of positive and negative reward, the animal's motivational strength to obtain those rewards and the duration of the relevant emotional states. Each need state can be assessed on a scale that ranges from maximum frustration to maximum satisfaction of that need. To assess these needs we must use information about environmental conditions, and empirical information from ethology and physiology (including production and patho-physiology). These can be regarded as different perspectives that provide relevant information to assess a state of need. In addition a subjective, psychological perspective can also be identified that specifies the nature of the emotional states. The psychological perspective and the various perspectives constituted by the empirical sciences are different perspectives on the same phenomenon, the state of need. The problem of OWA is to show how welfare, as defined from this psychological perspective at the conceptual level, can be assessed at the explanatory and operational levels (see Stafleu *et al.*, 1996) while using only empirical information. We suggest that the various scientific paradigms allow the formulation of assessment rules that can be used for this purpose.

Scientific measures relevant for OWA include feral data, preference tests (time budgets and choice experiments), operant techniques (including demand curves), measures of aversion and suffering, measures of the consequences of deprivation on behaviour, stress-physiology, pathology and production. These measures indicate what animals normally, naturally or experimentally are inclined to approach or avoid, how strong their preferences are, and how well animals are able to adapt or cope. Problems exist with the interpretation of all measures (Rushen, 1991; Mason & Mendl, 1993; Dawkins, 1998). For example, feral data may be criticised in that nature may be romantic but cruel (Dawkins, 1980); what animals chose may not always be what is best for their health (Duncan & Dawkins, 1983); and coping animals may still be suffering (Mendl, 1991). Problems of interpretation have been considered difficult to resolve. However, such problems typically involve cases where different scientific paradigms are in conflict with each other with respect to OWA.

We suggest taking a consensus-oriented approach for OWA. Despite the fact that much remains to be discovered, much knowledge that is relevant for OWA has been collected over the last decades. OWA concerns the attempt to make the best possible assessment based on the knowledge that is available (Bracke *et al.,* 1999b). We believe the available knowledge is sufficient to allow a reasonably accurate assessment. Similarly, despite difficulties in the interpretation of all scientific paradigms, each paradigm can be expected to capture at least part of the truth. This allows the formulation of assessment rules in relation to each paradigm. Every assessment rule includes a 'prima facie' clause indicating that the rule is valid for the most part and other things being equal. For example, prima facie, the more natural the behaviour, the better welfare (e.g. Wemelsfelder, 1997). Similarly for the other paradigms including predictability and controllability (Wiepkema, 1982, 1987; Wiepkema & Koolhaas, 1993), fitness (e.g. Fraser & Broom, 1990) and consumer demand theory (Dawkins, 1983).

However, from the conclusion above that OWA should be in accordance with welfare assessment as performed by the animals themselves it follows that of all scientific paradigms the study of preferences of animals takes a special place. For OWA we must answer the questions 'what do animals want, i.e. what do they find rewarding and/or aversive?' and 'How important is the satisfaction of these wants or needs for them?' Other scientific paradigms, e.g. studies of natural behaviour and stress-physiology have a more supportive function in that they provide additional information about proximate needs.

For every 'prima facie' assessment rule we also expect to find exceptions, which become evident when rules conflict. For example, the argument that nature may also be cruel constitutes a conflict between one rule that says that natural conditions indicate good welfare, and the second rule that says that disease indicates poor welfare. Nature is cruel when it subjects animals to disease. Such a conflict between assessment rules can be resolved when, in accordance with our definition of welfare, the primacy of the animal's emotional states is recognised: nature is good provided the animal doesn't experience poor health. Further specifications can be expected. For example, poor health indicates poor welfare, but an abdominal tumour may not be associated with negative emotional states. If so, again, the assessment rule should be refined. This revision of assessment rules into more and more specific rules may become very complicated and difficult. At some point we may have to stop formulating more and more specific rules. At such a point, these rules can be used as heuristic rules for OWA. As heuristic rules they will allow a most reasonable assessment of welfare despite the fact that some assessment errors will inevitably be made. Since large numbers of attributes are involved in OWA the use of heuristic rules may be the most rational approach to OWA until further research can provide a more complete set of specific assessment rules.

Types of needs

It is beyond the scope of this chapter to specify the different set points of animals, also because they tend to be very (species, age, sex) specific. However, we will attempt to specify which needs farm animals have.

Animals have many control systems that are designed to obtain or maintain a certain (local) goal or set point. Not all control systems are equally relevant for welfare, because they are not all equally associated with emotional states. Emotional states especially arise when the attention of the whole animal and a close monitoring of the efficiency of responses is required. We will call needs associated with such systems 'cognitive'. Other systems are more under autonomic regulation. Examples include many processes at cellular and tissue level, but also the immune system, the regulation of heart rate and respiration. These autonomic systems are either largely internally organised or only use rather stable environmental factors (such as oxygen). They don't require additional emotional states for normal regulation. However, when autonomic regulation fails, emotional states do occur even in these systems, e.g. by general symptoms of fatigue or illness. It follows that for welfare both types of control are relevant, but they are not relevant to the same degree. The tolerance for deviations between actual state (Istwert) and set point (Sollwert) is generally much lower where emotional states are involved which deal with (more fluctuating) environmental events. It is those needs that have an association with emotional states that are especially relevant and functional for welfare. However, although some needs are more important than other needs, they cannot be classified into necessities and luxuries, because their importance ranges over a continuum and because their relative importance may vary according to the circumstances.

Needs can also be classified into appetitive (e.g. hunger, thirst, sex) and aversive systems (e.g. fear and aggression) (cf. Toates, 1986). Appetitive needs have a special subclass of needs: the ethological needs. Ethological needs are those needs where the performance of behaviour is intrinsically rewarding, rather than, or in addition to, the attainment of some functional end-point that is normally associated with the performance of that behaviour. For example, rooting of pigs is itself rewarding, even when the normal consequences associated with this behaviour, i.e. food, is provided ad lib. Ethological needs concern activities that are essential in the environment of evolutionary adaptation (EEA). They are regulated by being positively rewarding, for example because the ultimate goal is beyond the cognitive capacities of the animal or because it would be disadvantageous to stop the behaviour in the absence of immediate functional consequences. According to Toates (1995) it is now generally accepted that animals indeed are motivated to perform certain species-specific behaviours (however see also Baxter, 1983) and that reward value is associated with the ability to perform these behaviours.

The above distinctions (appetitive-aversive, cognitive-autonomic and ethological) result in the following classification of needs. Appetitive cognitive needs include food, water, sex, rest and social contact. This class shows overlap with appetitive ethological needs such as exploration, play and body care related needs. Appetitive autonomic needs include thermoregulation and respiration. Aversive autonomic needs include health and no injury. Fear is an aversive cognitive need. This classification is tentative and provides an ordering principle, rather than an absolute classification.

Needs have also been classified into those needs that are largely internally motivated (e.g. food, water, ethological needs) and those that are largely externally motivated (e.g. aggression, predator avoidance). However, it is now generally recognised that all needs have both internal and external factors. Like the internal-external distinction the distinctions we use in our classification (appetitive-aversive; cognitive-autonomic; ethological) do not create mutually exclusive categories, but are differences that vary over a continuum and often (if not always) include elements of both extremes. Even within needs some elements may fit in one class, while other elements fit better in another class. For example, vasoconstriction and vasodilatation as part of thermoregulation are under autonomic control while nestbuilding, which is also part of thermoregulation, may be an ethological need. Although our classification is only tentative, we will use it as stepping stones that should not obstruct flexibility in the assessment procedure.

Interviews with experts

In order to perform actual OWA we have built a prototype welfare-model for pigs (more specifically for pregnant sows; Bracke *et al.*, 1999a). In this model we assessed overall welfare using a list of needs. To explore the degree of consensus for this model 21 experts from 8 different countries were interviewed about their concept of welfare and about needs. Together 11 Dutch experts, 9 experts from other European countries and one North-American expert were included with expertise in the fields of general farm animal welfare, fundamental ethology, physiology, veterinary science and experimental psychology (fields listed in the order of importance). All experts were scientists.

Issues that specifically concern weighting of welfare components were explicitly excluded from these interviews, because our aim was to explore the degree of consensus for the assumption that logically precedes weighting, namely using needs for OWA. We wanted to know whether our concept of welfare was in accordance with expert opinion, and whether OWA may be performed as a function of need states.

Three questions were asked about the concept of welfare and two questions were asked about needs.

1. How do you define welfare?
2. What type of welfare definition do you favour, in terms of measurable parameters or in terms of feelings?
3. Can welfare be assessed objectively?
4. What are the components of welfare and what is your opinion about the prototype need-list (which was presented to the expert, cf. Table 1)?
5. Do you believe OWA based on an assessment of needs is the proper way to proceed?

In response to the first question relatively few experts cite definitions from the literature: five experts cited three definitions. A common characteristic of the definitions as stated by the experts is that they all make reference to biological functioning. Furthermore, the importance of subjective feelings showed as follows: 18 out of 21 experts used mental terminology, 2 experts refused to give a definition of welfare and only 1 expert completely refrained from using mentalistic terminology in his first stated definition. However, when asked what type of definition was preferred (question 2), 9 experts favoured a definition in terms of measurable parameters over subjective feelings; 8 experts preferred a combination of both feelings and measurable parameters, and only 4 experts favoured feelings. 18 Experts answered that welfare can be measured at least in part objectively, while 3 experts stated that welfare cannot be measured objectively (question 3). This suggests that, while subjective feelings are considered important for welfare conceptually, scientists, as a group, believe welfare can be assessed objectively.

Table 1. Overview of need lists for welfare assessment (author, publication year, focal species and focal attention). The lists are ordered hierarchically in two levels. The first-level items are stated in bold. As much as possible, corresponding terms have been put on the same row. '-' indicates that an aspect, which is found in other lists, appears to be missing in the present list. Dotted lines indicate clusters of needs, namely appetitive cognitive, appetitive ethological, appetitive autonomic, aversive autonomic and aversive cognitive needs (see text).

BRACKE *et al.* *Prototype list* *Pigs* *Needs*	BRACKE *et al.* *Revised list* *Pigs* *Needs*	FRASER *1983* *Pigs, sheep, cattle* *Maintenance*	BAXTER & BAXTER *1984* *Pigs* *Needs*	SCHLICHTING & SMIDT *1989* *Pigs* *Behaviour systems*	SUNDRUM *et al.* *1994* *Farm animals* *Influencing areas*	TAYLOR *et al.* *1995* *Poultry* *Maintenance*
	Ingestion	**Ingestion**		**Ingestion**	**Intake**	**Ingestion**
Food	Food	Feed	**Hunger**	Feed		
Water	Water	Drink	**Thirst**	Drink		
Rest	**Rest**	**Rest**	**Sleep**	**Rest**	**Rest**	**Rest**
Social contact	**Social contact**	**Association**	**Sociality**	**Social**	**Social**	**Social**[3]
Reproduction	**Reproduction**					
Sexual	Sexual	-	**Sex**	**Sexual**	-	
Nest building	Nest building	-	Nest building	-	-	
Maternal	Maternal	-	**Maternal**	-	-	
Move	**Kinesis**	**Kinesis**[1]	**Living space**	**Locomotion**	**Locomotion**	**Kinesis**
Exploration	**Exploration**	**Exploration**	**Neophilia/recreation**	**Exploration**	**Comfort/explore**	**Exploration**
	Explore novelty			Explore novelty		
Learn				Nibble		
Root	Forage (root)			Root		
Play	**Play**			Play		
Body care	**Body care**	**Body care**	**Skin comfort**	**Comfort**		**Body care**
Groom	Groom, scratch	Groom		Body care		
Wallow	Wallow			Wallow		
		Thermoregulation		Thermoregulation		
		Comfort-seeking				
Eliminate	**Evacuation**	Evacuation	-	**Evacuation**	**Evacuation**	**Elimination**
-	**? Territiorialism**	**Territorialism**[2]				**Territoriality**
	? Stimulation			**Abnormal behaviour**		
Thermocomfort	**Thermoregulation**		**Thermoregulation**			**Thermoregulation**
-	**Respiration**	-	**Respiration**	-	-	**Respiration**
Health	**Health**	-	**Health**		**Hygiene**	-
	No illness					
No pain	No injury					
No fear	**Safety**	**Reactivity**	**Predictability and controllability**	**Fight/flight**	-	**Self-protection**
	no danger					
	no aggression					

[1] Fraser (1983) includes locomotion, play and stretching into 'Kinesis'; [2] Fraser (1983) includes individual space, home range and feeding range into 'Territorialism'.
[3] Taylor *et al.* (1995) include hierarchy formation, allelomimetic behaviour, peer bonding, reproductive and maternal behaviour in the social maintenance need.

The components of welfare were discussed with the help of the prototype list of needs as specified in the first column of Table 1. The interviewer explained that this list was not intended as a hierarchical ordering of behavioural elements, but that its function was to 'break down' the complex problem of welfare into manageable chunks which could support OWA. This implies that each component in the list must be necessary for assessing welfare overall. For example, health, food, water, and thermocomfort are not the only needs, because there may be a social welfare problem. So, this social component must be added to the list, etc.

Responses to this list were diverse. It was said to be a standard list, but when asked further several remarks were obtained. One remark was that the list was not uniform in that it puts incompatible terms on the same level, such as health, mental terms (e.g. 'no fear', 'no pain') and behavioural systems. A common denominator would be preferable. Furthermore, the concept of a hierarchy of needs was challenged on the grounds that components will overlap and that every hierarchy is necessarily artificial. Furthermore, experts differed with respect to the classification of needs and their definition. For example, some experts classified rooting as part of the need for food, but other experts classified it as a separate need. While several experts identified the need for stimulation as a separate need that refers to environmental complexity and novelty, the frustration of which can be expressed as apathy (as in Wemelsfelder, 1993), other experts argued that the need for stimulation reduces to other needs, such as exploration, locomotion and social contact. Despite these differences a consensus area could also be identified. For example, all experts included the needs for food, water, rest, social contact and thermocomfort.

At the end of the interview we asked the expert to state his/her opinion about the suggestion to assess overall welfare as a function of need states (question 5). Most experts (17 out of 21) had a positive attitude toward this suggestion; 2 experts showed a neutral attitude, saying that it may be possible to do so, and 2 experts tended to be sceptical about this approach.

With respect to expert opinion about needs, we conclude that, although not without difficulties and opportunities for further improvement, broad consensus exists for a scientific approach to OWA based on needs.

List of needs

For the purpose of actual OWA it is necessary to have a specified list of needs. Table 1 shows various lists of needs (Fraser, 1983; Baxter & Baxter, 1984; Schlichting & Smidt, 1989; Sundrum *et al.,* 1994; Taylor *et al.,* 1995). This table shows clear differences between authors. Most often 'left out' are items concerning respiration, territorialism, health and items related to reproduction. In addition, the hierarchical organisation differs; e.g. play is part of kinesis for Fraser (1983) and part of exploration for Schlichting & Smidt (1989). Furthermore, different terms are used to denote overlapping concepts, for example 'body care' and 'comfort'. Conversely, similar terms may be used for partly diverging concepts. For example, Fraser's list contains territorialism, which includes various aspects of space, namely individual space, home range and feeding range. Others do not include territorialism in their list, but have incorporated these aspects into other components such as social contact, locomotion and ingestion. These differences illustrate the importance of standardisation and unifying definitions in the field of OWA. However, maybe even more important for OWA is that Table 1 also identifies underlying consensus. It confirms that these authors believe needs are important constituents of welfare and that the list of needs for OWA includes ingestion (food and water), thermoregulation, rest, social contact, kinesis, exploration and body care/comfort. In addition to these needs, needs related to fear/avoidance and sex are also well established (Toates, 1986).

We made some minor revisions in our prototype list of needs for actual OWA in the case of pigs (see Table 1). The revised list does no longer contain the subjective terms 'no fear' and 'no pain'. As a common denominator we have chosen motivational systems. Although further revision may prove necessary, this list provides the starting point for development of a tool to assess the overall welfare-status of pigs. It contains the elements that we believe to be necessary to assess the overall welfare status of pigs.

The revised list includes the following needs: ingestion (including the need for food and water), rest, social contact, reproduction-related needs (sex, nest building and maternal care), kinesis, exploration (including exploration of novelty, foraging and play), body care, evacuation, thermocomfort, respiration, health (including no injuries or pain) and safety (including no danger and no aggression).

Specific for pigs are behavioural elements such as rooting, nest building, wallowing and the ability to separate the resting from the elimination area. For application to other species or to specific subgroups (e.g. pregnant sows or growing pigs) especially the ethological needs require modification. Below, we will discuss the needs that are relevant for welfare assessment in pregnant sows.

Body-care concerns the behavioural elements of scratching (grooming) and wallowing in pigs. Evacuation concerns the eliminative behaviour that is specific to pigs as compared to other farm animals such as cattle and poultry, namely to have elimination areas separate from the resting area.

The need to explore concerns the active behavioural processes by which an animal assimilates information about its environment. Exploration is especially evoked by mild disparity between sensory input and stored representations or expectations. (A larger disparity results in fear and avoidance, which we have classified under 'safety'.) Two major components of exploration are the need to explore novelty and the need to forage (rooting in pigs). Foraging is appetitive feeding behaviour. As such it could be argued to be part of the need for food. However, scientific data concerning contrafreeloading, where animals have been shown to work for food even when ad lib food is available, support a classification of foraging as a separate need. Since recent evidence suggests that foraging may be part of the need to explore or to gather information (Bean *et al.*, 1998), we have classified rooting as part of the need to explore. Play has also been subsumed under exploration, because a main function of play involves learning (e.g. skills).

The need to move, kinesis, may not be controlled by a separate control system. Instead it may be argued that it is part of various other needs, e.g. exploration. However, since there seems universal consensus that space and the ability to move are important components of welfare, we incorporate the need for kinesis as a separate functional element in the need list.

In line with other authors we include respiration as a separate need (cf. Table 1). Baxter & Baxter (1984) defined respiration as the need 'to prevent the sow feeling asphyxiated or choked' (p. 283). This need has its own control centre in the brain and accordingly may classify as a separate need. However, it may also be subsumed under 'safety' or under 'health' as noxious stimuli (e.g. NH_3, CO_2, H_2S, dust levels), because it is probably largely under autonomic control, meaning that only gross deviations are relevant for welfare.

The need for health is the need to be free from disease, i.e. the absence of clinical symptoms or pathological anatomical abnormalities. It is included as a separate need because it is related to a semi-behavioural system, namely 'sickness behaviour' (Hart, 1988). Health is clearly associated with welfare-relevant emotional states. In addition, when activated, sickness behaviour must clearly compete for time and motor output with other behaviour systems such as feeding, sex or the avoidance of danger. Often sickness behaviour takes priority indicating that combating the disease is important for the animal (and its welfare). Like other motivational systems sickness behaviour is functional for survival (Hart, 1988). It also involves learning processes. Examples of such learning include food aversion learning, self-narcotisation to alleviate pain and so-called antidotal thirst to alleviate sickness (reviewed in Toates, 1986, e.g. p. 76).

The need for health includes specific illnesses and injuries. The injury sub-component captures the 'no pain' item in the prototype list. Pain and fear are related motivational systems, but 'whereas the fear system is responsible for motivating escape from a dangerous location, the pain motivational system determines the behaviour of resting to allow recuperation.' (Toates, 1986, p. 154).

The fear system has been renamed as the need for safety. This need is associated with the flight/fight/fright syndrome that serves to maintain the integrity of the whole body against potential disturbance and damage. It does not imply that the goal is absolutely no fear. Exposure to mild or moderate fear is even thought to be beneficial (Jones, 1997). It seems to be a feature of various types of environmental stimuli that too much as well as too little stimulation may be suboptimal for need satisfaction (Fraser *et al.,* 1975, p. 655). Other examples of this phenomenon include temperature, food and social contact. The terms 'safety' and 'fear' are used here to denote only one component of welfare. In a wider sense, where safety would include aspects of health, ingestion, thermoregulation, etc., it could be interpreted to cover (almost) the entire field of welfare. In this chapter, safety denotes only one of the classes of things that motivate animals.

Not included in our list for pigs are territorialism, predictability and control, abnormal behaviour and stimulation. Territorialism is an ethological need that does not apply to pigs because pigs do not defend a territory, although they do live in home ranges (Graves, 1984). For pigs we will subsume the aspects that are related to territorialism under other needs such as kinesis and exploration. Abnormal behaviour (Schlichting & Smidt, 1989) and predictability and control (Baxter & Baxter, 1984) are not separate needs, but seem to be more general indicators of welfare problems. Even if the underlying motivational basis is not always well understood (e.g. Rushen *et al.,* 1993), they generally subsume under other needs. Similarly, for stimulation we prefer to subsume it under the need to explore (novelty).

As this discussion illustrates, drawing up a list of needs requires making decisions about issues that have not been fully resolved. It seems necessary to compare various need lists and to examine their practical implications. However, for actual OWA it is necessary to have a specified list. By making choices explicit we hope these issues will be re-actualised, which in turn may lead to improved welfare assessment.

Conclusions

This chapter deals with the question how OWA can be performed on a scientific basis. We suggested to perform OWA based on an assessment of needs. A needs-based approach allows welfare to be concerned with what matters from the animal's point of view, while at the same time allowing a scientific approach. This is because the term 'need' has both subjective and objective elements in its meaning. Needs were defined as the states of the animal's motivational systems, which specify the animal's proximate goals. The concept of emotional states plays a functional role, both in channelling various kinds of input to produce an efficient response (the causation of behaviour) as well as in constituting the animal's welfare status. However, emotional states do not provide an immediate operational tool for OWA. By contrast, the concept of needs provides a more useful approach to assess welfare. A needs-based approach provides the stepping stones to organise welfare-relevant attributes. It also ensures that welfare is assessed *overall*.

OWA requires taking into account all available scientific evidence. The various scientific paradigms concerned with welfare all provide relevant data. In addition, these paradigms may allow the formulation of various assessment rules. These rules should be specified as much as possible and apparent conflicts between them should be a reason to specify the rules in more detail. At some point these assessment rules may have to be used as heuristic rules for operational OWA.

Interviews with experts confirmed that feelings are an important element in the concept of welfare, but they also confirmed that they believe welfare can be assessed objectively and that OWA is probably best performed based on an assessment of needs.

For actual OWA it is necessary to specify a list of needs. A list of needs for pigs was formulated. It includes needs in relation to ingestion, rest, social contact, reproduction, kinesis, exploration, body-care, evacuation, thermoregulation, respiration, health and safety. This list is not final, but it can be used as a starting point to perform actual OWA in the case of pregnant sows.

Acknowledgements

This work was funded by the Technology Foundation (STW) (through a joint grant from the Netherlands Organisation for Scientific Research and the Netherlands Ministry of Agriculture, Nature Management and Fisheries) and the Dutch Society for the Protection of Animals. We thank Dr. Frans Stafleu and Dr. Willem Schouten for their comments.

References

Baxter, M.R., 1983. Ethology in environmental design for animal production. *Applied Animal Ethology* 9: 207-220.

Baxter, M.R. & S.H. Baxter, 1984. How to delineate the sow's welfare needs? *Annales Recherches Véterinaire* 15: 281-286.

Bean, D., G. Mason & M. Bateson, 1998. Why do animals work for food when food is free? Using starlings to test the information hypothesis for contrafreeloading. In: I. Vessier & A. Boissy (Eds), Proceedings of the 32nd Congress of the International Society for Applied Ethology. Chaumeil Repro, Clermont-Ferrand, p. 107.

Bracke, M.B.M., J.H.M. Metz, A.A. Dijkhuizen & B.M. Spruijt, 1999a. The development of a decision support system to assess the welfare status of pregnant sows. In: Th. Wensing (Ed), Production Diseases in Farm Animals: 10th International Conference 1998. Wageningen Pers, Wageningen, p. 235.

Bracke, M.B.M., J.H.M. Metz & B.M. Spruijt, 1999b. Overall animal welfare assessment reviewed. Part 1: Is it possible? *Netherlands Journal of Agricultural Science* 47: 279-291.

Bracke, M.B.M., J.H.M. Metz & B.M. Spruijt, 1999c. Overall animal welfare reviewed. Part 2: Assessment tables and schemes. *Netherlands Journal of Agricultural Science* 47: 293-305.

Broom, D.M., 1998. Welfare, stress, and the evolution of feelings. *Advances in the Study of Behavior* 27: 371-403.

Broom, D.M. & K.G. Johnson, 1993. Stress and Animal Welfare. Chapman & Hall, London, 211 pp.

Cabanac, M., 1971. Physiological role of pleasure. *Science* 173: 1103-1107.

Dawkins, M.S., 1980. Animal Suffering: The Science of Animal Welfare. Chapman and Hall, London, 149 pp.

Dawkins, M.S., 1983. Battery hens name their price: Consumer demand theory and the measurement of ethological 'needs'. *Animal Behaviour* 31: 1195-1205.

Dawkins, M.S., 1993. Through our Eyes only. The Search for Animal Consciousness. W.H Freeman and Company, Oxford, 192 pp.

Dawkins, M.S., 1998. Evolution and animal welfare. *The Quarterly Review of Biology* 73: 305-328.

Duncan, I.J.H., 1993. Welfare is to do with what animals feel. *Journal of Agricultural and Environmental Ethics 6, Supplement* 2: 8-14.

Duncan, I.J.H., 1996. Animal welfare defined in terms of feelings. *Acta Agricultura Scandinavica Section A, Animal Science, Supplement* 27: 29-35.

Duncan, I.J.H. & M.S. Dawkins, 1983. The problem of assessing "well-being" and "suffering" in farm animals. In: D. Smidt (Ed), Indicators relevant to Farm Animal Welfare. Martinus Nijhoff, The Hague, pp. 12-24.

Duncan, I.J.H. & J.C. Petherick, 1991. The implications of cognitive processes for animal welfare. *Journal of Animal Science* 69: 5017-5022.

Fraser, A.F., 1983. The behaviour of maintenance and the intensive husbandry of cattle, sheep and pigs. *Agriculure, Ecosystems and Environment* 9: 1-23.

Fraser, A.F. & D.M. Broom, 1990. Farm Animal Behaviour and Welfare, 3rd edition. Baillière Tindall, London, 437 pp.

Fraser, D. & I.J.H. Duncan, 1998. 'Pleasures', 'pains' and animal welfare: Towards a natural history of affect. *Animal Welfare* 7: 383-396.

Fraser, D., J.S.D. Ritchie & A.F. Fraser, 1975. The term 'stress' in a veterinary context. *British Veterinary Journal* 131: 653-662.

Fraser, D., D.M. Weary, E.A. Pajor & B.N. Milligan, 1997. A scientific conception of animal welfare that reflects ethical concerns. *Animal Welfare* 6: 187-205.

Graves, H.B., 1984. Behaviour and ecology of wild and feral swine (Sus Scrofa). *Journal of Animal Science* 58: 482-492.

Hart, B.L., 1988. Biological basis of the behavior of sick animals. *Neuroscience and Biobehavioral Reviews* 12: 123-137.

Hurnik, J.F., 1993. Ethics and animal agriculture. *Journal of Agricultural and Environmental Ethics 6, Supplement* 1: 21-35.

Hurnik, J.F. & H. Lehman, 1985. The philosophy of farm animal welfare: A contribution to the assessment of farm animal well-being. In: R.-M. Wegner (Ed), Second European Symposium on Poultry Welfare. German Branch of the World's Poultry Science Association, Celle, pp. 255-266.

Hurnik, J.F. & H. Lehman, 1988. Ethics and farm animal welfare. *Journal of Agricultural Ethics* 1: 305-318.

Jones, R.B., 1997. Chapter 6. Fear and distress. In: M.C. Appleby & B.O. Hughes (Eds), Animal Welfare. CAB International, Oxon, pp. 75-87.

Mason, G. & M. Mendl, 1993. Why is there no simple way of measuring animal welfare? *Animal Welfare* 2: 301-319.

McFarland, D., 1989. Problems in Animal Behaviour. Longman Scientific & Technical, Longman House, Burnt Mill, Harlow, Essex, 158 pp.

Mendl, M., 1991. Some problems with the concept of a cut-off point for determining when an animal's welfare is at risk. *Applied Animal Behaviour Science* 31: 139-146.

Overton, D.A., 1991. A historical perspective on drug discrimination. *Nida Research Monograph* 116: 5-24.

Nagel, T., 1974. What is it like to be a bat? *Philosophical Review* 83: 435-450.

Rollin, B.E., 1990. The Unheeded Cry: Animal Consciousness, Animal Pain and Science. Oxford University Press, Oxford, 308 pp.

Rushen, J., 1991. Problems associated with the interpretation of physiological data in the assessment of animal welfare. *Applied Animal Behaviour Science* 28: 381-386.

Rushen, J., A.B. Lawrence & E.M.C. Terlouw, 1993. The motivational basis of stereotypies. In: A.B. Lawrence & J. Rushen (Eds), Stereotypic Animal Behaviour: Fundamentals and Applications to Welfare. CAB International, Wallingford, pp. 41-64.

Schlichting, M. & D. Smidt, 1989. Subkriterium Tierverhalten. In: C. Fuchs, J. Gartung & J. Hoeges (Eds), Haltungssysteme für Mastschweine. Vergleich, Bewertung, Verbesserungsansätze. KTBL-Schrift 335. KTBL e.V., Darmstadt, pp. 71-88.

Stafleu, F.R., F.J. Grommers & J. Vorstenbosch, 1996. Animal welfare: Evolution and erosion of a moral concept. *Animal Welfare* 5: 77-84.

Sundrum, A., R. Andersson & G. Postler (Eds), 1994. Der Tiergerechtheitsindex-200/1994 - Ein Leitfaden zur Beurteilung von Haltungssystemen für Rinder, Kälber, Legehennen und Schweine. Verlag Köllen, Bonn, 211 pp.

Taylor, A.A., J.F. Hurnik & H. Lehman, 1995. The aplication of Cost-Benefit Dominance analysis to the assessment of farm animal quality of life. *Social Indicators Research* 35: 313-329.

Toates, F., 1986. Motivational Systems. Cambridge University Press, Cambridge, 188 pp.

Toates, F., 1995. Stress: Conceptual and Biological Aspects. Wiley, New York, 339 pp.

Webster, J., 1995. Animal Welfare: A Cool Eye Towards Eden. Blackwell, Oxford, 273 pp.

Wemelsfelder, F., 1993. Animal boredom: Toward an empirical approach of animal subjectivity, PhD. University of Leiden, Leiden, 91 pp.

Wemelsfelder, F., 1997. Life in captivity: Its lack of opportunities for variable behaviour. *Applied Animcl Behaviour Science* 54: 67-70.

Wiepkema, P.R., 1982. On the identity and significance of disturbed behaviour in vertebrates. In: W. Bessei (Ed.), Disturbed Behaviour in Farm Animals. Hohenheimer Arbeiten 121. Verlag Eugen Ulmer, Stuttgart, pp. 7-17.

Wiepkema, P.R., 1987. Behavioural aspects of stress. In: P.R. Wiepkema & P.W.M. van Adrichem (Eds), Biology of Stress in Farm Animals: An Integrative Approach. Martinus Nijhoff, Dordrecht, pp. 113-133.

Wiepkema, P.R. & J.M. Koolhaas, 1993. Stress and animal welfare. *Animal Welfare* 2: 195-218.

Chapter 6. Expert opinion about the welfare status in the seven main housing systems for pregnant sows[1]

Abstract

In interviews with eleven pig experts the main housing systems for pregnant sows were identified as tethering (T), individual housing in stalls (IS), group housing with stalls (GS), trickle feeding or Biofix (B), electronic sow feeding (ESF), and outdoor housing with huts (O). The family pen system (Fam) was added as a reference system.

The experts were asked to give a welfare score for each housing system. The two individual housing-systems (mean scores: T=1.8; IS=2.3) scored significantly lower than more intensive indoor group-housing systems (GS=5.4; B=5.3; ESF=6.2), and these scored lower than the more extensive systems (O=8.0; Fam=9.1; ANOVA, $P < 0.001$). Furthermore, T ranked lower than IS in the Sign test ($P = 0.008$).

The most important aspects for welfare assessment were space, substrate, feeding-related agonism and social parameters such as group size and group stability. Three different models were constructed to calculate welfare scores from the arguments given by the experts. When represented graphically the results seem comparable to the expert scores, although two of the three models differed significantly from the expert scores using analysis of variance. These results indicate that pig experts are able to perform overall welfare assessment in a rational way that allows modelling and that there is a consensus underlying welfare assessment. These outcomes provide support for the further development of a decision support system to assess farm animal welfare on a scientific basis.

Keywords: animal welfare, pigs, decision support system, model.

