

# Estimating Indirect Genetic Effects: Precision of Estimates and Optimum Designs

Piter Bijma<sup>1</sup>

*Animal Breeding and Genomics Centre, Wageningen University, Marijkeweg 40, 6709PG Wageningen, The Netherlands*

Manuscript received July 1, 2010

Accepted for publication August 10, 2010

## ABSTRACT

Social interactions among individuals are abundant both in natural and domestic populations. Such social interactions cause phenotypes of individuals to depend on genes carried by other individuals, a phenomenon known as indirect genetic effects (IGE). Because IGEs have drastic effects on the rate and direction of response to selection, knowledge of their magnitude and relationship to direct genetic effects (DGE) is indispensable for understanding response to selection. Very little is known, however, of statistical power and optimum experimental designs for estimating IGEs. This work, therefore, presents expressions for the standard errors of the estimated (co)variances of DGEs and IGEs and identifies optimum experimental designs for their estimation. It also provides an expression for optimum family size and a numerical investigation of optimum group size. Designs with groups composed of two families were optimal and substantially better than designs with groups composed at random with respect to family. Results suggest that IGEs can be detected with ~1000–2000 individuals and/or ~250–500 groups when using optimum designs. Those values appear feasible for agriculture and aquaculture and for the smaller laboratory species. In summary, this work provides the tools to optimize and quantify the required size of experiments aiming to identify IGEs. An R-package SE.IGE is available, which predicts SEs and identifies optimum family and group sizes.

**S**Ocial interactions among individuals, such as competition and cooperation, are fundamental to evolution by natural selection (DARWIN 1859; HAMILTON 1964; WILSON 1975; FRANK 1998; KELLER 1999; CLUTTON-BROCK 2002). Most research on social interactions among individuals has focused on the fitness cost and benefit of such interactions. This focus on fitness effects seems to originate from an interest in the factors determining evolutionary success of populations and from the observation of altruistic and cooperative behaviors, which seem to contradict the intuitive notion of “survival of the fittest.”

Effects of social interactions, however, are not limited to fitness. There is increasing evidence that not only fitness, but also trait values of individuals are affected by genes in other individuals (reviewed in WOLF *et al.* 1998; MCGLOTHLIN and BRODIE 2009). Such effects are known as indirect genetic effects (IGE; CHEVERUD and MOORE 1994; MOORE *et al.* 1997) or associative effects (GRIFFING 1967; MUIR 2005). An IGE is a heritable effect of one individual on the trait value of another individual. A well-known example in mammals is the maternal-genetic effect of a mother on preweaning growth rate of her offspring (WILLHAM 1963; MOUSSEAU

and FOX 1998). The IGE modeling approach has its roots in quantitative genetic models for maternally affected traits, which are widely used in agriculture (DICKERSON 1947; WILLHAM 1963; CHEVERUD 1984; KIRKPATRICK and LANDE 1989; LYNCH and WALSH 1998), but is increasingly applied to other traits, both in animals and plants (*e.g.*, MUIR 2005; PETFIELD *et al.* 2005; MUTIC and WOLF 2007; BERGSMA *et al.* 2008; CAPP and CANTET 2008; BLEAKLEY and BRODIE 2009; CHENOWETH *et al.* 2010).

IGEs may have drastic effects on the rate and direction of response to natural or artificial selection (GRIFFING 1967, 1976, 1981a,b; MOORE *et al.* 1997; WOLF *et al.* 1998; AGRAWAL *et al.* 2001; BIJMA and WADE 2008; MCGLOTHLIN *et al.* 2010). GRIFFING (1967), for example, showed that IGEs may cause traits to respond in the direction opposite to selection. This theoretical expectation has been confirmed empirically in selection experiments (MUIR 1996, 2005). BIJMA *et al.* (2007a) showed theoretically that IGEs may substantially increase the heritable variance determining a population's potential to respond to selection. Moreover, IGEs alter the impacts of multi-level selection and relatedness among individuals on response to selection (GRIFFING 1976, 1981a,b; AGRAWAL *et al.* 2001; BIJMA and WADE 2008; MCGLOTHLIN *et al.* 2010). The above clearly illustrates that knowledge of IGEs and their relationship to direct genetic effects (DGEs) is indispensable for understanding response to natural or artificial selection.

