OPEN

Heredity (2014) 112, 197–206

www.nature.com/hdy

ORIGINAL ARTICLE

Indirect genetic effects and kin recognition: estimating IGEs when interactions differ between kin and strangers

SW Alemu[1,2], P Berg[1,3], L Janss[1] and P Bijma[2]

Social interactions among individuals are widespread, both in natural and domestic populations. As a result, trait values of individuals may be affected by genes in other individuals, a phenomenon known as indirect genetic effects (IGEs). IGEs can be estimated using linear mixed models. The traditional IGE model assumes that an individual interacts equally with all its partners, whether kin or strangers. There is abundant evidence, however, that individuals behave differently towards kin as compared with strangers, which agrees with predictions from kin-selection theory. With a mix of kin and strangers, therefore, IGEs estimated from a traditional model may be incorrect, and selection based on those estimates will be suboptimal. Here we investigate whether genetic parameters for IGEs are statistically identifiable in group-structured populations when IGEs differ between kin and strangers, and develop models to estimate such parameters. First, we extend the definition of total breeding value and total heritable variance to cases where IGEs depend on relatedness. Next, we show that the full set of genetic parameters is not identifiable when IGEs differ between kin and strangers. Subsequently, we present a reduced model that yields estimates of the total heritable effects on kin, on non-kin and on all social partners of an individual, as well as the total heritable variance for response to selection. Finally we discuss the consequences of analysing data in which IGEs depend on relatedness using a traditional IGE model, and investigate group structures that may allow estimation of the full set of genetic parameters when IGEs depend on kin.
Heredity (2014) **112,** 197–206; doi:10.1038/hdy.2013.92; published online 30 October 2013

Keywords: social interactions; indirect genetic effects; kin recognition; kin; stranger

INTRODUCTION

Social interactions among individuals are common in both wild and domestic populations, and in animals, plants and microorganisms (Frank, 2007). With social interactions, the trait value of an individual may be affected by genes in other individuals, a phenomenon that works out as indirect genetic effects (IGEs; Griffing, 1967, 1976; Moore *et al.*, 1997; Wolf *et al.*, 1998). An IGE is a heritable effect of one individual on the trait value of another individual (reviewed in Wolf *et al.*, 1998; Bijma, 2011a). A well-known example is the maternal genetic effect of a mother on preweaning growth rate of her offspring (Willham, 1963; Falconer, 1965; Kirkpatrick and Lande, 1989).

IGEs may have significant effects on the rate and direction of response to selection, and can substantially increase or decrease heritable variation in a trait (Griffing, 1967; Moore *et al.*, 1997; Bijma and Wade, 2008; McGlothlin and Brodie III, 2009; Wilson *et al.*, 2011; Bijma, 2011b). Thus, knowledge of IGEs is essential for understanding response to selection in socially affected traits. The magnitude of IGEs can be estimated using linear mixed models that include a direct genetic effect for the individual producing the record, and an IGE for each of its social partners (Arango *et al.*, 2005; Muir, 2005; Bijma *et al.*, 2007). This approach has been used both in agricultural populations of animals and plants (see, for example, Muir, 2005; Cosa e Silva *et al.*, 2013), and in natural populations (see, for example, Wilson *et al.*, 2011). Bijma (2010a) showed that estimation of genetic parameters for IGEs in group-structured populations can be optimized by placing two families in each group. Such schemes are an attractive breeding design because they also yield a relatively high response to selection (Odegard and Olesen, 2011).

In the linear mixed model commonly used to estimate IGEs (Muir, 2005), it is assumed that an individual expresses the same IGE on each of its social partners, irrespective of whether a partner is its family member or an unrelated individual. Kin selection theory, however, predicts that individuals behave more cooperatively towards their relatives, because this increases their inclusive fitness (Hamilton, 1964). Hence, IGEs expressed on kin may differ systematically from those expressed on strangers; they may differ not only in average level, but also show incomplete correlation. Empirical evidence indeed suggests that kin recognition and preferential behaviour towards kin are widespread in both animals and plants (see, for example, Holmes and Sherman, 1982; Hepper, 1986; Olsen, 1989; Dudley and File, 2007; Biedrzycki and Bais, 2010), and at least four mechanisms for kin recognition have been described (Tang-Martinez, 2001; Mateo and Holmes, 2004; Mateo, 2004; Coffin *et al.*, 2011).

When individuals express a different IGE on kin versus strangers, estimated breeding values for direct and indirect effects from the common linear mixed model are incorrect, and selection based on those estimates will yield suboptimal response. Moreover, when IGEs are estimated from groups composed of strangers (see, for example, Ellen *et al.*, 2008), the resulting estimates may not accurately reflect

[1]Department of Molecular Biology and Genetics, Aarhus University, Tjele, Denmark; [2]Animal Breeding and Genomics Centre, Wageningen University, Wageningen, The Netherlands and [3]NordGen, Nordic Genetic Resource Center, Ås, Norway
Correspondence: SW Alemu, Department of Molecular Biology and Genetics, Aarhus University, Tjele DK-8830, Denmark.
E-mail: setegnw.alemu@agrsci.dk
Received 18 April 2013; revised 22 August 2013; accepted 19 August 2013; published online 30 October 2013

the IGEs that occur in the relevant natural or domestic populations, which may consist of kin groups. In natural populations, limited dispersal often leads to interactions among relatives (Hamilton, 1964), whereas in livestock populations such as domestic pigs, groups often contain a number of family members (Chen *et al.*, 2008). Thus, a potential difference between IGEs on kin vs strangers is relevant for both livestock and natural populations. The current statistical methods for estimating IGEs, however, ignore the dependency of IGEs on relatedness.

Here we propose a model for traits affected by IGEs that differ between kin and strangers, investigate whether genetic parameters of that model are statistically identifiable and develop statistical models to estimate those parameters. First we show that the full set of genetic parameters is not identifiable when IGEs differ between kin and strangers. Subsequently, we developed a reduced model, and showed that the reduced model can estimate meaningful linear combinations of the genetic parameters. In the Discussion, we consider population structures that may allow estimating the full set of genetic parameters.

QUANTITATIVE GENETIC MODEL

Trait model

This section introduces the trait model when IGEs differ between kin and strangers. We consider here a population stratified into groups of n members each, where interactions occur within groups. We consider the scheme that is optimal for the estimation of IGEs in the absence of kin recognition (Bijma, 2010a). In this scheme, each group is composed of members of two families, each family contributing $n/2$ individuals. Generalization of results to other group structures is addressed in the Discussion.

In traditional quantitative genetics, the phenotypic value of individual i is the sum of a heritable component, A_i, known as breeding value, and a nonheritable residual, E_i (Falconer and Mackay, 1996; see Table 1 for a notation key),

$$P_i = A_i + E_i \tag{1}$$

with IGEs that do not depend on relatedness, the phenotype of an individual stems from two components: a direct effect originating from the individual itself, and the sum of indirect effects originating from each of its $n-1$ group mates (Griffing, 1967),

$$P_i = A_{D,i} + E_{D,i} + \sum_{j=1}^{n-1} A_{S,j} + \sum_{j=1}^{n-1} E_{S,j} \tag{2}$$

where i denotes the focal individual, j a group mate, $A_{D,i}$ the direct genetic effect (DGE) of i, $E_{D,i}$ the corresponding non-heritable direct effect, $A_{S,j}$ the IGE of group mate j and $E_{S,j}$ the corresponding non-heritable indirect effect (subscript S, suggesting 'social', is used to denote indirect effects instead of a subscript I, to avoid confusion of i with I; Equation 2 is known as a variance component model of IGEs, as opposed to a trait-based model. See McGlothlin and Brodie III, 2009 for a comparison of models). Equation 2 contains two kinds of genetic effects, direct effects, A_D, and indirect effects, A_S. Hence, fitting Equation 2 involves the estimation of three genetic variance components; $\sigma^2_{A_D}$, $\sigma_{A_{DS}}$ and $\sigma^2_{A_S}$ (throughout, σ^2 denotes a variance and σ a covariance).