Introduction

Overall welfare assessment is a highly desired goal of policy makers in the Western countries. This has resulted in the development and political employment of practical welfare-index systems for on-farm use in certain countries, for example in Austria (Bartussek, 1988 and 1997) and Germany (Sundrum *et al.*, 1994). However, these authors also acknowledge that the scientific basis should be improved (Bartussek, 1997; Sundrum, 1997). On the other hand, several authors have stated that the task of performing objective overall welfare assessment is complex and may be too difficult for science (Rushen & de Passillé, 1992; Fraser, 1995; Dawkins, 1997). While individual humans generally have little problems in forming private opinions about the welfare status of animals, finding an objective basis appears to be more difficult. Authors differ on how to define welfare and, while there seems to be consensus that

1 Paper by Bracke, M.B.M., J.H.M. Metz, B.M. Spruijt & A.A. Dijkhuizen, 1999. Overall welfare assessment of pregnant sow housing systems based on interviews with experts. *Netherlands Journal of Agricultural Science* 47: 93-104.

for welfare assessment many factors should be taken into account, the problem of weighting different aspects remains open for study.

In an attempt to meet the demand for a more objective welfare-assessment tool we have started to develop a computerised decision support system for assessment of the welfare status of pigs (Bracke *et al.*, 1997; Bracke *et al.*, accepted). Such a system takes a description of a housing system as input and uses scientific knowledge collected in a database to yield output in the form of a welfare score. Such a system also requires a formalised procedure to perform overall welfare assessment in an explicit way, but no standard exists for such a task. In this study we take a group of scientific experts as the standard to collect quantitative data concerning welfare assessment which can be used in the development of a decision support system. In particular the objective of this study was to establish input-output relationships between housing and welfare; to determine which are the most important attributes of housing systems of pregnant sows with respect to welfare; and to assess the perspectives of modelling of welfare assessment. For this purpose we asked pig experts to give welfare scores for the main housing systems for pregnant sows and we modelled their arguments to assess whether subjective welfare scores can be recalculated from attributes of the housing systems.

Materials and methods

Interviews

Eleven scientists from six West-European countries, who had all been involved in the development of welfare friendly housing systems for pigs, were interviewed in independent sessions. They performed the following tasks:

1. Identify the most important housing systems for pregnant sows.
2. Compare the answers given with the following predefined list: tethering (T); individual housing in stalls (IS); group housing with stalls (GS); group housing with trickle feeding, i.e. Biofix (B); group housing with transponders, i.e. electronic sow feeding (ESF); outdoor housing with arcs/huts (O); and the family pen system (Fam). Descriptions of these housing systems can be found in the booklets of the Pig Welfare Advisory Group (Anon., 1997b), EU Scientific Veterinary Committee Report (Anon., 1997a), Svendsen & Svendsen (1997) and Backus (1997). Some references for the family pen (Fam) system include Stolba (1981) and Wechsler (1996). The Fam system, while not being a competitive production system, was included as a reference system. In relation to Fam it was pointed out to the expert that we were interested in the welfare status of the pregnant sows only. It was specified that we were not interested in the welfare of the piglets in this system, nor in the economic, labour or environmental values of any of the housing systems, unless they affected the overall welfare status of the pregnant sows.
3. Give an overall welfare score for typical-example farms of each housing system on a scale from 0 (worst) to 10 (best). A typical example was explained as being a most typical farm rather than as one with particularly good or bad welfare conditions. In addition, the expert was asked to give scores only for systems he/she knew well. As a result each expert specified his/her own systems within the limits of the main types of housing systems. This procedure allowed highly variable attributes such as management or stockmanship to be specified and it ensured maximally realistic conditions while still allowing generalisation of the results.
4. Justify each score. Each expert was asked to explain all scores including why the best and worst scores were not 0 and 10 respectively. This was done by listing the positive and negative attributes for each system and by explaining differences between adjacent scores.

In addition to the scores for the main housing systems, the experts were free to add variations of housing systems by specifying new scores, for example for the same ESF system with and without straw.

Modelling

Modelling proceeded as follows. A welfare value[1] is assigned to each attribute level in the list of welfare-relevant attributes. An example of such an attribute is 'straw quantity'. Its levels may be 'no straw', 'some straw' and 'deep straw'. Only one attribute level can be true for a given housing system. All levels within each attribute receive a welfare value between 0 (worst) to 10 (best) according to their welfare rank. Weighting factors, which apply across attribute levels, were not used in this study, i.e. within each model all levels weighed equally. For every housing system, overall welfare scores are calculated in an additive way as the average of all welfare values of all attributes (i.e. the sum of welfare values divided by the total number of attributes, cf. Table 4).

For example, an expert may explain the low score for one housing system by 'lack of exercise' as compared to 'spacious field' for the other, better housing system. Accordingly, pen space is being identified as a relevant attribute, with 'lack of exercise' being the worst and 'spacious field' being the best-ranked level. Corresponding welfare values are 0 and 10 respectively. Any intermediate levels, as specified by the expert, receive intermediate welfare values proportionally.

There were four methods to generate welfare scores. One consisted of the personal welfare scores as obtained directly from each expert. In addition, three models were constructed. In model A the set of attribute levels as mentioned by an expert were used to calculate scores for the housing systems as specified by the same expert. This was done for all experts to generate the set of data for model A, which is a within-expert model. Model B is an across-expert model, in which the different attribute-level lists of all experts were summarised and organised by removing doubles and by regrouping levels in a logical way. For example, when one expert makes a distinction between the levels 'straw is present' and 'no straw', and another expert between 'some straw' and 'deep straw', the integrated list contains 'no straw', 'some straw' and 'deep straw' (with welfare values of 0, 5 and 10 respectively). In this way all logical distinctions made by the experts were retained in model B. Model C consists of a selection of the more important attributes in model B as indicated by the degree these had been used by the experts to explain their welfare scores.

To test effects of housing, experts and expert-housing interactions on the expert scores analysis of variance (ANOVA) was used (Anon., 1993). The models were also tested using ANOVA to see whether any of the scores generated by the three models differed significantly from the scores given by the experts. ANOVA requires the data to be normally distributed. This was confirmed by evaluation of normal plots of residuals. The null hypothesis was that there is no difference, implying that all four methods (the expert scores and the models) generate equivalent welfare scores for the main housing systems for pregnant sows. A one-sided Sign test was used to compare ranks of the housing systems that did not differ significantly in the ANOVA test.

1 The term 'welfare value' is equivalent to the term 'attribute score' used in later chapters.

Results

The experts generally agreed that the main housing systems for pregnant sows are tethered (T), individual housing in stalls (IS), group housing with feeding and/or lying stalls (GS), trickle feeding or Biofix (B), transponder or electronic sow feeder systems (ESF) and outdoor housing with arcs/huts (O). Other variants include indoor floor feeding and various trough-feeding systems. The Biofix system, where food is delivered into the trough at a controlled rate (e.g. Olsson *et al.*, 1986), was the least well-known of the six major housing systems in that three experts felt insufficiently familiar with the Biofix system to give a welfare score.

Significant differences (ANOVA, $P < 0.001$) were found between the mean expert-scores for individual, confined housing (T, IS) versus loose, group housing (all other systems), and between generally more intensive indoor group housing (GS, B, ESF) versus more extensive group housing (O, Fam). The first group (T, IS) scored worst; the latter (O, Fam) scored best for overall welfare (Table 1). No differences were found between the mean scores for T and IS, between GS, B and ESF, or between O and Fam. The T system was ranked significantly worse for welfare than the IS system (one-sided Sign test, $P = 0.008$). No further significant differences in ranks between the indoor group housing systems or between the O and Fam system were found with the Sign test.

Table 1. Mean scores (from 0 worst to 10 best) obtained from 11 pig experts for the 7 main housing systems for pregnant sows, and welfare scores as calculated by 3 models (see text).

	T	IS	GS	B	ESF	O	Fam	l.s.d.	F probability
Expert scores*	1.75^a	2.34^a	5.37^b	5.27^b	6.19^b	8.02^c	9.14^c	1.182	<0.001
Model A	2.83^a	3.42^a	6.00^b	5.74^b	5.84^b	7.90^c	9.10^d	1.127	<0.001
Model B*	3.18^a	3.56^a	5.39^b	5.73^b	6.41^c	7.99^d	8.69^e	0.360	<0.001
Model C*	2.50^a	2.81^a	4.69^b	4.97^b	5.09^b	8.89^c	9.71^d	0.645	<0.001

T: tether; IS: individual housing in stalls; GS: group housing with stalls; B: Biofix; ESF: electronic sow feeder; O: outdoor; Fam: Family Pen. Means within a row lacking the same superscript character differ ($P<0.05$). * A significant expert effect was found by ANOVA ($P < 0.05$).

Table 2 shows various measures indicating the relative importance of the attributes which experts mentioned when explaining their personal welfare scores. The main criteria to add variants of housing systems by specifying additional overall scores were straw versus no straw (8 experts did so); stable versus dynamic groups (4 experts); roughage versus no roughage (4 experts); and housing systems with different space allowances (4 experts). Other variants include GS systems with and without a social resting area; O systems with and without nose rings; and O systems on proper and improper soil types. On average 12.2 attributes were mentioned per expert to justify the scores (range 7-20). All experts mentioned space, straw, and feeding sequence. Other often mentioned attributes include social stability; the ability to separate functional areas (resting and dunging area; resting and feeding area); the provision of roughage; hygiene and cleanliness; and social resting possibilities. With social resting is meant the ability for sows to rest with two or more together unrestricted by, for example, lying stalls. Several attributes were used in a non-dichotomous way, which means they took more than two (yes/no) levels. The most important one was space (9 experts). Other quantified attributes include straw quantity, separate functional areas and social resting.

Table 2. Number of experts (n=11) that, while explaining their personal welfare scores for the main housing systems for pregnant sows, used levels falling into any of the listed attribute classes (a); (b) as (a) but including only non-dichotomous level-ranges (see text); (c) as (a) but including only attributes which experts used to specify scores in addition to the scores for the main housing systems.

Attribute class	Number of experts		
	(a)	(b)	(c)
Space	11	9	4
Straw, substrate	11	2	8
Simultaneous feeding	11	1	0
Social stability and mixing	8	2	4
Separate functional areas	7	2	1
Roughage, bulk	7	1	4
Social resting	7	1	3
Hygiene & cleanliness	7	0	2
Agonism while feeding & individual feeding level	6	1	2
Group size	6	1	1
Space structure and agonism	5	1	0
Climate	5	0	0
Tether injuries	5	0	0

Modelling resulted in three series of calculated welfare scores. These models were based on data from each individual expert, applied only within experts (model A); based on the integrated list of all attribute levels as mentioned by all experts (model B); and on a short version of this integrated list (model C). Both model B and model C generated different scores for different experts, because the experts had specified the exact descriptive properties of the housing systems somewhat differently. Model C includes the attributes pen space, substrate (straw), feeding sequence (simultaneous feeding), social stability, separation of functional areas and food quantity (roughage). In addition, the integrated list of model B also includes the proportion of time the animals can move around freely, social resting, group size, degree of individual feeding, food-intake protection, climate (separated into two: 'when hot' and 'when cold'), air quality, flooring aspects (separated into 'rest surface' and 'walking surface'), grooming facilities, the ability to avoid conspecifics and the degree of social variation (i.e. having different age groups in one pen as in Fam). Hygiene and cleanliness were incorporated in both climate and rest-surface attributes.

Whereas there was no significant difference between the mean expert scores for the O and Fam systems, in all three models the calculated scores for these housing systems differed significantly from each other. For model B a significant difference was also found between the calculated score for B (lower) and ESF (higher) (Table 1).

Figure 1 also shows that the three models tend to follow the expert scores for the 7 main housing systems. With statistical analysis the failure to find a significant difference between a model and the expert scores indicates statistical similarity between the model and the expert scores. When the expert scores are taken as a standard, failure to find a significant difference between a model and the expert scores, indicate that it is better than those models which do differ. When only the scores for the seven main housing systems were tested, model A and model B failed, and only model C, the shortened version of the integrated model, was not significantly different from the expert scores. However, when all scores were tested, i.e. including those added by the experts to specify variations of the main housing systems, then only model A failed. Model B and C did not fail, i.e. the scores calculated by these models did not differ significantly from the expert scores (Table 3).

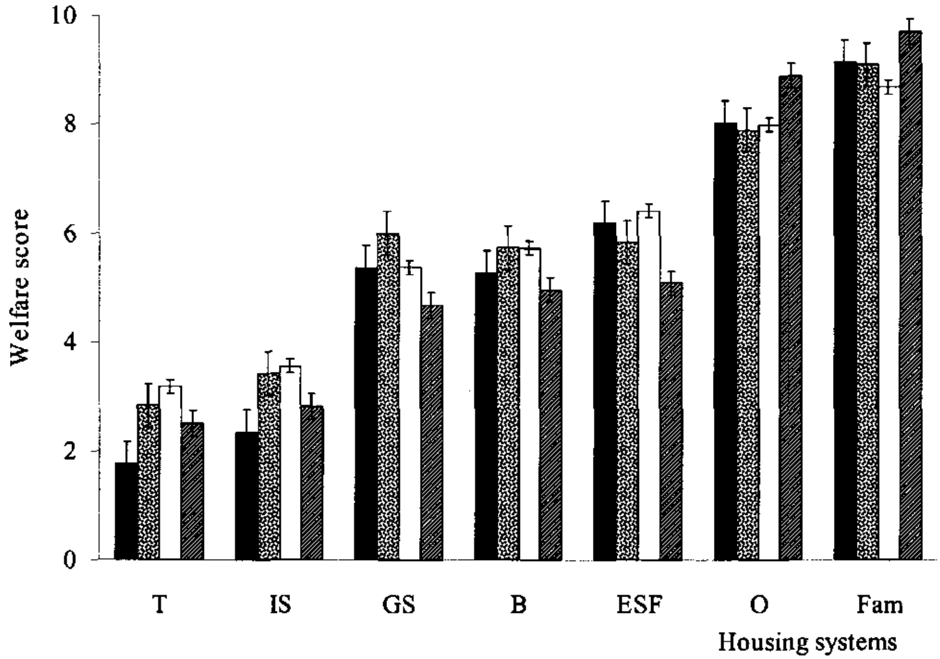


Figure 1. Mean welfare scores and their standard errors for the 7 main housing systems for pregnant sows (n = 11 experts). For every housing system the first column represents the mean expert score. The next three columns represent model A, B and C (see text) with which welfare scores are calculated. T: tether; IS: individual housing in stalls; GS: group housing with stalls; B: Biofix; ESF: electronic sow feeder; O: outdoor; Fam: Family Pen.

Table 3. Comparison of expert scores with three welfare assessment models (ANOVA). Means within a row lacking a common superscript character differ (P<0.05).

	Expert scores	Model A	Model B	Model C	l.s.d.	F probability
Main housing systems, scores	0.000^{a}	0.356^{b}	0.376^{b}	0.0412^{a}	0.3071	0.021
All housing systems, scores	0.000^{ab}	0.352^{c}	0.209^{ae}	-0.303^{bf}	0.3180	0.001

Discussion

The list of main housing systems we obtained by asking experts is similar to the one mentioned by Broom (1989). In Table 4 we have given a tentative distribution of welfare scores over the main housing systems as the general picture emerging from our interviews. We chose a numerical scale (0-10) based on welfare ranks within attributes. For example, increasing the amount of space per pen was generally considered to be increasingly better for welfare. When pen space is similar for T and IS, for GS and B, and for ESF and Fam we get a welfare value distribution as shown in Table 4 in the first row. The overall scores are calculated as the average score (without giving any attribute level additional weight).

Svendsen & Svendsen (1997) produced a similar table, but with plusses and minuses and they did not include the O and Fam system. They also did not compute overall welfare scores. However, when their plusses and minuses are added overall scores are found which are similar to our findings (Table 1 and Figure 1). The scores calculated from Svendsen & Svendsen (1997) would be (in number of plusses): T=IS=1.5; GS as cubicles=4; B=6.5; ESF=5 plusses. They include other variations of the GS system (with a social resting area, deep straw or straw-bedded kennel) which get higher scores (up to 10.5 plusses). We also

*Table 4. Tentative welfare distribution table. Within attributes (rows) a welfare value is assigned between 10 (best) and 0 (worst) on the basis of ranking the levels of the corresponding housing systems. Intermediate ranks get intermediate welfare values in a proportional way. Attribute levels may vary considerably within a given housing system, especially those marked with *. Attributes included in the table and rankings of attribute levels are tentatively based on the interviews with pig experts. Overall welfare is calculated as the average of the attribute scores. Mean subjective welfare scores of experts are also given. T: tether; IS: individual housing in stalls; GS: group housing with stalls; B: Biofix; ESF: electronic sow feeder; O: outdoor; Fam: Family Pen.*

Attribute	T	IS	GS	B	ESF	O	Fam
Space per pen	0	0	3.3	3.3	6.7*	10	6,7
Space per sow	0	0	3.3*	3.3	3.3	10	6,7
Straw, substrate	0	0	5*	0*	5*	10	10
Simultaneous feeding	10	10	10	10	0*	10	10
Social stability and mixing	5	5	0*	0	0	0	10
Separate functional areas	0	0	3.3*	3.3	6.7	10	10
Roughage, bulk	0	0	5*	0*	5*	10	10
Social resting	0	0	0*	10	10	10	10
Hygiene & cleanliness	0	0	5	10	10	5	10
Agonism while feeding & individual feeding level	10	10	10	0	10	0	10
Group size	0	0	10*	10	5*	10*	10
Climate (when cold)	10	10	5*	10	10	0	5
Climate (when hot)	0	0	5	5	5	10	5
Tether injuries	0	10	10	10	10	10	10
Overall score	2.5	3.2	5.4	5.4	6.2	7.5	8.8
Expert score	1.75[a]	2.34[a]	5.37[b]	5.27[b]	6.19[b]	8.02[c]	9.14[c]

found GS, defined as group housing with stalls, to be the system with the widest range (expert scores 2-8). Included are systems with feeding stalls and a communal resting area, but also systems with cubicles, i.e. feeding-lying stalls. GS systems with a substantial amount of space, substrate and a separate resting area are among the best rated housing systems. The Fam system in its revised form (e.g. Wechsler, 1996) also can be classified as such a system. However, this system has additional features including strictly separated functional areas for resting, activity/rooting and dunging, and a stable family structure of four sows, which are housed together with a boar during the breeding period and together with their offspring until these are about five months of age. Such features mimic the natural environment and are generally considered to be beneficial for welfare. Fam was included as a reference system. It received an average score of 9.14. The O system received the next highest score (8.02). The mean scores for the Fam and O system did not differ significantly. The one-sided Sign test also failed to identify the Fam system as ranking significantly better than the O system (P = 0.145). However, the O system was generally (but not by all experts) described as a system without nose rings, on pasture, with a proper soil type, sufficient and well-insulated arcs and a mud pool for wallowing. We suspect there may have been a tendency to take the better O farms as typical examples rather than the typical or 'average' ones.

The single most important finding of this chapter may be that the experts, who were all pig-welfare scientists that have been involved in the design of welfare friendly housing systems, give significantly higher scores to group housing systems as compared to the individual, confined systems (T and IS). On our scale from 0 to 10 T and IS are clearly 'poor' welfare systems, whereas GS and B, and to a lesser extent also ESF, get scores close to the midline of the scale (5.5). This indicates that these systems are assessed as neither very poor nor very good welfare systems. However, we must warn against drawing overhasty conclusions based on the present findings. The scores obtained for the different housing systems are relative to each other and the scores contain no explicit reference to which scores would be acceptable. Furthermore, variation between farms within a certain type of housing systems and between individual animals may be much larger than between housing systems (e.g. Rushen & de Passillé, 1992; Signoret & Vieuille, 1996). In particular, the best individual housing systems are almost certainly better than the worst group housing systems. Our integrated model B also suggests that individual housing may reach similar scores as the group housing systems when non-social parameters are optimised. However, even though our models allow predicting the effects of changing such attribute levels on the overall welfare status, they have not been properly tested to make such predictions.

In this chapter we have identified the most important attributes to assess the welfare status of pregnant sows in relation to housing, firstly by limiting the amount of time an expert had to justify his/her welfare scores, and secondly by counting the number of experts that use similar attributes. However, several caveats can be identified.

Firstly, experts may have attached different weighting factors to different attributes.

Secondly, overlap exists between various items in the list of attribute classes (Table 2). For example, between straw and roughage and between space, group size and group stability. It is, therefore, not correct to conclude that straw (mentioned by all experts) is more important than roughage (mentioned by seven experts, Table 2), because straw is also a form of roughage. However, it is not possible to construct a mutually exclusive list of attributes. For example, straw is also used for resting and rooting, and therefore cannot be lumped together with the roughage category.

Thirdly, the experts only evaluated typical examples of the main housing systems. This may have obstructed identification of certain attributes such as disease levels or stockmanship, which may vary especially within housing systems rather than between housing systems.

Fourthly, certain attributes seem to be missing. In accordance with Rushen and de Passillé (1992) our list contains both design criteria (environmental factors) and performance criteria (animal factors). Agonism was the most important performance criterion. Other animal factors such as stereotyped-behaviour levels were rarely mentioned, but do seem to be important. The limited time available for each interview (up to two hours) forced each expert to focus on only the most important attributes. Furthermore, 'missing' animal factors may be predictable from environmental factors that were specified by the experts. Examples are stereotyped-behaviour levels, which depend on feeding regime and a restricted space allowance (Terlouw *et al.*, 1991), and simultaneous feeding, which is an indicator of the ability to synchronise activities.

Fifthly, our results only identify opinions of a group of pig-welfare scientists. It remains to be shown that other scientifically trained professionals, such as veterinarians, have similar opinions and that the opinions we found are valid. This may be done by identifying their scientific basis and the consequences for the animals involved with respect to the whole range of needs the animals have (cf. Baxter & Baxter, 1984; de Koning, 1984; Bracke *et al.*, 1997).

In principle the present findings support the development of a decision support system for welfare assessment in several ways:

Firstly, they create anchor points to evaluate system performance by indicating major input-output relationships, i.e. between the descriptions of the main housing systems for pregnant sows and their respective overall welfare score.

Secondly, our findings indicate that there may be a substantial degree of consensus in overall welfare assessment, at least among welfare scientists; more so than would be expected from the diversity of literature concerning welfare definitions and animal welfare assessment (e.g. Rushen & de Passillé, 1992; Mason & Mendl, 1993). Our experts all seem to choose similar attributes to assess welfare, which indicates that they were talking about the same thing. No fundamentally different perspectives on welfare seemed to be involved as, for example, in constructing the integrated model (B) no cases of logical inconsistencies were found between experts (with one expert ranking a certain attribute-level better than another attribute-level and another expert ranking in reversed order). Differences were found between experts: not all experts mentioned exactly the same properties. One example concerned the amount of space. One expert specified that the amount of space in the family pen was more than strictly necessary for optimal welfare. This expert did not rate the extra space as allocated in the O system as contributing to welfare. Several other experts, however, used the extra space in the O system as compared to the Fam system to explain having given a higher welfare score to the O system. However, we believe it would be a mistake to focus on such differences at the expense of the broad consensus that is evident from our findings (Table 1).

Thirdly, overall welfare assessment seems to allow itself to be modelled. Therefore, the development of a formalised procedure to assess welfare may be feasible. Quite simple modelling ignoring weighting factors and interactions already seems to capture major aspects of reasoning about welfare. More sophisticated models may be constructed, where these factors or factors concerning the hierarchy of needs are being taken into account.

In this chapter we have presented a novel approach to the problem of overall welfare assessment. This method involves interviewing experts and employing an indirect way of collecting information about the importance of housing attributes by asking experts to rank and give scores for housing systems. Such an approach may help scientists to progress from pure subjectivity in overall welfare assessment through intersubjectivity toward the ideal of 'objectivity'. Respectively, these stages are characterised by personal welfare scores, subjectively agreed-upon group welfare-scores and welfare scores based on rational decision making based on all known facts including what is known scientifically. Inherent in the notion of overall welfare is the postulate that positive aspects of animal environments may compensate negative aspects to a certain degree. Such a quantified approach to welfare may therefore also help to get beyond cut-off point thinking (cf. Mendl (1991) for a critique of cut-off points) and its associated legislative objective in terms of means-prescriptions. Overall welfare assessment may eventually allow legislation to be formulated in terms of goal prescriptions in which minimum overall welfare standards can be defined generically and where individual farmers are allowed to reach such standards in farm specific ways.

Acknowledgements

This work was funded by the Technology Foundation (STW) and the Dutch Society for the Protection of Animals (NVBD).

References

Anonymous, 1993. Genstat 5 Release 3 Reference Manual, 1993. Oxford University Press, Oxford, 796 pp.

Anonymous, 1997a. The Welfare of Intensively kept Pigs. Scientific Veterinary Committee, Brussels, 199 pp.

Anonymous, 1997b. Booklets of the Pig Welfare Advisory Group: No 1 Introduction of Sows into Groups. 5 pp.; no 2 Non-straw or Low Straw Systems for Housing Dry Sows. 4pp.; no 4 Cubicles and Free-Access Stalls. 8 pp.; no 5 Yards and Individual Feeders. 8 pp.; no 6 Yard or Kennels with Short Stall Feeders (Trickle Feeding or Wet Feed Systems). 8 pp.; no 7 Yards or Kennels with Floor Feeding. 10 pp.; no 8 Outdoor Sows. 8 pp.; no 9 Electronic Sow Feeders (ESF). 16 pp. MAFF Publications, London.

Backus, G.B.C. (Ed.), 1997. Comparison of four housing systems for non-lactating sows (In Dutch). Report P 1.171. Research Institute for Pig Husbandry, Rosmalen, 176 pp.

Bartussek, H., 1988. Haltung. In: Haiger, A., R. Storhas & H. Bartussek (Eds), Natur-gemässe Viehwirtschaft. Verlag Ulmer, Stuttgart, pp. 147-248.

Bartussek, H., 1997. Animals Needs Index ANI (Tiergerechtheitsindex TGI-35L) for assessment of animals well being in housing systems for Austrian proprietary products and legislation. In: J.A.M. Arendonk (Ed.), Book of Abstracts of the 48th Annual Meeting of the European Association for Animal Production. Wageningen Pers, Wageningen, pp. 190.

Baxter, M.R. & S.H. Baxter, 1984. How to delineate the sow's welfare needs. *Annales de Recherches Véterinaires* 15: 281-286.

Bracke, M.B.M., J.H.M. Metz & A.J. Udink Ten Cate, 1997. Assessment of animal welfare in husbandry systems. In: J.T. Sørensen (Ed.), Livestock Farming Systems – More than food production - Proceedings of the fourth International Symposium on Livestock Farming Systems, Wageningen Pers, Wageningen, pp. 231-237.

Bracke, M.B.M., J.H.M. Metz, A.A. Dijkhuizen & B.M. Spruijt, 2001. Decision support system for assessing farm animal welfare: Strategy and prototype. *Journal of Agricultural and Environmental Ethics*: Accepted.

Broom, D.M., 1989. The assessment of sow welfare. *Pig Veterinary Journal* 22: 100-111.

Dawkins, M.S., 1997. D.M.G. Wood-Gush Memorial Lecture: Why has there not been more progress in animal welfare research? *Applied Animal Behaviour Science* 53: 59-73.

Fraser, D., 1995. Science, values and animal welfare: Exploring the 'inextricable connection'. *Animal Welfare* 4: 103-117.

Koning, R. de, 1984. An environmental model for sows: Materializing physical and behavioural needs. *Annales de Recherches Véterinaires* 15: 303-307.

Mason, G. & M. Mendl, 1993. Why is there no simple way of measuring animal welfare? *Animal Welfare* 2: 301-319.

Mendl, M., 1991. Commentary. Some problems with the concept of a cut-off point for determining when an animal's welfare is at risk. *Applied Animal Behaviour Science* 31: 139-146.

Olsson, A.-Ch., M. Andersson, D. Rantzer, J. Svendsen & T. Hellström, 1986. Group housing of sows in gestation: Comparison of computer-controlled individual feeding system with a group feeding system based on biological fixation (In Swedish with tables and summary in English). Report 51. Swedish University of Agricultural Sciences, Department of Farm Buildings, Lund, 221 pp.

Rushen, J. & A.M.B. de Passillé, 1992. The scientific assessment of the impact of housing on animal welfare: A critical review. *Canadian Journal of Animal Science* 72: 721-743.

Signoret, J. & C. Vieuille, 1996. Effectiveness and limitations of physiological versus ethological criteria to assess the welfare of pigs in relation to the housing system. *Pig News and Information* 17: 115-121.

Stolba, A., 1981. A family system in enriched pens as a novel method of pig housing. In: Anonymous (Eds) Alternatives in Intensive Husbandry Systems. Proceedings, Universities Federation for Animal Welfare, Potters Bar, pp. 52-67.

Svendsen, J. & L.S. Svendsen, 1997. Intensive (commercial) systems for breeding sows and piglets to weaning, *Livestock Production Science* 49: 165-179.

Sundrum, A., 1997. Assessing livestock housing conditions in terms of animal welfare – Possibilities and limitations. In: J.T. Sørensen (Ed.), Livestock Farming Systems – More than Food Production. Proceedings of the fourth International Symposium on Livestock Farming Systems, Wageningen Pers, Wageningen, pp. 238-246.

Sundrum, A., R. Andersson & G. Postler (Eds), 1994. Der Tiergerechtheitsindex-200/1994 - ein Leitfaden zur Beurteilung von Haltungssystemen für Rinder, Kälber, Legehennen und Schweine. Verlag Köllen, Bonn, 211 pp.

Terlouw, E.C.M., A.B. Lawrence & A.W. Illius, 1991. Influences of feeding level and physical restriction on development of stereotypies in sows. *Animal Behaviour* 42: 981-991.

Wechsler, B., 1996. Rearing pigs in species-specific family groups. *Animal Welfare* 5: 25-35.

Chapter 7. Description of the 'final' decision support system: Model structure and weighting procedure[1]

Abstract

The problem of how to objectively assess the overall welfare status of animals under farming conditions has contributed to an ongoing debate that has hampered actual decision making on animal welfare. For this reason we constructed a model based on the assumed hierarchical organisation of the animals' needs for overall welfare assessment in the case of pregnant sows. This sow welfare (SOWEL) model is implemented in a computer-based decision support system that takes a description of a housing and management system as input and produces a welfare score as output. A formalised procedure was used to construct the model on the basis of available scientific knowledge. The model contains 37 attributes that describe the welfare-relevant properties of housing and management systems. In the decision support system these attributes are linked to scientific statements and a list of needs to provide a scientific basis for welfare assessment. Weighting factors that represent the relative importance of the attributes, are derived from the scientific statements about the various welfare performance criteria that have been measured by scientists. The welfare score is calculated as the weighted average score. All information in the decision support system is stored in tables in a relational database such that newly available knowledge and insights can be incorporated to refine the model. The model has been developed in line with several existing models, but it differs from these models in that it is the first to provide a formalised procedure to explicate the reasoning steps involved in welfare assessment based on available scientific knowledge.

Keywords: Farm animal welfare assessment, index, pigs, housing systems, management.

Introduction

Farm animal welfare has been debated for over thirty years. Welfare proved to be a complex problem (Dawkins, 1997), and scientists have called for more basic research (e.g. Rushen and de Passillé, 1992; Barnard and Hurst, 1996; Dawkins, 1997). An alternative approach is to acknowledge the availability of a large amount of knowledge about the biological needs of the animals. The question is how we could use the available knowledge to assess the overall welfare status of farm animals, but without having the illusion that such an assessment can presently reach the status of a scientifically proven fact. Overall welfare assessment generally concerns what matters to animals from their point of view, integrated over its various facets (the animals' needs), over longer periods of time (e.g. months or years rather than days) and larger numbers of animals (e.g. entire housing systems rather than individual animals). Although a solution to this problem may have considerable utility in solving ethical or political issues, the primary focus here is to address a factual issue, namely to find an accurate

1 Paper by Bracke, M.B.M., B.M. Spruijt, J.H.M. Metz & W.G.P. Schouten. Decision support system for overall welfare assessment in pregnant sows A: Model structure and weighting procedure. Submitted to *Journal of Animal Science*.

and comprehensive description of the animals' level of welfare given the current state of science.

Much conceptual and factual work has already been done. Many studies have considered specific aspects of welfare in the different species of farm and laboratory animals. Causal models of behaviour and stress with relevance for welfare have been constructed (e.g. Lorenz, 1978; Wiepkema, 1987; McFarland, 1989; Hughes and Duncan, 1988; Toates and Jensen, 1991). But only few models have been developed specifically for overall welfare assessment (reviewed in Bracke et al., 1999b). These models either provide a rough theoretical outline (e.g. Mellor and Reid, 1994) or they directly assign points to easily identifiable attributes of the housing system such as the so-called 'Tiergerechtheitsindex' (cf. van den Weghe, 1998; Bartussek, 1999a). These models, however, are in need of a transparent scientific basis (e.g. Sundrum, 1997).