<sup>1</sup>Address for correspondence: Animal Breeding and Genomics Centre, Wageningen University, P.O. Box 338, 6700AH Wageningen, The Netherlands. E-mail: piter.bijma@wur.nl

The magnitude of IGEs and their contribution to heritable variance can be estimated empirically using mixed models that are common in the field of animal breeding (MUIR and SCHINCKEL 2002; VAN VLECK and CASSADY 2004, 2005; MUIR 2005; ARANGO *et al.* 2005; BIJMA *et al.* 2007b; CHEN *et al.* 2008; MCGLOTHLIN and BRODIE 2009). Such models are empirically powerful, because they allow estimating IGEs without knowledge of the underlying mechanisms and without the need to observe the interactions *per se*. Mixed models including IGEs can be fitted to field data, using restricted maximum likelihood estimation (PATTERSON and THOMPSON 1971; KRUIK 2004; MUIR 2005; BIJMA *et al.* 2007b). Using field data on domestic pigs, for example, BERGSMA *et al.* (2008) found that IGEs contributed more than half of heritable variance in growth rate.

At present, knowledge of the magnitude of IGEs and their relationship to DGEs is limited. More knowledge is required to better understand the relevance of IGEs for evolution by natural selection and for genetic improvement in livestock, aquaculture, and plants. Very little is known, however, of the factors determining the standard errors (SE) of the estimated genetic parameters for IGEs, which hampers the optimization of experimental designs. Although the variance of IGEs can be estimated when groups of interacting individuals are composed at random with respect to family, it is unknown whether such designs are efficient. Moreover, it has been shown that genetic covariance components are not identifiable when groups consist of entire families, when families are allocated to groups in a systematic manner (BIJMA *et al.* 2007b), or when fixed group effects are fitted (CANTET and CAPPA 2008).

Here I present equations for the SE of the estimated variances, covariance, and correlation of DGEs and IGEs. Accuracy of the prediction equations is evaluated using stochastic simulations. Subsequently, the prediction equations are used to identify optimum experimental designs for estimating IGEs.

## BACKGROUND

This section briefly summarizes the quantitative genetic theory of IGEs and may be skipped by readers familiar with this theory. An IGE is a heritable effect of one individual on the trait value of another. IGEs can be modeled in two alternative ways (reviewed in MCGLOTHLIN and BRODIE 2009): first, using a trait-based model, where an individual's IGE is a direct consequence of its trait value (FALCONER 1965; MOORE *et al.* 1997) and second, using a variance-component model, where the genetic variance in trait value is partitioned into a direct component due to the focal individual's genotype, and an indirect component due to the genotypes of its interactants (WILLHAM 1963; GRIFFING 1967). Here I use the variance-component model, because it allows identification of the full

variance due to IGEs without requiring knowledge of the trait(s) or mechanisms underlying them.

Consider a population consisting of groups of  $n_w$  members each, where social interactions occur among the members of a group. Then the trait value of focal individual  $i$  may be modeled as the sum of a direct effect rooted in the focal individual itself,  $P_{D,i}$ , and the sum of the indirect effects,  $P_{S,j}$ , of each of its  $n_w - 1$  group mates,

$$P_i = P_{D,i} + \sum_{j=1}^{n_w-1} P_{S,j}. \quad (1)$$

Hence,  $i$  denotes the focal individual,  $j$  one of its  $n_w - 1$  group mates, D direct effects, S indirect effects, and  $\sum_{j=1}^{n_w-1}$  the sum taken over the  $n_w - 1$  group mates of focal individual  $i$ . (S, suggesting "social," is used as subscript for indirect effects, since I may be confused with  $i$  or 1. See Table 1 for a notation key.)

