117. Analysis of social behaviors in large groups: simulation and genetic evaluation

Z. Wang*, H.P. Doekes and P. Bijma

Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; zhuoshi.wang@wur.nl

Abstract

Harmful social interactions, such as injurious feather pecking in poultry and tail biting in swine, reduce both animal welfare and efficiency. While these traits are heritable, application of breeding is still limited due to the lack of proper genetic models and precise phenotyping methods for large groups. In the near future, large scale longitudinal data on social interactions will become available thanks to developments in computer vision and artificial intelligence. Here we present models to simulate and analyze such data, which are an extension of the classic social genetic model. Latent traits were defined to represent the tendency of individuals to be engaged in behavioral interactions, distinguishing performer and recipient effects. Binary interaction records were simulated and subsequently analyzed using generalized linear mixed models. Results show that high accuracies of estimated breeding values can be obtained (0.4-0.7), despite the low observed-scale heritability of the binomial trait (0.05-0.2). We conclude that our model can be promising for breeding value estimation for social traits in large groups.

Introduction

Livestock regularly engage in behavioral interactions, and modern group-housing systems may increase the occurrence of harmful social behaviors, such as injurious feather pecking in poultry and tail biting in swine. To reduce the occurrence of harmful social behaviors, management strategies have been proposed (Van Krimpen *et al.*, 2016; De Vries *et al.*, 2004). However, the application of breeding strategies to behavioral traits is still limited.

One challenge of applying breeding strategies is that the social behavior of an individual may depend not only on the effect of the genes of the individual itself (known as the direct genetic effect, DGE), but also on that of its group mates (known as the indirect genetic effect, IGE). In the last two decades, genetic models for social traits have been developed to estimate DGE and IGE (Muir, 2005; Bijma *et al.*, 2007; Ellen *et al.*, 2014). These models assume that each individual has the opportunity to affect the phenotype of each of its pen mates equally. Therefore, these models are particularly useful in small groups, in which animals have ample interactions with all pen mates and the strength of social interaction can be assumed similar between all individuals.

For large groups the assumption of equal interaction does not hold anymore. In recent years, social network analysis has been successfully applied to large animal populations (Krause *et al.*, 2021; Radersma 2021). Social network analysis requires detailed records of the social interactions, including time, physical position, duration, and performer and recipient identity. Recording every interaction event between individuals in a large population used to be time- and labour-demanding, but is gradually becoming feasible due to developments in animal detection and tracking technologies. With these technologies, distinguishing performer and recipient effect is possible. Therefore, the terms performer and recipient effect will be used in the rest of this paper instead of DGE and IGE. It is expected that in the near future these detailed records of social interaction will become available. However, there is still a lack of methods to translate such data into estimates of individual performer effect and recipient effect.

R.F. Veerkamp and Y. de Haas (eds)

DOI: 10.3920/978-90-8686-940-4_117, © Z. Wang *et al.* 2022

In this study, we: (1) extended the social effect model for genetic analysis of large-scale longitudinal data on animal social behaviors; and (2) evaluated the effect of number of individuals and number of social interactions on accuracy of breeding values (BV) estimated with this model. First, latent traits were defined to represent the tendency of individuals to be engaged in behavioral interactions, distinguishing performer and recipient effects. Second, social interactions with known performer and recipient were simulated under various population settings based on an assumed genetic structure. Last, statistical models were applied to estimate BV of social traits from the simulated records, and to evaluate the accuracies of the EBV.

Materials & methods

Trait definition. For each individual, two latent traits were defined to represent its liability to be engaged in a social interaction as a performer (trait α) or a recipient (trait β). Each trait consisted of two parts: a heritable effect (the BV) and a permanent environmental effect. The latter was used to account for similarity between repeated records on the same individual.

$$P_{\alpha,i} = \mu + A_{\alpha,i} + Ep_{\alpha,i} \tag{1}$$

$$P_{\beta,i} = \mu + A_{\beta,i} + Ep_{\beta,i} \tag{2}$$

where $P_{\alpha,i}$ and $P_{\beta,i}$ were the individual's normally-distributed liability for each trait, $A_{\alpha,i}$ and $A_{\beta,i}$ were the individual's BVs, and $Ep_{\alpha,i}$ and $Ep_{\beta,i}$ were the permanent environmental effects. For the results presented here, we assumed that both traits were independent (genetic correlation of 0) and had the same genetic variance ($\sigma_{A\alpha}^2 = \sigma_{A\beta}^2$), and that the permanent environmental variances were equal to the corresponding genetic variances.