A step forward would be made if we had a formalised procedure of the reasoning steps involved in overall welfare assessment. This procedure must show how the available knowledge (Rushen, 1991; Dawkins, 1997) can be used to select the welfare-relevant attributes of housing systems, and how to weight them into an overall judgement (Sandøe and Simonsen, 1992; Sandøe, 1997, Bracke et al., 1999a). In search for such a formalised procedure we chose the pregnant sow as a case to develop a model for welfare assessment at the housing-system level. The domain of housing and management systems to which the model applies is defined as the set of all present and future systems for pregnant sows that could be regarded as actual or potential production systems under stabilised conditions. Pregnant sows were chosen because they are farmed under a wide range of housing conditions, including very intensive ones that have raised public concern and that resulted in many detailed studies examining the welfare of sows (cf. Scientific Veterinary Committee, 1997).

The starting point was to assess welfare overall using a list of 11 (welfare) needs of pregnant sows (Bracke et al., 1999c). These needs are instrumental in assessing the animals' emotional states from what is known about their behaviour and physiology. States of need frustration and satisfaction normally result from certain general categories of external stimuli or events to which the animals react with a concerted set of coping responses that tend to increase fitness (Bracke et al., 1999c; Spruijt et al., in press). For instance, all kinds of danger activate a cascade of actions that together make up the stress response, which helps the animal to cope with the situation. Such responses can be measured by scientific methods, and can be related to the attributes that describe the welfare-relevant properties of housing and management systems. Animals themselves integrate the positive and negative aspects of their environment into an overall state of well-being (Spruijt, in press). We hypothesise that the scientific information about the attributes and needs can be used to assess the welfare status with an additive weighting procedure that distinguishes between more and less important attributes (necessities and luxuries, Dawkins, 1990).

In order to facilitate the handling of large amounts of rather complex information, the model is embedded in a decision support system, i.e. a computer-based information system (Turban, 1995). We developed this system mainly to identify the underlying procedure, i.e. the reasoning steps involved in overall welfare assessment. For the development of the decision support system we used the so-called Evolutionary Prototyping Method (Turban, 1995; Bracke et al., accepted; Bracke et al., in press): an initial prototype was constructed and improved in repeated updating versions based in part on interviews with experts to fill in gaps in our knowledge (Bracke et al., 1999c; Bracke et al., 1999d). This led to the present 'final' version. The term 'evolutionary' here means that the decision support system and the welfare model are designed to be flexible, in that they allow the incorporation of new information and insights when these become available.

The objective of this chapter is to describe the decision support system, the welfare model for pregnant sows, and the procedure to select and weight attributes of housing and management systems on the basis of available scientific knowledge. Validation using expert opinion will be the topic of the subsequent chapter.

Components in the decision support system

Relational database

The model in the decision support system takes a description of a housing and management system as input and produces a welfare score as output. To describe a housing system the user must fill out an input form, specifying the attributes of the housing system such as space allowance, social conditions, and substrate availability. Descriptions of the seven main housing systems for pregnant sows are incorporated into the decision support system as a benchmark for providing input as well as for interpreting output.

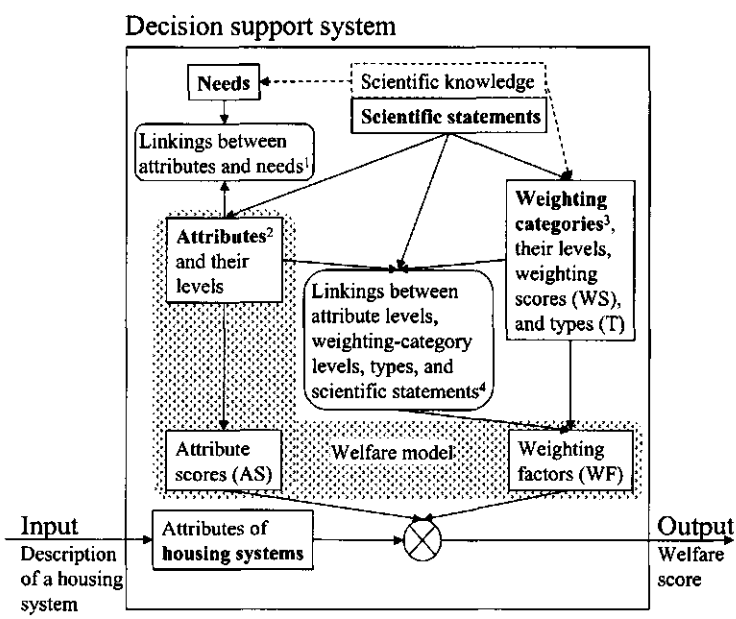


Figure 1. Decision support system for farm-animal welfare assessment implemented as a relational database with linked tables. Primary tables are shown in bold (see text). The welfare model covers the dotted area. Dotted lines indicate implicit (not formalised) relationships in the present decision support system. [1] Links between attributes and needs are illustrated in Figure 2; [2] The attributes are shown in Table 1; [3] The weighting categories are shown in Table 2; [4] Links between attribute levels, weighting-category levels, types, and scientific statements, which are used to determine weighting factors, are illustrated in Table 3.

The decision support system is implemented in Microsoft Access 97, which is a relational database that stores information in tables that are linked to each other (Date, 1995; Bracke et al., 1999b). Calculations are performed on the information in these tables using operators and so-called queries that allow combining the information from the tables by selecting specified data sets from them. The five primary tables in the decision support system contain scientific statements (1), a list of needs (2), attributes (3), weighting categories (4), and housing systems described by their attributes (5) respectively. Two secondary tables contain respectively links between attributes and needs, and links between attribute levels, weighting-category levels, types and scientific statements (Figure 1). The welfare model is constructed from the information contained in the first four primary tables. It consists of attributes, their levels, attribute scores and weighting factors. When the attributes of a housing system are described in the fifth primary table, the model assigns attribute scores and weighting factors, and calculates a welfare score as a weighted average of the attribute scores.

The knowledge base of the decision support system is filled with 352 statements from 12 different sources. These statements are all relevant for overall welfare assessment in pregnant sows. They cover the various relevant scientific disciplines such as ethology, stress-physiology, and veterinary science. The main sources are the EU-report on pig welfare (Scientific Veterinary Committee, 1997; 100 statements), Fraser and Broom (1990; 90 statements), Whittemore (1998; 50 statements), Hill and Sainsbury (1995; 35 statements), and Appleby and Hughes (1997; 29 statements). Together the statements contain 211 references with 1984 being the median publication year.

Modelling of attributes

Attributes are descriptors of housing and management systems. Attributes have two or more levels that specify the properties of the housing system. The model contains 37 attributes that together determine the overall degree of need satisfaction and frustration of the animals (Table 1). Some model attributes are environment-based (e.g. 'space per pen'), others are animal-based (e.g. 'health & hygiene status'), or management-related (e.g. 'mixing management'). The model attributes have between two and eight levels (3.9 on average), which are mutually exclusive, discrete classes that describe the (welfare-relevant) properties of housing systems. For instance, the attribute 'space per pen' has eight levels, which range from '1-1.5 m^2' to '> 6250 m^2' per enclosure.

We formulated a first set of attributes based on our own expertise, which derived from a study of the scientific literature and interviews with experts (Bracke et al., 1999c; 1999d). We then refined this set using the procedure described below. The procedure has three sub-procedures, which ensure that the attributes are qualified for welfare assessment as intended with the model. For this purpose, each attribute must relate to welfare as perceived from the animal's point of view, apply to the model's domain of housing systems, have an explicit scientific basis, and allow the calculation of welfare using a generic calculation rule. Together the attributes must cover welfare overall with a minimal degree of overlap. The three sub-procedures concern the relationship between the attributes on the one hand, and the housing systems (1), the concept of welfare (2), and the scientific basis (3) for welfare assessment on the other hand.

Table 1. List of attributes in the model (sorted according to their weighting factor, WF), the identification number (ID), the presumed best and worst level for welfare (excluding minimum-requirement levels), and the number of distinctly ranked levels (L). Column 'MR' identifies (with 'x') attributes that have a minimum-requirement level. Column 'Effect' shows the maximum number of welfare points with which the attribute can affect the overall (relative) welfare score.

ID	Attribute	Best level	Worst level	L	MR	WF	Effect
1	Space per pen	> 6250 m^2	1-1.5 m^2	8		25.8	1.73
2	Health & hygiene status	high	below average	3	x	23.8	1.60
3	Feeding level	elevated	low	4	x	23.0	1.54
4	Exposure to cold	e.g. heated unit	e.g. outdoor	3	x	22.0	1.48
5	Foraging & bulk	grazing/browsing	concentrates only	4		21.8	1.46
6	Space per sow	>10 m^2	1-1.5 m^2	4	x	21.4	1.44
7	Social stability	family groups	dynamic	4		20.4	1.37
8	Social contact	3-7 sows	auditory isolation	7		18.6	1.25
9	Food agonism	free-range/stalls	competitive feeding	4	x	17.8	1.19
10	Rooting substrate	> 5 cm	nose rings	4		15.4	1.03
11	Transport & penning	no	yes	3		14.8	0.99
12	Handling & fear	pleasant	unpleasant	3	x	14.6	0.98
13	Pain	no/little	some	3		13.2	0.89
14	Synchronisation	simultaneous	sequential	4		13.0	0.87
15	Water availability	ad lib	no water after feeding	2	x	12.6	0.85
16	Separate rest-elimination area	separate	not	4		11.8	0.79
17	Exposure to heat	minimal	e.g. only deep straw	3	x	11.2	0.75
18	Scratching	post present	e.g. tethered	3		10.4	0.70
19	Resting comfort	soft	hard	2		9.4	0.63
20	Air quality	as outdoor	poor	4	x	9.4	0.63
21	Mixing management	no (family)	e.g. small pen on slats	6		9.0	0.60
22	Food rationing	individual	cannot top up	3		8.0	0.54
23	Activity rhythm	bifasic	not bifasic	3		8.0	0.54
24	Food palatability	high	low	3		7.2	0.48
25	Movement comfort	grip	slippery (slats/wet)	2	x	7.0	0.47
26	Nestbuilding (resting nest)	yes (can)	no	3		7.0	0.47
27	Space to rest & eliminate	plenty	little	3		6.2	0.42
28	Social obstructions	none	possible	3	x	5.8	0.39
29	Novelty per week	at least 3 events	0-0.25 events	3		5.4	0.36
30	Visually isolated areas	at least 2	0	3		5.2	0.35
31	No. of food items	3	1	3		5.0	0.34
32	Light	day-light	dark	3		5.0	0.34
33	Huddling	yes (can)	no	2		4.8	0.32
34	Wallowing	mud pool	no	2		4.6	0.31
35	View & cover at rest	yes	no	3		4.4	0.30
36	Separate feeding-elimination area	yes	no	3		3.8	0.25
37	Separate rest-feeding area	yes	no	2		2.4	0.16

The first sub-procedure involved designing the levels of each attribute to be mutually exclusive and together exhaustive to cover the model's domain, which includes the wide range of farm types suited for agricultural production. As a result, all housing systems in the domain can be described with exactly one, and never more, nor less, than one level of each attribute. This ensures also that a generic calculation rule, i.e. calculating welfare as a weighted average score, can be used such that any welfare advantage ascribed to a housing system accrues to all the systems with the same descriptive property, and only to them.

Not only are the levels of each attribute mutually exclusive, but the model attributes are also as much as possible defined to be mutually exclusive in order to avoid double counting. For example, although the attributes 'space per pen' and 'movement comfort' may appear to overlap, they have been defined in the model with specific reference to space for locomotion, and grip provided by the floor, respectively.

The attributes and levels necessarily derive from the current state of housing for pregnant sows and the current state of science. Since the model's domain also includes novel housing systems with attributes that cannot be described at present, the levels of each attribute can be interpreted as equivalence (as-if) classes. For example, 'space per pen' has '> 6250 m^2' as its best level. Equivalent conditions may be achieved with a novel housing concept, such as regular exercise on a treadmill. Although research will first have to confirm this before it can

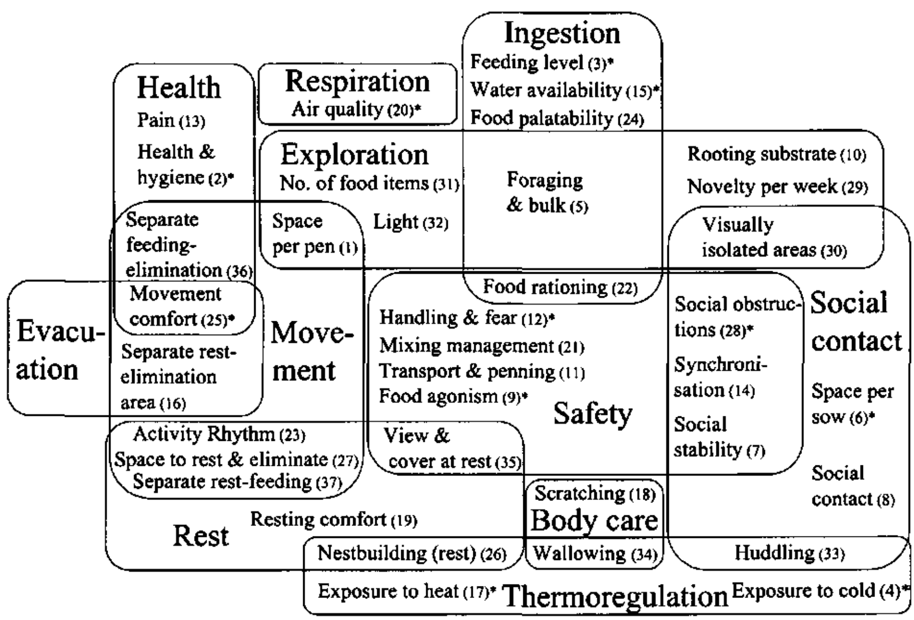


*Figure 2. Diagram showing how the attributes in the model are linked to the 11 (welfare) needs that cover pregnant sow welfare overall (needs from Bracke et al., 1999c). Attribute identification numbers of Table 1 are included between brackets. * Identifies attributes that contain a minimum-requirement level for welfare.*

be incorporated formally into the science-based model, the model can presently be used to evaluate the impact of such research on the overall welfare status by allowing attribute levels to be interpreted as equivalence classes, i.e. when the treadmill is regarded as equivalent to '> 6250 m^2' of pen space.

The second sub-procedure involved the linking of each attribute to at least one of the 11 needs, which we had previously formulated as the main aspects of welfare as perceived from the animal's point of view (Bracke et al., 1999c; Figure 2). For example, the attribute 'feeding level' is linked to the 'ingestion' need, because it covers the energy level of the feed and the body condition of the sows to indicate how hungry the sows are. The linking of attributes to needs ensures that each attribute in the model is relevant for welfare from the animal's point of view. It also ensures that welfare is covered overall, because all needs have at least one attribute, and it ensures a proportional distribution of attributes across the needs, because there is no major over- or under-representation of any single need as judged by the number and weights of the attributes assigned to it (cf. Figure 2; method derived from Streiner and Norman, 1995, p. 21). Furthermore, the attribute-need links help to define the attributes, and to minimise the overlap between them to avoid double counting. Note that the model calculates welfare directly from the attributes. Therefore, overlap of needs, as shown in Figure 2, is not a problem, but overlap of attributes would result in double counting. For example, the attributes 'space per pen' and 'space per sow' appear to overlap. However, 'space per pen' is defined as space per pen for the needs 'movement' and 'exploration', whereas 'space per sow' is defined as space for the 'social contact' need (cf. Figure 2). In this way, linking attributes to different needs helps to reduce the overlap between them. This example illustrates, also, that an attribute such as 'space per pen' may be multifunctional, i.e. linked to more than one need, but only when a conflict between the different functions (needs) cannot arise. The two requirements, no overlap and no conflict of functions, resulted in the disqualification of certain physical attributes such as 'straw' and 'soil' as attributes in our model.

'Straw' may serve as an illustration. Straw is multifunctional and its functions can conflict, because when it is hot, straw may be negative for thermal comfort, while it remains being positive for exploration (Fraser, 1985). The solution to distinguish between 'straw for thermal comfort' and 'straw as rooting substrate' would solve the problem of conflicting functions, but it runs into the problem of creating overlap with other attributes such as 'soil as rooting substrate'. This is because without scientific evidence to the contrary, we presume that what matters to sows as regards their need to root is that they have rooting substrate (attribute 10), not whether that substrate is straw or soil. (In fact we did collect scientific statements that allow making a distinction between different substrates, but this is irrelevant for the present illustration.) The solution is to identify the attributes 'straw' and 'soil' as factors that affect the state (level) of one or more model attributes. Influencing factors, as we will call such attributes that affect model attributes, were listed in the description of each model attribute in the decision support system (together with a reference to the scientific statements, see below). This means that our functional decomposition of welfare is as follows. For welfare assessment we de-composed welfare into 11 needs, which were de-composed into 37 attributes in the model (Figure 2). In turn, the model attributes received influencing factors, but the relationships between model attributes and influencing factors was only listed descriptively, and was not (yet) formalised in separate tables in the relational database. This awaits further data modelling.

In the third sub-procedure the model attributes (through their levels) were linked to the scientific statements in the database. These statements had previously been selected from the literature as being relevant to distinguish between housing systems on welfare grounds. The procedure requires that all statements are linked to at least one attribute level. This provides a scientific basis to the model, because it specifies the meaning of the attributes (1), weights them (2), and identifies the attribute levels (3).

To specify the meaning of the attributes on average 4.9 scientific statements are used per attribute. These statements specify the influencing factors that affect the model attribute and its levels. For example, the level 'ad lib water availability' (attribute 15) is specified by scientific statements about the amount drunk and the drinking speed of sows under ad lib conditions (between 17 and 21.5 litres per day; 3 litres per minute; Hill and Sainsbury, 1995, p. 235). In Table 1 short descriptions are given of the attributes and their best and worst level. The full description of all 37 attributes in the model adds up to 7,500 words, i.e. on average over 200 words per attribute. Such descriptions also include a specification of other attributes in the model from which the described attribute must be distinguished (to avoid overlap).

For attribute weighting on average 25 scientific statements were used per attribute. Scientific statements can be used to determine the weight of an attribute, because they provide evidence of relationships between the attribute and welfare performance criteria, which we have classified into weighting categories. For example, 'pigs will work for access to earth ... (Hutson and Haskell, 1990)' (Scientific Veterinary Committee, 1997, p. 69) is a scientific statement which says that (but not how much) pigs will work for rooting substrate. This gives some weight to the attribute. Further weight is 'added' by the statement that 'Matthews and Ladewig (1994) produced demand curves for access to bedding material and found that they [the demand curves for bedding, MB] were only second to food in extent of demand' (Scientific Veterinary Committee, 1997, p. 69; cf. bottom of Table 3; the weighting procedure will be described in more detail in the next sections).

For identification of attribute levels on a scientific basis the scientific statements were used to formulate each level with reference to at least one scientific argument. For example, the distinction between two of the levels of attribute 'rooting substrate', namely 'no substrate' and 'nose rings in the presence of substrate', is based on the statement that pigs with nose rings are likely to be thwarted more than pigs without substrate (statement from Scientific Veterinary Committee, 1997, p. 50). A conclusive scientific basis is also required for the ranking of the levels within each attribute (from the best to the worst level). When the scientific evidence was not conclusive, one or more levels were disqualified until all attribute levels and their rankings had a scientific basis.

Since each scientific statement that has been collected, is linked to at least one attribute or attribute level, this third sub-procedure ensures that all concepts encountered in the scientific statements are incorporated into the set of 37 model attributes. Both the linking of the attributes to needs and to scientific statements, therefore, contribute to ensuring that the model assesses welfare overall, rather than only partially.

Weighting categories and types

The procedure described above shows how the attributes have been modelled for overall welfare assessment based on the domain of housing systems, the list of needs, and the scientific statements. We will now further explain the scoring and weighting procedures, in particular how the responses of the animals, as measured by science and described in scientific statements, were used for weighting the attributes using so-called weighting categories and types.

If we succeeded in attribute modelling, then we can regard each attribute as an additive component of welfare with its own scale. The levels of an attribute identify the points on this attribute scale, where each level receives an attribute score (AS). In our model the worst level of each attribute received a score of 0, the best level received as score of 1, and any intermediate levels received intermediate attribute scores in direct proportion to their ranks. As a result an attribute with three levels receives attribute scores of 0, 0.5, and 1. An attribute with four levels has 0, 0.33, 0.67, and 1 as its attribute scores.

For weighting of the attribute scores across attributes we used a list of 12 weighting categories that were linked to the attribute levels and their scientific statements. The weighting categories classify welfare performance criteria, which have been measured in the various welfare disciplines, namely veterinary science (with the weighting categories 'pain' and 'illness'), evolutionary biology ('survival' and 'fitness'), stress-physiology ('HPA', i.e. hypothalamic-pituitary-adrenocortical axis, and 'SAM', i.e. sympathetic-adrenal-medullary activation), and ethology ('aggression', 'abnormal behaviour', 'frustration and avoidance', 'natural behaviour', 'preferences' and 'demand'; see Table 2, also for a brief explanation of each weighting category). Weighting categories may be regarded as the independent variables of empirical research, while the attributes in the model are the dependent variables. Scientific statements describe a relationship between the two kinds of variables. In virtue of these relationships weighting categories can be used to weight the attributes: the more scientific evidence is available showing the positive and negative welfare consequences of an attribute, the higher its weighting factor (which will be defined more formally below). Most weighting categories define a negative contribution to welfare. Positive contributions are made by the weighting categories of 'natural behaviour', 'preferences' and 'demand'.

Each weighting category has been assigned four levels that have weighting scores (WS) attached to them. These weighting scores are positive for positive weighting categories and negative for negative categories (see Table 2, column 'Range of WS').

One of the four weighting-category levels conveys a very high weighting score (set at 10,000 in the decision support system). When this level is assigned to an attribute (based on a scientific statement), this attribute turns into a minimum requirement for welfare, i.e. the welfare status is low, no matter what else is true about the housing system. The weighting score of 10,000 results in an overall welfare score for the housing system as a whole of less than 0.5 on a scale from 0 to 10.

The three remaining levels per weighting category assign weighting scores of either 1, 2, and 3 points, or of 1, 3, and 5 points, which have been assigned based on the dimensions of intensity, duration and incidence (Willeberg, 1991). With these dimensions we judged 'pain', 'illness' and 'HPA' to be the most negative weighting categories. The most positive one was 'demand' (see Table 2). It follows that when we compare the two positive weighting categories 'demand' (with weighting scores 1, 3, and 5) and 'preferences' (with weighting scores 1, 2, and 3), we find two pairs of levels with equal weights, namely those with a weighting score of 1, and those with a weighting score of 3. The basis for this normative judgement lies in the above mentioned dimensions of intensity, duration, and incidence (cf. also Anon., in press).

Table 2. Weighting categories as used in the weighting procedure of the model with their brief description, the range of weighting scores covered by the levels of each weighting category (range of WS), and the number of scientific statements in the database linked to each category (No. of stats).

Weighting category	Brief description	Range of WS	No. of stats
Pain	Evidence of pain including lameness and skin lesions, e.g. from aggression.	-1 to -5	26
Illness	Evidence of health problems, including increased mortality, but excluding lameness, skin lesions, and specific survival aspects.	-1 to -5	43
Survival	Evidence of reduced survival related to physiological requirements (other than through specific health problems), e.g. longevity, deprivation of food or water, and a poor climate.	-1 to -5	7
Fitness	Evidence of decreased fitness (that is likely to indicate negative affect), including (re)production effects, but excluding specific survival aspects related to physiological necessities, HPA, and illness.	-1 to -3	71
HPA	Evidence of activation of the HPA (hypothalamic-pituitary-adrenocortical) axis indicative of distress.	-1 to -5	36
SAM	Evidence of SAM (sympathetic-adrenal-medullary) activation (indicative of negative affect), e.g. increased heart rate and (nor)adrenaline levels.	-1 to -3	10
Aggression	Evidence of increased aggression excluding skin lesions (cf. pain).	-1 to -3	52
Abnormal behaviour	Evidence of disturbed behaviour such as oral stereotypies, apathy, and disturbed sexual behaviour.	-1 to -3	34
Frustration and avoidance	Evidence of blocked behaviour or deprivation including willingness to work to avoid a treatment.	-1 to -3	62
Natural behaviour	Evidence of (potential positive reward from) behaviour as seen in (semi)natural conditions including time budgets and species specificity of that behaviour.	+1 to +3	35
Preferences	Evidence from preference tests and behaviour under other than natural circumstances, including rebound effects and anticipation.	+1 to +3	131
Demand	Evidence that animals spend effort to obtain a commodity, especially using operant conditioning.	+1 to +5	33

For weighting we did not only determine which weighting-category levels, and associated weighting scores (WS), were predicated by the scientific statements about each attribute level. We also registered the specific type (T) of the weighting category. Types identify the quality or nature of the scientific measurement. Types provide a further differentiation of the (type of) welfare performance criteria that have been classified in each weighting category. Examples of 'types' are 'duration' of the pain weighting-category, 'a measure of cortisol' of the HPA category, and 'stereotypic behaviour' (versus, for example 'abnormal sexual behaviour') of the weighting category 'abnormal behaviour' (see also Table 3). As will be explained in the next section, the number of unique types per weighting category is functional for weighting, because it represents the different types of scientific argument for weighting within each weighting category.

Calculation rules

The welfare model calculates a welfare score on a scale from 0 to 10 as a weighted average score from attribute scores (AS) and weighting factors (WF; cf. Figure 1). Important to note is the distinction between weighting scores (WS), which are properties of weighting-category levels, weighting factors, which are properties of entire attributes, and 'weights', which will be used below to denote properties of attribute levels.

Weighting factors (WF) express the importance of an attribute compared to the other attributes in the model. The importance of each attribute can be deduced from the difference between the welfare 'weight' of its best and worst levels, i.e. those levels that have attribute scores (AS) of 1 and 0 respectively. Weighting factors are calculated from the table that contains the links between the attribute levels, the weighting-category levels, types, and scientific statements using the following procedure.

For the best level and for the worst level of each attribute (AL_{best} and AL_{worst} respectively) the maximum weighting score, and the total number of distinct types are determined per weighting category.

Table 3 illustrates the procedure for the best level, '> 5 cm', of attribute 10 'rooting substrate'. This attribute has four levels. The best level specifies more than 5 cm substrate depth that is available for at least 4 hours per day. Two intermediate levels specify 'less than 5 cm at least 2 hours per day', and 'no rooting substrate' respectively. Its worst level is 'nose rings' in the presence of rooting substrate.

To determine the 'weight' of AL_{best}, firstly, the maximum weighting scores per weighting category ('natural behaviour', 'preferences' and 'demand') are determined (3, 3, and 3 respectively, marked with * in Table 3). Secondly, the number of unique types per weighting category is counted (1, 5, and 1 respectively, marked with # in Table 3). Thirdly, per weighting category the maximum weighting score is increased with 0.2 times its number of unique types (resulting in 3.2, 4, and 3.2 respectively). The factor of 0.2 is, somewhat arbitrarily, chosen to reduce the contribution of the number of distinct types. Five distinct types (as for the weighting category 'preferences' in Table 3) with a factor of 0.2 add 1 point to the weighting score (WS) of the maximum weighting-category level (3+1=4). Fourthly, the weighting points are summated over the weighting categories (3.2 + 4 + 3.4). This results in a 'weight' of 10.4 for the level '> 5 cm' (AL_{best}) of attribute 'rooting substrate'.

The rationale for the procedure is that the impact of any novel scientific statement mainly derives from its ability to identify a new weighting category (establishing a new welfare performance criterion) or from its ability to increase the weighting score of a weighting category (e.g. the work by Matthews and Ladewig (1994) described in S13 increased the WS of 'demand' from WS = 1 to WS = 3). Statements that fail to do either of these two things do not contribute to the 'weight' of the attribute level, unless they identify a novel type. For example, statement S6 in Table 3 'adds' the type 'anatomy and cognition' (pigs have a sensitive snout). By itself this statement conveys a 'preference' weighting score of 1, which does not contribute to this attribute level's 'weight' since a weighting score of 3 is already established by S12 (pigs have a strong preference to root). However, it does 'add' a novel

Table 3. Illustration of part of the table used to calculate weighting factors. The whole table contains links between attribute levels, weighting categories (Wcat), weighting-category levels with weighting scores (WS), types, and scientific statements (with their identification number, ID). The table here shows the links for the attribute level '> 5 cm', which is the best level (AL_{best}) of attribute 10 'rooting substrate'. The source from which each statement is quoted is also given.

Wcat	WS	Type	ID	Scientific statement
Natural behaviour	2	Duration#	S1	'Pigs [spend] a substantial period of the day foraging … when observed under semi-natural conditions (Hsia and Wood-Gush, 1983; Stolba and Wood-Gush, 1989).' (Scientific Veterinary Committee, 1997, p. 88).
Natural behaviour	3*	Duration	S2	'Much food searching is performed by rooting…' (Scientific Veterinary Committee, 1997, p. 15).
Preferences	1	#	S3	'…straw provides an outlet for … rooting …' (Scientific Veterinary Committee, 1997, p. 32).
Preferences	1	Using it #	S4	'They root in earth …' (Scientific Veterinary Committee, 1997, p.69).
Preferences	1	Using it	S5	'If sows cannot root sufficiently, i.e. 2 hours per day, then they will try to compensate by other exploratory behaviour (Sambraus 1982; Sambraus, 1993).' (Walter and Postler, 1994. p. 62).
Preferences	1	Cognition and anatomy#	S6	'The pig derives satisfaction from seeking feed and is well-equipped to do so with a high sense of taste and smell and a sensitive snout.' (Whittemore, 1998. p. 152).
Preferences	2	Using it	S7	'Wood-Gush and Beilharz (1983) showed that pigs in bare environments use earth a lot.' (Fraser and Broom, 1990. p. 365).
Preferences	2	Balanced#	S8	'Earth floors are much preferred to concrete floors … in studies by van Rooijen (1980, 1981, 1982) where floor preference was balanced against social attraction.' (Fraser and Broom, 1990. p. 365).
Preferences	2	Duration#	S9	'Pigs can be occupied by rooting material for up to 6 hours per day (Grauvogl 1989).' (Walter and Postler, 1994. p. 69).
Preferences	2	Using it	S10	'As opposed to other materials wood chips maintains its attractivity over months, as it provides back pressure during rooting.' (Walter and Postler, 1994. p. 69).
Preferences	2	Duration	S11	'With ample feed and ample straw for manipulation and rooting, pigs spend about 15% of their time in eating-related and rooting-activities (compared to 2% in stalls and tethers).' (Whittemore, 1998. p. 152).
Preferences	3*		S12	'Rooting is a behaviour which pigs have a strong preference to perform (Hutson 1989) …' (Scientific Veterinary Committee, 1997, p. 50).
Demand	1	#	S4	'…pigs will work for access to earth … (van Rooijen 1980, 1981, 1982, Hutson and Haskell 1990).' (Scientific Veterinary Committee, 1997, p. 69).
Demand	3*		S13	'Matthews and Ladewig (1994) produced demand curves for access to bedding materials and found that they were second only to food in extent of demand.' (Scientific Veterinary Committee, 1997, p. 69).

* Identifies the maximum weighting scores per weighting category; # identifies the unique types (T) per weighting category. (Note that empty cells are also counted.)

type ('anatomy and cognition') to the 'weight' of the attribute level '> 5cm' substrate, although it is only counted in a moderated way, namely by a factor 0.2.

The 'weight' of the best level (AL_{best}) of an attribute expresses the degree to which it is (relatively) positive for welfare, based on available scientific knowledge. The 'weight' of the worst level (AL_{worst}) expresses the degree to which that level is (relatively) negative for welfare. Therefore, the same procedure that was used for AL_{best} is used for AL_{worst}, except that the minima of the (largely negative) weighting scores are used, and that the number of unique types are subtracted, rather than summated. In the decision support system we obtained a 'weight' of –5 for the worst level 'nose rings' of the attribute 'rooting substrate'.

The weighting factor (WF) for an attribute as a whole is the difference between the 'weights' of its best and worst level. The weighting factor for the attribute 'rooting substrate' is 15.4 (= 10.4 - (-5)). In our model weighting factors for the 37 attributes range between 2.4 and 25.8 (Table 1, column WF).

More formally, the weighting factor (WF_i) of the i-th attribute in de model is calculated as

$$WF_i = \left(\sum_{wc} \left(\underset{wcl}{Max}(WS_{wcl}) + 0.2 \cdot NT_{wc} \right) \right)_{AL_{i,best}} - \left(\sum_{wc} \left(\underset{wcl}{Min}(WS_{wcl}) - 0.2 \cdot NT_{wc} \right) \right)_{AL_{i,worst}} \qquad (1)$$

where $AL_{i,worst}$ is the worst level, and $AL_{i,best}$ is the best level of the i-th attribute; WS_{wcl} is the weighting score assigned to the attribute level based on a scientific statement; wc identifies the weighting categories linked to the attribute level; wcl identifies the weighting-category levels (to which WS have been assigned) within one weighting category; NT_{wc} is the number of unique types per weighting category assigned to the attribute level.