Both direct and indirect effects may be decomposed into an additive genetic (*i.e.*, heritable) component,  $A$ , and a remaining nonheritable component,  $E$  (GRIFFING 1967),

$$P_i = A_{D,i} + E_{D,i} + \sum_{j \neq i}^{n_w-1} A_{S,j} + \sum_{j \neq i}^{n_w-1} E_{S,j}, \quad (2)$$

where  $A_{D,i}$  is the DGE of the focal individual,  $E_{D,i}$  the corresponding nonheritable direct effect,  $A_{S,j}$  the IGE of group mate  $j$ , and  $E_{S,j}$  the corresponding nonheritable indirect effect.

On the population level, the magnitude of DGEs and IGEs is measured by their (co)variances. Hence, interest is in the variance of DGE,  $\sigma_{A_D}^2$ , the variance of IGE,  $\sigma_{A_S}^2$ , and their covariance,  $\sigma_{A_{DS}}$ . Moreover, interest is in the total heritable variance arising from the joint effect of DGEs and IGEs. The total heritable impact of an individual's genes on the mean trait value of the population is given by the individual's total breeding value (TBV; BIJMA *et al.* 2007a)

$$\text{TBV}_i = A_{D,i} + (n - 1)A_{S,i}. \quad (3)$$

Note that, in contrast to the trait value, the TBV in Equation 3 is entirely a heritable property of the focal individual. It is a generalization of the classical breeding value and is the heritable property relevant for response to selection in traits affected by IGEs. The total heritable variance in the trait due to both DGEs and IGEs, *i.e.*, the magnitude of the heritable differences among individuals determining the potential of the population to respond to selection, equals the variance in TBVs among individuals (BIJMA *et al.* 2007a),

$$\sigma_{\text{TBV}}^2 = \sigma_{A_D}^2 + 2(n_w - 1)\sigma_{A_{DS}} + (n_w - 1)^2\sigma_{A_S}^2. \quad (4)$$

The  $\sigma_{\text{TBV}}^2$  expresses heritable variance in absolute units. Interpretation is often facilitated by expressing

**TABLE 1**  
**Notation key**

Symbol	Meaning
DGE, IGE	Direct genetic effect, indirect genetic effect
$T, N, n$	Total number of individuals, number of families, family size, $T = Nn$
$i, j$	Focal individual, group mate of focal individual
$n_w, n_g, n_{gf}$	Group size, number of groups, number of groups per family, $T = n_g n_w$
$P_{D,i}, P_{S,i}$	Direct effect of $i$ , indirect effect of $i$
$A_{D,i}, A_{S,i}$	Direct genetic effect of $i$ , indirect genetic effect of $i$
$E_{D,i}, E_{S,i}$	Direct nongenetic effect of $i$ , indirect nongenetic effect of $i$
$TBV_i, \tau^2$	Total breeding value of $i$ (Equation 4), relative heritable variance (Equation 5)
$\sigma_{A_D}^2, \sigma_{A_S}^2$	Variance of DGE among individuals, variance of IGE among individuals
$\sigma_{A_{DS}}, r_g$	Covariance between DGEs and IGEs, correlation between DGEs and IGEs
$\sigma_{E_{DS}}, r_E$	Covariance and correlation between nongenetic direct and indirect effects
$\sigma_P^2, \sigma_{TBV}^2$	Phenotypic variance among individuals, variance of TBVs among individuals
$k, l, m$	Focal family, index for records within family, effective no. records per family
$z_{kl}$	$l$ h effective record on family $k$ for use in ANOVA (Equation 6)
$f_k, e_{kl}$	Effect of $k$ th family, residual of $l$ th record on family $k$ , $z_{kl} = f_k + e_{kl}$ .
$\sigma_z^2, \sigma_f^2, \sigma_c^2$	Full variance of records, between-family variance, residual variance
$n_{opt,x}$	Optimum family size for parameter $x$
SE, SE( $\hat{x}$ )	Standard error, standard error of estimate of $x$
$d$	Degree of dilution of indirect effects with group size (Equation 18)
$\sigma_{P_{S,2}}^2$	Variance of indirect effects in groups of two individuals
$\sigma_{A_{S,2}}^2$	Variance of IGEs in groups of two individuals

Throughout, hats ( $\hat{\phantom{x}}$ ) denote estimates, whereas symbols without hats refer to true values.

heritable variance relative to phenotypic variance. In classical theory, heritability measures heritable variance relative to phenotypic variance,  $h^2 = \sigma_A^2 / \sigma_P^2$  (FALCONER and MACKAY 1996). By analogy, in case of IGEs we may define the ratio of total heritable variance over phenotypic variance (BERGSMAN *et al.* 2008),

$$\tau^2 = \frac{\sigma_{TBV}^2}{\sigma_P^2}. \tag{5}$$

A comparison of  $h^2$  and  $\tau^2$  reveals the proportional contribution of IGEs to heritable variance.