For interaction between individuals i and j (see section 'Simulations' for conditions related to their proximity), the probability of i performing the social interaction towards j is given by:

$$p_{ij} = \text{logistic}(P_{\alpha,i} + P_{\beta,i}) = 1/\,[1+\exp(-P_{\alpha,i} - P_{\beta,i})] \tag{3}$$

Thus, the logistic function was used to rescale ($P_{\alpha,i} + P_{\beta,i}$) from the real number domain of the liability to the probability domain of 0 to 1. Binary interaction records were then generated by sampling a random number from a Bernoulli distribution, where 1 meant the interaction took place, while 0 meant it did not take place.

Simulations. For social interactions to occur, animals need to be in each other's proximity. In this study, we simulated the dynamic physical position of animals through agent-based modelling, using three behaviors that were alternated: eating (at feeders in the pen), resting and walking. Only when animals encountered each other, a 0/1 interaction record was generated as described above. We generated 2000 individuals from 100 sires and 400 dams, and each dam produced 5 offspring. Simulations were run for 20 pens, each with 100 random individuals, and until a total of 10,000 interactions took place in a pen. Given a simulated mean interaction probability of 1% (E(p) = 0.01), this implied a total of ~1,000,000 encounters per pen.

Finding realistic input values. Observed-scale heritabilities for binomial traits related to animal behavior are typically in the range of 0.05 to 0.20. Kjaer and Sørensen (1997), for example, found heritabilities in this range for number of pecks performed, based on an average of ~25 interactions per individual. To find realistic input values for our simulations, we simulated a dataset where individuals performed on average 25 pecks, then estimated the observed-scale heritability with ordinary linear mixed models, and tuned the input value for the genetic variance until observed-scale heritability was either 0.05, 0.1 or 0.2 (Table 1).

Breeding value estimation. To estimate BV for the performer effect (α) and the recipient effect (β), a generalized linear mixed model with a logit link-function and a binomial distribution was fitted in ASReml 4.1 (Gilmour *et al.*, 2015):

$$\text{logit}(p_{ij}) = \mu + Pen_k + A_{\alpha,i} + A_{\beta,i} + Ep_{\alpha,i} + Ep_{\beta,i} \tag{4}$$

where p_{ij} is the probability that i performs a social interaction towards j, Pen_k is a fixed pen effect, A is the BV and Ep is the permanent environment effect.

Results

Table 1 shows the genetic variances (σ_A^2) that correspond to observed scale heritabilities of total number of interactions performed (h_o^2) of 0.02, 0.05 and 0.1. These genetic variances correspond to interaction probabilities of top and bottom ranking individuals (± $2\sigma_A$) that differed by a factor of 1.5 for $h_o^2 = 0.05$, 2.5 for $h_o^2 = 0.10$, and 5.5 for $h_o^2 = 0.20$. Hence, despite the low observed-scale heritability, the genetic differences between individuals in their tendency to engage in social interactions were very large.

To examine the relationship between the number of individuals and the accuracy of EBV, we included 5, 6, …, up to 20 pens in the genetic analysis (thus 500, 600, …, 2,000 individuals). When 500 individuals were included, the GLMM yielded an accuracy of EBV of 0.30, 0.36 and 0.44, given heritabilities of 0.05, 0.10 and 0.2, respectively (Figure 1a). When all the 2,000 individuals were included in the analysis, the accuracies were 0.47, 0.59 and 0.71.

To investigate the effect of the number of interactions on accuracy of EBVs, the GLMM was fitted by including the first 5, 10, 20, 50 or 100 thousand interaction records. Accuracy increased with an increasing number of records because more observations of sibs were included, but benefit was small above 20 thousand records (Figure 1b).

Table 1. Simulated genetic variances (σ_A^2) required to obtain a certain observed-scale heritability (h_o^2), the corresponding m ± $2\sigma_A$ range of expected interaction probabilities, and accuracy of EBV yielded by the ordinary linear mixed model, using data from all 20 pens.

h_0^2	σ_A^2	$\bar{p}(A_a=\mu-2\sigma_A)$	$\bar{p}(A_a=\mu+2\sigma_A)$	Accuracy LMM
0.05	0.012	0.008	0.012	0.173
0.1	0.038	0.006	0.015	0.195
0.2	0.170	0.004	0.022	0.241