Attribute scores are calculated for each level of an attribute on a scale from 0 to 1 proportional to their rank as

$$AS_{i,j} = \frac{NL_i - RL_{i,j}}{NL_i - 1} \qquad (2)$$

where $AS_{i,j}$ is the attribute score of the j-th level of the i-th attribute in the model; NL_i is the total number of levels of attribute i; $RL_{i,j}$ is the rank number of the j-th level of the i-th attribute, where levels are ranked for welfare from 1 for $AL_{i,best}$ to NL_i for $AL_{i,worst}$; $RL_{i,j} \in [1, NL_i]$. For example, an attribute with five levels has AS of 1, 0.75, 0.5, 0.25, and 0 for its levels ranked 1 to 5.

The 'absolute' welfare score for a housing system (AW_h, on scale 0 to 1) is calculated in our model as the weighted average of the attribute scores, i.e. as the sum of the attribute scores (AS) multiplied by the weighting factor (WF) of each attribute in the model, divided by the total sum of WF:

$$AW_h = \frac{\sum_{i=1}^{m}\left(AS_{i,h} \cdot WF_i\right)}{\sum_{i=1}^{m} WF_i} \qquad (3)$$

where $AS_{i,h}$ is the attribute score (between 0 and 1) of the level of attribute i that represents the (welfare-relevant) property of the housing system (h); WF_i is the weighting factor of the i-th attribute; m is the total number of attributes in the model i.e. 37; $i \in [1,m]$.

The 'absolute' scores (AW_h) represent the ratio of all positive and negative aspects of a housing system. AW_h covers the domain of logically possible housing systems, which is much wider than encountered in reality. Therefore, we transform the scores (linearly) to a scale that covers the actual systems. In this chapter we use the seven main housing systems for pregnant sows as a benchmark to define the (relative) welfare scores (W_h) on a scale from 0 to 10. These systems include tethered housing, individual housing in stalls, electronic sow feeding (ESF), group housing with free-access stalls, Biofix (i.e. trickle feeding), outdoor housing with huts on pasture, and the Family Pen system. A brief description and a reference for each system are given in Table 4. Together these systems cover most of the variation

Table 4. Reference housing systems for pregnant sows (including their main source of reference).

System	Main characteristics
Tethered	Individual housing; girth tethered; 1.2 m^2/sow; trough fed; partly slatted floor; no straw; indoors in thermocontrolled building (Baxter, 1984).
Individual stalls	Individual housing in stalls; 1.3 m^2/sow; trough fed; partly slatted floor; no straw; indoors in thermocontrolled building (Backus, 1997).
ESF	Group of 30 sows; mixed twice per pregnancy; 60 m^2/pen; 1 electronic sow feeder (ESF); partly slatted floor; no straw; indoors in thermocontrolled building (Backus, 1997).
Free-access stalls	Stable group of 12 sows with stalls for feeding and resting; 26 m^2/pen; partly slatted floor; no straw; indoors in thermocontrolled building (Backus, 1997).
Biofix	Stable group of 6 sows; 15 m^2/pen; trickle feeding (Biofix); partly slatted floor; no straw; indoors in thermocontrolled building (Backus, 1997).
Outdoor huts	Dynamic group of 10 sows; 5000 m^2/enclosure; floor feeding; wallowing pool; outdoors on pasture with isolated huts with straw (Pig Welfare Advisory Group, 1997).
Family Pen	Family group of 4 sows with offspring; 108 m^2/pen; fed from a trough with head partitions; straw provided in straw racks; rooting area with peat or bark; 'furnished' pen with rooting area, nesting area, activity area and dunging passage; open-fronted building (Stolba, 1981).

between attributes in the domain of current housing systems (Bracke et al, 1999d). Score transformation results in the worst of these seven systems (tethered system with an AW_h = 0.39 = AW_{min}) receiving a score of 0 and the best system (the Family Pen with AW_h = 0.74 = AW_{max}) receiving a score of 10. For score transformation the following formula is used:

$$W_h = \frac{10 \cdot (AW_h - AW_{min})}{(AW_{max} - AW_{min})} \tag{4}$$

where W_h is the relative welfare score for a housing system ($_h$, on a scale from 0 to 10); AW_h is its 'absolute' welfare score (scale 0 to 1); AW_{min} and AW_{max} are the 'absolute' scores for the worst and best reference systems respectively.

Although most housing and management systems are expected to fall within the range of 0 to 10, in theory systems can get a much higher and a much lower score. Housing systems with AW_h = 0 and AW_h = 1 would receive a W_h of −11.1 and +17.4 respectively. These are logically possible systems with only positive or only negative aspects. In reality, however, housing systems tend to have both positive and negative attributes. This is also true for the best (Family Pen) and worst (tethered) housing system in our set of seven reference system, but the proportion of positive attributes is larger for the Family Pen system.

Besides providing a benchmark to define the range of the relative scale (W_h), the seven main systems for pregnant sows also provide reference points to interpret welfare scores calculated for other housing systems.

Some model calculations

Compound attributes

In the section 'modelling of attributes' we explained that some features of housing systems have been identified as 'influencing factors'. With the decision support system we can quantify the welfare impact of such 'influencing factors' and other attributes that affect one or more model attributes. We will refer to such influencing factors as 'compound attributes'. For this, we must first specify the relationships between the compound attribute and the model attributes. Its welfare impact (which is a weighting factor) is then calculated as the difference in welfare points between its best and worst level using the relative welfare scale between 0 and 10.

We calculated the welfare impacts of several compound attributes, such as stockmanship, behavioural restrictions, and outdoor access, to serve as examples.

Stockmanship was the most important compound attribute, with a welfare impact of 9.4 points. To calculate this impact 'stockmanship' was defined to include specific management-related attributes (e.g. food rations, social stability and mixing management) as well as regimes for providing basic needs such as food, water, thermal comfort and an adequate health status. Also important compound attributes were behavioural restrictions (7.7 welfare points), and space quality (6.3). The separation of functional areas and outdoor access per se were relatively unimportant, with 1.6 and 0.8 welfare points respectively. However, it should be noted that we used a narrow definition of 'outdoor access per se', i.e. we used only its impact on the model attributes 'air quality', 'visually isolated areas', and 'light', and we disregarded the extra space that is usually provided with it. This shows that the welfare impact of an attribute may be much affected by the way it is defined in terms of the attributes in the model.

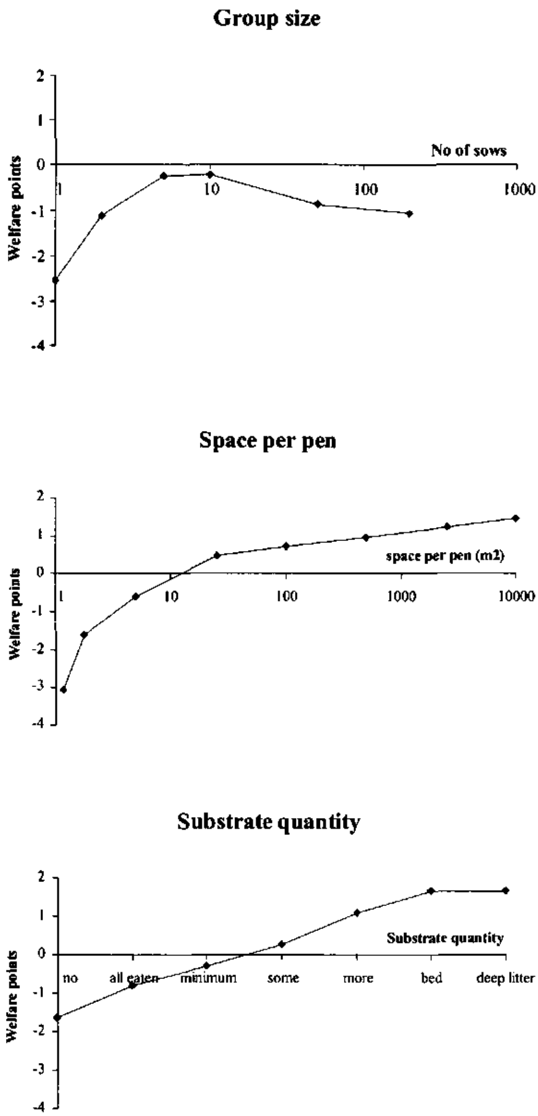


Figure 3. Welfare impact of three compound attributes. 'Group size' was defined in terms of the model attributes 1, 7, 8, 22, and 33; 'space per pen' in terms of the attributes 1, 6, 16, 27, 36, and 37; and 'substrate quantity' in terms of the attributes 5, 10, 19, 26, 29, and 31. Note that for 'group size' and for 'space per pen' a logarithmic scale is used. The 0-point on the y-axis derives from the positive and negative weighting scores (cf. Table 2) of the underlying model attributes.

The model also allows the calculation of dose-response relationships between the separate levels of a compound attribute and welfare, expressed in welfare points. These relationships need not be linear, due to the differential effects of scientific evidence in relation to each level. The welfare points are calculated from the welfare 'weights' of its levels. Since different 'amounts' of scientific evidence are available for the 'weights' of the different levels of a compound attribute, its dose-response curve will often not be linear. Figure 3 shows the dose-response relationships for the compound attributes 'group size', 'space per pen' and 'substrate quantity'. These attributes have welfare impacts of 2.3, 4.5, and 3.3 respectively. 'Substrate quantity' shows a substantial increase when little substrate is provided. With larger amounts of substrate the contribution to welfare is tailing off. 'Space per pen' shows a sharp increase at low space allowances and the curve tails off at larger space allowances. This compound attribute has an impact of 4.5 points. It is defined in relation to the model attribute with the same short descriptor 'space per pen' (which has an effect of only 1.7 welfare points, Table 1), but it also relates to other model attributes such as 'space per sow' and separate functional areas (involving the attributes 6, 16, 27, 36, and 37). For 'group size' we presumed a constant space allowance of 2 m^2 per sow and feeding stalls. Its curve shows an optimum value for small groups. Very large groups have a sub-optimal value for welfare, mainly because pigs have a natural tendency to live in small groups (weighting category 'natural behaviour').

The optimum value for small groups remains below the 0 line in Figure 3, because we presumed 2 m^2 per sow and normal husbandry conditions as regards mixing frequency in relation to group size. As a result, small groups received positive points for social contact (attribute 8), but negative points for relatively small pens (at a space allowance of 2 m^2 per sow) and for the mixing frequency of once per pregnancy (attribute 7). The model contains more negative weightings than positive ones, because more knowledge is available about negative welfare performance criteria than about positive ones (cf. column 'WS' in Table 2).

The curves in Figure 3 show how the levels of compound attributes affect welfare according to the model. They summarise the state of the knowledge base underlying the scores. They can be used to identify shortcomings in the knowledge base, and to make cost-benefit analyses of measures that are intended to improve welfare.

Comparison with other models

We compared the results of our model with results from three models for welfare assessment in pregnant sows derived from the literature. These are Fraser's Behavioural Deprivation Index (BDI, Fraser, 1983) and two versions of the Tiergerechtheitsindex (TGI), i.e. one version by Sundrum et al. (1994), and one version by Bartussek (1999b). The models were interpreted to apply at the housing-system level in order to allow comparison with our model.

The output of the different models is shown in Figure 4. The Spearman rank correlation coefficients (Rho) between each pair of models were all larger than 0.85 ($P < 0.05$). Kendall's coefficient of concordance for the four models is 0.94 ($P = 0.001$).

The models largely agree as to which are the best systems (Outdoor huts and Family Pen), and which are the worst systems (Tethered and Individual housing in stalls). More variation is found in the ranking of the mid-welfare systems. Nevertheless, the general pattern for the seven systems appears to be a-a-b-b-b-c-c (with $a < b < c$), where individual intensive systems obtain a substantially lower score than intensive group-housing systems indoors without substrate, which receive a substantially lower score than more extensive group housing systems with substrate and outdoor access. This is also what we have found in a previous study where we consulted pig-welfare scientists (Bracke et al., 1999d). As intended, our model correlates well with these previous findings (Spearman's Rho is 0.86, $P < 0.05$).

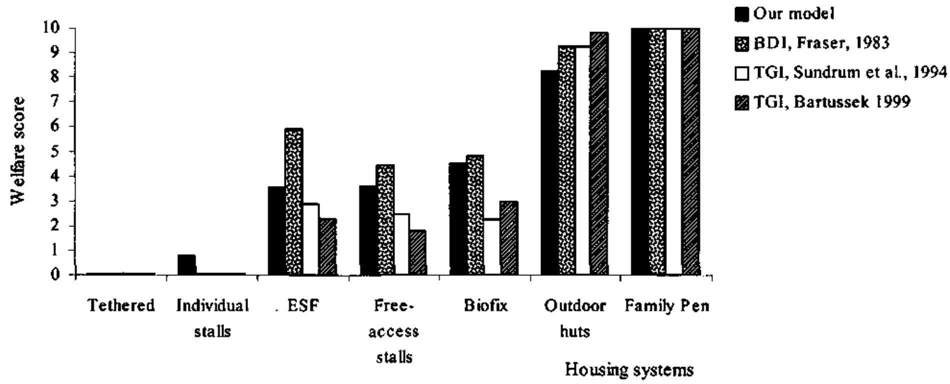


Figure 4. Our welfare model compared with three models derived from the literature. For each model the best and worst system were set at welfare scores of 0 and 10 respectively. Housing systems are sorted according to the scores calculated with our model. ESF: electronic sow feeder system.

Flexibility and some 'sensitivity analysis'

The model is embedded in a computer-based decision support system, which is designed to be flexible, i.e. to be adaptable when new knowledge about welfare becomes available. The flexibility of the decision support system covers a range of aspects. An experienced user may want to modify (add, delete or change) scientific statements, the calculation procedure, the attributes in the model, descriptions of housing systems, weighting categories, or the list of needs. He (or she) may also want to produce a number of different versions of the model and compare the output. The user may also construct his own model, or insert a model derived from the literature, as we have done for the three models described above. Such flexibility allows a quantitative assessment of criticism. We will illustrate this point with reference to Table 3, which illustrates the weighting procedure.

An objection may be that it is arbitrary to weight the number of types by a factor of 0.2. With the decision support system we can calculate the effect of changing this factor. We compared the present results with model variants where this factor was set at 0 and 1 respectively. The largest deviation in overall welfare scores for any of the seven housing systems was only 0.23 points on the 10-point scale.

Would it, then, be legitimate to exclude this factor from the calculation procedure altogether, e.g. to reduce its complexity? We may even ask whether the whole weighting procedure may be redundant, because the correlation between the weighted and unweighted versions of the model is relatively high (Spearman's Rho is 0.89, $P < 0.01$). Theoretically, the correlation may be as high as 0.99, using an equation from Gulliksen (1950). However, even if weighting was found to be empirically redundant (Dalkey, 1975; Wainer, 1976), it certainly is not redundant from a conceptual point of view (Streiner and Norman, 1995, p. 86). Since our goal was to assess welfare as much as possible in accordance with knowledge of the biology of the animals, we cannot assess welfare without a weighting procedure, as we must obviously distinguish between the more and the less important attributes for welfare (cf. luxuries and necessities, e.g. Dawkins, 1990). Similarly, we cannot discard the weighting of the number of types in the calculation procedure, because new knowledge that 'adds' a new type to an attribute level conceptually gives it a somewhat higher weight. In other words, setting the 0.2 factor at 0 is illegitimate.

Another objection to Table 3 is that the weighting categories ('natural behaviour, 'preferences' and 'demand') largely overlap. Maybe they should be regarded as different levels of one weighting category.

We identified these categories as three different weighting categories, because they involve three different paradigms of welfare research. With little effort (less than 15 minutes) we produced a version of the model that summarised the three weighting categories into one. The maximum effect was a (downward) deviation of only 0.41 points for the system 'Outdoor huts'. Larger effects were found when these specifically behavioural weighting categories were deleted completely from the weighting procedure. This resulted in a maximum deviation of 1.8 points (for 'Outdoor huts').

These calculations illustrate the flexibility of the decision support system, and its use in evaluating points of criticism in a quantitative way. Especially when large effects are found, closer evaluation and possible upgrading of the model are warranted.

Discussion

In this section we will discuss the quality of the model following the steps taken to construct and apply it. This will illustrate that we now have a formal procedure for transparent, step-by-step overall welfare assessment that may be used for systematic reasoning and debating about animal welfare.

This chapter shows that welfare can be assessed without involving any ethical questions, although it does involve normative issues (cf. Fraser, 1995), especially in dealing with uncertainties in knowledge. For integrated welfare assessment many decisions were needed to identify the 352 scientific statements, 145 attribute levels, 11 needs, 12 weighting categories and the many links between them. In doing so, many hidden assumptions had to be made explicit, e.g. the assignment of weighting scores made explicit how we weighted the weighting-category levels. Most of these decisions were somewhat arbitrary (cf. Hurnik, 1988), but we showed how our model can be used to determine the impact of such arbitrary choices by doing a sensitivity analysis. Our technical solution is not the most elegant, but it works: we can calculate welfare scores that correspond for the most part with other models as well as with previously obtained expert opinion (Bracke et al., 1999d). This means the model is ready for a more explicit validation test, even though the model can always be improved further. The decision support system, in which the model is embedded, is developed to accommodate this course of 'evolution'. The value of such a system is that it may contribute to increase the objectivity and intersubjectivity in welfare assessment. Welfare assessment on the basis of the biological needs of animals and the findings from empirical research guides the way to an assessment of animal welfare as perceived from the animal's point of view, even though such assessment must necessarily remain an assessment performed by humans (Bekoff et al., 1992).

Model construction

For model construction it is important to demarcate the scope of the model. Our model concerns assessment of the welfare status of pregnant sows in relation to their housing and management system based on available scientific knowledge. The model's domain includes the wide range of present and future housing and management systems for pregnant sows. The model is designed primarily to distinguish between different types of housing systems, rather than between individual farms within systems.

A list of hierarchically organised needs was used to break up the concept of overall welfare into manageable chunks. Further functional decomposition resulted in the formulation of a list of weighted attributes, which identify the welfare-relevant properties of housing systems through the relationship between the attributes and welfare performance criteria (weighting categories) as specified in the scientific statements. These attributes are the proverbial apples and oranges that are 'added' into the mixed fruit basket of overall welfare.

We carefully selected attributes covering all aspects of welfare while minimising the (conceptual) overlap between them. We did not minimise (empirical) correlations between attributes. The average inter-attribute correlation was considerable (Crohnbach's Alpha was 0.79 for the 37 attributes and the seven reference housing systems, see e.g. Nunnally, 1970). As a result, when we change only one attribute of a housing system there will often be multiple effects in the model. This was illustrated above for influencing factors and compound attributes. Empirical correlations between the model attributes may contribute to the rather stable model performance when weighting factors are changed. For example, when all weighting factors are set at 1 the Spearman rank correlation coefficient with the original model is still 0.89 ($P < 0.01$).

Attribute scores for the levels of an attribute were determined in proportion to their welfare rank. An objection may be that the model transforms ordinal scales to interval scales. Although for setting the relative distance between the attribute levels we could have used a procedure similar to the one we used for the weighting of attributes as a whole, we used the simplifying assumption of proportionality, because it is expected to have only minor effects on the overall scores as it concerns only a fraction of the attribute effect shown in Table 1.

Minimum-requirement levels identify those properties of a housing system below which welfare is poor no matter what. Because we used discrete levels, these minimum requirements have relatively sharp cut-off points. In reality, the cut-off points are rather fuzzy. Techniques such as fuzzy logic (e.g. Bardossy and Duckstein, 1995) with membership curves that fuzzify the cut-off points may be used to improve this aspect of the model.

For weighting beyond minimum requirements we identified a number of weighting categories. The correlation between the weighting factor of an attribute and this attribute having a minimum-requirement level attached to it is low but significant (Spearman rank correlation is 0.38, $P < 0.05$; data from Table 1). The effect is mitigated by the fact that some attributes with a high weighting factor have their minimum-requirement level covered by another attribute, e.g. 'space per sow' in the case of 'space per pen'. An other reason why an attribute may have a high weighting factor without a minimum-requirement level is that it may affect welfare positively rather than negatively, as is the case for e.g. 'social contact', which is defined to exclude the aspect of agonism.

The weighting categories identify the various welfare performance criteria that have been measured in the various welfare-science disciplines. Some of these disciplines have been disputed, e.g. preference testing (Dawkins, 1983), and stress-physiology (Rushen, 1991). Our aim was not to settle these disputes, but to find a way that allowed using the relevant findings from each of these disciplines. The attributes in the model often specify environment-based design criteria. Our model is the first to spell out how environment-based attributes can be used for overall welfare assessment, namely in virtue of their effects on welfare performance criteria as described in the scientific statements. Previous approaches have either focussed predominantly on identifying and measuring performance criteria (e.g. Broom and Johnson, 1993, and many others), or they have focussed on modelling environment-based attributes without explicitly showing how already available knowledge is taken into account (e.g. TGI indexes mentioned above). Our approach may provide a bridge between the comprehensiveness of the latter approach with the scientific credibility of the former.

To calculate welfare scores we used an additive calculation rule. As a result, the contribution of an attribute to welfare is constant and independent of the state of the other attributes. However, interactions such as between space quantity and quality, and between feeding level and thermal comfort, would seem to be relevant. For the most part we reduced the impact of interactions by carefully describing the attributes. For example, the attribute 'exposure to cold' has included in its description an influencing factor 'feeding level'. We presumed that the effects of interactions, which remained despite careful description, are negligible. This assumption may be false, but incorporating interactions in the model would have increased its complexity. The basic assumption of additive calculation seems warranted, as it is commonly used (cf. Keeny and Raiffa, 1976; Huirne and Hardaker, 1998). Models of stress often use multiple stressors that are supposed to act in an additive way. Furthermore, empirical evidence is available that stressors may act additively (Webster, 1995, p. 120) and that compensation is possible. For example, studies with chicks (McFarlane et al., 1989; McKee and Harrison, 1995) and with growing pigs (Hyun et al., 1998a; Hyun et al., 1998b) have demonstrated additive effects with multiple stressors. It has also been shown that stress can be compensated by reward (e.g. van den Berg et al., 1999, 2000; Spruijt et al., in press). Pedersen et al. (1998) showed that positive handling of pregnant gilts could reduce the negative stress-physiological consequences of tethered housing. A biological basis for summation and compensation may lie in the capacity of animals to adapt. This capacity is substantial, but has its limits (e.g. Broom and Johnson, 1993; Barnard and Hurst, 1996). Our welfare model incorporates a two-tiered approach, using minimum requirements to set limits, but otherwise allowing compensation between welfare-relevant attributes.

Model application and utility

When the properties of a housing system from within the model's domain are known, the user may be able to determine attribute levels and calculate a welfare score in as little as five minutes. However, at present our decision support system requires an experienced user. Firstly, because he needs to understand the scientific statements and the model's weighting procedure. Secondly, because the attributes have been designed to avoid overlap, which resulted in somewhat technical definitions of attributes, weighting categories, and needs. Finally, because the decision support system requires expertise in using the software (MS Access). However, the decision support system includes assessments of the seven main housing systems for pregnant sows. This facilitates the application of the model, because it provides points of reference for model application. Further research will be needed to make the model suitable for laymen use.

At this point we cannot determine what difference in welfare scores would represent a definite difference in welfare. We estimate that this may be a difference of as much as 2 points on the scale from 0 to 10, as a retest showed an average absolute difference of 0.71 points with a range of 0.1 to 1.7. Comparison with other models (Figure 4) showed even larger differences, especially for the mid-welfare systems, but the present data do not support an evaluation of which is the better model.

Validation can take essentially three forms: (sensitivity) analysis, expert opinion, and empirical research. In this chapter we have shown some results of analytical sensitivity analysis, comparison with other models, and previous interviews with experts. The next chapter will discuss an extended validation using expert opinion. Empirical validation will be difficult because it requires extensive data collection on farms to allow comparison at the housing-system level. Further validation of the model will result from using it in practice and from upgrading it periodically.

The main strength of our work is that it shows how pregnant-sow welfare can be quantified using a systematic and transparent procedure covering all reasoning steps from basic assumptions, specifications of the model's domain, and available welfare knowledge, all the way to the interpretation of the welfare scores.

Conclusions

According to our study the main steps of a procedure to assess animal welfare are:

1. Identify the model's domain, i.e. determine the category of animals and the set of housing systems to which the model applies.
2. Define welfare and de-compose it into functional elements, i.e. a list of the biological needs relevant for the model's domain.
3. Collect the 'facts', in the form of the relevant scientific statements.
4. Determine the set of welfare-relevant attributes, which specify the different welfare advantages and disadvantages of housing systems, in relation to the model's domain, the list of needs, and the scientific statements.
5. Weight the attributes relative to each other based on welfare performance criteria as described in the scientific statements using the dimensions of intensity, duration, and incidence.
6. Sum up, i.e. determine the overall welfare status based on the weighted attributes of the housing systems, while using the main housing systems in the domain to serve as a benchmark to interpret the results.

With this procedure an operational decision support system to assess the welfare status of pregnant sows in relation to their housing and management system based on available scientific knowledge, has been developed. The overall welfare status is expressed as a score between 0 and 10, based on what is known about the biological needs of the animals. The decision support system shows how farm animal welfare can be assessed in an explicit and transparent way, with the flexibility to accommodate new insights about welfare assessment when these become available.

Acknowledgements

This work was funded by the Technology Foundation (STW) (through a joint grant from the Netherlands Organisation for Scientific Research and the Netherlands Ministry of Agriculture, Nature Management and Fisheries), and the Dutch Society for the Protection of Animals.

References

Anonymous. 2001. Scientists' assessment of the impact of housing and management on animal welfare. J. Appl. Anim. Welfare Sci. In press.

Appleby, M. C., and B. O. Hughes. (Eds). 1997. Animal Welfare. CAB International, Oxon.

Backus, G. B. C. (Ed.). 1997. Comparison of four housing systems for non-lactating sows (In Dutch). Report P.1.171. Research Institute for Pig Husbandry, Rosmalen.

Bardossy A., and L. Duckstein. 1995. Fuzzy Rule-based Modelling with Applications to Geophysical, Biological and Engineering Systems. CRC press, Boca Raton, FL.

Barnard, C. J., and J. L. Hurst. 1996. Welfare by design: The natural selection of welfare criteria. Animal Welfare 5:405-433.

Bartussek, H. 1999a. A review of the animal needs index (ANI) for assessment of animal's well-being in the housing systems for Austrian proprietary products and legislation. Livest. Prod. Sci. 61:179-192.

Bartussek, H. 1999b. Tiergerechtheit für leere und tragende Sauen. TGI 35L/1999-Zuchtsauen. BAL Gumpenstein.

Bekoff, M., L. Gruen, S. E. Townsend, and B. E. Rollin. 1992. Animals in science: Some areas revisited. Anim. Behav. 44:473-484.

Berg, C. L. van den, I. Kitchen, M. A. F. M. Gerrits, B. M. Spruijt, and J. M. van Ree. 1999. Morphine treatment during juvenile isolation increases social activity and opioid peptides release in the adult rat. Brain Res. 830:16-23.

Berg, C. L. van den, J. M. van Ree, and B. M. Spruijt. 2000. Morphine attenuates the effects of juvenile isolation in rats. NeuroPharmacology 39:969-976.

Bracke, M. B. M., B. M. Spruijt, and J. H. M. Metz. 1999a. Overall animal welfare assessment reviewed. Part 1: Is it possible?. Neth. J. Agric. Sci. 47:279-291.

Bracke, M. B. M., J. H. M. Metz, and B. M. Spruijt. 1999b. Overall animal welfare reviewed. Part 2: Assessment tables and schemes. Neth. J. Agric. Sci. 47:293-305.

Bracke, M. B. M., B. M. Spruijt, and J. H. M. Metz. 1999c. Overall animal welfare reviewed. Part 3: Welfare assessment based on needs and supported by expert opinion. Neth. J. Agric. Sci. 47:307-322.

Bracke, M. B. M., J. H. M. Metz, B. M. Spruijt, and A. A. Dijkhuizen. 1999d. Overall welfare assessment of pregnant sow housing systems based on interviews with experts. Neth. J. Agric. Sci. 47:93-104.

Bracke, M. B. M., J. H. M. Metz, A. A Dijkhuizen, and B. M. Spruijt. 2001. Decision support system for assessing animal welfare: Strategy and prototype. Agric. Envir. Ethics. Accepted.

Bracke, M. B. M., J. H. M Metz, and B. M. Spruijt. 2001. Development of a decision support system to assess farm animal welfare. Proceeding of the International Workshop on Assessment of Animal Welfare at Farm or Group Level, 27-28 August in Copenhagen. Acta Agric. Scand. Sect. A. Animal Sci. Suppl. In press.

Broom, D. M., and K. G. Johnson. 1993. Stress and Animal Welfare. Chapman & Hall, London.

Dalkey, N. C. 1975. A Delphi study of factors affecting the quality of life. In: H. A. Linstone, and M. Turoff (Eds) The Delphi Method: Techniques and Applications. pp 387-401. Addison-Wesley, New York.

Date, C. J. 1995. An Introduction to Database Systems (6th Ed.). Addison-Wesley, Reading, MA.

Dawkins, M. S. 1983. Battery hens name their price: Consumer demand theory and the measurement of ethological 'needs'. Anim. Behav. 31:1195-1205.

Dawkins, M. S. 1990. From an animal's point of view: Motivation, fitness, and animal welfare. Behav. Brain Sci. 13:1-61.

Dawkins, M. S. 1997. D. M. G. Wood-Gush Memorial Lecture: Why has there not been more progress in animal welfare research? Appl. Anim. Behav. Sci. 53:59-73.

Fraser, A. F. 1983. The behaviour of maintenance and the intensive husbandry of cattle, sheep and pigs. Agric. Ecosyst. Environ. 9:1-13.

Fraser, D. 1985. Selection of bedded and unbedded areas by pigs in relation to environmental temperature and behaviour. Appl. Anim. Behav. Sci. 14:117-126.

Fraser, D. 1995. Science, values and animal welfare: Exploring the 'inextricable connection'. Animal Welfare 4:103-117.

Fraser, A. F., and D. M. Broom. 1990. Farm Animal Behaviour and Welfare (3rd Ed.). Baillière Tindall, London.

Gulliksen, H. O. 1950. Theory of Mental Tests. Wiley, New York.

Hill, R. J., and D. W. B. Sainsbury. 1995. The Health of Pigs. Longman Scientific & Technical, Harlow.

Hughes, B. O., and I. J. H. Duncan. 1988 The notion of ethological 'need', models of motivation and animal welfare. Anim. Behav. 36:1696-1707.

Huirne, R. B. M., and J. B. Hardaker. 1998. A multi-attribute utility model to optimise sow replacement decisions. Eur. Rev. of Agic. Econ. 25:488-505.

Hurnik, J. F. 1988. Welfare of farm animals. Appl. Anim. Behav. Sci. 20:105-117.

Hyun Y., M. Ellis, G. Riskowski, and R. W. Johnson. 1998a. Growth performance of pigs subjected to multiple concurrent environmental stressors. J. Anim. Sci. 76:721-727.

Hyun Y., M. Ellis, and R. W. Johnson. 1998b. Effects of feeder type, space allowance and mixing on the growth performance and feed intake pattern of growing pigs. J. Anim. Sci. 76:2771-2778.

Keeney, R. L., and H. Raiffa. 1976. Decisions with Muliple Objectives: Preferences and Value Tradeoffs. Wiley, New York.

Lorenz, K. Z., 1978. Vergleichende Verhaltensforschung: Grundlagen der Ethologie. Springer, Wien.

Matthews, L.R., and J. Ladewig. 1994. Environmental requirements of pigs measured by behavioural demand functions. Anim. Behav. 47:713-719.

McFarland, D., 1989. Problems of Animal Behaviour. John Wiley, New York.

McFarlane, J. M., S. E. Curtis, R. D. Shanks, and S. G. Carmer. 1989. Multiple concurrent stressors in chicks. 1. Effect on weight gain, feed intake, and behaviour. Poult. Sci. 68:501-509.

McKee, J. S., and P. C. Harrison. 1995. Effects of supplemental asorbic acid on the performance of broiler chickens exposed to multiple concurrent stressors. Poult. Sci. 74:1772-1785.

Mellor, D. J., and C. S. W. Reid. 1994. Concepts of animal well-being and predicting the impact of procedures on experimental animals. In: R. M. Baker, G. Jenkin, and D. J. Mellor (Eds) Improving the Well-being of Animals in the Research Environment, Proceedings. pp 3-18. ANZCCART, Sidney.

Nunnally, C. J. Jr. 1970. Introduction to Psychological Measurement. McGraw-Hill, New York.