**PREDICTION OF STANDARD ERRORS**

This section presents expressions for the SE of estimated values of  $\sigma_{A_D}^2, \sigma_{A_S}^2, \sigma_{A_{DS}}, \sigma_{TBV}^2$ , and  $\tau^2$ , for two experimental designs.

**Experimental designs:** Two experimental designs were investigated; groups were either composed at random with respect to family, or groups were composed of members of precisely two families. For both designs there was a one-way classification of families, where  $N$  unrelated families each contribute  $n$  individuals, yielding a total of  $T = Nn$  individuals with records. Genetic relatedness within family,  $r$ , is the same for all families. The  $T$  individuals are allocated to  $n_g$  groups, each of size  $n_w$  so that  $T = n_g n_w$ .

*Random group composition:* In this design, the individuals making up a group are sampled at random with respect to family. Hence, there is no relatedness among

group mates, other than by chance, so that genotypes of group mates are independent.

*Two families per group:* In this design, each group is composed of members of two distinct families, each family contributing  $\frac{1}{2} n_w$  individuals (Table 2). Hence, group size is an even number, and each family contributes to  $n_{gf} = n / (\frac{1}{2} n_w)$  groups. Thus each family is combined with  $n_{gf}$  distinct other families. This design leads to a block structure, with  $n_{gf} + 1$  families per block. Each of the  $n_{gf} + 1$  families within a block is combined with each of the other  $n_{gf}$  families in that block, each combination occurring precisely once. In the stochastic

**TABLE 2**

**Example of the design with groups composed of two families**

Family A	Family B				
	2	3	4	5	6
1	2/2	2/2	2/2	2/2	2/2
2		2/2	2/2	2/2	2/2
3			2/2	2/2	2/2
4				2/2	2/2
5					2/2

Group size equals  $n_w = 4$ . Each group consists of members of two families, each family contributing 2 individuals. Family size equals  $n = 10$ , so that each family is combined with  $n_{gf} = 10 / (\frac{1}{2} \times 4) = 5$  other families. Thus a block consists of 6 families, each being combined with each of the 5 other families in that block. Hence, there are  $n_{gb} = n_{gf} (n_{gf} + 1) / 2 = 5 \times 6 / 2 = 15$  groups per block,  $n_{gb} \times n_w = 15 \times 4 = 60$  individuals per block, and  $n_b = T / 60$  blocks.

simulations, design parameters were chosen such that the number of blocks was an integer number.

**Expressions for SEs:** Prediction equations for the SE of estimated genetic parameters are based on one-way ANOVA, partitioning the full variance into a between- and within-family component. Thus records were modeled as

$$z_{kl} = f_k + e_{kl}, \tag{6}$$

where  $z_{kl}$  is the  $l$ th effective record on family  $k$ ,  $f_k$  is the effect of the  $k$ th family, and  $e_{kl}$  is the residual. The definition of an effective record  $z_{kl}$  will depend on the experimental design and on the genetic parameter of interest. In general, the  $z_{kl}$  refers to the variable that provides information on the genetic parameter of interest in the experimental design considered (Appendix B). There are  $m$  effective records per family and  $Nm$  effective records in total.