With different interactions among kin versus strangers, two types of IGEs may be distinguished: IGEs on kin versus IGEs on strangers. In our population structure, where $n/2$ members of each family make up a group, the trait model becomes:

$$P_i = A_{D,i} + E_{D,i} + \sum_{j=1}^{\frac{n}{2}-1} A_{S_f,j} + \sum_{j=1}^{\frac{n}{2}-1} E_{S_f,j} + \sum_{k=1}^{n/2} A_{S_u,k} + \sum_{k=1}^{n/2} E_{S_u,k} \tag{3}$$

where j denotes a family member of i, k a member of the other family in the group, $\frac{n}{2}-1$ the number of group mates of i from its own family, $n/2$ the number of group mates of i from the other family, subscript 'S_f' denotes IGEs on family members and subscript 'S_u' denotes IGEs on members of the other, unrelated, family (u indicating 'unrelated'). Equation 3 contains three genetic effects: direct effects, A_D, IGEs on family members, A_{S_f}, and IGEs on strangers, A_{Su}. Hence, fitting Equation 3 involves the estimation of six genetic variance components: three variances $\sigma^2_{A_D}$, $\sigma^2_{A_{S_f}}$ and $\sigma^2_{A_{S_u}}$ and three covariances $\sigma_{A_{D,S_f}}$, $\sigma_{A_{D,S_u}}$ and $\sigma_{A_{S_f,S_u}}$. The genetic correlation between an individual's IGE on kin and its IGE on strangers, $r_{A_{S_f}A_{S_u}} = \sigma_{A_{S_f},A_{S_u}}/(\sigma_{A_{S_f}}\sigma_{A_{S_u}})$, reflects the difference between IGEs on kin and strangers. Equation 3 does not explicitly include a potential difference in the mean value of the IGE on kin vs strangers, because this has little consequences for the estimation of genetic parameters. Nevertheless, such a difference is relevant in statistical data analysis, and can be accommodated easily in the fixed effects part of the model (see Discussion).

Table 1 Notation key[a]

Symbol	*Meaning*
i,j,k,x,l,m	Subscript to denote an individual.
P_i	Observed trait value of an individual.
$P_{D,i}, P_{S_f,i}, P_{S_u,i}$	Direct effect of i, indirect effect of i to kin, indirect effect of i to stranger.
$\sigma^2_P, \sigma^2_{P_D}, \sigma^2_{P_{S_f}}, \sigma^2_{P_{S_u}}$	Phenotype variance among individuals, unobserved phenotype variance on self, kin and strangers.
$\sigma^2_{A_D}, \sigma^2_{A_{S_f}}, \sigma^2_{A_{S_u}}$	Variance of DGEs among individuals, variance of IGEs on kin among individuals, variance of IGEs on strangers among individuals.
$\sigma^2_{A_{T_f}}, \sigma^2_{A_T}$	Variance of family breeding value among individuals, variance of total breeding value among individuals.
$\sigma^2_{E_D}, \sigma^2_{E_{S_f}}, \sigma^2_{E_{S_u}}$	Variance of direct environment among individuals, variance of indirect environment on kin among individual, variance of indirect environment on strangers among individual.
$\sigma_{A_{D,S_f}}, r_{A_{D,S_f}}, \sigma_{A_{D,S_u}}, r_{A_{D,S_u}}$	Covariance between DGEs and IGEs to kin, correlation between DGEs and IGEs to kin, covariance between DGEs and IGEs to strangers and correlation between DGEs and IGEs to strangers.
$\sigma_{A_{S_f,S_u}} r_{A_{S_f,S_f}}$	Covariance between IGEs to kin and IGEs to strangers, correlation between IGEs to kin and IGEs to strangers.
$\sigma_{E_{D,S_f}}, r_{E_{D,S_f}}$	Covariance and correlation between nongenetic direct and nongenetic indirect on kin.
$\sigma_{E_{D,S_u}}, r_{E_{D,S_u}}$	Covariance and correlation between nongenetic direct and nongenetic indirect on strangers.
$\sigma_{E_{S_f,S_u}}, r_{E_{S_f,S_u}}$	Covariance and correlation between nongenetic in direct on kin and nongenetic indirect on strangers.
r, ρ, n, σ^2_g	Relatedness among individual in a group, residual correlation of family member in a group, group size, variance of nonfamily member in a group.

Abbreviations: DGE, direct genetic effect; IGE, indirect genetic effect.
[a]Throughout the text and in the tables, hats (^) denote estimates, whereas symbols without hats refer to true values.

Total breeding value and heritable variation

This section presents the heritable variation available for response to selection in a trait when IGEs differ between kin and strangers.

Irrespective of the trait model, response to selection in any trait can be expressed as

$$R = \Delta\bar{A}_T = \imath\rho\sigma_{A_T}, \tag{4}$$

where R is the genetic change in mean trait level from one generation to the next because of selection, $\Delta\bar{A}_T$ the change in mean total breeding value (A_T) of the population, $\imath$ the intensity of selection, ρ the accuracy of selection and σ_{A_T} the s.d. in total breeding value (Bijma, 2011a); an equivalent expression in terms of a selection gradient can also be found there, and may be more appropriate for natural populations). In the context of Equation 4, the accuracy of selection is the correlation between an individual's value for the selection criterion and its total breeding value (this definition applies to any selection criterion; see Falconer and Mackay, 1996 for further explanation of

the 'accuracy of selection'). The total breeding value represents the average impact of an individual's genes on the mean trait value of the population, and is a generalization of the traditional breeding value to account for IGEs and to allow modelling of so-called emergent traits (Bijma, 2011b). Thus, analogous to the classical breeding value, the total breeding value represents an individual's value for response to selection. As illustrated in Equation 4, in which $\imath$ and ρ are standardized parameters, the s.d. in total breeding value represents the intrinsic potential of a population to respond to selection.

For any trait model, the total breeding values follow from the genetic mean of the population (Bijma, 2011b). From Equation 3, the genetic mean of the trait value for our population structure equals

$$\bar{P}_A = \overline{A_D} + (\tfrac{1}{2}n-1)\overline{A_{S_f}} + \tfrac{1}{2}n\,\overline{A_{S_u}}.$$

Therefore, following Bijma (2011b), an individual's total breeding value is the sum of its DGE, $1/2n-1$ times its IGE on family members, and $1/2n$ times its IGE on strangers,

$$A_{T,i} = A_{D,i} + (\tfrac{1}{2}n-1)A_{S_f,i} + \tfrac{1}{2}n\,A_{S_u,i}. \tag{5}$$

Taking the variance of the total breeding value yields an expression for the heritable variation available for response to selection,

$$\begin{aligned}\sigma^2_{A_T} \;=\; & \sigma^2_{A_D} \\ & + \;(n-2)\,\sigma_{A_D,S_f} \;+\; (\tfrac{1}{2}n-1)^2\sigma^2_{A_{S_f}} \\ & + \;n\sigma_{A_D,S_u} \;+\; n\left(\tfrac{1}{2}n-1\right)\sigma_{A_{S_f},S_u} \;+\; \tfrac{1}{4}n^2\sigma^2_{A_{S_u}}.\end{aligned} \tag{6}$$

Note that $\sigma^2_{A_T}$ does not reflect the additive genetic component of phenotypic variance, but the heritable variation that determines the potential of a population to respond to selection (see Equation 4 and Bijma, 2011b).

An individual's total breeding value can be partitioned into a family component, A_{T_f}, which summarizes all its heritable effects on family members (including the direct effect on itself) and is considered the family breeding value here, and a non-family component, A_{T_u}, the non-family breeding value. This partitioning will be used below, where the family components of the total breeding value will be grouped for reasons of statistical identifiability. With each family contributing, $1/2n$ group members

$$A_{T,i} = A_{T_f,i} + A_{T_u,i}, \tag{7a}$$

$$A_{T_f,i} = A_{D,i} + (\tfrac{1}{2}n-1)A_{S_f,i}, \tag{7b}$$

$$A_{T_u,i} = \tfrac{1}{2}nA_{S_u,i}. \tag{7c}$$

Taking the variances of Equation 7 yields

$$\sigma^2_{A_T} = \sigma^2_{A_{T_f}} + 2\sigma_{A_{T_f}T_u} + \sigma^2_{A_{T_u}} \tag{8a}$$

$$\sigma^2_{A_{T_f}} \;=\; \sigma^2_{A_D} + \;(n-2)\sigma_{A_D,S_f} + (\tfrac{1}{2}n-1)^2\sigma^2_{A_{S_f}}, \tag{8b}$$

$$\sigma^2_{A_{T_u}} \;=\; \tfrac{1}{4}n^2\sigma^2_{A_{S_u}}, \tag{8c}$$

$$\sigma_{A_{T_f}T_u} \;=\; \tfrac{1}{2}n\sigma_{A_D,S_u} + \tfrac{1}{2}n\left(\tfrac{1}{2}n-1\right)\sigma_{A_{S_f},S_u}. \tag{8d}$$

VARIANCE COMPONENT ESTIMATION

Genetic parameters can be estimated using a linear mixed model including correlated random genetic effects, the so-called animal model (Henderson, 1953, 1975; Lynch and Walsh, 1998). The classical animal model includes DGEs only, but can be extended with IGEs (Muir, 2005).