Pedersen, V., J. L. Barnett, P. H. Hemsworth, E. A. Newmann, and B. Schirmer. 1998. The effects of handling on behavioural and physiological responses to housing in tether-stalls among pregnant sows. Animal Welfare 7:137-150.

Pig Welfare Advisory Group. 1997. Booklet No. 8 Outdoor Sows. MAFF Publications, London.

Rushen, J. 1991. Problems associated with the interpretation of physiological data in the assessment of animal welfare. Appl. Anim. Behav. Sci. 28:381-386.

Rushen, J., and A. M. B. de Passillé. 1992. The scientific assessment of the impact of housing on animal welfare: A critical review. Can. J. Anim. Sci. 72:721-743.

Sandøe, P. 1997. Ethics. In: M. C. Appleby, and B. O. Hughes (Eds). Animal Welfare. pp 3-17. CAB International, Oxon.

Sandøe, P., and H. B. Simonsen. 1992. Assessing animal welfare – Where does science end and philosophy begin? Animal Welfare 1:257-267.

Scientific Veterinary Committee. 1997. The welfare of intensively kept pigs. Report to the Directorate General XXIV of the European Commission. Adopted 30. September 1997. Doc. XXIV/ScVc/0005/97, Scientific Veterinary Committee, Animal Welfare Section, Brussels.

Spruijt, B. M. 2001. How the hierarchical organization of the brain and increasing cognitive abilities may result in consciousness. Anim. Welfare. In press.

Spruijt, B. M., R. van den Bos, and F. T. A. Pijlman. 2001. A concept of welfare based on reward evaluating mechanisms in the brain: Anticipatory behaviour as an indicator for the state of reward systems. Appl. Anim. Behav. Sci. In press.

Streiner, D. L., and G. R. Norman. 1995. Health Measurement Scales: A Practical Guide to Their Development and Use (2nd Ed.). Oxford University Press, Oxford.

Stolba, A. 1981. A family pen system in enriching pens as a novel method of pig housing. In: Anonymous (Eds) Alternatives in Intensive Husbandry Systems. Proceedings. pp 52-67. Universities Federation for Animal Welfare, Potters Bar.

Sundrum, A., R. Andersson, and G. Postler (Eds). 1994. Der Tiergerechtheitsindex-200/1994 - Ein Leitfaden zur Beurteilung von Haltungssystemen für Rinder, Kälber, Legehennen und Schweine. Verlag Köllen, Bonn.

Sundrum, A. 1997. Assessing livestock housing conditions in terms of animal welfare – Possibilities and limitations. In: J. T. Sørensen (Ed.) Livestock Farming Systems - More than Food Production. pp 238-246. Wageningen University Press, Wageningen.

Toates, F. M., and P. Jensen, 1991. Ethological and psychological models of motivation - Towards a synthesis. In: J. A. Meyer, and S. Wilson (Eds) From Animals to Animats. pp 194-205. MIT Press, Cambridge, MA.

Turban, E. 1995. Decision Support Systems and Expert Systems (4th Ed.). Prentice-Hall, Englewood Cliffs, NJ.

Wainer, H. 1976. Estimating coefficients in linear models: It don't make no nevermind. Psychol. Bull. 83:213-217.

Walter, J., and G. Postler. 1994. Tiergerechtheitsindex Sauen und Mastschweine. In: A. Sundrum, R. Andersson, and G. Postler (Eds) Der Tiergerechtheitsindex-200/1994 - Ein Leitfaden zur Beurteilung von Haltungssystemen für Rinder, Kälber, Legehennen und Schweine. pp 56-72. Verlag Köllen, Bonn.

Webster, J. 1995. Animal Welfare: A Cool Eye Towards Eden. Blackwell, Oxford.

Weghe, S. van den. (Ed.). 1998. Beurteilung der Tiergerechtheit von Haltungssystemen. KTBL-Schrift 377. KTBL, Darmstadt.

Whittemore, C. T. 1998. The Science and Practice of Pig Production (2nd Ed.). Blackwell, Oxford.

Wiepkema, P. R. 1987. Behavioural aspects of stress. In: P. R. Wiepkema, and P. W. M. van Adrichem (Eds) Biology of Stress in Farm Animals: An Integrative Approach. pp 113-133. Martinus Nijhoff, Dordrecht.

Willeberg, P. 1991. Animal welfare studies: Epidemiological considerations. Proc. Meet. Soc. Vet. Epid. Prev. Med. London, 76-82.

Chapter 8. Validation of the 'final' decision support system by expert opinion[1]

Abstract

This chapter examines the validity of a model that is embedded in a computer-based decision support system to assess the welfare status of pregnant sows in housing and management systems. The model was constructed using a formalised procedure to identify and weight welfare-relevant attributes of housing systems in relation to the animal's needs, and evidenced by scientific statements collected in a database. The model's predictions about welfare scores for 15 different housing systems and weighting factors for 20 attributes were compared with expert opinion, which was solicited using a written questionnaire for pig-welfare scientists.

The experts identified tethering and individual housing in stalls as low welfare systems. The group of mid-welfare systems contained indoor group-housing systems and an individual-housing system with additional space and substrate. The five best systems were all systems with outdoor access and the provision of some kind of substrate such as straw. The highest weighting factors were given for the attributes 'social contact', 'health & hygiene status', 'water availability', 'space per pen', 'foraging & bulk', 'food agonism', 'rooting substrate', 'social stability', and 'movement comfort'.

The degree of concordance among the experts was reasonable for welfare scores of housing systems, but low for weighting factors of attributes. Both for welfare scores and weighting factors the model correlated significantly with expert opinion (Spearman's Rho: 0.92, $P < 0.001$, and 0.72, $P < 0.01$ respectively).

The results support the validity of the model and its underlying procedure to assess farm-animal welfare in an explicit and systematic way based on available scientific knowledge.

Keywords: Farm animal welfare assessment, model, index, pigs, housing systems, management.

Introduction

In the previous chapter (Bracke et al., submitted) we described the computer-based decision support system with the model for overall welfare assessment in the case of pregnant sows.

The development of the model started with the hypothesis that the overall welfare status is a function of the different need states of the animals. It allows compensation between the negative aspects of need frustration and the positive aspects of need satisfaction (Spruijt, in press; Spruijt et al., in press). We used a scientific conceptual framework (Bracke et al., 1999c) which was recently confirmed to be a common framework among welfare scientists (Anon., in press), and we used interviews with experts to fill in gaps in our knowledge. In interviews we confirmed the idea to assess welfare based on needs (Bracke et al., 1999c). We

1 Paper by Bracke, M.B.M., J.H.M. Metz, B.M. Spruijt & W.G.P. Schouten. Decision support system for welfare assessment in pregnant sows B: Validation by expert opinion. Submitted to *Journal of Animal Science*.

also showed what was expert opinion for the seven main housing systems for pregnant sows, and how expert reasoning can be modelled from attributes of housing systems to overall welfare scores (Bracke et al., 1999b; 1999d). We did not consult experts about their reasoning to select and weight attributes, but used the conceptual framework, the list of needs, and factual scientific statements (about what scientists had found in empirical research) to construct our model using a formalised procedure. With this procedure we try to capture the experts' implicit reasoning process from welfare performance criteria (i.e. weighting categories based on the dimensions of intensity, duration and incidence; Willeberg, 1991; Anon., in press) to weighting factors for attributes, which are then used to calculate overall welfare scores. With the model both weighting factors of attributes and welfare scores of housing systems can be calculated as scores on a scale between 0 (lowest) and 10 (highest).

In the previous chapter we showed that our model produces overall welfare scores for the seven reference systems that are in accordance with previous expert consultation (Bracke et al., 1999d), and with several models derived from the literature. In this chapter we attempt to validate the model further by comparing it with expert opinion.

The decision support system is designed to be adaptable, i.e. new insights can be incorporated when these become available. Validation of such a 'dynamic' model will, therefore, be an ongoing process with a dual purpose: firstly, to establish that the model is doing what it is intended to do, and secondly, to identify points for further improvement.

A major problem is that there is no golden standard to assess the overall welfare status. For empirical validation a variety of measures would have to be used, but there is no generally accepted way of how to combine these measures into one overall judgement (Fraser, 1995). Since the overall welfare status cannot be measured directly, we previously described overall welfare assessment as the attempt to make the best possible assessment based on available scientific knowledge (Bracke et al., 1999a). This suggests that experts may be used as a standard.

Expert opinion has been used in the social sciences, in health index systems (e.g. Streiner and Norman, 1995), and in other areas such as the assessment of global climatic changes (van Lenthe et al., 1996). In the field of animal welfare experts have been consulted for conceptual purposes (e.g. Duncan and van Putten, 1983) as well as for assessment purposes (e.g. den Ouden, 1996; Beyer, 1998; Bracke et al., 1999d). Since the model in the decision support system and the underlying formalised assessment procedure were constructed in order to perform similar to, i.e. to complement rather than to replace, expert consultation, expert opinion seems to be the best available, though not entirely independent, standard for welfare assessment. We suggest that if the model were found to be in accordance with expert opinion, this could be regarded as a confirmation that experts implicitly, as does our model explicitly, use an assessment of welfare based on scientific findings and the biological needs of the animals.

The main aim of this chapter, therefore, is a 'validation' of the welfare model using expert opinion.

Materials and Methods

Expert consultation

In February 2000 we sent a written questionnaire to 29 pig-welfare experts from 14 different nationalities. They were all scientists, members of the International Society for Applied Ethology, and known for their work on pig welfare. The questionnaire contained two main parts. (See the appendix to this chapter.) One part concerned the weighting of 20 welfare-relevant attributes. The other part concerned the welfare assessment of 15 housing systems for pregnant sows. Part order was reversed for half of the experts.

One part of the questionnaire contained 20 randomly presented attributes, which were a selection from the attributes in the model in order to limit the length of the questionnaire. The attributes were described by their levels, which we had ranked from best to worst, but we had omitted all minimum-requirement levels, such as a poor health status or very slippery floors, because these convey an overriding weight making welfare 'poor' (i.e. low), no matter what else is true about the other attributes.

For each attribute we asked the experts to assign a score between 0 and 10 expressing its weighting factor relative to the other attributes in the list. The least important attribute was to be given a weighting factor of 0 and the most important attribute a weighting factor of 10. We also asked to verify the ranking of the levels, and to indicate whether any of the levels, in the expert's opinion, constituted a minimum requirement for welfare. Finally, we asked to give one confidence score (scale 0 to 10) for the entire set of weighting factors to express the expert's confidence in the validity of his/her own set of weighting factors.

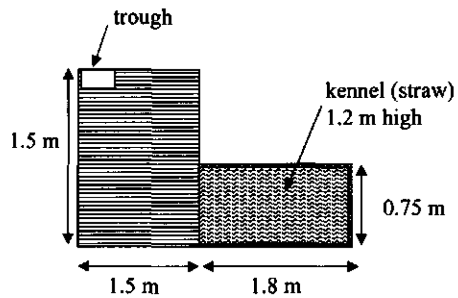
The set of 15 housing systems in the other part of the questionnaire included the seven main housing systems for pregnant sows, which are the reference systems that had been used for model development, and which had been evaluated in previous interviews (Bracke et al., 1999d). In addition, eight 'novel' systems were added to test the model's predictions (Table 1).

*Table 1. Housing systems for pregnant sows (ID: identification number, ranked according to our model from low to high welfare). Included in the description is the main reference used to define the system. The seven reference systems that were used to construct the model have been marked with *.*

ID	System	Main Characteristics
1*	Tethered	Individual housing; girth tethered; 1.2 m^2/sow; trough fed; partly slatted floor; no straw; indoors in thermocontrolled building (Baxter, 1984).
2*	Individual housing in stalls	Individual housing in stalls; 1.3 m^2/sow; trough fed; partly slatted floor; no straw; indoors in thermocontrolled building (Backus, 1997).
3*	ESF	Group of 30 sows; mixed twice per pregnancy; 60 m^2/pen; 1 electronic sow feeder (ESF); partly slatted floor; no straw; indoors in thermocontrolled building (Backus, 1997).
4*	Free-access stalls	Stable group of 12 sows with stalls for feeding and resting; 26 m^2/pen; partly slatted floor; no straw; indoors in thermocontrolled building (Backus, 1997).
5*	Biofix	Stable group of 6 sows; 15 m^2/pen; trickle feeding (Biofix); partly slatted floor; no straw; indoors in thermocontrolled building (Backus, 1997).
6	Stalls with resting area	Stable group of 5 sows with free-access stalls for feeding; resting in a kennel; 14 m^2/pen; solid floor; no straw; general purpose building (Ober and Blendl, 1969).
7	Ad lib	Stable group of 10 sows; 20 m^2/pen; fed ad libitum; partly slatted floor; no straw; indoors in thermocontrolled building.
8	Floor feeding	Dynamic group of 80 sows; 223 m^2/pen; automatic dump feeding; deep straw; general purpose building (Pig Welfare Advisory Group, 1997a).
9	Zigzag	Individual housing in a pen; 3.6 m^2/pen; trough fed; partly slatted floor; straw in kennelled resting area; general-purpose building (Brent, 1986).
10	Hurnik-Morris	Groups of 6 sows; 12 m^2/pen; fed simultaneously per group in an ESF, three times daily; partly slatted floor; daily some straw (which is all eaten); indoors in thermocontrolled building (Morris and Hurnik, 1990).
11*	Outdoor huts	Dynamic group of 10 sows; 5000 m^2/enclosure; floor feeding; wallowing pool; outdoors on pasture with isolated huts with straw (Pig Welfare Advisory Group, 1997b).
12	Huts with stalls	Stable group of 8 sows; 36 m^2/pen; 9.6 m^2 hut with straw bed for resting; fed in feeding stalls outdoors; outdoor exercise yard (Rist, 1989).
13*	Family Pen	Family group of 4 sows with offspring; 108 m^2/pen; fed from a trough with head partitions; straw provided in straw racks; rooting area with peat or bark; 'furnished' pen with rooting area, nesting area, activity area and dunging passage; open-fronted building (Stolba, 1981).
14	ESF with pasture	Group of 30 sows; 96 m^2/pen for resting; 1 hectare of pasture; 1 ESF; straw bed (Bokma and Houwers, 1988).
15	Semi-natural	Family groups of 4 sows with offspring; many hectares of semi-natural environment; adequate food; no predators; no supervision (Stolba and Wood-Gush, 1989).

9
individual housing with straw
general purpose building
1 sow

3.6 m^2/pen
[3.6 m^2/sow]
1.35 m^2 strawed kennel (for resting)
2.3 m^2 slatted floor
1.35 m^2/sow for resting
straw
twice per week new straw
no mixing (individual housing)



manual feeding
fed once daily
fibrous concentrates (7% CF), straw
ad lib water
water in the trough

welfare score: .... 90% range: ± .... point(s)

nose contact with neighbouring sows and visual contact trough barred front gate
[slats]
1 ear tag, docked, 3 vacc/pregn

main arguments:

Figure 1. Example of a card used in the questionnaire describing housing system 9 'Zigzag' (Brent, 1986).

In the questionnaire the housing systems were described in a standardised way on cards (cf. Figure 1). For all systems a moderate climate, suitable soil conditions, suitable building construction, and stabilised conditions with animals that are used to the system were presumed. The experts were asked first to rank the cards, which were presented in a randomised order, and then to give a relative welfare score for each of these systems on a scale from 0 for the worst system to 10 for the best system. For each system we also asked to state the main arguments for the score and an indication of what we called the '90% range'. This range indicates uncertainty about the welfare score. It is expressed as the number of points above and below the welfare score that covers 90% of the farms with the system. The 90% range indicates uncertainty due to differences in familiarity of the expert with the different systems, and due to differences between farms, e.g. between stockmen and other factors not described on the card.

Finally, for the worst and the best systems only (with relative welfare scores of 0 and 10 respectively), we asked to give an 'absolute' welfare score, also on a scale from 0 to 10, where all conceivable production systems defined the scale, rather than only the ones we had described on the cards. In the remainder of this chapter the term 'scores' will refer to relative welfare scores unless specified otherwise.

The response rate of this enquiry was 79% (n = 23; 12 different nationalities). The average time needed to complete the questionnaire was 125 minutes, with 44 minutes for the assessment of the attributes, and 81 minutes for the assessment of the housing systems.

Of the 23 experts that had assigned weighting factors one expert was excluded, because he had used a scale from 7 to 10, which resulted in many extreme values upon transformation to a scale from 0 to 10.

Statistics

Nonparametric statistics were used (cf. Siegel and Castellan, 1988), because the scores were bound between 0 and 10, the variance in welfare scores for housing systems and weighting factors was not constant, and the scores were not always normally distributed.

Spearman rank correlation coefficients (Rho) were used to compare the various sets of scores. In particular, Rho was used to determine the correlation between our model and the median expert scores. Rho was also used to determine the model's performance compared to the experts individually using the median expert scores as a reference. Since the set of Rho's of the individual experts as correlated with median expert scores was not normally distributed, we expressed the performance of our model as the percentile level, where a value of, for example, 75 indicates that the model's performance is at least as good as 75 percent of all the individual experts.

Kendall's coefficient of concordance (W) was used to examine the degree of consensus between the experts, and between the experts and the model.

The Friedman two-way analysis of variance by ranks and the post-hoc multiple comparisons test were used to determine whether there were significant differences in ranks between housing systems and between attributes (Siegel and Castellan, 1988, p. 174-183). Since the Friedman test cannot handle missing values, for the set of welfare scores data from 2 experts, and for the set of weighting factors data from 3 experts had to be excluded listwise because they each had one missing value. However, since this resulted in discarding many valuable data, we used overall medians, rather than the Friedman mean ranks to represent expert opinion (with n = 22 experts for welfare scores and n = 23 for weighting factors).

The data were analysed using SPSS 10.0 (SPSS, 1999).

Results

Attributes

Table 2 lists the attributes sorted according to the weighting factors calculated with our model.

Differences between attributes were found (Friedman test, $P < 0.001$), but multiple comparisons identified only 51 significant differences out of 190 combinations, i.e. 27% of all cases (cf. Table 2).

According to the experts the most important attributes were the attributes 'social contact' (median score: 8.94 on a scale from 0 to 10), 'health & hygiene status' (8.57), 'water availability' (8.54), 'space per pen' (8.00), 'foraging & bulk' (7.89), 'food agonism' (7.89), 'rooting substrate' (7.57), 'social stability' (7.14), and 'movement comfort' (7.14). The least important attributes were 'light' (3.88), 'visually isolated areas' (3.75), 'food palatability' (3.17), 'nestbuilding (resting nest)' (2.68), 'scratching' (2.36), and 'wallowing' (1.38) (Table 2).

Table 2. Weighting factors for 20 welfare-relevant attributes sorted according to the weighting factors calculated with our model (transformed to a scale between 0 and 10). Missing ID numbers represent attributes in the model that were not included in the questionnaire. The column 'Experts' gives the median weighting factor given by the experts (n = 22). Also presented are the Friedman mean ranks and significant differences from the Friedman multiple comparisons test using data from all experts who did not have missing values (n = 19).

ID	Attribute	Model	Experts	Friedman mean rank	Friedman multiple comparison[1]
1	Space per pen	10.00	8.00	13.42	abc
2	Health & hygiene status	9.06	8.57	13.68	ab
4	Exposure to cold	8.21	5.00	8.32	bcd
5	Foraging & bulk	8.11	7.89	15.29	a
7	Social stability	7.45	7.14	13.66	ab
8	Social contact	6.60	8.94	15.82	a
9	Food agonism	6.23	7.89	14,58	ab
10	Rooting substrate	5.09	7.57	13.37	abc
14	Synchronisation	3.96	5.00	9.00	abcd
15	Water availability	3.77	8.54	15.79	a
16	Separate rest-elimination area	3.40	5.00	8.95	abcd
18	Scratching	2.74	2.36	5.32	d
19	Resting comfort	2.26	5.00	9.32	abcd
20	Air quality	2.26	5.63	10.05	abcd
24	Food palatability	1.23	3.17	6.21	d
25	Movement comfort	1.13	7.14	13.21	abc
26	Nestbuilding (resting nest)	1.13	2.68	6.55	cd
30	Visually isolated areas	0.28	3.75	6.55	cd
32	Light	0.19	3.88	5.66	d
34	Wallowing	0.00	1.38	5.26	d

[1] Attributes that do not share a single character differ significantly ($P < 0.05$).

Figure 2 shows a boxplot of the scores obtained from the experts as well as the predictions made with our model.

The variance around the median weighting-factor scores was considerable. The interquartile ranges vary from 1.7 for attribute 19 ('resting comfort') to 5.7 for attribute 2 ('health & hygiene status'), with a median interquartile range of 3.5 (cf. Figure 2).

The degree of concordance among the experts is significant, but low (Kendall's W is 0.43, $P < 0.001$), and the experts themselves were also only moderately confident that their own weighting factors represented the actual importance of the attributes as they gave a median confidence score of 7.00 on a scale from 0 to 10 (interquartile range: 5.0 to 8.0).

The model contributed positively to the concordance, because when we included its scores as if the model were an additional expert, the concordance increased slightly (new W is 0.44).

Our model correlated moderately with the median expert scores (Spearman's Rho is 0.73, $P < 0.001$). With this Rho the model performed at the 55 percentile level, i.e. 55% of the experts had a lower correlation coefficient, and 45% had a higher correlation than our model.

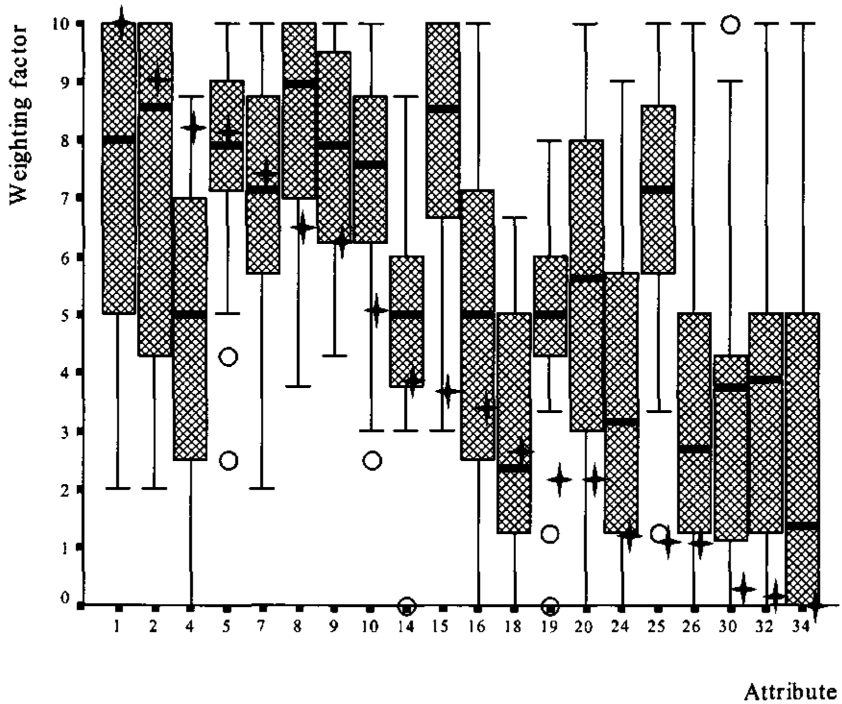


Figure 2. Boxplot of weighting factors per attribute (cf. Table 2) as assigned by the experts, showing the median score, the quartiles (box), and the whiskers, i.e. those values that are not outliers. Outliers are identified as 'O' for >= 1.5 (but <= 3) times the interquartile range. Stars (✧) represent weighting factors calculated with the model.

Large absolute differences between the model and the experts were found for several attributes. Our model gave lower scores for attribute 25 'movement comfort' (-6.0 points; 0 percentile of the expert scores for this attribute), attribute 15 'water availability' (-4.8 points; 5 percentile), and attribute 19 'resting comfort' (-3.37 points; 10 percentile). For attribute 4 'exposure to cold' our model gave a higher score (3.2 points; 95 percentile).

Two further, more qualitative indicators of the model's quality are the number of minimum requirements and changed level rankings. The model's attributes were presented to the experts without minimum-requirement levels and with their levels sorted as done in the model from best to worst. Therefore, if the experts would fully 'agree' with the model, they would never assign a minimum requirement, nor change the level ranking.

Minimum requirements were assigned legitimately in only 2.1 % of the cases (8 out of 378 instances assigned by 5 experts). It concerned the attributes 'space per pen' (3 times), 'health & hygiene status' (twice), 'air quality' (once), 'social contact' (once) and 'water availability' (once).

The level rankings were changed only 42 times out of 399 cases (11%). The attributes 'social contact', 'rooting substrate', 'exposure to cold', and 'social stability' were changed most often (9, 7, 5, and 6 times respectively). Few patterns of changed level rankings could be identified, but these did not help to identify obvious conceptual errors in the model, although they could be used as a starting point for further analysis.

Housing systems

One expert assessed the housing systems very differently from all other experts. He accounts for seven outlier values in Figure 3. Principal Components Analysis gave a factor score of 3.8 for this expert, which may be considered extreme (SPSS, 1999; p. 334). Although this expert may be considered to be an outlier, his main arguments and weighting factors presented a consistent view. He highly valued protection during feeding and the control of feed intake (e.g. using feeding stalls and/or an electronic sow feeding system (ESF)), and placed a low value on the provision of space and rooting substrate such as straw and pasture. As this was a consistent view of only one out of 23 experts (with only minor effects on the overall scores), we decided not to exclude this expert from the datatset.

Table 3 lists the housing systems according to the welfare scores calculated with our model. The experts gave the tethered system (1) and individual housing in stalls (2) the lowest median welfare scores (0.00 and 0.56 respectively). Group-housed indoor systems (3, 4, 5, 6, 7, 8 and 10), and one individual housing system with additional space and substrate (Zigzag, 9) scored considerably higher (between 4.00 and 6.00; Table 3).

Table 3. Welfare scores for 15 housing systems sorted according to the scores calculated with our model (transformed to a scale between 0 and 10). 'ID' is the system's identification number, which is equivalent to its rank according to our model. The column 'Experts' gives the median welfare score given by the experts (n = 23). Also presented are the Friedman mean ranks and significant differences from the Friedman multiple comparisons test using data from all experts who did not have missing values (n = 21). Column '90% Range' lists median 90% range scores assigned by the experts themselves estimating variability in their scores.

ID	System	Model	Experts	Friedman mean rank	Friedman multiple comparison[1]	90% Range
1	Tethered	0.00	0.00	1.19	a	1.0
2	Individual housing in stalls	0.66	0.56	2.21	a	1.0
3	ESF	2.96	4.00	5.83	ab	2.0
4	Free-access stalls	3.00	4.33	5.74	ab	1.0
5	Biofix	3.74	4.00	5.71	ab	1.0
6	Stalls with resting area	3.92	5.00	7.90	bcde	1.0
7	Ad lib	4.03	5.00	7.19	bcd	1.0
8	Floor feeding	4.22	4.44	7.12	bc	2.0
9	Zigzag	4.31	4.00	5.76	ab	1.0
10	Humik-Morris	5.09	6.00	9.21	bcdef	2.0
11	Outdoor huts	6.86	8.00	11.95	def	2.0
12	Huts with stalls	7.12	7.78	11.57	cdef	2.0
13	Family pen	8.31	8.89	12.74	f	2.0
14	ESF with pasture	9.89	9.00	13.38	f	2.0
15	Semi-natural	10.00	9.50	12.48	ef	3.0

[1] Housing systems that do not share a single character differ significantly ($P < 0.05$).

The five systems with the highest scores were all systems with outdoor access including the Family Pen system (which uses open fronted buildings). Their scores range from 7.78 for system 12 'Huts with stalls' to 9.50 for system 15 'Semi-natural'.

Differences in welfare scores between housing systems were found (Friedman, $P < 0.001$), and multiple comparisons identified 47 significant differences (where $P < 0.05$) out of 105 combinations, i.e. 45% of all cases.

No significant differences between tethered and individual housing in stalls could be detected with the Friedman test, but the Sign test was significant ($P < 0.001$).

Figure 3 shows a boxplot of the (relative) welfare scores obtained from the experts as well as predictions made by our model. This Figure suggests that according to the experts the set of 15 housing systems can be divided into roughly three groups: 2 low-, 7 or 8 mid-, and 5 high-welfare systems, with system 10 'Hurnik-Morris' being an intermediate mid- to high-welfare system.

However, the distinctions between low- and mid-welfare systems, and between mid- and high-welfare systems were not confirmed with the Friedman test (Table 3). This may be due to the fact that for the Friedman test two experts with missing values had to be excluded, and

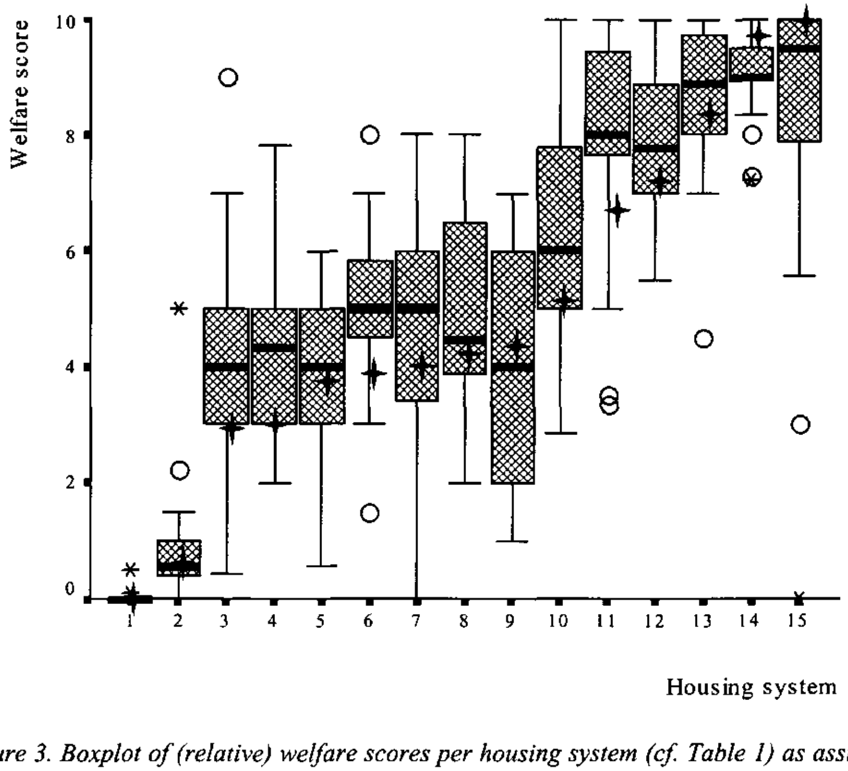


Figure 3. Boxplot of (relative) welfare scores per housing system (cf. Table 1) as assigned by the experts, showing the median score, the quartiles (box), and the whiskers, i.e. those values that are not outliers. Outliers are identified as 'O' for >= 1.5 and '' for >= 3 times the interquartile range. Stars (✧) represent welfare scores calculated with our model.*

that it only considers ranks and does not take the distance between scores into account. ANOVA with bonferroni correction, which does take distances into account, did show significant differences between low- and mid-welfare, and between mid- and high-welfare systems, while system 10 'Hurnik-Morris' was identified as an intermediate system between mid- and high-welfare.

The variance around the median expert scores was smaller than for the weighting factors of the attributes. The interquartile ranges for the 15 housing systems range from 0.0 for the tethered system to 4.0 for system 9 'Zigzag', with a median interquartile range of 2.00. Especially the low-welfare systems, 1 and 2, had low interquartile ranges.

The experts themselves assigned a median 90% range score of 2.00, i.e. an uncertainty margin of 2 welfare points above, and two points below the welfare score. The median 90% range scores correlated positively with the median welfare scores (Spearman's Rho was 0.73, $P < 0.05$), pointing to the fact that relatively little is known about welfare in supposedly welfare-friendly housing systems.

The absolute scores given by the experts differed considerably from the scores calculated with the model. With the model we calculated a range of absolute scores from 3.9 to 8.1 for the tethered and semi-natural 'systems' respectively (calculated from scores from Bracke et al., submitted). These scores fall just within the ranges of absolute scores given by the experts, which were 0.0 to 4.0 for the worst (tethered) system, and 8.0 to 10.0 for the best systems. For the tethered system and individual housing in stalls the experts gave median absolute scores of 0.00 and 1.00 respectively. The semi-natural 'system' received the highest absolute score (9.00), and the 'ESF with pasture' system had the second highest absolute score (8.83).

For the experts the absolute scores for the set of 15 housing systems range from 0.00 to 9.00, while their relative scores range from 0.00 to 9.50. The relative scores, therefore, cover as much as 92 % of the experts' absolute welfare scale. By contrast, the relative scores calculated with our model cover only 42 % of the model's absolute welfare scale. This difference between the experts and the model is not very important. It may be due to the fact that our model is based solely on the logical possibilities to alter welfare within the model's domain of housing systems, and, therefore, ignores aspects of practical feasibility and moral acceptability, which may have affected expert opinion.