*Additive genetic variances:* From one-way ANOVA, the SE of an estimated genetic variance equals (Appendix A)

$$SE(\hat{\sigma}_A^2) \approx \frac{1}{r} \sqrt{\frac{2}{N-1} \left[ \sigma_f^4 + \frac{2\sigma_f^2\sigma_c^2}{m} + \frac{\sigma_c^4}{m(m-1)} \right]}, \tag{7}$$

where  $\sigma_f^2$  denotes the between-family variance and  $\sigma_c^2$  the residual variance,

$$\sigma_c^2 = \sigma_z^2 - \sigma_f^2. \tag{8}$$

Equations 7 and 8 are used to predict  $SE(\hat{\sigma}_{Ab}^2)$ ,  $SE(\hat{\sigma}_{As}^2)$ , and  $SE(\hat{\sigma}_{TBV}^2)$ . Application of those equations requires knowledge of  $\sigma_z^2$  and  $\sigma_f^2$ , which depend on the parameter and experimental design of interest, and are given in Tables 3 and 4.

*Direct-indirect additive genetic covariance:* The SE of the estimated genetic covariance,  $SE(\hat{\sigma}_{Abs})$ , was obtained from a general expression for the variance of an estimated covariance (Appendix A),

$$SE(\hat{\sigma}_{Abs}) = \sqrt{\frac{SE(\hat{\sigma}_{Ab}^2) SE(\hat{\sigma}_{As}^2)}{2} + \frac{\sigma_{Abs}^2}{N-1}}. \tag{9}$$

The values of  $SE(\hat{\sigma}_{Ab}^2)$  and  $SE(\hat{\sigma}_{As}^2)$  required for Equation 9 follow from Equations 7 and 8.

*Relative genetic variance:* The SE of the ratio of total heritable variance over phenotypic variance is given by

$$SE(\hat{\tau}^2) \gg \frac{SE(\hat{\sigma}_{TBV}^2)}{\sigma_P^2}, \tag{10}$$

where  $SE(\hat{\sigma}_{TBV}^2)$  follows from Equations 7 and 8. This approximation assumes that  $\sigma_P^2$  is accurately known, or equivalently that  $Var(\hat{\sigma}_P^2) \approx \sigma_P^4$  and  $Cov(\hat{\sigma}_{TBV}^2, \hat{\sigma}_P^2) \approx 0$ .

*Additive genetic correlation:* The SE of the estimated correlation between DGEs and IGEs,

TABLE 3

Components of SEs with random group composition

Parameter	Expression
All VC	$m = n$ $\sigma_P^2 = \sigma_{Pd}^2 + (n_w - 1) \sigma_{Ps}^2$ $Cov_w = 2\sigma_{Pds} + (n_w - 2) \sigma_{Ps}^2$
$\hat{\sigma}_{Ab}^2$	$\sigma_z^2 = \sigma_P^2$ <span style="float:right;"><math>\sigma_f^2 = r\sigma_{Ab}^2</math></span>
$\hat{\sigma}_{As}^2$	$\sigma_z^2 = \frac{\sigma_P^2 + (n_w - 2)Cov_w}{n_w - 1}$ <span style="float:right;"><math>\sigma_f^2 = r\sigma_{As}^2</math></span>
$\hat{\sigma}_{TBV}^2$	$\sigma_z^2 = n_w [\sigma_P^2 + (n_w - 1)Cov_w]$ <span style="float:right;"><math>\sigma_f^2 = r\sigma_{TBV}^2</math></span>

SEs for the estimated genetic variance of interest follow from substituting  $\sigma_z^2$  and  $\sigma_f^2$  for that parameter into Equations 7 and 8. See also Equations 19 and 20, and see Appendix B for a derivation.

$$\hat{r}_g = \frac{\hat{\sigma}_{Abs}}{\hat{\sigma}_{Ab} \hat{\sigma}_{As}} \tag{11}$$

was approximated by using a Taylor-series expansion (Appendix A), yielding

$$SE(\hat{r}_g) \approx \sqrt{\frac{Var(\hat{\sigma}_{Abs})}{\sigma_{Ab}^2 \sigma_{As}^2} + \frac{\sigma_{Abs}^2 Var(\hat{\sigma}_{Ab} \hat{\sigma}_{As})}{\sigma_{Ab}^4 \sigma_{As}^4}}, \tag{12}$$

where  $Var(\hat{\sigma}_{Abs})$  is given by the square of Equation 9. Equation 12 ignores covariances among the components of  $\hat{r}_g$ . The  $Var(\hat{\sigma}_{Ab} \hat{\sigma}_{As})$ , required for Equation 12, was calculated as (Appendix A)

$$Var(\hat{\sigma}_{Ab} \hat{\sigma}_{As}) \approx \left[ \frac{Var(\hat{\sigma}_{Ab}^2)}{4\sigma_{Ab}^2} + \sigma_{Ab}^2 \right] \left[ \frac{Var(\hat{\sigma}_{As}^2)}{4\sigma_{As}^2} + \sigma_{As}^2 \right] - \sigma_{Ab}^2 \sigma_{As}^2. \tag{13}$$