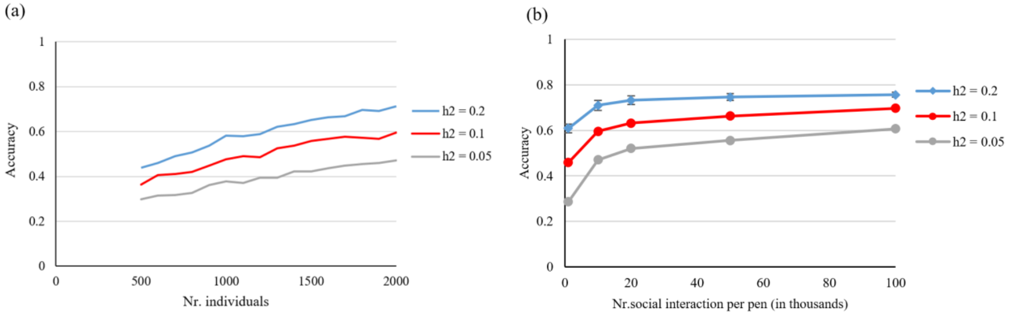

Figure 1. Accuracy of EBV from the GLMM for (a) different numbers of individuals involved and (b) different numbers of observed interactions.

Discussion

Here we developed an individual-based simulation method to generate social interaction events, and a GLMM to estimated breeding values for performer and recipient effect from the resulting data. Results showed that our method yielded a high accuracy of EBV (of 0.47, 0.59 and 0.71 for 2,000 individuals with h_o^2 of 0.05, 0.1 and 0.2, respectively) in comparison to an ordinary LMM (accuracies of 0.17, 0.20 and 0.24), and that genetic differences in the tendency to express a behavior can be large despite a low observed-scale heritability (Table 1).

This simulation shows that, even though the genetic variance (σ_A^2) was the same as the permanent environment variance (σ_{Ep}^2) on the liability scale, the observed-scale heritability can still be very low (e.g. 0.05). This low heritability is due to the random sampling of the event, which introduces extra variance in the observed records.

Our results indicate that the breeding values can be accurately estimated for social interaction traits even if the observed level heritability is very low. Similar results were found when both traits were positively or negatively correlated (results not shown). Note that in these simulations, the maximum information used for BV estimation was own performance, 4 full sibs and 15 half sibs. The accuracy was quite high with this amount of information, and we expect that higher accuracy can be achieved with genomic prediction. This high accuracy can partly be explained by the repeated measures of the trait. In our simulations, one individual could interact with 99 pen mates and was on average engaged in 200 interactions and 20,000 encounters. This means that the breeding value of each individual was expressed 20,000 times. While this may seem a large amount of data, one can easily imagine that an animal performs 10 tail bites or feather pecks a day in total. In that case, our data would correspond to one month of recording, which is shorter than the typical duration of, e.g. the fattening period in pigs or the laying period in laying hens.

In summary, we presented methods to simulate and analyse behavioral data, which distinguishes between performer and recipient effects, and is applicable to large-scale longitudinal data on animals kept in large groups, which will become available in the near future.

Acknowledgements

This work was financially supported by the Dutch Research Council (NWO-TTW).

References

Bijma P., Muir W.M., & Van Arendonk, J.A. (2007). Genetics, 175(1), 277-288. https://doi.org/10.1534/genetics.106.062711

Brinker T., Bijma P., Visscher J., Rodenburg T. B., and Ellen E. D. (2014). Genetics Selection Evolution, 46(1), 1-10. https://doi.org/10.1186/1297-9686-46-33

DeVries T. J., Von Keyserlingk M. A. G., & Weary D. M. (2004). Journal of dairy science, 87(5), 1432-1438. https://doi.org/10.3168/jds.S0022-0302(04)73293-2

Ellen E.D., Rodenburg T. B., Albers G. A., *et al.*(2014). Frontiers in genetics, 5, 377. https://doi.org/10.3389/fgene.2014.00377

Gilmour A. R., Gogel B. J., Cullis B. R., Welham S., and Thompson R. (2015). ASReml user guide release 4.1 structural specification. Hemel hempstead: VSN international ltd.

Kjaer J. B., and Sørensen P. (1997). British Poultry Science, 38(4), 333-341. https://doi.org/10.1080/00071669708417999

Krause H. V., Baum K., Baumann A., and Krasnova H. (2021). Media Psychology, 24(1), 10-47. https://doi.org/10.1080/15213269.2019.1656646

Muir, W. M. (2005). Genetics 170(3):1247-1259.

Radersma R. (2021). Methods in Ecology and Evolution, 12(1), 42-53. https://doi.org/10.1111/2041-210X.13499

Van Krimpen M. M., Leenstra F., Maurer V., and Bestman M. (2016). Journal of Applied Poultry Research, 25(1), 129-138. https://doi.org/10.3382/japr/pfv048