For the set of all 15 housing systems the concordance between the experts in assigning relative welfare scores was 0.73 (Kendall's W, $P < 0.001$), which increased to 0.79 ($P < 0.001$) when the one 'outlier' expert was excluded. For the various sub-sets of housing systems W varied from 0.19 for the five high-welfare systems to 0.82 for the seven reference systems (all $P < 0.01$; Table 4). When our model was added as if it were an additional expert, W increased slightly (by about 0.01 point).

The model correlated well with expert opinion. For the set of 15 housing systems, Spearman's Rho was 0.92 ($P < 0.01$, Table 4). For the eight novel systems, which were those systems that had not been used to develop the model, Rho was 0.78 ($P < 0.05$).

In order to examine whether the model was also capable of making more fine-grained distinctions, we calculated Rho for the eight mid-welfare systems, and for the five high-welfare systems separately. For these sub-sets Rho was 0.50 (not significant), and 0.90 ($P < 0.05$) respectively.

For the eight novel systems and for the eight mid-welfare systems our model performed as an 'average' expert in that its correlation coefficient was at the 45 and 50 percentile level of how the individual experts correlated with the median expert scores. For all 15 systems and for the five high-welfare systems our model compared well to the individual experts, in that it performed at the 70 and 80 percentile level respectively (Table 4).

Table 4. Kendall's coefficient of concordance among the experts (W) and Spearman rank correlation coefficients (Rho) between our model and expert opinion for various sets of housing systems and for the set of 20 attributes. The percentile levels express the % of individual experts that correlate less well with the median expert scores than does our model.

Dataset	Kendall's W	Spearman's Rho	Percentile
15 housing systems	0.73**	0.924**	70
8 novel systems	0.59**	0.778*	45
7 reference systems	0.82**	0.937**	60
8 mid-welfare systems	0.24**	0.503	50
5 high-welfare systems	0.19**	0.900*	80
20 attributes	0.43**	0.729**	55

* $P < 0.05$; ** $P < 0.01$

We did not find large differences between our model and the median expert scores. The largest difference was found for system 4 'Free-access stalls' (-1.34 points difference, with the model score lying at the 25 percentile level of the expert scores for this system).

The expert data can be used to estimate a confidence zone for scores obtained with the model. These zones represent measurement error, which for practical purposes is often centred around the scores (Nunnally, 1970, p. 554-556). Two factors, which affect the confidence zone, are the standard error and the correlation between the model and the experts. Spearman's Rho is 0.92 ($P < 0.01$). As indicators of the 'standard error' we have the median interquartile range around the experts' welfare scores (which is 2.00), the median 90% range (which is also 2.00), and the maximum difference in welfare scores between the model and the experts (which is 1.34 welfare points for system 4 'Free-access stalls'). A confidence zone of about 3.00 (i.e. 1.5 points above and below the score calculated with the model) would, therefore, seem to be proper. This is substantial, and defines the scope for potential improvements in further research.

Discussion

Consensus among experts

We believe that we have found a sufficient degree of consensus among experts to use the results to 'validate' the model. Most importantly, we found clear significant differences between housing systems and between attributes. Furthermore, for both welfare scores and weighting factors Kendall's coefficient of concordance among the experts was highly significant ($P < 0.001$). Finally, with a much larger group of different experts we confirmed in the present study the previous findings (Bracke et al., 1999d) for three levels of welfare for the seven main housing systems for pregnant sows.

These observations lead us to conclude that we succeeded in probing the degree of consensus among experts using relative scores on a scale from 0 to 10, and that the results provide a frame of reference for welfare assessment in pregnant sows.

This is not to deny that differences exist between experts. Expert opinion is not fail-safe (cf. Siegel and Castellan, 1988, p. 271). Despite the fact that we consulted a substantial proportion of the world's senior pig-welfare scientists, some welfare aspects may have been missed. For example, one expert systematically identified wrong slat directions of the floor as a main argument for assigning welfare scores in our questionnaire. To our knowledge this has never been published, and, therefore, could not have been included in our science-based model, but if it were shown, it could be incorporated.

We believe that the degree of concordance among the experts was acceptable for welfare scores (Kendall's W 0.73, $P < 0.001$), because it was within the range of what McDowell and Newell (1987, p. 32) reported as typical findings (0.65 to 0.90) in the field of health measurement scaling, even though it falls below what they would call satisfactory (0.85). However, for the weighting of attributes Kendall's W was only 0.43 ($P < 0.001$), which is low. This low concordance for attribute weighting is what we expected on theoretical grounds, and this was the reason why in the model we separated the calculation of weighting factors from assigning attribute scores (Bracke et al., 1999b). The low concordance relates to the larger variance of weighting factors (median interquartile range of 3.5 compared to 2.0 for welfare scores) and the lower percentage of significant differences found (Friedman test: 27% for attributes versus 45% for housing systems). The experts were also only moderately confident about the validity of their weighting factors as indicated by their median confidence score of 7.00, and they took only half as much time for weighting the 20 attributes compared to assessing the 15 housing systems (44 versus 81 minutes). Assessing housing systems may have seemed to be the more complex task because it is more comprehensive, but it is also more in accordance with the scientists' profession. Logical errors were also encountered most evidently in the part on attribute weighting, e.g. some experts assigned weighting factors lower than 10 to minimum-requirement attributes, and one expert commented to have assigned a weighting factor of 6 to an attribute that contained both an important component with weight 9 and an unimportant component with weight 1. (An overall weighting factor of at least 9 would have been correct.)

Our finding that welfare assessment of housing systems has a higher degree of concordance than the weighting of attributes points towards a problem as well as a solution for the drafting of animal welfare legislation. The problem is that legislation focuses on prescribing many minimum-requirement attributes. A potential solution could be to focus instead on prescribing a minimum overall welfare status.

The experts assigned the highest weighting factors to the attributes 'social contact', 'health & hygiene status', 'water availability', 'space per pen', 'foraging & bulk', 'food agonism', 'rooting substrate', 'social stability', and 'movement comfort'. This list is largely in accordance with welfare priorities set by a group of 22 welfare scientists who independently contributed to a larger paper on farm animal welfare assessment (Anon., in press). There, the main design criteria for pregnant sows were preliminarily identified as 'space (quantity and quality)', 'substrate', 'social contact', and 'social stability (mixing)' (which were all classified as most important), while the main welfare performance criteria were 'abnormal behaviour', 'aggression', 'behavioural restrictions', and 'health problems'.

Our study was the first to express weighting factors of attributes and welfare scores of housing systems as scores on a scale from 0 to 10. This allows a more quantitative assessment of what constitutes a substantial improvement for overall welfare, and what is only of minor importance. The relative scale used by the experts to assess housing systems covered as much as 92% of their absolute scale. This implies that the data set included in their opinion both truly high and truly low welfare systems, which the experts appear to have classified into three main groups (see Figure 3): low-, mid- and high-welfare systems.

In accordance with previous findings (Bracke et al., 1999d), the experts gave very similar welfare scores to tethered housing and individual housing in stalls (median scores of 0.00 and 0.56 respectively), with the latter system often receiving a slightly higher score (Sign test, $P < 0.01$). In other words, the experts indicate that individual housing in stalls constitutes an improvement for welfare compared to tethered housing, but this improvement is not substantial.

A more substantial step seems to be made going from individual stalls (0.56) to group housing (with scores of 4.00 and up). However, we may ask whether group housing is necessary to improve welfare substantially. The experts identified social contact as the most important attribute (score 8.94). However, one of the mid-welfare systems was the Zigzag system, which is an individual housing system with substrate and additional space. Therefore, group housing cannot be a necessary condition for welfare, and other attributes such as space and substrate for pregnant sows seem to be able to compensate, at least partly, for the lack of social contact in individual housing systems. This, again, identifies a problem for legislation or for any other attempt to capture welfare solely in terms of cut-off points. Welfare is a quantitative and multifactorial state, which allows at least some compensation between positive and negative attributes. All housing systems contain a mixture of such attributes, and it is the balance between them that results in a high or low welfare-status of its animals.

A final implication of the experts' opinion about welfare is that natural conditions are not normative for welfare. In the group of high-welfare systems the Semi-natural 'system' was scored as the best system (median absolute score: 9.00). This 'system', however, did not provide fully natural conditions in that predation, starvation and extreme weather were all presumed to be absent. It is hard to imagine that these conditions would reduce, rather than improve the overall welfare status. The second best system was 'ESF with pasture', which contains an electronic sow identification system for computer controlled sow feeding (absolute score: 8.82). The Hurnik-Morris system was the most high-tech system. It was found to be intermediate between mid- and high-welfare systems (absolute score: 6.00), while several more natural systems received a lower score. These three systems show that the experts do not just equate welfare with the provision of natural conditions, but suggest that their views are more in accordance with an assessment based on the fulfilment of the animals' needs as is done for the first time in an explicit way in our model.

Model validation

Our model was designed to provide overall welfare assessment with a scientific basis. For this we aimed to formalise the expert's implicit reasoning steps to assess welfare based on available scientific knowledge. In the development process of the model we used information collected from experts about their views on welfare in the seven main housing systems for pregnant sows to act as a reference. The present study confirmed that the model calculates welfare scores for these systems in accordance with expert opinion. Spearman's Rho was 0.94 ($P < 0.01$). We also found a lower, but significant correlation between the model and the experts for the eight novel housing systems (Rho was 0.78, $P < 0.05$). This indicates that the model can predict the welfare status for systems that have not been used to construct the model. A cautionary note, however, is that all housing systems in this study were derived from the literature. We have not shown that the model can predict welfare scores for truly novel concepts in pregnant sow housing and husbandry systems.

We also examined the model's correlation with expert opinion for the whole set of 15 housing systems, and for the mid- and high-welfare systems separately (Table 4). Except for mid-welfare systems significant correlations were found, and the performance of the model was at least at the 45 percentile level of the performance of the individual experts.

Further evidence why we think we were successful in modelling the experts' implicit reasoning process was found for the weighting of attributes, where we found also a significant correlation (Rho was 0.73, $P < 0.001$), and a performance of the model at the 55 percentile level. Furthermore, when the model was added as if it were an additional expert, both for the different sets of housing systems and for the weighting of attributes Kendall's coefficient of concordance (slightly) increased, indicating that the model was a 'good' expert.

In general the differences between the scores predicted with the model and the median expert scores were relatively small for the housing systems, but for some attributes, such as 'water availability', 'resting comfort', and 'exposure to cold', the discrepancy was large. For attribute 'movement comfort' our model even predicted as much as 6 points below expert opinion. Contrary to our model the experts may have included space aspects into the attribute 'movement comfort', which was a separate attribute in our model. For some reason three experts also changed the level rankings of this attribute, thereby preferring more slippery floors to less slippery ones (sic). Further research would be needed to determine which are the reasons for such differences in weighting between the model and the experts. One reason may be that in the model attributes were technically defined so as to reduce the overlap between them, whereas the experts may have used the 'concept' of an attribute in a more integrated way. More detailed descriptions of the attributes with techniques like fuzzy logic could possibly be useful to clarify the fuzzy borders between different attributes, and between different levels of an attribute, especially those involving minimum requirements.

Despite some differences, we believe that our model performed sufficiently well in predicting median expert scores as regards overall welfare scores and weighting factors of attributes. Therefore, we may well have succeeded in making the underlying reasoning process explicit. Our model is the first to derive overall welfare scores from scientific statements in a formalised way. This has two main advantages. It makes a quantitative welfare assessment possible, which has considerable utility, e.g. to evaluate measures for improved welfare. It also makes a structured assessment possible, where hidden assumptions and intermediate steps are made explicit. It enforces generic rules such that personal biases are controlled to a large extent, because it is not possible to make exceptions for individual cases unless there is general knowledge available that justifies making such an exception.

The model still requires an experienced user (Bracke et al., submitted), and we have not tested its inter- and intra-observer reliability. We have only validated the model with expert opinion, which, though not a truly independent measure, is the best available standard at present. Empirical validation would be welcome, but it is presently not known how to measure the overall welfare status directly or indirectly. Nevertheless, future scientific measurements can be used to further 'validate' the model in an ongoing process. When such new knowledge is also incorporated in the decision support system to upgrade the model, the decision support system could remain the best possible assessment based on available scientific knowledge. Eventually, such a system may exceed human performance as it continuously integrates information. When this decision support system were generalised to other species, its value may increase even further, because not every species has experts as in the case of pigs.

Conclusions

Two main conclusions can be drawn. Firstly, we found a considerable degree of consensus among pig welfare scientists, when assessing housing systems for pregnant sows as regards the overall welfare status of the sows. We also found a certain, be it less pronounced, degree of consensus when assigning weighting factors to attributes. These findings support the conclusion that welfare scientists are able to formulate an identifiable position and perspective on animal welfare. Maybe even more important, however, is the finding that our model correlated well with expert opinion, especially for overall welfare scores. This supports the validity of our model and its underlying assessment procedure, which makes explicit the scientists' reasoning steps when assessing animal welfare. Our work shows how integrated farm animal welfare assessment can be performed in a structured and transparent way based on available scientific knowledge.

Acknowledgements

This work was funded by the Technology Foundation (STW) (through a joint grant from the Netherlands Organisation for Scientific Research and the Netherlands Ministry of Agriculture, Nature Management and Fisheries), and the Dutch Society for the Protection of Animals.

References

Anonymous. 2001. Scientists' assessment of the impact of housing and management on animal welfare. J. Appl. Anim. Welfare Sci. In press.

Backus, G. B. C. (Ed.). 1997. Comparison of four housing systems for non-lactating sows (In Dutch). Report P.1.171. Research Institute for Pig Husbandry, Rosmalen.

Baxter, S. 1984. Intensive Pig Production. Environmental Management and Design. Granada, London.

Beyer, S. 1998. Konstruktion und Überfrüfung eines Bewertungskonzeptes für pferdehaltende Betriebe unter dem Aspekt der Tiergerechtheit, PhD. Wissenschaflicher Facherlag, Giessen.

Bokma, Sj., and H. W. J. Houwers. 1988. Studiereis Zweden. Report P.3.20. Research Institute for Pig Husbandry, Rosmalen.

Bracke, M. B. M., B. M. Spruijt, and J. H. M. Metz. 1999a. Overall animal welfare assessment reviewed. Part 1: Is it possible?. Neth. J. Agric. Sci. 47:279-291.

Bracke, M. B. M., J. H. M. Metz, and B. M. Spruijt. 1999b. Overall animal welfare reviewed. Part 2: Assessment tables and schemes. Neth. J. Agric. Sci. 47:293-305.

Bracke, M. B. M., B. M. Spruijt, and J. H. M. Metz. 1999c. Overall animal welfare reviewed. Part 3: Welfare assessment based on needs and supported by expert opinion. Neth. J. Agric. Sci. 47:307-322.

Bracke, M. B. M., J. H. M. Metz, B. M. Spruijt, and A. A. Dijkhuizen. 1999d. Overall welfare assessment of pregnant sow housing systems based on interviews with experts. Neth. J. Agric. Sci. 47:93-104.

Bracke, M. B. M., B. M. Spruijt, J. H. M. Metz, and W. G. P. Schouten. Decision Support System for Overall Welfare Assessment in Pregnant Sows A: Model structure and Weighting Procedure. Submitted to J. Anim. Sci.

Brent, G. 1986. Housing the Pig. Farming Press, Ipswich.

Duncan, I. J. H., and G. van Putten. 1983. Abnormal Behaviours in Farm Animals. Commission of the European Communities, Brussels.

Fraser, D. 1995. Science, values, and animal welfare: Exploring the 'inextricable connection'. Animal Welfare 6:187-205.

Lenthe, J. van, L. Hendrickx, C. A. J. Vlek, and W. Biesiot. 1996. Using decision analysis for the integrated assessment of the climate change problem, IVEM report 87. IVEM, Groningen.

McDowell, I., and C. Newell. 1987. Measuring Health: A Guide to Rating Scales and Questionnaires. Oxford University Press. Oxford.

Morris, J. R., and J. F. Hurnik. 1990. An alternative housing system for sows. J. Anim. Sci. 70:957-961.

Nunnally, J. C. Jr. 1970. Introduction to Psychological Measurement. McGraw-Hill, New York.

Ober, J., and D. M. Blendl. 1969. Schweineställe: Plannung, Bau, Einrichting (6th Ed.). BLV Verlagsgesellschaft, München.

Ouden, M. den. 1996. Economic modelling of pork production-marketing chains, PhD. Wageningen Agricultural University, Wageningen.

Pig Welfare Advisory Group. 1997a. Booklet No. 7. Yards or Kennels with Floor Feeding. MAFF Publications, London.

Pig Welfare Advisory Group. 1997b. Booklet No. 8. Outdoor Sows. MAFF Publications, London.

Rist, M. 1989. Artgemässe Nutztierhaltung: Ein Schritt zum wesensgemässen Umgang mit der Natur (2nd Ed.). Verlag Freies Geistesleben, Stuttgart.

Spruijt, B. M. 2001. How the hierarchical organization of the brain and increasing cognitive abilities may result in consciousness. Anim. Welfare. In press.

Spruijt, B. M., R. van den Bos, and F. T. A. Pijlman. 2001. A concept of welfare based on reward evaluating mechanisms in the brain: Anticipatory behaviour as an indicator for the state of reward systems. Appl. Anim. Behav. Sci. In press.

SPSS. 1999. SPSS ® Base 9.0 Applications Guide. SPSS inc., Chicago, IL.

Siegel, S., and N. J. Castellan, Jr. 1988. Nonparametric Statistics for the Behavioral Sciences (2nd Ed.). McGraw-Hill, New York.

Streiner D. L., and G. R. Norman. 1995. Health Measurement Scales: A Practical Guide to Their Development and Use (2nd Ed.). Oxford University Press, Oxford.

Stolba, A. 1981. A family pen system in enriching pens as a novel method of pig housing. In: Anonymous (Eds) Alternatives in Intensive Husbandry Systems. Proceedings. pp 52-67. Universities Federation for Animal Welfare, Potters Bar.

Stolba, A., and D. G. M. Wood-gush. 1989. The behaviour of pigs in a semi-natural environment. Anim. Prod. 48:419-425.

Willeberg, P. 1991. Animal welfare studies: Epidemiological considerations. Proc. Meet. Soc. Vet. Epid. Prev. Med. London, 76-82.

Appendix: Questionnaire used to validate the welfare model

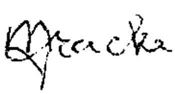
Dear Colleague,

Enclosed you will find my questionnaire on welfare assessment. This questionnaire is the final part of my PhD project on modelling of animal welfare and the development of a decision support system to assess farm animal welfare. With the help of my supervisors, Jos Metz, Berry Spruijt and Willem Schouten I have constructed a model to assess the welfare status of pregnant sows based on available scientific knowledge. This questionnaire aims to validate the model. Also, we want to examine the degree of consensus concerning welfare assessment among welfare scientists.

We want to test whether there is an agreement between expert judgement and our model for a range of different housing systems. Since the model uses weighted attributes of housing systems, our second research question concerns the correspondence between expert weightings and the weightings used in the model. For these reasons this questionnaire contains two distinct parts: assessment of housing systems and assessment of the weighting factors of attributes of housing systems. Because the answers may be affected by the order in which these parts are answered, we would like you to complete these parts in the order in which you will encounter them in your version of the questionnaire.

All answers will be treated confidentially and anonymously. The aggregated results are intended for publication. We are grateful that you agreed to participate.

Sincerely yours,



Marc Bracke

Assessment of housing systems

I have described 15 housing systems for pregnant sows. For each housing system a list of specifications and a floor plan of a typical pen is included (see the set of cards below; the first card contains the legend that explains the list of specifications and the patterns used in the floor plan).

We are interested in your opinion about the welfare status of pregnant sows in these systems and ask you to give a welfare score for each housing system on a scale from 0 (worst) to 10 (best). A number of considerations should be kept in mind:

- The score should only concern the welfare status of the pregnant sows. We are not interested in aspects related to labour, environmental or economic issues. Nor should the welfare score represent what you find morally or ethically acceptable. Rather, the welfare status should represent, as much as possible, your opinion about how the pregnant sows perceive their quality of life in that system. The welfare of pigs other than the pregnant sows, such as offspring in the Family Pen system, should be ignored. Welfare should be assessed at the housing-system level, i.e. covering whole herds/groups of animals and covering longer periods of time, e.g. months.
- We presume stabilised conditions, i.e. all systems should be interpreted as being run by 'average' farmers who have several years of experience of working with the system. We also presume a moderate outdoor climate and suitable soil types for outdoor housing, as well as suitable building construction for all systems.
- The score should represent the 'average' or typical farm with this system. When you feel information is missing, you should give a score for the most typical case while stating the assumptions you made.
- You are also asked to give, for each housing system, a rough indication of the *90% range* (90% confidence interval) around the welfare score, indicating the degree of uncertainty involved in welfare assessment. One source of uncertainty concerns individual differences between farms, such as stockmanship or the exact type of building (in as far as this is not specified on the card). These individual differences may result in differences in welfare. The 90% range is intended to cover 90% of the farms that have the system as specified on the card. Another source of variation that should be covered by the 90% range concerns uncertainty in the weighting of aspects of welfare. You may also be more familiar with some housing systems than with other systems, and have a more definite opinion about the welfare status in more familiar systems.

  The 90% range is expressed in welfare points (of which there are 10 on the scale from 0 to 10). These bounds may result in scores below 0 or above 10. E.g. the worst system with a score of 0 may have a 90% range of $\pm$ 2 points.

  In short, a welfare score of 3 with a 90% range of 1 point means that 90% of the farms with the system as specified on the card are expected to have a welfare score between 2 (3-1) and 4 (3+1), taking into account the variability between individual farms and the uncertainty involved in welfare assessment.

In summary, this is the procedure we ask you to follow for the <u>assessment of housing systems</u>:

1. *Make a preliminary ranking of the cards by sorting them. When you find this helpful, you may write or mark positive and negative welfare aspects on the card.*

2. *For each system state on the card*
   a. *a relative welfare score between 0 (worst) and 10 (best), where you are allowed to use decimals to express small but definite differences. You may want to use pairwise comparison (between housing systems which have similar ranks or are otherwise similar) to determine your final score for each system. Let the best and worst systems be 10 and 0 respectively. Missing values are permitted and preferred to random guesses.*
   b. *the 90% score-range*
   c. *your main arguments for assigning the welfare score.*

3. *Give an absolute score to your worst and best housing system. In 2a these systems received a score of 0 and 10 respectively on a relative scale. On the absolute scale 0 and 10 represent the worst and best conceivable systems that are still operational production systems.*

   The best system, i.e. number .... gets an absolute score of ....

   The worst system, i.e. number .... gets and absolute score of ....

4. *How long did you need to complete this part of the questionnaire?*

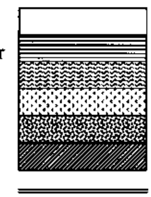
   It took me approximately .... minutes.

5. *Do you have any other remarks you wish to make on this part of the questionnaire?*

Legend
1. Main identifying characteristic (not unique)
2. Building type
3. Sows/pen
   other pigs/pen
4. Space (m^2/pen)
   m^2/sow
5. Space (m^2 of pen compartments)
   (the resting surface is listed first)
6. Resting area/sow
7. Substrates & nose rings
   substrate provision (novelty)
8. Mixing (frequency, mixing pen)
   mixing place
9. Feeding system

   frequency of feeding
   food types
10. Water (ad lib/restricted)
    number and type of water points
11. Wallowing pool
12. Scratch post & trees
13. Safety & social contact
14. Mucking out (hygiene & human contact)
15. Sources of pain (ear tags, docked tails & number of vaccinations/pregnancy

Concrete floor
Concrete slatted floor
Straw
Pasture
Soil/peat
Not part of the pen
Solid wall
ESF = electronic sow feeder



Welfare score between 0 and 10 for the typical farm with this system.

welfare score: .... 90% range: ± .... point(s)

main arguments:

Give the main arguments for the welfare score as compared to the ranked nearest-neighbour systems.

Number of points indicating uncertainty concerning the welfare score, and covering 90% of the farms with that system.

I
girth tethered
thermocontrolled building
1 sow/pen

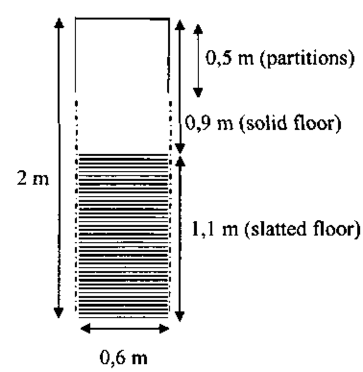
1.2 m^2/pen
[1.2 m^2/sow]
0.54 m^2 solid floor
0.66 m^2 slatted floor

no rooting substrate

no mixing (individual housing)

individual feeding in stalls (trough fed)
automatic feeding, triggered by hand
fed twice daily
concentrates only (pellets)

restricted water (i.e. only at feeding times)
water nipple



welfare score: .... 90% range: ± .... point(s)

main arguments:

[slats]
tethered, 1 ear tag, docked, 3 vacc/pregn

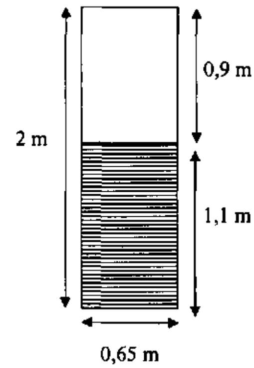
2
individual stall
thermocontrolled building
1 sow/pen

1.3 m^2/pen
[1.3 m^2/sow]
0.6 m^2 solid floor
0.7 m^2 slatted floor

no rooting substrate

no mixing (individual housing)

individual feeding in stalls (trough fed)
automatic feeding, triggered by hand
fed twice daily
concentrates only (pellets)
restricted water (i.e. only at feeding times)
water nipple



welfare score: .... 90% range: ± .... point(s)

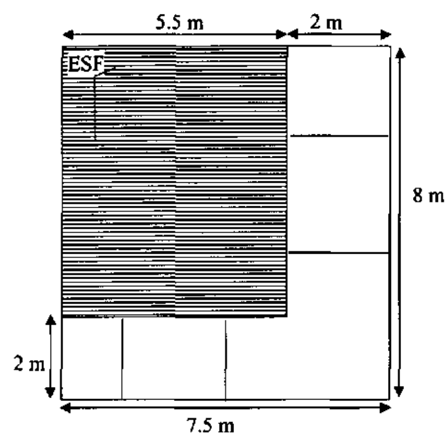
main arguments:

[slats]
1 ear tag, docked, 3 vacc/pregn

3
ESF
thermocontrolled building
30 sows/pen

60 m^2/pen
2 m^2/sow
30 m^2 solid floor
30 m^2 slatted floor
1 m^2 /sow resting area
no rooting substrate

2 times mixed/pregnancy
uses a separate mixing pen (for 3 hours)
1 ESF for 30 sows
[ESF, electronic sow feeding]
feeding start at 4 p.m.
concentrates only (pellets)
water ad lib
nipple in the feeder and 3 drinking cups



welfare score: .... 90% range: ± .... point(s)

main arguments:

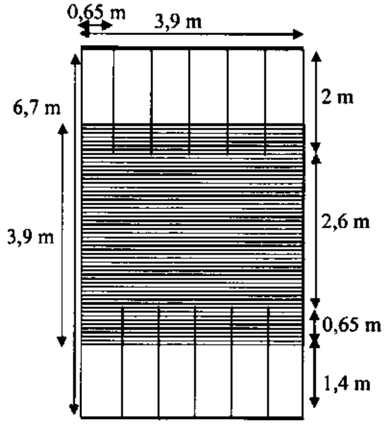
[slats]
1 ear tag, docked, 3 vacc/pregn

4
free-access stalls
thermocontrolled building
12 sows/pen

26 m²/pen
2.2 m²/sow
11 m² solid floor
15 m² slatted floor
1.3 m² per stall for resting
no rooting substrate

stable group (1x mixed for grouping)
uses a separate mixing pen (for 3 hours)
fed in the stalls (locked-in during feeding)
automatic feeding, triggered by hand
fed twice daily
concentrates only (pellets)
water ad lib
nipples in the trough and a drinking cup

[slats]
1 ear tag, docked, 3 vacc/pregn



welfare score: .... 90% range: ± .... point(s)

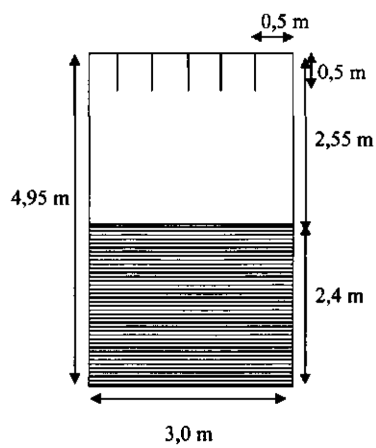
main arguments:

5
Biofix
thermocontrolled building
6 sows/pen

15 m²/pen
2.5 m²/sow
7.7 m² solid floor
7.2 m² slatted floor
1 m²/sow resting area
no rooting substrate

stable group (1x mixed for grouping)
uses a separate mixing pen (for 3 hours)
Biofix, trickle feeding
automatic feeding, triggered by hand
fed twice daily
concentrates only (pellets)
water ad lib
nipples in the trough and a drinking cup

[slats]
1 ear tag, docked, 3 vacc/pregn



welfare score: .... 90% range: ± .... point(s)

main arguments:

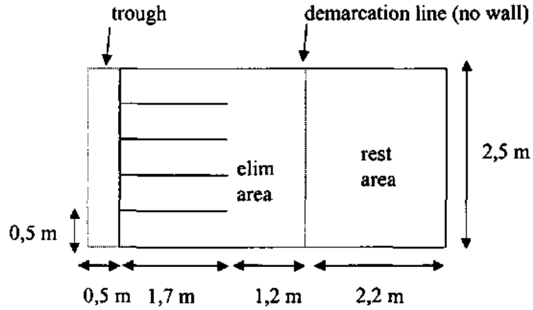
6

feeding stalls & group resting area
general purpose building with kennels
5 sows/pen

14 m²/pen
2.8 m²/sow
5.5 m² kennel with solid floor
3 m² solid floor for elimination
1.1 m²/sow for resting (in a kennel)
no rooting substrate

stable group (1x mixed for grouping)
uses a separate mixing pen (for 3 hours)
feed stalls (locked-in during feeding)
manual feeding
fed twice daily
concentrates only (pellets)
ad lib water
drinking cup

mechanically scraped
1 ear tag, docked, 3 vacc/pregn



welfare score: .... 90% range: ± .... point(s)

main arguments:

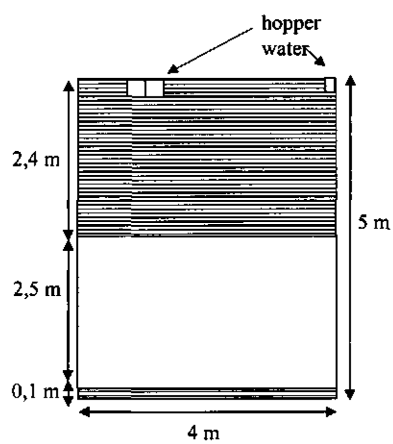
7

ad lib feeding
thermocontrolled building
10 sows/pen

20 m²/pen
2 m²/sow
10 m² solid floor
10 m² slatted floor
1 m²/sow resting area
no rooting substrate

stable group (1x mixed for grouping)
uses a separate mixing pen (for 3 hours)
ad lib hopper with 2 feeding places
[self-feeding, ad lib
[ad lib, 24 hours/day available]
high fibre food (14% CF), pellets
ad lib water
water nipple

[slats]
1 ear tag, docked, 3 vacc/pregn



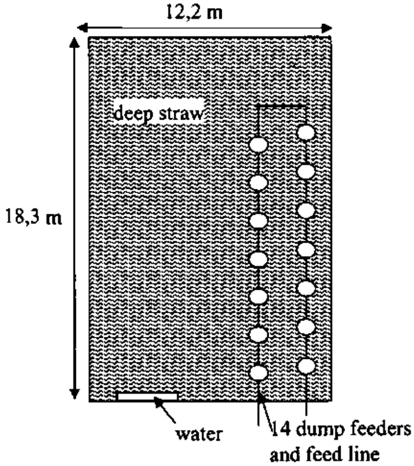
welfare score: .... 90% range: ± .... point(s)

main arguments:

8
dump feeding in a large group
general purpose building
80 sows/pen

233 m²/pen
2.8 m²/sow
233 m² deep straw yard

1.5-2 m²/sow resting area
deep straw
once per week new straw
6 times mixed per pregnancy
mixed in the pen
dump feeding
automatic, triggered by hand
fed twice daily
concentrates (pellets), straw
ad lib water
water trough



welfare score: .... 90% range: ± .... point(s)