The values of  $Var(\hat{\sigma}_{Ab}^2)$  and  $Var(\hat{\sigma}_{As}^2)$ , required for Equation 13, follow from the square of Equation 7.

**Accuracy of predicted SEs:** To evaluate the precision of predicted SEs, experimental populations were stochastically simulated, variance components were estimated using ASReml (GILMOUR *et al.* 2006), and realized SEs of estimates were calculated from the variance among the estimates of 1000 replicates (Appendix C). Accuracies of predicted SEs were evaluated for a range of input parameters (Table 5). A basic scheme with 50 families of 100 members each was used as default. In the basic scheme, direct and indirect effects contributed equally to phenotypic variance, heritabilities of direct and indirect effects were 0.3, and total heritable variance was 60% of phenotypic variance ( $T^2 = 0.6$ ).

Results show that predicted SEs are accurate, both for groups composed at random (Table 6) and for groups composed of two families (Table 7), except for the genetic correlation between DGEs and IGEs. When the



**TABLE 4**  
**Components of SEs with group composed of two families**

Parameter	Expression	
All VC	$m = 2n/n_w$ $\sigma_P^2 = \sigma_{P_b}^2 + (n_w - 1)\sigma_{P_s}^2 + 2r(\frac{1}{2}n_w - 1)[\sigma_{A_{bs}} + (\frac{1}{2}n_w - 1)\sigma_{A_s}^2]$ $\text{Cov}_{w,\text{fam}} = 2\sigma_{P_{bs}} + (n_w - 2)\sigma_{P_s}^2 + r[\sigma_{A_b}^2 + 2(\frac{1}{2}n_w - 2)\sigma_{A_{bs}} + (\frac{1}{2}n_w - 2n_w + 3)\sigma_{A_s}^2]$ $\text{Cov}_{w,\text{nonfam}} = 2\sigma_{P_{bs}} + (n_w - 2)\sigma_{P_s}^2 + 2r(\frac{1}{2}n_w - 1)[\sigma_{A_{bs}} + (\frac{1}{2}n_w - 1)\sigma_{A_s}^2]$ $\text{Var}(\bar{P}_{\text{fam}}) = \frac{\sigma_P^2 + (\frac{1}{2}n_w - 1)\text{Cov}_{w,\text{fam}}}{\frac{1}{2}n_w}$	
$\hat{\sigma}_{A_b}^2$	$\sigma_z^2 = (1 + \phi^2)\text{Var}(\bar{P}_{\text{fam}}) - 2\phi\text{Cov}_{w,\text{nonfam}}$ $\phi = \frac{\frac{1}{2}n_w - 1}{\frac{1}{2}n_w}$	$\sigma_f^2 = r\sigma_{A_b}^2$
$\hat{\sigma}_{A_s}^2$	$\sigma_z^2 = \frac{4\text{Var}(\bar{P}_{\text{fam}})}{n_w^2}$	$\sigma_f^2 = r\sigma_{A_s}^2$
$\hat{\sigma}_{\text{TBV}}^2$	$\sigma_z^2 = \frac{4}{n_w} \left[ \sigma_P^2 + (\frac{1}{2}n_w - 1)\text{Cov}_{w,\text{fam}} + \frac{1}{2}n_w\text{Cov}_{w,\text{nonfam}} \right]$	$\sigma_f^2 = r\sigma_{\text{TBV}}^2$

SEs for the estimated genetic variance of interest follow from substituting  $\sigma_z^2$  and  $\sigma_f^2$  for that parameter into Equations 7 and 8. See also Equations 19 and 20, and see Appendix B for a derivation.