Full model

The full model includes DGEs, IGEs on family members and IGEs on strangers,

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_D\mathbf{a}_D + \mathbf{Z}_{S_f}\mathbf{a}_{S_f} + \mathbf{Z}_{S_u}\mathbf{a}_{S_u} + \mathbf{Wg} + \mathbf{e}, \tag{9}$$

where $\mathbf{b}$ is a vector of fixed effects with incidence matrix $\mathbf{X}$, $\mathbf{a}_D$ is a vector of DGEs with incidence matrix $\mathbf{Z}_D$ linking observation on individuals to their own DGE, $\mathbf{a}_{S_f}$ is vector of IGEs on family members with incidence matrix $\mathbf{Z}_{S_f}$ linking observations on individuals to the IGEs of their group mates belonging to the same family, $\mathbf{a}_{S_u}$ is vector of IGEs on strangers with incidence matrix $\mathbf{Z}_{S_u}$ linking observations on individuals to the IGEs of their group mates belonging to the other family, $\mathbf{g}$ is a vector of random group effects, with $\mathbf{g} \sim N(\mathbf{0}, \mathbf{I}_g\sigma^2_g)$ and incidence matrix $\mathbf{W}$ linking records to groups, and $\mathbf{e}$ is a vector of residuals with $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}_e\sigma^2_e)$, where $\mathbf{I}$ is an identity matrix. The covariance structure of the genetic terms is

$$\begin{bmatrix}\mathbf{a}_D \\ \mathbf{a}_{S_f} \\ \mathbf{a}_{S_u}\end{bmatrix} \sim MVN(\mathbf{0},\ \mathbf{C}\otimes\mathbf{A})$$

$$\text{where } \mathbf{C} = \begin{bmatrix}\sigma^2_{A_D} & \sigma_{A_D,S_f} & \sigma_{A_D,S_u} \\ \sigma_{A_D,S_f} & \sigma^2_{A_{S_f}} & \sigma_{A_{S_f},S_u} \\ \sigma_{A_D,S_u} & \sigma_{A_{S_f},S_u} & \sigma^2_{A_{S_u}}\end{bmatrix},$$

where $\otimes$ indicates the Kronecker product of matrices, and $\mathbf{A}$ is a matrix of additive genetic relationships between individuals, the so-called numerator relationship matrix (Henderson, 1985).

When fitting the full model, the results showed that there are multiple parameter combinations that give the same likelihood. Hence, using this model, the genetic parameters are statistically nonidentifiable. In particular, results showed that the variance of IGEs on strangers, $\sigma^2_{A_{S_u}}$, is identifiable, but that the variance components referring to interactions between family members, $\sigma^2_{A_D}$, $\sigma_{A_D,A_{S_f}}$ and $\sigma^2_{A_{S_f}}$, are fully confounded. We investigated why this occurs and found that there are only five informative genetic covariances in the data, but six genetic parameters to estimate (see Appendix A). Thus, when IGEs differ between kin and strangers, it is not possible to estimate all six genetic parameters from group-structured data. This is not a problem of the estimation method, but a property of the data structure and occurs when group composition with respect to family is the same for all groups (see Discussion and Appendix A). Thus, the data structure that is optimal for estimating the variance of IGEs that do not depend on kin renders the estimation of kin-dependent IGEs impossible. In the Discussion, we consider alternative schemes that may allow estimating all parameters of the full model. Note that the variance structure given above for the residual of Equation 9 ignores the distinction between indirect effects on kin vs strangers. However, as the full model is nonidentifiable, we did not further investigate this issue.

Reduced model

Because the full model was not identifiable, we investigated a reduced model, aiming to estimate part of the genetic parameters or meaningful linear combinations. As the full model indicated that the effects because of the focal family were fully confounded, we fitted only a single term for the family of the focal individual. Therefore, the reduced model was

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_D\mathbf{a}_F + \mathbf{Z}_{S_u}\mathbf{a}_{S_u} + \mathbf{Wg} + \mathbf{e} \tag{10}$$

where $\mathbf{a}_F$ is a vector of genetic effects due to the family of the focal individual and $\mathbf{Z}_D$ is the incidence matrix for direct genetic effects as in the full model (Equation 9). Hence, with respect to the genetic terms, the only difference between the full and reduced model is that the $\mathbf{Z}_{S_f}\mathbf{a}_{S_f}$ term is omitted in Equation 10; the other genetic terms are the same. However, as omitting the $\mathbf{Z}_{S_f}\mathbf{a}_{S_f}$ will change both the estimates and the interpretation of the 'direct' genetic effects, we write $\mathbf{Z}_D\mathbf{a}_F$ in Equation 10, where subscript F suggests 'family', rather than

$\mathbf{Z}_D\mathbf{a}_D$. The covariance structure of the genetic terms in Equation 10 is

$$\begin{bmatrix}\mathbf{a}_F\\ \mathbf{a}_{S_u}\end{bmatrix} \sim MVN(\mathbf{0}, \mathbf{C}_r \otimes \mathbf{A})$$

where $\mathbf{C}_r = \begin{bmatrix}\sigma^2_{A_F} & \sigma_{A_F,S_u}\\ \sigma_{A_F,S_u} & \sigma^2_{A_{S_u}}\end{bmatrix}$.

The **Wg** term is as in Equation 9. The covariance structure for the residual term is

$$\text{var}(\mathbf{e}) = \mathbf{R}\sigma^2_e, \tag{11}$$

where $\mathbf{R}_{ii}=1$, $\mathbf{R}_{ij}=\rho$ when i and j are group mates from the same family, and $\mathbf{R}_{ij}=0$ otherwise. Hence, this structure allows for a covariance between residuals of group mates belonging to the same family. Thus, when individuals are ordered by group and by family within group, then **R** is block-diagonal, with blocks of size $n/2$, diagonal elements equal to 1, off-diagonals of blocks equal to ρ, all other off-diagonals equal to zero, and two blocks per group, one for each family. Appendix B shows that this residual variance structure together with the random group effect corresponds to the nongenetic variance structure generated by the assumed true model (Equation 3). Thus, the $\mathbf{Wg}+\mathbf{e}$ in Equation 10 accounts for the variance structure generated by the term $E_{D,i}+\sum E_{S_{f,j}}+\sum E_{S_{u,k}}$ in Equation 3.

Investigation of Equation 10 showed that there are five informative genetic covariances in the data to estimate three genetic parameters, indicating that the model in Equation 10 is identifiable. To investigate the interpretation of the genetic estimates from the reduced model, we derived their expectation, assuming that the data are generated by the model given in Equation 3 (Appendix A). With

$$A_{F,i} = A_{D,i} + (\tfrac{1}{2}n-1)A_{S_f,i}, \tag{12a}$$

and

$$A_{S_u,i} = A_{S_u,i}, \tag{12b}$$

it follows that

$$\text{E}(\hat{\sigma}^2_{A_F}) = \sigma^2_{A_D} + (n-2)\,\sigma_{A_D S_f} + \left(\tfrac{1}{2}n-1\right)^2\sigma^2_{A_{S_f}}, \tag{12c}$$

$$\text{E}(\hat{\sigma}_{A_{F,S_u}}) = \sigma_{A_D S_u} + \left(\tfrac{1}{2}n-1\right)\,\sigma_{A_{S_f,S_u}}, \tag{12d}$$

$$\text{E}(\hat{\sigma}^2_{A_{S_u}}) = \sigma^2_{A_{S_u}}. \tag{12e}$$

Equations 12c–e sum up all the variance components considered in the true model. Equation 12e shows that the reduced model yields an estimate of the variance of IGEs on strangers. Moreover, combining Equations 12c–e with the decomposition of the total breeding value into a family and a non-family component given in Equations 7 and 8 above shows that the reduced model yields estimates of the family and non-family genetic parameters,

$$\hat{\sigma}^2_{A_{T_f}} = \hat{\sigma}^2_{A_F}, \tag{13a}$$

$$\hat{\sigma}_{A_{T_f}A_{T_u}} = \tfrac{1}{2}n\hat{\sigma}_{A_F,S_u} \tag{13b}$$

$$\hat{\sigma}^2_{A_{T_u}} = \tfrac{1}{4}n^2\hat{\sigma}^2_{A_{S_u}}. \tag{13c}$$

Thus, the variance of the total breeding value can be obtained from the reduced model as

$$\hat{\sigma}^2_{A_T} = \hat{\sigma}^2_{A_F} + n\hat{\sigma}_{A_F,S_u} + \tfrac{1}{4}n^2\hat{\sigma}^2_{A_{S_u}}. \tag{13d}$$

Thus, the reduced model allows the estimation of the total heritable variation, even though not all the underlying parameters are identifiable.

Equation 13b refers to the covariance between the family breeding value and the non-family breeding value. This is a meaningful linear combination, as it expresses the covariance between genetic effects on kin (including self) and those on strangers. If this covariance is positive, members from different families are cooperative, whereas a negative value indicates competition between families.