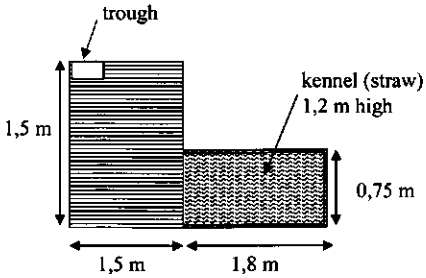
main arguments:

[deep straw, removed 2x/year]
1 ear tag, docked, 3 vacc/pregn

]

9
individual housing with straw
general purpose building
1 sow

3.6 m²/pen
[3.6 m²/sow]
1.35 m² strawed kennel (for resting)
2.3 m² slatted floor
1.35 m²/sow for resting
straw
twice per week new straw
no mixing (individual housing)



manual feeding
fed once daily
fibrous concentrates (7% CF), straw
ad lib water
water in the trough

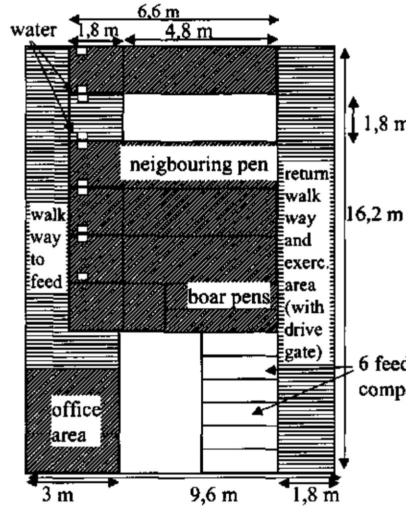
welfare score: .... 90% range: ± .... point(s)

nose contact with neighbouring sows and
visual contact trough barred front gate
[slats]
1 ear tag, docked, 3 vacc/pregn

main arguments:

10
ESF with 'simultaneous' feeding
thermocontrolled building
6 sows/pen

12 m²/pen
2 m²/sow (excluding walk ways)
8.7 m² solid floor
3.3 m² slatted floor
1.3 m²/sow resting area (solid floor)
little straw
daily 0.25 kg straw/sow, which is all eaten
stable group (1x mixed for grouping)
uses a separate mixing pen (for 3 hours)
ESF with 6 feeding compartments
all sows in a pen feed simultaneously
fed 3 times/day
concentrates (pellets), some straw
ad lib water
2 drinking cups



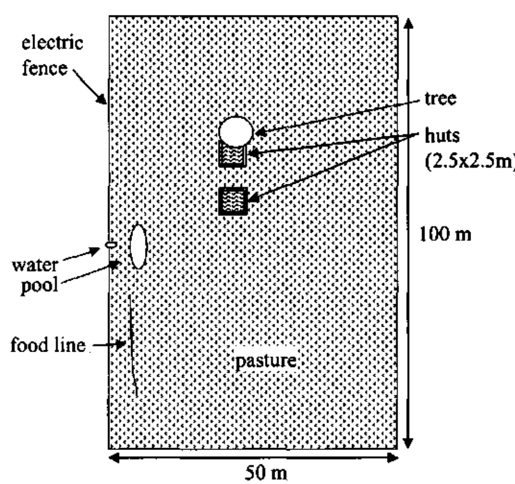
welfare score: .... 90% range: ± .... point(s)

main arguments:

[slats]
electric driving gate, 1 tag, docked, 3 vacc

11
outdoor housing with huts
2 insulated huts (6.3 m² each) on pasture
10 sows/pen

5000 m²/pen
500 m²/sow
12.5 m² straw bed in the huts
4988 m² pasture (with grass)
1.25 m²/sow resting area in huts
pasture, straw, no nose rings
twice per week new straw
10 times mixed/pregnancy
mixed at pasture
line feeding
manual feeding
fed once daily
concentrates (rolls/cobs), grass, straw
water ad lib
water in a trough
wallowing pool
1 tree



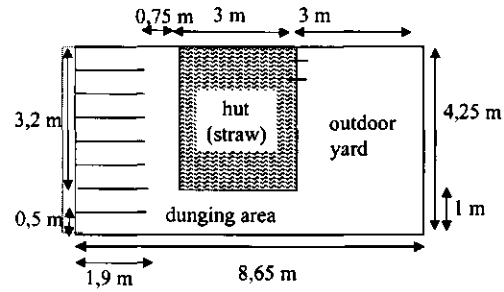
welfare score: .... 90% range: ± .... point(s)

main arguments:

[pasture]
electric fence, 1 ear tag, docked, 3 vacc/pregn

12
huts with feeding stalls
insulated hut with straw
8 sows/pen

36 m^2/pen
4.5 m^2/sow
9.6 m^2 straw bed in the hut (for resting)
19 m^2 sold floor outdoors
1.2 m^2/sow for resting
straw bed
3 x/w. new straw & daily mucking out
stable group (1x mixed for grouping)
mixed in the pen
feed stalls outdoors (locked-in at feeding)
manual feeding
fed twice daily
concentrates (pellets), straw
ad lib water
drinking cup



welfare score: .... 90% range: ± .... point(s)

main arguments:

daily mucking out manually
1 ear tag, docked, 3 vacc/pregn

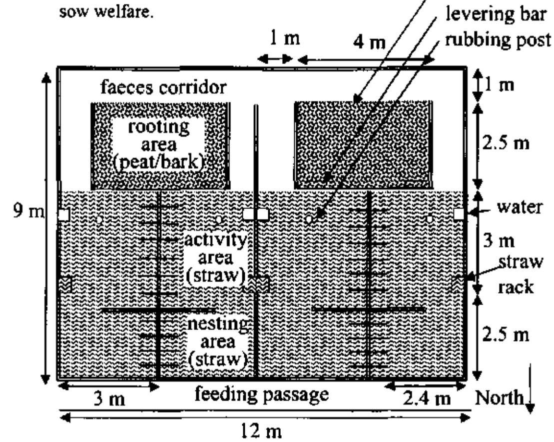
13
family pen
open-front building, facing south
4 sows, 1 replacement gilt &
offspring (36 pigs of ± 40 kg)
108 m^2/pen
27 m^2/sow, 2.6 m^2/pig
28 m^2 straw bed (nest); 20 m^2 peat/bark
22 m^2 solid floor & 29 m^2 some straw
5.6 m^2/sow; 0.7 m^2/pig resting area
peat/bark, straw
daily fresh straw in racks & mucking
no mixing ('family' groups)

fed from a trough with head partitions
manual feeding
fed once daily
liquid food (concentrates), straw
ad lib water
4 water troughs

scratch post

daily mucking out manually
1 ear tag, docked, 3 vacc/pregnancy

Note: please assess only pregnant sow welfare.



welfare score: .... 90% range: ± .... point(s)

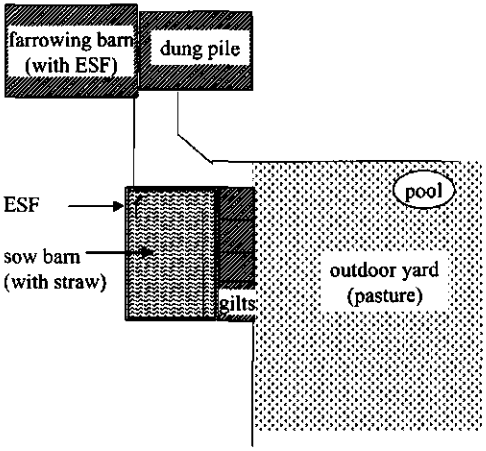
main arguments:

14
ESF, 'family' group & pasture
general purpose building
30 sows/pen

1 hectare outdoor yard
330 m²/sow
96 m² straw bed in the barn
1 ha. Pasture
3 m²/sow for resting
pasture, straw, no nose rings
once/week new straw
no mixing ('family'group)

2 ESFs for 30 sows
[ESF, electronic sow feeding]
feeding start at 8 pm
concentrates (pellets), grass, straw
ad lib water
water from the ESF & a trough
wallowing pool

[pasture]
1 ear tag, docked, 3 vacc/pregn



welfare score: .... 90% range: ± .... point(s)

main arguments:

15
feral pigs under free-range conditions
no built shelter (only natural habitat)
groups of about 4 sows with offspring

many hectares

many hectares of forrest and pasture

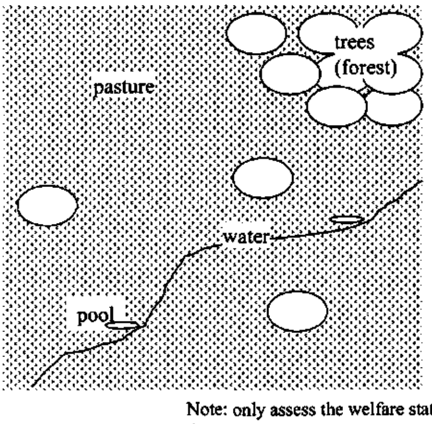
resting in self-made resting nests
pasture, forest, no nose rings

no mixing (family groups)

natural environment with adequate food
[not fed by humans]

grass, roots, insects, etc.
natural sources of water

wallowing pool
[scratch posts as in nature]
no predators; no supervision (only infrequent 'harvesting' of pigs)
[forest]
no ear tags, not docked, no vaccin.



Note: only assess the welfare status of the pregnant sows

welfare score: .... 90% range: ± .... point(s)

main arguments:

Weighting of welfare relevant attributes and levels

Housing and management systems can be described in terms of their properties or attributes. Each attribute can have two or more levels that are mutually exclusive. For example, the attribute 'presence of substrate' can have the levels 'much substrate is present', 'little substrate' or 'no substrate present'. The levels of an attribute can be ranked according to what is better for welfare (e.g. 'much substrate' – 'little substrate' – 'no substrate'). In addition, attributes differ in importance for welfare (e.g. 'substrate' may be less important than 'social contact').

In the list below, some of the attributes we consider relevant for welfare assessment have been specified, together with their levels. Within each attribute we have ranked the levels from best to worst. You are asked to verify this ranking and to assign a weighting factor for the attribute as a whole as defined by all its (specified) levels. Some considerations must be kept in mind:

- Weighting factors depend on the levels that have been specified and the degree to which they matter to the animals. Considerations about how difficult it may be for humans to provide these things for the sows in practice should be ignored.
- Weighting factors should be assigned as much as possible in isolation of each other. When you feel interactions are important, you should give the 'average' weighting factor.
- Some descriptions of levels may be vague. You should make an assessment based on the most reasonable interpretation that could be formulated. However, levels specifying minimum requirements are not included in the listing. Since it is pointless to assign a weighting factor to attributes containing minimum requirements (e.g. poor health, poor climate, very high agonism levels), these have been removed for this exercise. Should you encounter levels that specify minimum requirements in your opinion, you should give the attribute a weighing factor of 10 and mark the minimum-requirement level with 'MR'.
- The attributes and the levels should be interpreted to concern long-term conditions and to apply at group level, i.e. we are interested in weightings for assessment of the overall welfare status as described in the part on assessment of housing systems.

In summary, this is the procedure we ask you to follow for weighting of attributes:

1. *For each attribute, indicate on the form 'List of attributes and their levels' whether you agree with the ranking of the levels of that attribute (see also the example on the next page).*

2. *Assign a weighting factor between 0 and 10 to each attribute, taking into account only the range of levels as provided. Please, use the entire scale from 0 to 10. You are allowed to use decimals, and missing values are preferred to random guesses.*

You are also asked to answer the following general questions:

3. *On a scale from 0 (little or not at all) to 10 (very much), the degree to which you are confident that your weightings represent the actual importance of the attributes for the animals is ……………*

   Remarks (if any):

4. *How long did you need to complete this part of the questionnaire?*

   It took me approximately ……….. minutes.

5. *Do you have any other remarks you wish to make on this part of the questionnaire?*

Example of how to mark your answer for weighting of attributes.

In the far right-hand column labelled 'attribute weights', you should state your weighting factor for the attribute as a whole (between 0 and 10), e.g. 7 in this example.
When you agree with the ranking of the levels, you should tick the column 'ranking of levels' (see option 1 below), otherwise you should indicate the alternative ranks (see option 2 below). Where you wish to do so, you may state your main considerations in the space below each attribute.

ID	Attribute and attribute levels	ranking of levels	attribute weights
1	Substrate much substrate no substrate some substrate	√	7
2	Substrate much substrate no substrate some substrate	 1 3 2	7

List of attributes and their levels

ID	Attribute and attribute levels	ranking of levels	attribute weights
1	Space per pen for exploration & locomotion > 6250 m2/pen (can range) 1250-6250 m2/pen ('inspection' possible) 250-1250 m2/pen (can 'gallop') 50-250 m2/pen (can canter) 10-50 m2/pen (can walk) 2-10 m2/pen (can circle) 1.5-2 m2/pen (can turn, as state) 1-1.5 m2/pen (cannot turn, as state)		
2	Health & hygiene status high (e.g. free of scabies, Aujesky's disease etc.; high hygiene status; low lameness) average health status below average (e.g. reduced longevity)		
4	Thermocomfort as regards cold (flooring, heating & insulation) least cold, e.g. thermocontrolled building with isolated floor (or deep straw) intermediate cold, e.g. general purpose building some cold, e.g. outdoor housing in huts		

5	Foraging activity & bulk foraging such as grazing/browsing, 2-6 hrs/day bulky diet (e.g. straw intake 0,5 kg/sow/day)/ad lib feeding, but not grazing some fibre (e.g. 7% CF)/some increased feeding duration, e.g. Biofix concentrates only		
7	Social stability (mixing & individual confinement) (as an event) no mixing, i.e. group as stable as family groups stable group, i.e. mixed once (for group formation) sub-stable group (2 or 3x mixed/pregnancy) dynamic group (>= 4x mixed/pr.) individual confinement (as an event of social isolation)		
8	Social contact (group size & contact with neighbours) small group of 3-7 sows medium group of 8-25 sows large group of > 25 sows group of 2 sows (pair) *individual housing with visual & some tactile contact* individual housing with visual isolation (but auditory contact) individual housing with auditory isolation		
9	Food-related agonism protection as related to stall design and feeding system locked in individual stalls for feeding (excluding ESF systems) or plenty of space (as in free-range) fed from a trough with partitions/line feeding/ad lib/Biofix/<= 30 sows per ESF fed from a trough without partitions/floor feeding (not ad lib)/ > 30 sows per ESF		
10	Rooting ability as species specific behaviour (other than for finding food) substrate > 5cm depth, available >= 4 hrs/day, no nose ring rooting substrate, < 5 cm depth, >=2 hrs/day, no ring no rooting substrate, no ring nose rings in the presence of substrate		
14	Synchronisation and regularity of feeding and resting simultaneous and regular feeding & resting sequential feeding between pens/sequential in the pen but ad lib *sequential feeding in the pen*		
15	Water availability water ad lib at all times *water restricted (in time, in flow rate or in no. of pigs/drinker)* water is not available at or immediately after feeding		
16	Separation of resting & elimination area separate areas for resting and elimination no physical separation between resting and elimination areas		
18	Scratching scratch post allows full scratching (including scratching of the back) some scratching is possible, e.g. only vertical post scratching is limited, e.g. as when tethered (little or no post)		

19	Resting surface comfort soft resting surface (e.g. straw or earth) hard resting surface (e.g. concrete)		
20	Air quality (includes dust, NH_3, relative humidity, etc.) outdoor housing, e.g. huts as open-front with straw/outdoor access/general purpose building with straw as in average thermocontrolled buildings without straw elevated NH3/dust, e.g. thermocontrolled building with straw/limited ventilation		
24	Food palatability high palatability medium palatability low palatability		
25	Movement comfort (floor quality as regards slipping and injury risk) soil/straw/concrete providing grip deep litter/loose earth/mud with minor hindrance of movements somewhat slippery or injurious floor, e.g. slatted floor		
26	Nestbuilding (resting nest) as species specific behaviour (other than for thermocomfort) can build a resting nest no nest building possible		
30	Visually isolated pen-areas for exploration and areas for (social) retreat >= 2 visually isolated areas or much space (> 25 m2/sow) 1 visually isolated area 0 visually isolated areas		
32	Light during the activity period day-light-like light duration and intensity (resting area may be dark) little light		
34	Wallowing as species specific behaviour (other than for thermoregulation) wallowing pool with mud when the ambient temperature exceeds 18 C no pool		

Chapter 9. General discussion

Introduction

The overall aim of this thesis was to find a formalised, i.e. structured, transparent, yet flexible way to 'objectively' assess the overall welfare status of farm animals in relation to the housing and management system based on available (and undisputed) scientific knowledge. The approach was to use interviews with experts to fill in gaps in knowledge and to develop a computer-based decision support system (Turban, 1995) with a model in order to formalise the assessment of the overall welfare status in the case of pregnant sows. The overall aim includes biological and information-technology related aspects. The former concern the attempt to assess welfare overall (1), of farm animals at the housing-system level (2) and on a scientific basis (3). The latter include formalisation and flexibility (4) and user aspects and practical implications (5) of the decision support system.

Biological aspects

Overall welfare

The first sub-aim was to assess the animal's overall welfare status. For this we de-composed welfare into a list of needs ('behoeften' in Dutch) in order to break up the complex concept of overall welfare into manageable chunks. 'Needs' were used not only to refer to necessities for survival, but more widely as general states necessary for adequate behavioural, physiological, and psychological functioning. As such the concept of 'needs' is functional for welfare assessment, because it relates observable phenomena to welfare, defined as the animal's quality of life as perceived from the animal's point of view. Although the concept is not undisputed (Anon., in press), we found wide support for it in the literature and in interviews with experts (Chapter 5).

We chose not to use the concept of the five freedoms (Brambell Committee, 1965; FAWC, 1992; Webster, 1995), because this concept is more of an ethical than a biological concept, and because it strongly suggests that welfare is to be assessed using cut-off points. Cut-off points have been criticised (Mendl, 1991), and are not in accordance with the biological phenomenon that animals can adapt, and can compensate negative experiences with positive ones. However, Webster's practical interpretation (Webster, 1995) seems to be very similar to our approach, suggesting that the differences may be mostly semantic. With the decision support system one could 'mutate' the needs-based model into a freedoms-based one in order to examine what the consequences for welfare assessment are.

A list of needs was used to construct an operational welfare model for pregnant sows (see Chapter 7). Several versions of the 'Tiergerechtheitsindex (Animal's Needs Index, Bartussek, 1999; Sundrum *et al.*, 1994) also use a linking of needs and attributes of housing systems. We have now formalised the procedure to select and weight the model's attributes with the help of the list of needs and scientific statements collected in the knowledge base of the decision support system. To calculate a welfare score for a housing and management system, we used a simple weighted average calculation rule in accordance with existing models for overall welfare assessment (reviewed in Chapter 4), and health and psychological index systems (McDowell & Newell, 1987; Streiner & Norman, 1995). Our model allows compensation between positive and negative attributes of the housing system, but also includes minimum-requirement levels, which are cut-off points that result in a low overall welfare status, no matter what else is true about the housing systems.

It may be objected that this procedure did not result in an assessment of welfare as perceived from the animal's point of view (Bekoff *et al.*, 1992). It, indeed, remains a third-person perspective (Wemelsfelder, 1993; Lijmbach, 1998). However, if Wemelsfelder *et al.* (2000) are right that humans can reliably assess welfare, and if experts are knowledgeable humans, then we have come as close as we can get at present to an assessment of welfare from the animal's point of view, because our model was validated with expert opinion (see Chapter 8). However, without a major break-trough in cognitive neurobiology we cannot say to what extent we actually succeeded in modelling the way the animals themselves integrate their need states into an overall welfare status (Spruijt *et al.*, in press; Spruijt, in press). Our model only aimed to provide the best possible assessment given presently available knowledge.

Housing and management system

The second biological aspect was to assess the welfare status of farm animals in relation to their housing and management system.

We restricted our modelling to the case of pregnant sows with application at the housing-system level, rather than at the farm level. The 'system level' means that the model is designed to distinguish between various housing systems, i.e. types of farm, rather than between individual farms within systems. It reflects the fact that the model is calibrated to cover the wide domain of different, present and future, housing systems.

It may be objected that the 'system level' is a rather abstract and static concept, while welfare is a property of an individual animal at a particular moment in time, and, even for housing systems as a whole, conditions may vary over time.

Welfare is an internal property of an individual at a particular moment in time. This internal property is related to the external housing environment using the 'attribute' concept, which was formulated in close connection to the 'needs'. An integrated judgement was sought that would transcend the particular here and now, and that would apply to a larger number of animals over a longer period of time. This could be regarded as the 'farm level'. Our model applies 'at the housing-system level', which would appear to be even more general. However, we used typical farms as representatives of housing and management systems, and within these typical farms we calculated a welfare score for the typical, or otherwise 'average', animal.

Accordingly, the attributes in the model were not formulated directly at system level, i.e. across farms, but they were formulated to apply within farms. For example, the fact that outdoor systems mostly, i.e. across farms, have a wallowing pool, cannot be an attribute level in the model, but the fact that the pool is available mostly, e.g. for the most part of the day, can be, provided there are scientific arguments to identify it as a separate attribute level.

An assessment of welfare at the housing-system level, defined in this way, matches very well with the kind of general knowledge provided by biological science, but it requires that the user of the model can determine the 'static' properties of the system, also in relation to welfare. In this 'static' model dynamic events such as the periodic mixing of sows are described with static attribute levels such as 'no mixing', 'once mixing, and 'frequent mixing' during pregnancy. Such a model is less well suited to identify fluctuations in welfare over time. Theoretical problems, e.g. how to weight the suffering of the few today against the pleasures of the many tomorrow, remain to be resolved, and solutions should be incorporated in the model. Even without such refinement, however, the model is suited for further development for specific use, e.g. at the farm- or the group (pen) level.

Scientific basis

The third aspect was to 'objectively' assess welfare based on available scientific knowledge.

The word 'objectively' is in quotation marks, because a welfare status cannot be proven scientifically at present, neither empirically, nor by strict logical deduction (Chapter 3). We designed a prototype decision support system (Chapter 2). In relation to the issues raised in the prototype and subsequent versions of the decision support system we studied the literature in search for consensus issues (Chapters 3, 4 & 5), and we conducted interviews with experts to fill in gaps in knowledge, especially about the reasoning steps involved in welfare assessment (Chapters 5 & 6). Independent of the interviews, we collected scientific statements from literature reviews about pig welfare (e.g. Scientific Veterinary Committee, 1997) and downloaded these into a database, in the form of a decision support system, to construct the welfare model in a procedural way using needs and a scientific conceptual framework for welfare assessment (Chapter 5; Anon., in press).

It may be objected that this is not science, because we did not take scientific measurements, and we did not propose a proper welfare theory, but instead defined 'welfare' in terms reminiscent of animal psychology that has long been rejected by biological scientists.

Since scientific proof is not possible at present, our goal was to find a best possible assessment given the current state of biological science. In the absence of a comprehensive welfare theory (Haynes, 2000), we resorted to a conceptual framework that is common among welfare scientists, and which recognises that an animal is the product of evolution with cognitive and emotional systems that are designed to support its biological functioning (Anon., in press; Spruijt, in press). It allowed us to presume that animals have a welfare status, which is a common presumption in the field of applied ethology as well as in society at large. With the conceptual framework design criteria and performance criteria for welfare can be formulated. The framework suggests that design criteria for welfare, which are mostly environment-based attributes, can be assessed based on their relationship with welfare performance criteria, which in turn can be assessed based on the dimensions of intensity, duration, and incidence (Willeberg, 1991).

In the absence of empirical tools to measure welfare directly we constructed an IT-tool to make use of available knowledge. The scientific statements collected in the knowledge base were interpreted using the conceptual framework. We did this in an explicit way, and we made hidden assumptions explicit. We constructed a formalised procedure to make a welfare model, showing that welfare assessment is no longer a purely subjective, intuitive act, but 'objective', i.e. an act of rational decision making based on knowledge of the facts including aspects of uncertainty. We expressed this uncertainty as a margin of confidence, which was estimated to cover a range of 3 points centred around the calculated scores. Further refinement of the conceptual framework (as described in the Chapters 3, 4, 5 & Anon., in press), which may eventually lead to a welfare theory, will undoubtedly lead to insights that can be used to upgrade the welfare model.

Information-technology related aspects

Formalisation and flexibility

The next sub-aim of this thesis concerns formalisation and flexibility, i.e. to find a structured and transparent way to assess welfare based on scientific facts, while sufficiently flexible to deal with changes in knowledge. Flexibility is required, because welfare assessment has been, until now, an unstructured task, and new insights in the future can be anticipated to require adapting the procedure. This may happen, for example, when new scientific knowledge is generated, or when different users use the decision support system for different purposes.

For formalisation a whole 'new' vocabulary was needed. We defined terms like 'attributes', 'needs', 'weighting categories' and 'attribute scores'. Formalisation consisted of two parts: the selection of model attributes and the model's weighting procedure.

Attributes are defined in relation to the model's domain, the scientific statements and the needs, in order to ensure that the attributes can be used to calculate welfare scores with a generic calculation rule.

In the weighting procedure we calculate welfare scores (on a scale between 0 and 10) as a weighted average of the attribute scores. Attribute scores are assigned in the model based on the ranking of the levels within an attribute (which specify the properties of the housing system). This ranking is based on scientific arguments as found in the scientific statements and list of needs. The weighting factor of an attribute is based on an assessment of the magnitude of difference between its best and its worst level. The 'weight' of each attribute level is calculated from the weighting scores attached to the weighting-category levels, as well as from the number of distinct types of scientific arguments found in the scientific statements. Weighting categories are a classification of the different kinds of welfare performance criteria as measured by the various disciplines involved in welfare science. The weighting scores are assigned to the weighting-category levels based on the dimensions of intensity, duration, and incidence (Willeberg, 1991; Anon., in press).

Together the two parts of the procedure, to select attributes and to weight them, provide a holistic framework for overall welfare assessment, which applies general scientific knowledge in a generic way across housing and management systems.

It may be objected that the procedure is complex and at times conceptually weak, resulting, for example, in unnecessary concepts such as 'weighting categories' and 'types'.

We used the classification in weighting categories and types only after previous versions of the decision support system had produced counter-intuitive results. In these previous versions we have tried modelling welfare more directly in terms of intensity, duration, and incidence. We have also tried a strict deduction with if-then rules from scientific statements, and assessing welfare from intermediate need-state scores. These approaches failed, maybe because of our choice for a hierarchical representation formalism (welfare – needs – attributes – influencing factors) cast in the tables in a relational database. A Bayesian belief-network or other rule-based (if-then) expert system, possibly supplemented with fuzzy logic, may have suited better, but we feared that these would reduce the transparency and flexibility of the decision support system.

Further research may aim to improve or modify the weighting procedure, or to reduce the model's complexity, e.g. in order to make it available for the non-experienced user. One approach would be to build a shell around the model, such that it would no longer be required to describe the housing system in terms of the rather complex model attributes. Such a shell would take simple, physical attributes such as straw and soil as input, and use if-then rules to determine the levels of the attributes in the model to calculate welfare. An alternative approach would be to use empirical correlations to reduce complexity by reducing attributes and/or weighting factors. For conceptual reasons the attributes have been assigned weighting factors, because it makes sense to distinguish between luxuries and necessities for welfare (Dawkins, 1990). However, when there would be a high correlation between the weighted and unweighted versions of the model (as predicted on theoretical grounds by Gulliksen, 1950), it may be simpler to use the unweighted version. Similarly, some attributes in the model could be removed, because they are correlated under normal conditions with other attributes (e.g. 'rooting substrate' and 'foraging & bulk', because both are usually provided by straw). Both approaches to reduce the model's complexity by building an expert system shell, or by building a simple model that correlates well with the present model, but especially the latter approach, will have to pay the price of reduced conceptual validity of the model.

User aspects and practical implications

The last IT-related aspect concerns user aspects.

The development of the decision support system for farm animal welfare assessment had a proximate scientific aim, i.e. to show how overall welfare assessment can be performed in a formalised way based on available scientific knowledge (Chapters 7 & 8). Further research is needed to meet its ultimate aim. For other usages, e.g. when designing housing systems or evaluating welfare policies, decision makers would need the assistance of an experienced user to operate the model. Further research may improve various aspects of the model's userfriendliness (e.g. by reducing complexity), validity (e.g. inter- and intra-observer reliability, conceptual analysis, sensitivity analysis, empirical validation), and various application aspects, such as in relation to specific moral theories and political contexts, e.g. for the assessment of novel housing systems before they are allowed on the market, and for the welfare assessment of other species than pigs. Although this is a long list, such practical goals are now within reach because of our work, i.e. because an operational decision support system for welfare assessment in the case of pregnant sows on the basis of available scientific knowledge (scientific facts and a list of needs) has been constructed and validated with expert opinion.

The main practical implications of our work are the following. Both the model and the experts indicate that welfare is improved substantially only when considerable changes in the housing and management systems are made. As a result, the model may be used in the design of new housing systems especially when substantial changes are foreseen. For small changes the present model may be less well suited, in part because it was developed for a rather wide domain of housing systems. This placed practical restrictions on the degree to which fine-grained distinctions have been modelled at present. More scientific knowledge could be collected to improve the model in this respect. However, for some aspects such as the quality of the farmer's stockmanship scientific evidence may be lacking at present. Relevant knowledge that is based not on scientific evidence, but on practical experiences could also be incorporated, but this could reduce the scientific credibility of the knowledge base. A final practical implication of our work concerns implications for further research. Up to now, science has focussed mainly on comparing two levels of an attribute, e.g. straw versus no straw, and low fibre versus high fibre in the diet. However, scientists increasingly recognise the need to study factors in a more quantitative way, and to look at interactions between multiple factors (Kemp, 2000). Our model can be used to construct dose-response relationships (cf. Chapter 7), and to 'predict' effects of welfare design criteria on welfare performance criteria. In principle, these can be tested in empirical research in order to verify the theoretical assumptions made in the model or elsewhere, e.g. the model could be used to test the hypothesis by Hurnik (1988) that longevity is a good measure of the animals' overall welfare status.

We will end the discussion with a brief description of how the decision support system can actually be used to quantify welfare.

Firstly, the user must specify the attribute levels, i.e. the properties of the housing system, which he wants to assess. For this, the descriptions of the attribute levels of the 7 (or 15) housing systems may serve as a benchmark. When the user is interested only in some aspect of welfare, he (or she) may copy and paste a whole series of housing systems, and change only the relevant attributes.

Secondly, he must open the output screen(s) showing the overall welfare scores of the housing systems, and compare the scores with the scores of the reference systems. A first check would be to verify whether the calculated scores are in accordance with an extrapolation from the scores of related reference systems.

Thirdly, the user can reason backwards by asking how overall scores are built up from attribute scores and weighting factors, and how these in turn are based on weighting categories and scientific statements in the database. The user can also determine what the effects on welfare would be if he were to re-design the housing system, e.g. in order to reach a certain minimum overall welfare status as could be prescribed in legislation or in a product-quality control programme.

Finally, when the user disagrees with the model, he may take the developer's seat and change whatever he wants to adapt the model to his personal views, i.e. he may change scientific statements, weighting categories, needs, attributes, attribute scores, weighting factors and weighting scores. Some of these changes may have major consequences, e.g. when the weighting factors are changed, the model's connection to the scientific statements is broken, and when the attributes are deleted, welfare may no longer be assessed overall. With the changed, personal model new scores can be calculated for comparison with the original model. In this way the decision support system can be used to quantify the effect of different approaches to welfare as well as to quantify the welfare status in different (and related) housing systems for pregnant sows.

For application to other species somewhat more work is needed, including the collection of scientific statements about the species, the formulation of a mutated need tree with needs and attributes suited to the species, and an analysis of how each scientific statement relates to every attribute in the new model. Such work would not only extend the range of practical applications of the model, but may also provide further validation of the proposed welfare-assessment procedure.

Conclusions

We have explored a whole new field of welfare research, and proposed a formalised, i.e. structured, transparent, yet flexible procedure to 'objectively' assess the overall welfare status of farm animals based on available (and undisputed) scientific knowledge. In particular, we constructed a computer-based decision support system with a model for welfare assessment in pregnant sows. It allows the calculation of welfare scores, which express the overall welfare status of pregnant sows in relation to their housing and management system. For this the model consists of a list of welfare-relevant attributes that have been selected and weighted in a procedural way using a list of needs and scientific statements collected in the database. The welfare model was validated, in that it calculated welfare scores of housing systems and, to a lesser extent, weighting factors of attributes that were in accordance with expert opinion. This means we have made a small, but significant step towards improved actual decision making in the field of farm animal welfare, and to let animal welfare count.

References

Anonymous, 2001. Scientists' assessment of the impact of housing and management on animal welfare. *Journal of Applied Animal Welfare Science*: In press.

Bartussek, H., 1999. A review of the animal needs index (ANI) for the assessment of animals' well-being in the housing systems for Austrian proprietary products and legislation. *Livestock Production Science* 61: 179-192.