true  $r_g$  deviated considerably from zero (Alt. 1 through 5, and 7),  $\text{SE}(\hat{r}_g)$  was severely overpredicted. Errors in the predicted  $\text{SE}(\hat{r}_g)$  originated from correlations between the components of  $\hat{r}_g$  (Equation 11), which were ignored in the derivation (Appendix A). Hence, SEs of

$\hat{\sigma}_{A_{bs}}$ ,  $\hat{\sigma}_{A_b}^2$ , and  $\hat{\sigma}_{A_s}^2$  were predicted accurately, but  $\text{SE}(\hat{r}_g)$  was inaccurate due to correlations among  $\hat{\sigma}_{A_{bs}}$ ,  $\hat{\sigma}_{A_b}^2$ , and  $\hat{\sigma}_{A_s}^2$ . Moreover, the approximate SEs reported by AS-Reml (GILMOUR *et al.* 2006) were generally in good agreement with the empirical SEs from the simulations. In conclusion, prediction equations presented above are accurate, except for  $\text{SE}(\hat{r}_g)$  in cases the true  $r_g$  deviates considerably from zero.

**TABLE 5**

**Schemes used for validation of prediction equations**

Scheme	Deviation from basic scheme <sup>a</sup>
Alt. 1	$r_g = r_E = 0.7$
Alt. 2	$r_g = r_E = -0.7$
Alt. 3	$r_g = 0.7, r_E = -0.7$
Alt. 4	$r_g = -0.7, r_E = 0.7$
Alt. 5	$r_g = 0.7$
Alt. 6	$r_E = 0.7$
Alt. 7	$r_g = -0.7$
Alt. 8	$r_E = -0.7$
Alt. 9	$n_w = 2, n_g = 2,500, \sigma_{P_s}^2 = 1, T = 2450$
Alt. 10	$n_w = 10, n_g = 500, \sigma_{P_s}^2 = \frac{1}{9}$
Alt. 11	$h_S^2 = 0.1$
Alt. 12	$h_S^2 = 0.5$
Alt. 13	$\sigma_{P_s}^2 = \frac{1}{9}$
Alt. 14	$\sigma_{P_s}^2 = 1$
Alt. 15	$N = 20, n = 250$
Alt. 16	$N = 100, n = 50$
Alt. 17	$n = 40, T = 2,000$
Alt. 18	$n = 200, T = 10,000$
Alt. 19	$N = 100, n = 10, T = 1,000$
Alt. 20	$N = 200, n = 10, T = 2,000$

<sup>a</sup>Basic scheme:  $N = 50$  families,  $n = 100$  members per family, so that  $T = 5000$ ;  $n_w = 4$  individuals per group, so that  $n_g = 1250$  groups;  $\sigma_{P_b}^2 = 1, \sigma_{P_s}^2 = \frac{1}{3}, h_D^2 = \sigma_{A_b}^2/\sigma_{P_b}^2 = 0.3, h_S^2 = \sigma_{A_s}^2/\sigma_{P_s}^2 = 0.3, r_g = r_E = 0$ . For schemes with groups composed of two families,  $T$  was adjusted when needed to obtain an integer number of blocks.

**OPTIMUM DESIGNS**

**Optimum family sizes for  $\hat{\sigma}_{A_b}^2, \hat{\sigma}_{A_s}^2, \hat{\sigma}_{\text{TBV}}^2$ , and  $\tau^2$ :** ROBERTSON (1959) showed that the optimum family size for estimating an intraclass correlation,  $t = \sigma_f^2/\sigma_z^2$ , equals

$$n_{\text{opt},t} \approx 1/t. \tag{14}$$

Since (classical) heritability follows directly from the intraclass correlation,  $h^2 = t/r$ ,  $n_{\text{opt},t}$  also represents the optimum family size for estimating heritability. With respect to optimum family size, IGEs introduce two complications. First, with IGEs there are multiple genetic parameters, each of which may be of interest, and optimum family size may differ among parameters. Second, the parameters  $h_S^2 = \sigma_{A_s}^2/\sigma_P^2$  and  $\tau^2 = \sigma_{\text{TBV}}^2/\sigma_P^2$  are not true intraclass correlations, because the denominators of those ratios differ from the  $\sigma_z^2$  used in the ANOVA for those parameters (Tables 3 and 4). As a consequence, ROBERTSON'S  $n_{\text{opt},t}$  cannot be applied directly to  $h_S^2$  and  $\tau^2$ .