Appendix B shows that the expectations of the nongenetic variance components in Equation 10 are given by

$$E(\hat{\sigma}^2_g) = 2\sigma_{E_{D,S_u}} + (n-2)\sigma_{E_{S_f,S_u}} \tag{14a}$$

$$\text{E}(\hat{\sigma}^2_e) = \sigma^2_{E_D} + \left(\tfrac{1}{2}n-1\right)\sigma^2_{E_{S_f}} + \tfrac{1}{2}n\,\sigma^2_{E_{S_u}} - \sigma^2_g, \tag{14b}$$

$$\text{E}(\hat{\rho}) = \frac{2\sigma_{E_D E_{S_f}} + \left(\tfrac{1}{2}n-2\right)\sigma^2_{E_{S_f}} + \tfrac{1}{2}n\,\sigma^2_{E_{S_u}} - \sigma^2_g}{\sigma^2_e} \tag{14c}$$

Equations 14a–c show that the underlying nongenetic parameters are not uniquely identifiable, because there are only three estimable parameters (σ^2_g, σ^2_e and ρ) that are a function of six unknowns. This was expected, as it is also the case for models not distinguishing between IGEs on kin vs strangers (Bijma *et al.*, 2007).

Consequences of ignoring kin-dependent IGEs

This section investigates the bias in the estimated genetic parameters when IGEs differ between kin and strangers, although this is ignored in the statistical analysis. Thus, it is assumed that the true model generating the trait values is given by Equation 3 above, which distinguishes between IGEs on kin and strangers, whereas the statistical model used to estimate genetic parameters is the traditional direct–indirect mixed linear model (Muir, 2005),

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_D\mathbf{a}_D + \mathbf{Z}_S\mathbf{a}_S + \mathbf{Wg} + \mathbf{e}, \tag{15}$$

where $\mathbf{a}_S$ is vector of IGEs on group mates, not distinguishing between kin and strangers, and $\mathbf{Z}_S$ an incidence matrix linking observations on individuals to the IGEs of all their group mates. The $\mathbf{Z}_D\mathbf{a}_D$ and **Wg** are as in the full model (Equation 9), whereas the residual variance structure is as in the reduced model (Equation 11; see discussion below). Note that Equation 15 differs from the reduced model (Equation 10) because the term $\mathbf{Z}_S\mathbf{a}_S$ includes the IGEs of all n–1 group mates, and not only those belonging to the other family making up the group.

To investigate the bias resulting from fitting a conventional IGE model (Equation 15) to data in which IGEs differ between kin and strangers, we derived the expectations of the estimated breeding values and variance components produced by Equation 15 when data are generated by Equation 3. Those expectations follow from the informative covariances in the data that Equation 15 utilizes to estimate the genetic parameters, and can be obtained using the method of Bijma, 2010a; see Appendix C). Results showed that

$$\begin{aligned}E(\hat{\sigma}^2_{A_D})_{Eqn.15} = \sigma^2_{A_D} &+ \left(\tfrac{1}{2}n-1\right)^2\left(\sigma^2_{A_{S_f}} - 2\sigma_{A_{S_f}A_{S_u}} + \sigma^2_{A_{S_u}}\right)\\ &+ (n-2)(\sigma_{A_D A_{S_f}} - \sigma_{A_D A_{S_u}})\end{aligned} \tag{16a}$$

$$E(\hat{\sigma}_{A_{DS}})_{Eqn.15} = \sigma_{A_D A_{S_u}} + \left(\tfrac{1}{2}n-1\right)(\sigma_{A_{S_f}A_{S_u}} - \sigma^2_{A_{S_u}}) \tag{16b}$$

$$E(\hat{\sigma}^2_{A_S})_{Eqn.15} = \sigma^2_{A_{S_u}} \tag{16c}$$

$$E(\hat{\sigma}^2_{A_T})_{Eqn.15} = \sigma^2_{A_T}. \tag{16d}$$

These results show that the direct genetic variance and the direct–indirect genetic covariance estimated with the conventional linear model for IGEs (Equation 15) are biased when IGEs depend on relatedness. In other words, the estimate of $\sigma^2_{A_D}$ is biased because the right-hand side of Equation 16a differs from $\sigma^2_{A_D}$. Similarly,

difference of the right-hand side of Equation 16b from $\sigma_{A_{DS}}$ indicates bias of $\sigma_{A_{DS}}$. Moreover, the estimated indirect genetic variance from Equation 15 refers to the magnitude of IGEs expressed on strangers (Equation 16c). Surprisingly, despite the incorrect model assumptions, the traditional direct–indirect model yields an unbiased estimate of the total heritable variance (Equation 16d). Beware that results in Equations 16a–d are correct only if the residual covariance structure accounts for differences between indirect effects on kin and strangers, as given by Equation 11. Therefore, when the aim is to estimate total breeding values (TBVs) using the traditional direct–indirect mixed model of Muir (2005), this model should be implemented including a random group effect and the residual variance structure given in Equation 11 above.

We did not attempt to derive the expectations of estimated genetic parameters from the traditional direct–indirect model (Equation 15) when the residual variance structure is incorrect (that is, different from that given in Equation 11). The reason is that those expectations will depend not only on the assumed true genetic model (Equation 3), but also on the data structure. For example, in data consisting of many groups, the covariance between relatives in different groups will dominate the estimates, and incorrect covariances within groups may have little effect. In that case, estimates may be close to values given in Equation 16. On the other hand, when groups are fewer, information from the within-group (co)variances will become more important and results may deviate more from Equation 16.

Simulation

Methods. We used Monte Carlo simulation to validate the theoretical relationships between the true model, the reduced model and the traditional model presented above (Equations 12, 14 and 16). Data were generated under the model in Equation 3, and analysed using either the reduced model in Equation 10 or the traditional model in Equation 15, using the ASReml software (Gilmour *et al.*, 2006). A population of two discrete generations was simulated using R (R Development Core Team, 2011). No fixed effects were simulated. The base generation consisted of 100 sires and 1000 dams, which were unrelated. To produce the second generation, sires and dams of the first generation were mated at random, each sire being mated to 10 dams, and each dam producing 10 full-sib offspring. Individuals of the second generation were kept in 2500 groups of 4 individuals each, and each group consisted of two full-sib families, each family contributing two individuals. Table 2 shows the range of genetic parameters simulated. For each set of genetic parameters, estimates were averaged over 100 replicates. Details of the simulation are given in Appendix D.

Results. Table 3 shows a comparison between simulated and estimated values for $\sigma^2_{A_{T_f}}, \sigma_{A_{T_f}A_{S_u}}, \sigma^2_{A_{S_u}}$ and $\sigma^2_{A_T}$ from the reduced model for different magnitudes of IGEs. We used $\sigma^2_{P_{S_f}} = \sigma^2_{P_{S_u}}$ of either 50 or 25% of $\sigma^2_{P_D}$ to represent high or low indirect effects, and $\sigma^2_{A_{S_f}}, \sigma^2_{A_{S_u}}$ of either 10, 12.5, 20 and 25% of $\sigma^2_{P_{S_u}}, \sigma^2_{P_{S_f}}$, to represent high or low heritability of IGE, and a range of genetic correlations between direct effects, indirect effects on kin and indirect effects on strangers (Table 2). Results show close agreement between simulated and estimated values as proven by the relative error that is $\leqslant 5\%$ in all cases (those small errors originate from stochasticity among replicates, and do not indicate systematic bias). These results confirm the theoretical relationships between the full and reduced models presented in Equations 12 and 13. Thus, the reduced model yields unbiased genetic parameters of the family and non-family breeding values, and of the total breeding value. We also compared the estimated nongenetic components with their expectations given in Equation 14, showing close agreement (results not shown).

Table 4 shows a comparison between the theoretically expected values of the estimated variance components from the traditional model (Equation 16) and the empirical values estimated from the simulated data using the traditional model (Equation 15). Results confirm the theoretical expectation that the traditional direct–indirect model yields biased estimates of the direct genetic variance and the direct–indirect genetic covariance, but unbiased estimates of the genetic variance of IGEs on strangers and of the total genetic variance.