Bekoff, M., L. Gruen, S.E. Townsend & B.E. Rollin, 1992. Animals in science: Some areas revisited. *Animal Behaviour* 44: 473-484.

Brambell Committee, 1965. Report of the technical committee to enquire into the welfare of animals kept under intensive livestock husbandry systems. Command Report 2836. Her Majesty's Stationary Office, London.

Dawkins, M.S., 1990. From an animal's point of view: Motivation, fitness, and animal welfare. *Behavioral and Brain Sciences* 13: 1-61.

FAWC, 1992. FAWC updates the five freedoms. *Veterinary Record* 131: 357.

Gulliksen, H.O., 1950. Theory of Mental Tests. Wiley, New York.

Haynes, R.P., 2000. Do regulators of animal welfare need to develop a theory of psychological well-being? In: P. Robinson (Ed.), Two Systems – One world. EurSafe 2000 Preprints. 2nd Congress of the European Society for Agricultural and Food Ethics, 24-26 August 2000. Centre for Bioethics and Risk Assessment, Copenhagen, pp. 135-138.

Hurnik, J.F., 1988. Welfare of farm animals. *Applied Animal Behaviour Science* 20: 105-117.

Kemp, B., 2000. Het kan verkeren met adapteren. Inaugurele rede. Wageningen Universiteit, Wageningen, 23 pp.

Lijmbach, S., 1998. Animal subjectivity: A study into philosophy and theory of animal experience, PhD. Wageningen Agricultural University, Wageningen, 169 pp.

McDowell, I. & C. Newell, 1987. Measuring Health: A Guide to Rating Scales and Questionnaires. Oxford University Press, Oxford, 342 pp.

Mendl, M., 1991. Some problems with the concept of a cut-off point for determining when an animal's welfare is at risk. *Applied Animal Behaviour Science* 31: 139-146.

Scientific Veterinary Committee, 1997. The welfare of intensively kept pigs. Report to the Directorate General XXIV of the European Commission. Adopted 30. September 1997. Doc. XXIV/ScVc/0005/97, Scientific Veterinary Committee, Animal Welfare Section, Brussels, 199 pp.

Spruijt, B.M, 2001. How the hierarchical organization of the brain and increasing cognitive abilities may result in consciousness. *Animal Welfare*: In press.

Spruijt, B.M., R. van den Bos & F.T.A. Pijlman, 2001. A concept of welfare based on reward evaluating mechanisms in the brain: Anticipatory behaviour as an indicator for the state of reward systems. *Applied Animal Behaviour Science*: In press.

Streiner, D.L. & G.R. Norman, 1995. Health Measurement Scales: A Practical Guide to their Development and Use, 2nd edition. Oxford University Press, Oxford, 231 pp.

Sundrum, A., R. Andersson & G. Postler (Eds), 1994. Der Tiergerechtheitsindex-200/1994 - Ein Leitfaden zur Beurteilung von Haltungssystemen für Rinder, Kälber, Legehennen und Schweine. Verlag Köllen, Bonn, 211 pp.

Turban, E., 1995. Decision Support Systems and Expert Systems, 4th edition. Prentice-Hall, London, 887 pp.

Webster, J., 1995. Animal Welfare: A Cool Eye Towards Eden. Blackwell, Oxford, 273 pp.

Wemelsfelder, F., 1993. Animal boredom: Toward an empirical approach of animal subjectivity, PhD. University of Leiden, Leiden, 91 pp.

Wemelsfelder, F., E.A. Hunter, M.T. Mendl & A.B. Lawrence, 2000. The spontaneous qualitative assessment of behavioural expressions in pigs: First explorations of a novel methodology for integrative animal welfare assessment. *Applied Animal Behaviour Science* 67: 193-215.

Willeberg, P., 1991. Animal welfare studies: Epidemiological considerations. Proc. Meet. Soc. Vet. Epid. Prev. Med. London, pp. 76-82.

Glossary

Attribute	Descriptive variable or parameter of a housing and management system that covers a range of values, so-called attribute levels, that describe the properties of the system. Attributes include animal-based parameters such as 'health & hygiene status' and environment-based parameters such as 'space per pen' and 'rooting substrate'. Welfare-relevant attributes include the attributes in the welfare model and those attributes (also called 'influencing factors') that somehow influence one or more model attributes, e.g. 'straw presence' affects the model attributes 'rooting substrate' and 'foraging & bulk'.
Attribute level	Descriptive property or characteristic of a housing and management system. The levels that belong to one attribute are mutually exclusive, and for every housing system in the model's domain exactly one attribute level is true. Examples of levels of the attribute 'space per pen' are '> 6250 m^2' and '1-1.5 m^2'.
Attribute score	Score given to each level of an attribute in the model. Attribute scores are assigned on a scale (e.g. from 0 to 10) to the levels within each attribute.
Domain	Set of housing and management systems to which the model is designed to apply.
DSS	Decision support system. 'A DSS is an interactive, flexible, and adaptable CBIS [computer based information system, MB], specially developed for supporting the solution of a non-structured management problem for improved decision making' (Turban, E., 1995. Decision Support Systems and Expert Systems, 4th edition. Prentice-Hall, London, p. 84). A DSS for welfare assessment contains a knowledge base with scientific statements and a model base with one or more welfare models.
Housing system	Housing and management system. Type of farm, which can be identified by its attributes. The term 'housing system' denotes the whole system including the buildings, the farmer, and the animals living in the system. Examples of housing systems for pregnant sows are tethering of sows, the electronic sow feeding system and outdoor housing with huts on pasture. Our model is designed to apply at the system level, which means that it is designed to distinguish between different housing and management systems (i.e. types of farm), rather than between individual farms within systems.
Influencing factor	Attribute that affects the state (level) of one or more attributes in the model. For example, 'straw' is an influencing factor of the model attribute 'rooting substrate'.
Model	Assessment scheme, which takes a description of a housing system as input and produces a welfare score as output. Our welfare model contains a list of attributes with attribute levels and attribute scores, weighting factors and an additive (weighted average) calculation rule.

Need	Biological need, i.e. a necessity for adequate behavioural, physiological and presumably also psychological functioning, but which does not necessarily have to lead to mortality. Each need can take various states of satisfaction and frustration, which are states of an animal that may be conceived as degrees of deviation between Istwerte (actual external or internal states as perceived by the animal) and Sollwerte (setpoints or reference points) concerning the animal's main life-functions. A need may cover several control mechanisms (sub-needs), because it covers a set of behaviours and physiological responses, which can be activated by a certain class of stimuli and deactivated by a specific event or behaviour. Some examples of needs are the need for food, water, rest, thermal comfort, social contact, and exploration.
OWA	Overall-welfare assessment. Integrated evaluation of the welfare status of animals taking into account all components of their welfare status, such as needs and attributes.
Scientific statement	Proposition about a finding from empirical research, preferably published in a peer-reviewed publication or otherwise established as a 'fact' among welfare scientists. In the knowledge base of the decision support system we included only those scientific statements, which are not disputed and which appear to be relevant to distinguish between housing systems for pregnant sows on welfare grounds. Scientific statements typically specify some relationship between an environment-based attribute of a housing system and an animal-based welfare performance criterion that was empirically measured (see also weighting category).
Type	See weighting category.
Weighting category	Class of welfare performance criteria, generally related to some scientific paradigm to measure welfare. Examples of weighting categories are 'preferences', 'natural behaviour', and 'fitness'. In our model each weighting category has four levels with weighting scores (WS) attached to each level. One of these four weighting-category levels has a very high weighting score, which is used to define minimum requirements for welfare. The assignment of weighting scores is based on the dimensions of intensity, duration and incidence of the welfare-relevant condition.
Welfare (status)	State of an animal regarding its quality of life (absence of suffering and degree of comfort) as evaluated from the animal's point of view. In this thesis the welfare status is assessed using available scientific knowledge about the animal's biological functioning, i.e. about the degree of satisfaction and frustration of its needs.
WF	Weighting factor, numerical score assigned to an attribute to weight it relative to the other attributes in the model. The weighting factor score expresses the magnitude of difference between the best and worst level of the attribute as indicated by the scientific statements in the knowledge base of the decision support system.

Summary

This thesis aimed to find a formalised procedure to assess the welfare status of farm animals in their husbandry environment on the basis of available scientific knowledge about the animals' biological needs in a transparent and structured way, while sufficiently flexible to allow improvements when new insights would become available in the future. Techniques from information technology were used to handle the available knowledge and complex weighting procedure. Interviews with experts were used to validate the outcomes.

The case of pregnant sows was chosen to develop an operational welfare model, which takes a description of a housing and management system as input and produces a welfare score as output. To increase transparency and structure the model was implemented into a computer-based decision support system, which is a computer programme that contains the model in the model base and scientific statements collected from the literature in the knowledge base. The decision support system was developed according to the so-called Evolutionary Prototyping Method, where a first prototype was improved in repeated upgrading versions.

A first, very simple, prototype decision support system was constructed, in which a welfare score was calculated based on the biological needs of the animals. An evaluation showed that further development was practically feasible, but required a more firm theoretical basis, which was reviewed subsequently in order to formulate recommendations for the further development of the decision support system.

It was acknowledged that overall welfare assessment involves normative judgements, but these are inherent in science generally, and are not peculiar for animal welfare. Furthermore, overall welfare assessment does not try to solve moral or political acceptability questions. It concerns the attempt to provide the best possible answer to the question what the welfare status of farm animals in a certain housing and management system de facto is, given the current state of biological science.

Finally, the current state of welfare assessment tables and schemes (welfare models) was reviewed, and the needs-based approach to welfare assessment was discussed in interviews with experts. Although experts differ on details, wide support for welfare assessment based on the animals' biological needs was found, and a 'common' list of needs for the purpose of farm animal welfare assessment was formulated.

The further development of the decision support system went through several stages, which indicated that a frame of reference was needed. For this reason, some pig-welfare scientists were asked to identify the main housing systems for pregnant sows, and to give an overall welfare score for each system. The main systems and their welfare scores were (where the welfare status is indicated with different superscripts when the systems differ significantly using ANOVA, with a welfare status of a < b < c): tethered^{a}, individual housing in stalls^{a}, group housing with stalls^{b}, Biofix^{b} (i.e. trickle feeding), electronic sow feeding^{b}, outdoor housing with huts^{c}, and the Family Pen system^{c}. These systems, and their classification in low-, mid-, and high-welfare systems, were used as a benchmark for subsequent modelling, which resulted in a 'final' model.

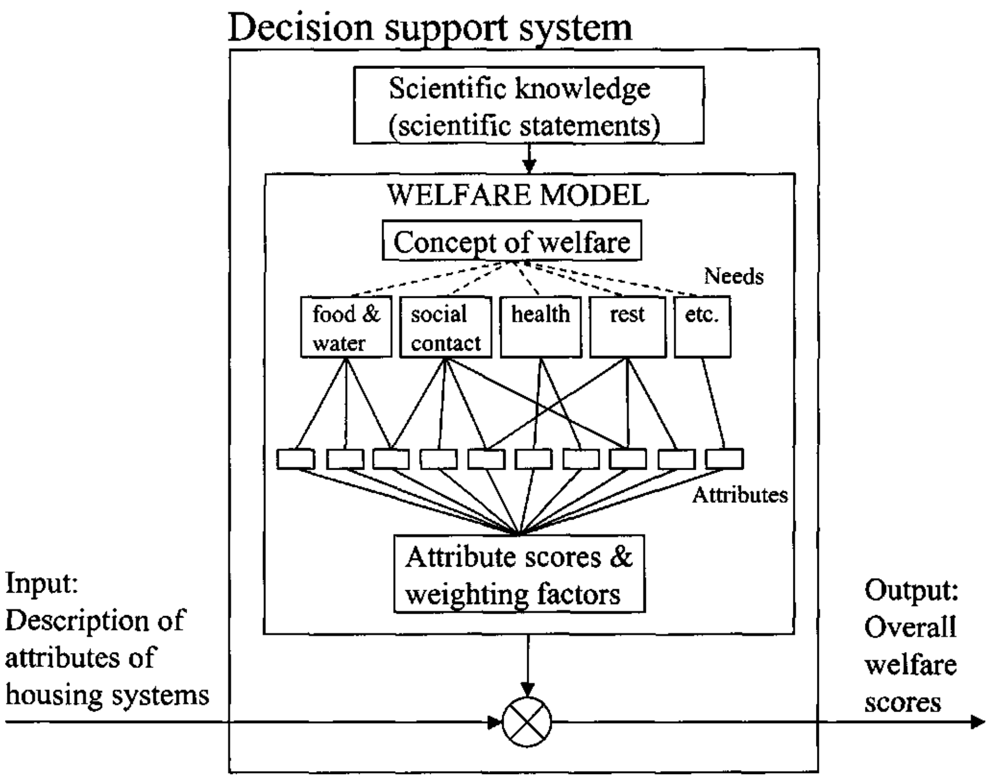
The decision support system with the 'final' sow welfare (SOWEL) model[1] is a so-called relational database where information, such as the descriptions of housing systems, attributes, needs and scientific statements, is stored in related tables. The model was constructed using a formalised procedure to capture the reasoning steps involved in overall welfare assessment. The procedure involves several steps. Firstly, the model's domain is identified, i.e. the category of animals and the set of housing systems to which the model applies are determined. Secondly, welfare is defined and de-composed into functional elements, i.e. a list of the biological needs relevant for the model's domain. Thirdly, the relevant scientific statements (the 'facts') are collected. Fourthly, the set of welfare-relevant attributes, which specify the different welfare advantages and disadvantages of housing systems, are determined in relation to the model's domain, the list of needs, and the scientific statements. Fifthly, the attributes are weighted relative to each other based on welfare performance measures as described in the scientific statements using the dimensions of intensity, duration and incidence. Finally, the information is summed up, i.e. the overall welfare status is determined based on the weighted attributes of the housing systems, while the main housing systems in the domain are used to serve as a benchmark to interpret the results.

The decision support system was used to calculate weighting factors for the attributes in the model as well as to calculate overall welfare scores for an extended set of housing and management systems. The predicted scores were used to validate the model using expert opinion. Expert opinion was solicited in a written questionnaire, in which experts were asked to assess 15 different housing systems, and to weight 20 different attributes. We confirmed the results of the previous interviews for the seven main housing systems, and found that the extended set of 15 housing systems was (also) classified roughly into low-, mid- and high-welfare systems. The two conventional, individual housing systems (tethered and individual housing in stalls) were classified as low-welfare systems, while the high-welfare systems in our data set all had outdoor access and rooting substrate. The attributes related to social contact, space and substrate were among the most important ones.

The results show that the experts were capable of assigning overall welfare scores. The model correlated well with expert opinion, but the degree of concordance among the experts was better for overall welfare scores than for the weighting of attributes. The predictions made by the model correlated significantly with expert opinion. Spearman's rank correlation coefficients were 0.92 ($P < 0.01$) for overall welfare assessment of housing systems and 0.73 ($P < 0.01$) for the weighting of attributes. In both cases the model's performance was equivalent to the performance of an average expert, i.e. half of the experts had higher correlations, while the other half had lower correlation coefficients. This validated the model with expert opinion, showing that information technology and interviews with experts can effectively be used to construct an operational, flexible, and adaptable decision support system to assess the overall welfare status of farm animals based on available scientific knowledge. Overall welfare assessment is no longer just a matter of personal (expert) opinion.

1 The model configuration in the decision support system is shown in the appendix to the summary.

Appendix to the summary



Structure of the 'final' model in the decision support system for integrated animal welfare assessment in the case of pregnant sows.

Samenvatting

Modelleren van dierenwelzijn: De ontwikkeling van een beslissingsondersteunend systeem om de welzijnsstatus van dragende zeugen in te schatten

In dit proefschrift is geprobeerd een geformaliseerde procedure te vinden om de welzijnsstatus van landbouwhuisdieren in te schatten op basis van beschikbare wetenschappelijke kennis van de biologische behoeften van de dieren, op een transparante en gestructureerde manier die voldoende flexibel is om verbeteringen door te kunnen voeren wanneer deze beschikbaar komen in de toekomst. Technieken uit de informatietechnologie werden gebruikt om de beschikbare kennis en complexe weegprocedure aan te kunnen. Interviews met deskundigen werden gebruikt om de uitkomsten te valideren.

De dragende zeugen dienden als voorbeeld om een operationeel model te ontwikkelen welke een beschrijving van een huisvestings- en managementsysteem als input heeft, en een welzijnsscore als output. Ten behoeve van de transparantie en structuur van het model werd het model geïmplementeerd in een beslissingsondersteunend computersysteem, dat wil zeggen een computerprogramma dat een model in de modelbank bevat en wetenschappelijke uitspraken uit de literatuur in een kennisbank. Het beslissingsondersteunend systeem was ontwikkeld volgens de zogenaamde 'Evolutionary Prototyping Method', waarbij een eerste prototype is gemaakt wat daarna stapsgewijs is verbeterd.

In het eerste, heel eenvoudig prototype van het beslissingsondersteunend systeem werd de welzijnsscore berekend op basis van de biologische behoeften van de dieren. Evaluatie van het prototype liet zien dat verdere ontwikkeling haalbaar was, maar dat daarvoor een betere theoretische onderbouwing wenselijk was. Een literatuur overzicht is daarom gemaakt met als doel aanbevelingen te doen voor de verdere ontwikkeling van het beslissingsondersteunend systeem.

Waardeoordelen spelen een rol bij het inschatten van welzijn, maar waardeoordelen zijn inherent in wetenschap in het algemeen, en niet specifiek voor dierenwelzijn. Bovendien probeert welzijnsinschatting geen vragen te beantwoorden betreffende de morele of politieke aanvaardbaarheid. Het betreft de poging om het best mogelijke antwoord te vinden op de vraag wat de welzijnsstatus van landbouwhuisdieren in een gegeven huisvestings- en managementsysteem de facto is, gegeven de huidige stand van de wetenschap van de biologie van het dier.

Tenslotte is een overzicht gegeven van tabellen en schema's (modellen) voor welzijnsinschatting, en de benadering van welzijn via de behoeften van dieren is besproken in interviews met deskundigen. Alhoewel de deskundigen van mening verschillen over details, bestaat er brede steun voor het idee om welzijn op basis van de biologische behoeften van dieren in te schatten. Een 'gemeenschappelijke' lijst van behoeften is geformuleerd om welzijn van landbouwhuisdieren te bepalen.

De verdere ontwikkeling van het beslissingsondersteunend systeem ging via verschillende stadia die aangaven dat een referentiekader nodig was. Daarom werd aan een beperkt aantal deskundigen op het gebied van het welzijn van varkens gevraagd om de belangrijkste huisvestingssystemen te benoemen, en een totaal score voor welzijn te geven voor elk systeem. De belangrijkste systemen en hun welzijnsscores waren (waarbij de welzijnsstatus is aangeduid middels een ander superscript wanneer de systemen significant verschilden in de ANOVA, waarbij de welzijnsstatus van a < b < c): aangebonden[a], individuele huisvesting in voerligboxen[a], groepshuisvesting met voerligboxen[b], Biofix[b] (d.w.z. een langzame dosering van het voer), krachtvoerstations[b], weidegang met hutten[c] en de Familie stal[c]. Deze systemen en hun classificatie in systemen met een lage, gemiddelde en hoge welzijnsstatus werden vervolgens gebruikt als ijkpunten voor de daaropvolgende modellering, die uiteindelijk leidde tot het huidige model.

Het beslissingsondersteunend systeem met het huidige welzijnsmodel voor zeugen (SOWEL, SOw WElfare) [1] is een zogenaamde relationele database waarin informatie, zoals de beschrijvingen van de huisvestingssystemen, attributen, behoeften en wetenschappelijke uitspraken, is opgeslagen in aan elkaar gerelateerde tabellen. Het model was gemaakt met een geformaliseerde procedure voor de redeneerstappen van welzijnsinschatting. De procedure beslaat verschillende stappen. Allereerst is het domein van het model geïdentificeerd, d.w.z. de diercategorie en de verzameling van huisvestingssystemen waarop het model van toepassing is zijn vastgesteld. Ten tweede is welzijn gedefinieerd en ontleed in functionele elementen, d.w.z. een lijst van biologische behoeften die relevant zijn voor het domein. Ten derde zijn de relevante wetenschappelijke uitspraken (de 'feiten') verzameld. Ten vierde is de verzameling van welzijnsrelevante voor- en nadelen van huisvestingssystemen bepaald in relatie tot het domein, de lijst met behoeften en de wetenschappelijke uitspraken. Ten vijfde zijn de attributen relatief ten opzichte van elkaar gewogen middels de welzijnsprestatiematen zoals beschreven in de wetenschappelijke uitspraken in relatie tot de dimensies intensiteit, duur en incidentie. Tot slot is de informatie 'opgeteld', d.w.z. de algehele welzijnsstatus is vastgesteld op basis van de gewogen attributen van de huisvestingssystemen, waarbij de belangrijkste huisvestingssystemen in het domein gebruikt zijn om als ijkpunten te dienen om de resultaten te kunnen interpreteren.

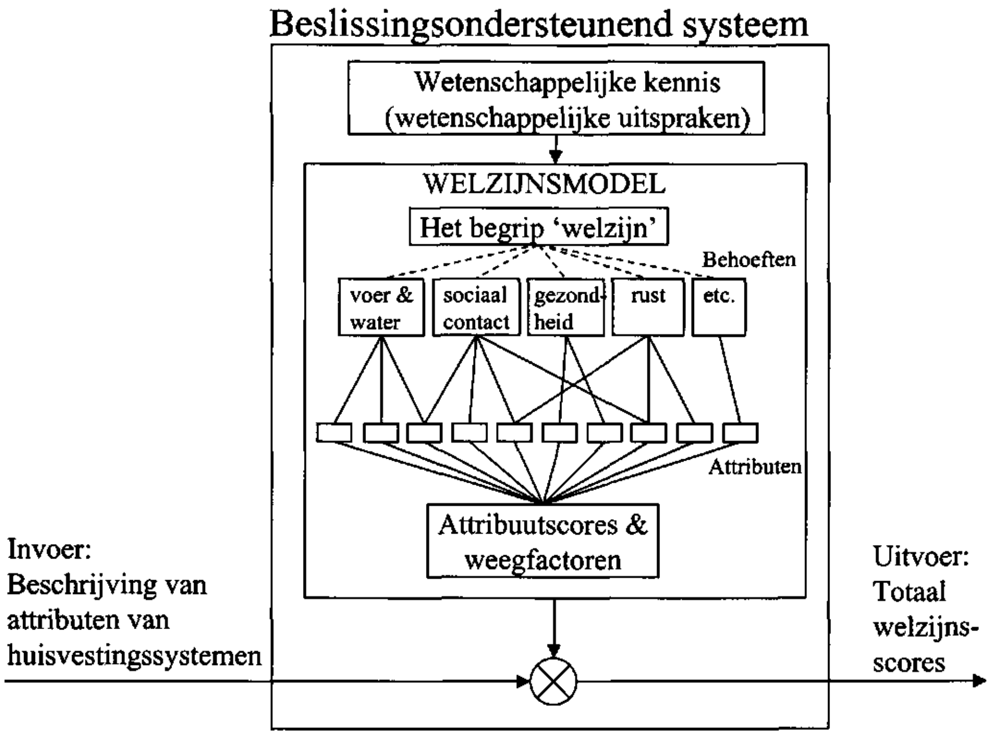
Het beslissingsondersteunend systeem is gebruikt om weegfactoren voor attributen in het model te berekenen, evenals om algehele welzijnsscores voor een uitgebreidere verzameling van huisvestings- en managementsystemen te berekenen. De voorspelde scores werden gebruikt om het model te valideren met de opinie van deskundigen. De deskundigenopinie werd verzameld met een geschreven vragenlijst, waarin aan de deskundigen werd gevraagd om 15 verschillende huisvestingssystemen in te schatten en om 20 verschillende attributen te wegen. We bevestigden de resultaten van eerdere interviews betreffende de zeven belangrijkste huisvestingssystemen, en vonden dat de uitgebreidere verzameling van 15 huisvestingssystemen (ook) werd geclassificeerd in systemen met een lage, gemiddelde en hoge welzijnsstatus. De twee conventionele, individuele huisvestingsystemen (aangebonden en individuele voerligboxen) werden geclassificeerd als systemen met laag welzijn, terwijl de systemen met hoog welzijn in onze dataset allemaal uitloop naar buiten en (wroet)substraat hadden. De belangrijkste attributen waren vooral de attributen die relateerden aan sociaal contact, ruimte en substraat.

De resultaten laten zien dat deskundigen in staat waren algehele welzijnsscores te geven. Het model correleerde goed met de deskundigen, maar de mate van overeenstemming tussen de deskundigen was beter voor algehele welzijnsscores dan voor de weging van de attributen. De voorspellingen van het model correleerden significant met de deskundigen. De Spearman rang-correlatiecoëfficiënten waren 0,92 ($P < 0{,}01$) voor algehele welzijnsscores voor huisvestingssystemen en 0,73 ($P < 0{,}01$) voor de weging van de attributen. In beide gevallen

1 Voor een schematische weergave van de modelconfiguratie in het beslissingsondersteunend systeem zie de bijlage bij de samenvatting.

was de prestatie van het model equivalent met die van de gemiddelde deskundige, d.w.z. de helft van de deskundigen had een hogere correlatie, terwijl de andere helft een lagere correlatiecoëfficiënt had. Dit valideerde het model met de opinie van deskundigen en liet zien dat informatietechnologie en interviews met deskundigen daadwerkelijk gebruikt kunnen worden om een operationeel, flexibel en adapteerbaar beslissingsondersteunend systeem te maken om de algehele welzijnsstatus van landbouwhuisdieren in te schatten op basis van beschikbare wetenschappelijke kennis. Welzijnsinschatting is niet langer slechts een kwestie van persoonlijke (deskundigen) opinie.

Bijlage bij de samenvatting



Structuur van het huidige model in het beslissingsondersteunend systeem voor geïntegreerde welzijnsinschatting bij dragende zeugen.

Curriculum Vitae

Marc Bracke was born on the 31st of December 1965 in Clinge, Zeeuws Vlaanderen, The Netherlands. In 1984 he finished his secondary education at the Jansenius Lyceum in Hulst. From 1984 to 1994 he studied Veterinary Medicine and Philosophy at the University of Utrecht. In 1989-1990 he studied philosophy at the University of Florida in Gainesville (USA, supported by a Fulbright scholarship and the Scrinerius Foundation). His majors in philosophy were animal ethics and the philosophy of (the animal) mind. His Master's thesis in Philosophy was entitled: 'Killing animals, or why no wrong is done to an animal when it is killed painlessly'.

In 1991-1992 he took a Master's course in Applied Animal Behaviour and Welfare at the University of Edinburgh (Scotland, UK, supported by the Dutch Society for the Protection of Animals). His MSc thesis was entitled: 'Can animals have a preference not to be killed? A study in cognitive ethology with observations on red deer, laying hens and mice'.

As of April 1994 he worked as a veterinarian in several mixed (large and small) animal practices in the Netherlands. In 1996 he started working at IMAG in Wageningen. As of 2001 he will be working at ID-Lelystad as sr-scientist of animal welfare.

Related publications

Anonymous, 2001. Scientists' assessment of the impact of housing and management on animal welfare. *Journal of Applied Animal Welfare Science*: In press.

Bracke, M.B.M., J.H.M Metz & A.J. Udink ten Cate, 1996. Modelling animal welfare by structural decomposition. In: I.J.H. Duncan et al. (Eds), Proceedings of the 30th International Congress of the International Society for Applied Ethology. Guelph, Canada, p. 81.

Bracke, M.B.M., J.H.M. Metz & A.J. Udink ten Cate, 1997. Assessment of animal welfare in husbandry system. In: J.T. Sørensen (Ed), Livestock Farming Systems - More than Food Production. EEAP Publication No. 89, Wageningen Pers, Wageningen, pp. 231-237.

Bracke, M.B.M. & J.H.M. Metz, 1997. Quantifying welfare for decision making purposes: An example. In: P. Koene & H.J. Blokhuis (Eds), Proceedings of the 5th European Symposium on Poultry Welfare. Ponsen & Looyen, Wageningen, p. 165-166.

Bracke, M.B.M. & J.H.M. Metz, 1997. The development of a prototype decision support system to assess the welfare status of pigs. Proceedings of the 31st International Congress of the International Society for Applied Ethology, Prague, p. 129.

Bracke, M.B.M, J.H.M. Metz, A.A. Dijkhuizen & B.M. Spruijt. 1998. The development of a decision support system to assess the welfare status of pregnant sows. In: Th. Wensing (Ed), Production Diseases in Farm Animals: 10th International Conference 1998, Utrecht 24-28 august, Wageningen Pers, Wageningen, p. 98.

Bracke, M.B.M., B.M. Spruijt & J.H.M. Metz, 1999. Overall animal welfare assessment reviewed. Part 1: Is it possible? *Netherlands Journal of Agricultural Science* 47: 279-291.

Bracke, M.B.M., J.H.M. Metz & B.M. Spruijt, 1999. Overall animal welfare reviewed. Part 2: Assessment tables and schemes. *Netherlands Journal of Agricultural Science* 47: 293-305.

Bracke, M.B.M., B.M. Spruijt & J.H.M. Metz, 1999. Overall animal welfare reviewed. Part 3: Welfare assessment based on needs and supported by expert opinion. *Netherlands Journal of Agricultural Science* 47: 307-322.

Bracke, M.B.M., J.H.M. Metz, B.M. Spruijt & A.A. Dijkhuizen, 1999. Overall welfare assessment of pregnant sow housing systems based on interviews with experts. *Netherlands Journal of Agricultural Science* 47: 93-104.

Bracke, M.B.M., J.H.M. Metz & B.M. Spruijt, 1999. Modelling of pig-expert reasoning to assess overall welfare of pregnant sows. In: K. Bøe, M. Bakken & B.O. Braastad (Eds), Proceedings of the 33rd International Congress of the International Society for Applied Ethology. Agricultural University of Norway, Ås, p. 144.

Bracke, M.B.M., J.H.M. Metz & B.M. Spruijt, 1999. Values in overall farm animal welfare assessment In: V. Beekman & F. Brom (Eds), First European Congress on Agricultural and Food Ethics, Preprints, March 4-6, Wageningen, pp. 69-72.

Bracke, M.B.M., 2000. Systemen voor integrale welzijnsbeoordeling in relatie tot beleid. In: E. Röder & R. van den Bos (Eds), Soorteigen gedrag van dieren-huisvestingsbeleid. Rapport van de workshop op 13 April. ICWD, Utrecht, p. 17-20.

Bracke, M.B.M., J.H.M. Metz & B.M. Spruijt, 2000. Welfare model to assess housing systems for pregnant sows. In: M.J.M. Tielen & M.Th. Voets (Eds), Proceedings of the

10th International Congress on Animal Hygiene, Volume 1. Animal Health Service, Boxtel, p. 377-380.

Bracke, M.B.M., J.H.M. Metz & B.M. Spruijt, 2001. Development of a decision support system to assess farm animal welfare. Proceedings of the International Workshop on Assessment of Animal Welfare at Farm or Group Level, 27-28 August in Copenhagen, *Acta Agriculturae Scandinavica Section A, Animal Science, Supplement*: In press.

Bracke, M.B.M., J.H.M. Metz, A.A. Dijkhuizen & B.M. Spruijt, 2001. Development of a decision support system for assessing farm animal welfare in relation to husbandry systems: Strategy and prototype. *Journal of Agricultural and Environmental Ethics*: Accepted.

The research work described in this thesis was carried out at the Institute of Agricultural and Environmental Engineering (IMAG). This work was funded by the Netherlands Organisation for Scientific Research (NWO) and the Netherlands Ministry of Agriculture, Nature Management and Fisheries trough the Priority Programme 'Grenzen aan welzijn en dierlijke productie'. Additional financial support was given by the Dutch Society for the Protection of Animals (NVBD). The project was under the supervision of the Technology Foundation (STW).

Bracke, M.B.M.
Modelling of animal welfare: The development of a decision support system to assess the welfare status of pregnant sows
Modelleren van dierenwelzijn: De ontwikkeling van een beslissingsondersteunend systeem om de welzijnsstatus van dragende zeugen in te schatten
Thesis Wageningen University – With Ref. – With summaries in English and Dutch
ISBN 90-5808-391-8

This thesis is also available as IMAG report 2001-02 (ISBN 90-5406-188-X) of the Institute of Agricultural and Environmental Engineering (IMAG), P.O. Box 43, 6700 AA Wageningen, The Netherlands.

All rights reserved. No parts of this book may be reproduced, stored in a retrieval system of any nature, in any form or by any means, electronical, mechanical, photocopying, recording, or otherwise, without the prior permission of IMAG, P.O. Box 43, 6700 AA Wageningen, The Netherlands.
© 2001

Printed in the Netherlands, by Grafisch Service Centrum van Gils B.V., Wageningen.