Table 2 Parameter values used for validation of reduced and traditional models

Scheme	*Deviation from basic scheme*[a]
Alt.1	$\sigma^2_{A_{S_f}} = 0.125, \sigma^2_{A_{S_u}} = 0.100$
Alt.2	$\sigma^2_{P_{S_f}} = \sigma^2_{P_{S_u}} = 0.25, \sigma^2_{A_{S_f}} = 0.050, \sigma^2_{A_{S_u}} = 0.063$
Alt.3	$\sigma^2_{P_{S_f}} = \sigma^2_{P_{S_u}} = 0.25, \sigma^2_{A_{S_f}} = 0.063, \sigma^2_{A_{S_u}} = 0.050$
Alt.4	$r_{A_D, A_{S_f}} = r_{E_D, E_{S_f}} = 0.5$
Alt.5	$r_{A_D, A_{S_f}} = r_{E_D, E_{S_f}} = 0.5, \sigma^2_{A_{S_f}} = 0.125, \sigma^2_{A_{S_u}} = 0.1$
Alt.6	$r_{A_D, A_{S_f}} = r_{E_D, E_{S_f}} = 0.5, \sigma^2_{A_{S_f}} = 0.050, \sigma^2_{P_{S_f}} = 0.25$
Alt.7	$r_{A_D, A_{S_f}} = r_{E_D, E_{S_f}} = 0.5, \sigma^2_{A_{S_f}} = 0.125, \sigma^2_{A_{S_u}} = 0.05, \sigma^2_{P_{S_f}} = \sigma^2_{P_{S_u}} = 0.25$
Alt.8	$r_{A_D, A_{S_f}} = r_{E_D, E_{S_f}} = -0.5$
Alt.9	$r_{A_D, A_{S_u}} = r_{A_{S_f}, A_{S_u}} = r_{E_D, E_{S_u}} = r_{E_{S_f}, E_{S_u}} = 0.5$
Alt.10	$r_{A_D, A_{S_u}} = r_{A_{S_f}, A_{S_u}} = r_{E_D, E_{S_u}} = r_{E_{S_f}, E_{S_u}} = 0.5, \sigma^2_{A_{S_f}} = 0.100$
Alt.11	$r_{A_D, A_{S_u}} = r_{A_{S_f}, A_{S_u}} = r_{E_D, E_{S_u}} = r_{E_{S_f}, E_{S_u}} = 0.5,$ $\sigma^2_{A_{S_f}} = 0.050, \sigma^2_{A_{S_u}} = 0.063, \sigma^2_{P_{S_f}} = \sigma^2_{P_{S_u}} = 0.25$
Alt.12	$r_{A_D, A_{S_u}} = r_{A_{S_f}, A_{S_u}} = r_{E_D, E_{S_u}} = r_{E_{S_f}, E_{S_u}} = 0.5, \sigma^2_{A_{S_f}} = 0.050,$ $\sigma^2_{P_{S_f}} = \sigma^2_{P_{S_u}} = 0.25$
Alt.13	$r_{A_D, A_{S_f}} = r_{A_D, A_{S_u}} = r_{A_{S_f}, A_{S_u}} = r_{E_D, E_{S_f}} = r_{E_D, E_{S_u}} = r_{E_{S_f}, E_{S_u}} = 0.5$
Alt.14	$r_{A_D, A_{S_f}} = r_{A_D, A_{S_u}} = r_{A_{S_f}, A_{S_u}} = r_{E_D, E_{S_f}} = r_{E_D, E_{S_u}} = r_{E_{S_f}, E_{S_u}} = -0.1$

[a]The basic scheme has $\sigma^2_{P_D} = 1$, $\sigma^2_{A_D} = 0.5$, $\sigma^2_{P_{S_f}} = \sigma^2_{P_{S_u}} = 0.5$, $\sigma^2_{A_{S_f}} = 0.125$ and $\sigma^2_{A_{S_u}} = 0.100$, and all correlations are zero. Alternative schemes only show parameters that deviate from the basic scheme.

Table 3 Errors in estimates for the reduced model

Scheme	*Error%*			
	$\hat{\sigma}^2_{A_{T_f}}$	$\hat{\sigma}_{A_{T_f}A_{S_u}}$	$\hat{\sigma}^2_{A_{S_u}}$	$\hat{\sigma}^2_{A_T}$
Basic	0	0	−1	0
Alt.1	1	0	0	1
Alt.2	−2	0	2	−1
Alt.3	−1	0	−2	−1
Alt.4	2	0	−1	−1
Alt.5	0	0	0	−1
Alt.6	−2	0	0	−2
Alt.7	−1	0	2	0
Alt.8	−1	0	1	0
Alt.9	1	1	2	1
Alt.10	−1	−2	−2	−2
Alt.11	0	1	2	1
Alt.12	1	1	0	1
Alt.13	−2	−4	−3	−3
Alt.14	1	5	−1	1

See Table 2 for a description of schemes. Error % = 100% × (estimated − simulated)/simulated. When the prediction equals the true value E[error%] ≈ 0. The expected absolute error equals E [|error%|] ∼2.5%, and E|error %| > 5% implies significant bias ($P<0.05$; two sided).

Table 4 Comparison of the expected (Equation 16) and empirical estimates for the traditional model

Scheme	*Error%*			
	$\hat{\sigma}^2_{A_D}$	$\hat{\sigma}_{A_D S}$	$\hat{\sigma}^2_{A_S}$	$\hat{\sigma}^2_{A_{TBV}}$
Basic	−1	−2	−2	−1
Alt.1	1	−1	0	1
Alt.2	0	1	0	−1
Alt.3	0	4	−2	0
Alt.4	0	−4	0	2
Alt.5	−2	0	−1	−2
Alt.6	−1	2	1	−1
Alt.7	0	0	0	0
Alt.8	1	2	−1	−3
Alt.9	0	2	2	1
Alt.10	−3	−4	0	−2
Alt.11	−1	−2	0	−1
Alt.12	0	2	0	1
Alt.13	−2	−5	−3	−3
Alt.14	1	3	0	−2

See Table 2 for description of schemes. Error % = 100% × (estimated − expected)/expected, where the expected values of estimates are taken from Equation 16. When prediction equals true value E[error%] ≈ 0. The expected absolute error equals E[|error%|] ∼ 2.5%, and Elerror %| > 5% implies significant bias (P < 0.05;two sided). Beware that expected values do not correspond to the simulated values (Equation 16).

DISCUSSION

We have proposed a quantitative genetic model and investigated methodology to estimate the genetic parameters of traits affected by IGEs when those IGEs differ systematically between kin and strangers. Results show that the full set of genetic parameters for the full model is not statistically identifiable. We also presented a reduced model that yields unbiased estimates of meaningful linear combinations of genetic parameters: the variance of the family breeding value, the covariance between family breeding value and IGEs on strangers and the variance of IGEs on strangers. The reduced model also provides estimates of the variance in total breeding value, and predictions of the total breeding values of individuals.

An interesting question is whether experimental designs exist that allow estimating all six genetic parameters of the full model (Equation 3). Our results show that this is not possible when pairs of individuals can be categorized into either kin or unrelated, each category shows a different IGE and group composition is the same for all groups. As long as group composition with respect to family is the same for all groups, this situation results in full confounding of the direct effect and the IGE on kin, irrespective of the composition of the groups (that is, 50/50, 25/75 and so on; Appendix A).

When differences in IGE originate from factors that usually go together with relatedness such as familiarity, rather than from relatedness *per se*, experimental designs that disconnect relatedness from those factors may allow estimation of the full set of genetic parameters. For example, when individuals recognize each other because of prior association (see Introduction), relatives who grow up together will recognize each other and adjust their behaviour, whereas relatives who grow up separately will interact similarly to unrelated individuals. This may, for example, occur in mammals such as grey mouse lemur (Kessler *et al.*, 2012) or rats (Hepper, 1983, 1986), where full siblings often grow up in the same litter, whereas paternal half siblings grow up in different environments. Our preliminary investigations show that all six genetic parameters are statistically identifiable in this situation when groups consist of a mix of full sibs, half sibs and unrelated individuals. A statistically more powerful approach may come from cross-fostering designs, where full siblings that grow up in different litters may interact as if they were unrelated. When cross-fostering is impossible and a mix of full and half siblings is unavailable, a solution may come from utilizing the variation in relatedness among pairs of full siblings, estimated using genome-wide genetic markers (Hill, 1993; Visscher *et al.*, 2006). However, as variation in relatedness among full siblings is limited, this approach will require large sample sizes.

When relatedness itself (as opposed to, for example, familiarity) is the causal factor underlying a difference in IGE, it would seem unlikely that the full set of genetic parameters can be identified. When individuals adjust their behaviour according to their relatedness to the recipient of the behaviour, as predicted by kin selection theory (Hamilton, 1964), any covariance between trait values of individuals is a function of relatedness and of genetic parameters of interest, which depends on this relatedness. This would seem to suggest full confounding.

However, variation in group composition seems to offer a solution. For example, having three different group compositions in a population may allow estimating all six genetic parameters. The first composition may have unrelated individuals only, the second may have two family members supplemented with unrelated individuals and the third may have three family members supplemented with unrelated individuals. From the first composition, $\sigma^2_{A_D}$, $\sigma_{A_D A_{S_u}}$ and $\sigma^2_{A_{S_u}}$ can be estimated using the traditional direct–indirect mixed model (Muir, 2005). Then, using the reduced model, $\sigma^2_{A_F}(n_f = 2) = \sigma^2_{A_D} + 2\sigma_{A_D S_f} + \sigma^2_{A_{S_f}}$ can be estimated from the second composition, and $\sigma^2_{A_F}(n_f = 3), = \sigma^2_{A_D} + 4\sigma_{A_D S_f} + 4\sigma^2_{A_{S_f}}$ can be estimated from the third composition, as well as $\sigma_{A_{S_f} S_u}$ and again $\sigma^2_{A_{S_u}}$. Then, as $\sigma^2_{A_D}$ is known from the first composition, this yields two equations with two unknowns, and thus can be solved yielding estimates of $\sigma_{A_D A_{S_f}}$ and $\sigma^2_{A_{S_f}}$. Moreover, the estimate of $\sigma_{A_{F,S_u}}$ from either the second or third composition can be used to obtain $\sigma_{A_{S_f,S_u}}$, because $\sigma_{A_D A_{S_u}}$ is known from the first composition (see Equation 12b). Then all six genetic parameters are estimated. Thus, variation in group composition with respect to family seems to allow estimating all six genetic parameters. Statistical power, however, may be very limited, and further complications may arise when IGEs depend on group size (Hadfield and Wilson, 2007; Bijma, 2010b), which we did not investigate here.

When IGEs depend on relatedness, the traditional direct–indirect mixed model that ignores this dependency yields biased estimates of the direct genetic variance and the direct–indirect genetic covariance, but an unbiased estimate of the variance in total breeding value. Thus, even though the full set of genetic parameters is not statistically identifiable, the total heritable variance and total breeding values can be estimated, either using the reduced model or the traditional model. This is an important result, because kin-dependent IGEs appear to be widespread in natural and domestic populations of both animals and plants (see Introduction).

The reduced model and traditional model are statistically equivalent, that is, yield the same maximum likelihood, but represent different linear combinations of the underlying parameters. The main difference is that the estimates of the reduced model are biologically meaningful in the context of kin-selection theory (Hamilton, 1964), as they separate the effects on kin (the family breeding value) from those on unrelated individuals. The correlation between the family breeding value and IGE on strangers, for example, measures the degree of competition or cooperation between families. With the

exception of the IGEs on strangers and the total breeding value, the estimates of the traditional model do not seem to have a clear biological meaning (Equation 16). Thus, the reduced model is preferable in terms of interpretation.

In this study, we have considered only the random effects; consequences of kin-dependent IGEs on the fixed effects to be included in the **Xb** term of the models have been ignored. When IGEs depend on relatedness, IGEs on kin vs strangers probably not only show incomplete correlation, but also differ systematically in level. In other words, individuals interacting primarily with kin probably receive more favourable IGEs than those interacting primarily with strangers, which creates a systematic difference in trait level between individuals interacting with different numbers of kin. This is not accounted for by the random effects in the model, because those are zero on average by construction. Hence, a fixed effect for the number of relatives an individual interacts with should be included in the model. This is similar to the inclusion of a fixed effect for the number of group mates when group size varies. Because estimation of a fixed effect with a few degrees of freedom is straight forward, we did not investigate this in detail. In our simulations, there was no need to account for such a fixed effect, because all individuals had the same number of kin and strangers among their group mates.

In animal and plant breeding, the focus is on improving the mean trait value of the population in the next generations. Theoretical studies have shown that group and kin selection methods utilize the total heritable variation for response to selection (Muir, 2005; Bijma *et al.*, 2007; Ellen *et al.*, 2008; McGlothlin *et al.*, 2010). This theoretical expectation is supported by results from selection experiments that have used group and/or kin selection without explicit reference to the total breeding value (Wade, 1976, 1977; Goodnight, 1985; Muir, 1996). Whether or not this result extends to the situation where IGEs differ between kin and strangers is interesting, but has not been investigated to our knowledge.

To optimize selection for traits affected by interactions among individuals, the ideal selection criterion is the TBV of selection candidates estimated using all available information. This is because response to selection equals the change in mean TBV from one generation to the next, so that maximizing the accuracy of estimated TBVs also maximizes response to selection. Because Equation 4 is generally valid, this result holds irrespective of whether or not IGEs depend on relatedness (Bijma, 2011b). Hence, the availability of kin and group selection methods does not make estimated TBVs superfluous. Moreover, knowledge of the total heritable variance quantifies the intrinsic potential of a population to respond to selection, and therefore provides a measure of efficiency for breeding schemes (Bijma, 2011b). The variance in TBV, therefore, is an important parameter for both optimizing individual selection decisions and evaluation of breeding schemes. This work has shown how the definition and estimation of the variance in TBV can be extended to schemes where IGEs differ between kin and strangers. This extension of variance in TBV to schemes where IGEs differ between kin and strangers may contribute to breeding plan design and application.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We acknowledge financial support for Piter Bijma by the Dutch Technology Foundation (STW), which is part of the Netherlands Organisation for Scientific Research (NWO), and support from the Danish Ministry of Food, Agriculture and Fisheries, Program Innovationsloven. SWA thanks Lei Zhou for his good inputs.

Arango J, Misztal I, Tsuruta S, Culbertson M, Herring W (2005). Estimation of variance components including competitive effects of large white growing gilts. *J Anim Sci* **83**: 1241–1246.

Biedrzycki ML, Bais HP (2010). Kin recognition in plants: a mysterious behaviour unsolved. *J Exp Bot* **61**: 4123–4128.

Bijma P (2010a). Estimating indirect genetic effects: precision of estimates and optimum designs. *Genetics* **186**: 1013–1028.

Bijma P (2010b). Multilevel selection 4: modeling the relationship of indirect genetic effects and group size. *Genetics* **186**: 1029–1031.

Bijma P (2011a). *Breeding for Social Interaction, for Animal Welfare*. In: Meyers RA (ed) Encyclopedia of Sustainability Science and TechnologySpringer Science and Business Media LLC: New York, pp 1–40.

Bijma P (2011b). A general definition of the heritable variation that determines the potential of a population to respond to selection. *Genetics* **189**: 1347–1359.

Bijma P, Muir WM, Ellen ED, Wolf JB, Van Arendonk JAM (2007). Multilevel selection 2: estimating the genetic parameters determining inheritance and response to selection. *Genetics* **175**: 289–299.

Bijma P, Wade MJ (2008). The joint effects of kin, multilevel selection and indirect genetic effects on response to genetic selection. *J Evol Biol* **21**: 1175–1188.

Chen CY, Kachman SD, Johnson RK, Newman S, Van Vleck LD (2008). Estimation of genetic parameters for average daily gain using models with competition effects. *J Anim Sci* **86**: 2525–2530.

Coffin HR, Watters JV, Mateo JM (2011). Odor-based recognition of familiar and related conspecifics: a first test conducted on captive Humboldt penguins (Spheniscus humboldti). *PLoS One* **6**: e25002.

Costa e Silva J, Potts BM, Bijma P, Kerr RJ, Pilbeam DJ (2013). Genetic control of interactions amongst individuals: contrasting outcomes of indirect genetic effects arising from neighbour disease infection and competition in a forest tree. *New Phytologist* **197**: 631–641.

Dudley SA, File AL (2007). Kin recognition in an annual plant. *Biology Lett* **3**: 435–438.

Ellen ED, Visscher J, van Arendonk JA, Bijma P (2008). Survival of laying hens: genetic parameters for direct and associative effects in three purebred layer lines. *Poult Sci* **87**: 233–239.

Falconer DS (1965). Maternal effects and selection response. In: Geerts SJ (ed) *Genetics Today: Proceedings of the XI International Congress on Genetics* vol. 3. Pergamon Press: New York, pp 763–774.

Falconer DS, Mackay TFC (1996). *Introduction to Quantitative Genetics, edn 4*. Longmans Green: Harlow, Essex, UK.

Frank SA (2007). All of life is social. *Curr Biol* **17**: R648–R650.

Gilmour AR, Gogel BJ, Cullis BR, Thomson R (2006). *ASReml User Guide Release 2.0*. VSN International Ltd: Hemel Hempstead, UK.

Goodnight CJ (1985). The influence of environmental variation on group and individual selection in a cress. *Evolution* **39**: 545–558.

Griffing B (1967). Selection in reference to biological groups. I. Individual and group selection applied to populations of unordered groups. *Aust J Biol Sci* **20**: 127–139.

Griffing B (1976). Selection in reference to biological groups. V. Analysis of full-sib groups. *Genetics* **82**: 703–722.

Hadfield JD, Wilson AJ (2007). Multilevel selection 3: modeling the effects of interacting individuals as a function of group size. *Genetics* **177**: 667–668.

Hamilton WD (1964). The genetical evolution of social behavior. *J Theor Biol* **7**: 1–16.

Henderson CR (1953). Estimation of variance and covariance components. *Biometrics* **9**: 226–252.

Henderson CR (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**: 423–447.

Henderson CR (1985). Best linear unbiased prediction using relationship matrices derived from selected base-populations. *J Dairy Sci* **68**: 443–448.

Hepper PG (1983). Sibling recognition in the rat. *Anim Behav* **31**: 1177–1191.

Hepper PG (1986). Kin recognition: functions and mechanisms. A review. *Biol Rev Camb Philos Soc* **61**: 63–93.

Hill WG (1993). Variation in genetic identity within kinships. *Heredity* **71**: 652–653.

Holmes WG, Sherman PW (1982). The ontogeny of Kin recognition in 2 species of ground-squirrels. *Am Zool* **22**: 491 517.

Kessler SE, Scheumann M, Nash LT, Zimmermann E (2012). Paternal kin recognition in the high frequency/ultrasonic range in a solitary foraging mammal. *BMC Ecol* **12**: 26.

Kirkpatrick M, Lande R (1989). The evolution of maternal characters. *Evolution* **43**: 485–503.

Lynch M, Walsh B (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates: Sunderland, MA.

Mateo JM (2004). Recognition systems and biological organization: the perception component of social recognition. *Ann Zool Fenn* **41**: 729–745.

Mateo JM, Holmes WG (2004). Cross-fostering as a means to study kin recognition. *Anim Behav* **68**: 1451–1459.

McGlothlin JW, Brodie III ED (2009). How to measure indirect genetic effects: the congruence of trait-based and variance-partitioning approaches. *Evolution* **63**: 1785–1795.
McGlothlin JW, Moore AJ, Wolf JB, Brodie ED III (2010). Interacting phenotypes and the evolutionary process. III. Social evolution. *Evolution* **64**: 2558–2574.
Moore AJ, Brodie ED III, Wolf JB (1997). Interacting phenotypes and the evolutionary process: 1. direct and indirect genetic effects of social interactions. *Evolution* **51**: 1352–1362.
Muir WM (1996). Group selection for adaptation to multiple-hen cages: selection program and direct responses. *Poultry Sci* **75**: 447–458.
Muir WM (2005). Incorporation of competitive effects in forest tree or animal breeding programs. *Genetics* **170**: 1247–1259.
Odegard J, Olesen I (2011). Comparison of testing designs for genetic evaluation of social effects in aquaculture species. *Aquaculture* **317**: 74–78.
Olsen KH (1989). Sibling recognition in juvenile Arctic charr, Salvelinus-alpinus (L). *J Fish Biol* **34**: 571–581.
R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, ISBN 3-900051-07-0, URL. http://www.R-project.org/
Tang-Martinez Z (2001). The mechanisms of kin discrimination and the evolution of kin recognition in vertebrates: a critical re-evaluation. *Behav Process* **53**: 21–40.
Visscher PM, Medland SE, Ferreira MAR, Morley KI, Zhu G, Cornes BK *et al.* (2006). Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *Plos Genet* **2**: 316–325.
Wade MJ (1976). Group selection among laboratory populations of Tribolium. *Proc Natl Acad Sci USA* **73**: 4604–4607.
Wade MJ (1977). Experimental-study of group selection. *Evolution* **31**: 134–153.
Willham RL (1963). The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics* **19**: 18–27.
Wilson AJ, Morrissey MB, Adams MJ, Walling CA, Guinness FE, Pemberton JM *et al.* (2011). Indirect genetics effects and evolutionary constraint: an analysis of social dominance in red deer, Cervus elaphus. *J Evol Biol* **24**: 772–783.
Wolf JB, Brodie ED III, Cheverud JM, Moore AJ, Wade MJ (1998). Evolutionary consequences of indirect genetic effects. *Trends Ecol Evol* **13**: 64–69.

APPENDIX A

This appendix shows that the full model is not statistically identifiable, whereas the reduced model is identifiable. Estimation of genetic parameters for direct and indirect genetic effects rests on covariances between phenotypes of relatives and of their social partners (Lynch and Walsh, 1998). Only covariances between relatives (or social partners) present in different groups contribute to the estimation of genetic parameters because within-group covariances are fully confounded with the nongenetic direct and indirect effects. The following, therefore, considers between-group covariances only.

Each group consists of members of two families. There are no genetic covariances between groups not sharing a family; hence those group combinations can be ignored. Then, when considering two groups having one family in common, there are three families in total; the common family, denoted F_1, and its partner family in each group, denoted F_2 and F_3. Before we derive covariance between individual, the individual's total breeding value, which is the total heritable impact of an individual's genes on the mean trait value of the population when interaction differ between kin and strangers, is given as:

$$A_{T,i} = A_{Di} + 2\left(n_f - 1\right)A_{S_f} + n_f A_{S_u} \tag{A1}$$

Taking the variance of the total breeding value yields

$$\sigma^2_{A_T} = \sigma^2_{A_D} + 2\left(n_f - 1\right)\sigma_{A_{D A_{S_f}}} + \left(n_f - 1\right)^2\sigma^2_{A_{S_f}} + 2n_f\left(n_f - 1\right)\sigma_{A_{S_f A_{S_u}}} + n^2\sigma^2_{A_{S_u}} \tag{A2}$$

When we have one family common in two groups, there are only three informative covariances. First, the covariance between the phenotypes of a member of F_1 in each group,

$$Cov(P_i, P_j \mid i, j \in F_1) = r\left[\sigma^2_{A_D} + 2(n_f - 1)\,\sigma_{A_D A_{S_f}} + (n_f - 1)^2\sigma^2_{A_{S_f}}\right], \tag{A3}$$

where r denotes relatedness between members of the same family, and n_f the number of members of F_1 in each group (assumed to be the same in both groups). Second, the covariance between a member of the common family (F_1) in the one group, and a member of a partner family in the other group (F_2 is considered here, but the result for F_3 is identical),

$$Cov(P_i, P_j \mid i \in F_1, j \in F_2) = r\left[n_f\,\sigma_{A_D A_{S_u}} + n_f(n_f - 1)\,\sigma_{A_{S_f} A_{S_u}}\right]. \tag{A4}$$

Third, the covariance between two members of the partner families (F_2 and F_3) in different groups is

$$Cov(P_i, P_j \mid i \in F_2, j \in F_3) = r\,n_f^2\,\sigma^2_{A_{S_u}} \tag{A5}$$

This equation shows that the variance of IGE on strangers is estimable. In total, however, these three equations contain six unknowns (the six genetic parameters to be estimated) and cannot be solved. Thus, the full model is not identifiable.

Equations A1 to A3 also show that the reduced model is identifiable, as they represent the informative covariances and there are only three genetic parameters to estimate. Moreover, Equations A1 to A3 imply that the expected values of the estimated genetic parameters of the reduced model are given by Equations 12a–c when $n_f = \frac{1}{2}n$.

APPENDIX B

This appendix shows the derivation of the nongenetic covariance structure generated by Equation 3 (Equations 11 and 14) and refers to the reduced model. Nongenetic covariances occur only among individuals within the same group. As the genetic model terms fully account for genetic covariances within groups, those can be ignored here.

There are three nongenetic parameters of interest: the covariance between group members of different families, the covariance between group members of the same family and the residual variance. Because all groups have the same composition, these parameters are the same for all groups. This leads to the block-diagonal residual variance structure given by Equation 11, which has a single residual variance, a covariance between group mates of the same family and a second covariance between group mates of different families.

Consider two group mates, say i and k. The group mates of i of its own family are denoted j, and those of the other family j'. Analogously, the group mates of k are denoted l and l'. Note that k is one of the individuals included in j and j', whereas i is one of the individuals included in l and l'. Then, the nongenetic covariance between the phenotypes of i and k is given by

$$\mathrm{cov}(P_i, P_k)_E = \mathrm{cov}\left(E_{D_i} + \sum_{j=1}^{\frac{n}{2}-1} E_{S_f,j} + \sum_{j'=1}^{\frac{n}{2}} E_{S_u,j'}, E_{D_k} + \sum_{l=1}^{\frac{n}{2}-1} E_{S_f,l} + \sum_{l'=1}^{\frac{n}{2}} E_{S_u,l'}\right). \tag{B1}$$

First consider this covariance when i and k are group members of different families, giving

$$\mathrm{cov}(P_i, P_k)_E = \quad 2\sigma_{E_{D,S_u}} + (n-2)\sigma_{E_{S_f,S_u}}. \tag{B2}$$

The first term arises because i affects k and vice versa, whereas the second term arises because i and k have $n-2$ group mates in common. In Equation 10, the nongenetic covariance between unrelated group mates equals the variance of the random group effect. Hence,

$$E(\hat{\sigma}_g^2) \;=\; 2\sigma_{E_{D,S_u}} + (n-2)\sigma_{E_{S_f,S_u}}$$

which is Equation 14a.

Next, consider the full nongenetic variance. From Equation B1, it follows that

$$\mathrm{var}(P)_E = \sigma^2_{E_D} + \left(\tfrac{1}{2}n-1\right)\sigma^2_{E_{S_f}} + \tfrac{1}{2}n\sigma^2_{E_{S_u}}. \tag{B4}$$

In Equation 10, the full nongenetic variance is the sum of the group variance and the residual variance. Hence, the residual variance in Equation 10 follows from subtracting the group variance from Equation B4, giving

$$\mathrm{E}(\hat{\sigma}_e^2) = \sigma^2_{E_D} + \left(\tfrac{1}{2}n-1\right)\sigma^2_{E_{S_f}} + \tfrac{1}{2}n\,\sigma^2_{E_{S_u}} - \sigma^2_g,$$

which is Equation 14b.

Finally, consider the covariance when i and k are group members of the same family,

$$\mathrm{cov}(P_i, P_k)_E = 2\sigma_{E_{D,S_f}} + \left(\tfrac{1}{2}n-2\right)\sigma^2_{E_{S_f}} + \tfrac{1}{2}n\sigma^2_{E_{S_u}} \tag{B5}$$

The first term arises because i affects k and vice versa, the second term arises because i and k have $\frac{1}{2}n-2$ group mates of their own family in common and the third term arises because i and k have $\frac{1}{2}n$ group mates of the other family in common. In Equations 10 and 11, the covariance between unrelated group mates is the sum of the group variance and the residual covariance, $\sigma^2_g + \rho\sigma^2_e$. Hence, the residual correlation follows from subtracting the group variance from Equation B5 and dividing by the residual variance, giving

$$\mathrm{E}(\hat{\rho}) = \frac{2\sigma_{E_D E_{S_f}} + \left(\tfrac{1}{2}n-2\right)\sigma^2_{E_{S_f}} + \tfrac{1}{2}n\,\sigma^2_{E_{S_u}} - \sigma^2_g}{\sigma^2_e}$$

which is Equation 14c.

APPENDIX C

This appendix shows the derivation of Equations 16a–d, being the expectations of genetic parameters when the traditional direct-–indirect model (Equation 15) is applied to data generated by the model in Equation 3. The derivation uses the method of Bijma (2010a).

Direct genetic variance

With two families per group, the information for estimating the direct genetic variance using Equation (15) comes from the variable $z_{kl} = \bar{P}_{kl} - \varphi\bar{P}_{k'l}$, in which $\varphi = (\frac{1}{2}n-1)/\frac{1}{2}n$, $\bar{P}_{kl}$ is the mean phenotype of the family of interest k in group l, and $\bar{P}_{k'l}$ is the mean phenotype of the other family k' in group l (Equations B15 and B16 in Bijma, 2010a; the z_{kl} is referred to as the 'effective record') and n is the group size. When the data are generated by Equation 3, the expectation of z_{kl} conditional on the family of interest k in group l equals

$$E[z_{kl} \mid k] = A_{D,k} + \left(\tfrac{1}{2}n-1\right)A_{S_f,k} - \left(\tfrac{1}{2}n-1\right)A_{S_u,k}, \tag{C1}$$

which depends not only on the DGE of family k, but also on the IGEs of family k on kin and strangers. The expected value of the estimated direct genetic variance follows from the variance of z, giving

$$E(\hat{\sigma}^2_{A_D})_{Eqn.15} = \sigma^2_{A_D} \;+\; \left(\tfrac{1}{2}n-1\right)^2\left(\sigma^2_{A_{S_f}} - 2\sigma_{A_{S_f}A_{S_u}} + \sigma^2_{A_{S_u}}\right) + \;(n-2)\left(\sigma_{A_D A_{S_f}} - \sigma_{A_D A_{S_u}}\right),$$

which is Equation 16a.

Indirect genetic variance

The information for estimating the indirect genetic variance using Equation (15) comes from the variable $z_{kl} = \bar{P}_{k'l}/(\frac{1}{2}n)$ (Equation B18 in Bijma, 2010a). When the data are generated by Equation 3, the expectation of z_{kl} conditional on the family of interest, equals

$$E[z_{kl} \mid k] = A_{S_u,k} \tag{C2}$$

Thus, the expected value of the estimated indirect genetic variance equals

$$E(\hat{\sigma}^2_{A_S})_{Eqn.15} = \sigma^2_{A_{S_u}}$$

which is Equation 16c.

Direct–indirect genetic covariance

From Equations C1 and C2, it follows that

$$E(\hat{\sigma}_{A_{DS}})_{Eqn.15} = \sigma_{A_D A_{S_u}} + \left(\tfrac{1}{2}n-1\right)\left(\sigma_{A_{S_f}A_{S_u}} - \sigma^2_{A_{S_u}}\right)$$

which is Equation 16b. Thus, when the IGE on kin is identical to the IGE on strangers, the second term becomes zero, and$E(\hat{\sigma}_{A_{DS}})_{Eqn.15} = \sigma_{A_{DS}}$.

Total heritable variation

The information for estimating the total heritable variance using Equation (15) comes from the variable $z_{kl} = \sum_{j=1}^{n} P_{kl,j}\,/(\frac{1}{2}n)$ (Equation B20 in Bijma, 2010a). When the data are generated by Equation 3, the expectation of z_{kl} conditional on the family of interest equals

$$E[z_{kl} \mid k] = \tfrac{1}{2}n\,\left[A_D + \left(\tfrac{1}{2}n-1\right)A_{S_f} + \tfrac{1}{2}n\,A_{S_u}\right]\;/\;\left(\tfrac{1}{2}n\right) = A_{T,k} \tag{C3}$$

Thus, the expected value of the estimated total heritable variance equals

$$E(\hat{\sigma}^2_{A_T})_{Eqn.15} = \sigma^2_{A_T},$$

which is Equation 16d.

APPENDIX D

This appendix shows details of the stochastic simulation. Breeding values of individuals in the base generation were simulated from the multivariate normal distribution

$$\begin{bmatrix} A_D \\ A_{S_f} \\ A_{S_u} \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{A_D} & r_{A_{D,S_f}}\sigma_{A_D}\sigma_{A_{S_f}} & r_{A_{D,S_u}}\sigma_{A_D}\sigma_{A_{S_u}} \\ r_{A_{D,S_f}}\sigma_{A_D}\sigma_{A_{S_f}} & \sigma^2_{A_{S_f}} & r_{A_{S_f,S_u}}\sigma_{A_{S_f}}\sigma_{A_{S_u}} \\ r_{A_{D,S_u}}\sigma_{A_D}\sigma_{A_{S_u}} & r_{A_{S_f,S_u}}\sigma_{A_{S_f}}\sigma_{A_{S_u}} & \sigma^2_{A_{S_u}} \end{pmatrix}\right].$$

To produce the second generation, sires and dams of the first generation were mated at random, each sire being mated to 10 dams, and each dam producing 10 full-sib offspring. Second-generation breeding values for all three genetic effects were simulated as $A_i = \frac{1}{2}A_{sire} + \frac{1}{2}A_{dam} + MS_i$, where MS denotes the Mendelian sampling term. The MS were simulated from the multivariate normal

distribution

$$\begin{bmatrix} MS_D \\ MS_{S_f} \\ MS_{S_u} \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{2}\sigma^2_{A_D} & \frac{1}{2}r_{A_D,S_f}\sigma_{A_D}\sigma_{A_{S_f}} & \frac{1}{2}r_{A_D,S_u}\sigma_{A_D}\sigma_{A_{S_u}} \\ \frac{1}{2}r_{A_D,S_f}\sigma_{A_D}\sigma_{A_{S_f}} & \frac{1}{2}\sigma^2_{A_{S_f}} & \frac{1}{2}r_{A_{S_f},S_u}\sigma_{A_{S_f}}\sigma_{A_{S_u}} \\ \frac{1}{2}r_{A_D,S_u}\sigma_{A_D}\sigma_{A_{S_u}} & \frac{1}{2}r_{A_{S_f},S_u}\sigma_{A_{S_f}}\sigma_{A_{S_u}} & \frac{1}{2}\sigma^2_{A_{S_u}} \end{pmatrix} \right].$$

Nongenetic effect were simulated only for the second generation, from

$$\begin{bmatrix} E_D \\ E_{S_f} \\ E_{S_u} \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{E_D} & r_{E_D,S_f}\sigma_{E_D}\sigma_{E_{S_f}} & r_{E_D,S_u}\sigma_{E_D}\sigma_{E_{S_u}} \\ r_{E_D,S_f}\sigma_{E_D}\sigma_{E_{S_f}} & \sigma^2_{E_{S_f}} & r_{E_{S_f},S_u}\sigma_{E_{S_f}}\sigma_{E_{S_u}} \\ r_{E_D,S_u}\sigma_{E_D}\sigma_{E_{S_u}} & r_{E_{S_f},S_u}\sigma_{E_{S_f}}\sigma_{E_{S_u}} & \sigma^2_{E_{S_u}} \end{pmatrix} \right].$$

Then, we calculated the phenotypes of the individual from second generation using the full model (Equation 3). Those phenotypic values were used to estimate the variance components. We simulated 100 replicates for each set of genetic parameter and the estimates were averaged over replicates.