Approximate expressions for optimum family sizes for  $\hat{\sigma}_{A_b}^2, \hat{\sigma}_{A_s}^2$ , and  $\hat{\sigma}_{\text{TBV}}^2$  can be obtained as follows. The  $\text{sdev}(\hat{t})$  originates mainly from uncertainty in  $\hat{\sigma}_f^2$  rather than  $\hat{\sigma}_z^2$ . Thus the optimum family size for estimating  $\sigma_f^2$

TABLE 6

Precision of predicted SEs for groups composed at random

Scheme	Error%					
	$\hat{\sigma}_{Ab}^2$	$\hat{\sigma}_{Abs}$	$\hat{\sigma}_{As}^2$	$\hat{\sigma}_{TBV}^2$	$\hat{T}^2$	$\hat{r}_g$
Basic	2	0	-1	1	5	1
Alt. 1	1	-11	-3	-2	3	146
Alt. 2	3	3	1	1	2	129
Alt. 3	4	0	0	2	9	121
Alt. 4	0	8	-1	-1	-1	47
Alt. 5	3	-6	-1	0	7	134
Alt. 6	1	-4	-1	-2	1	-5
Alt. 7	2	7	-1	0	1	85
Alt. 8	3	1	0	2	6	2
Alt. 9	-1	2	2	3	5	-2
Alt. 10	2	-1	0	3	5	-10
Alt. 11	2	0	-3	0	2	-7
Alt. 12	2	1	0	1	7	2
Alt. 13	2	0	-1	-1	2	-1
Alt. 14	2	2	0	2	7	0
Alt. 15	-2	1	-1	-2	3	0
Alt. 16	2	-3	4	2	5	-6
Alt. 17	-1	-1	4	2	5	-10
Alt. 18	3	-4	1	0	3	-5
Alt. 19	1	-1	4	4	6	-38
Alt. 20	0	-6	-1	1	2	-33

See Table 5 for description of schemes. Error% = 100% × (predicted – empirical)/empirical. Empirical SEs were based on stochastic simulations with 1000 replicates. With 1000 replicates, the relative SE of the empirical SE equals SE( $\hat{\sigma}$ )/ $\sigma \approx 1/\sqrt{2000} \approx 0.022$  (STUART and ORD 1994). Therefore, when predictions equal the true values, E[error %] = 0, and the SE of error% given the predicted value equals ~2.2%. Then the expected absolute error equals E[|error %|]  $\approx$  1.8%, and |error %| > 4.4% indicates significant bias ( $P < 0.05$ ; two-sided, not accounting for multiple testing).

is close to the optimum family size for estimating  $t$ . Hence, from Equation 14, the approximate optimum number of records per family for estimating a genetic variance equals  $m_{opt,\sigma^2} \approx \sigma_z^2/\sigma_f^2$ . Thus, for schemes with groups composed at random, where  $n = m$  (Table 3), approximate optimum family size equals

$$n_{opt,\sigma^2,random} \approx \sigma_z^2/\sigma_f^2, \tag{15}$$

and for schemes with groups composed of two families, where  $n = \frac{1}{2}n_w m$  (Table 4), approximate optimum family size equals

$$n_{opt,\sigma^2,2fam} \approx \frac{1}{2}n_w \sigma_z^2/\sigma_f^2. \tag{16}$$

These results can be applied for the parameter of interest by using the appropriate  $\sigma_z^2$  and  $\sigma_f^2$  from Table 3 or 4. Finally, under the assumptions given above, optimum family size for  $\tau^2$  is approximately equal to that for  $\sigma_{TBV}^2$ ,

$$n_{opt,\tau^2} \approx n_{opt,\sigma_{TBV}^2}. \tag{17}$$

To evaluate precision of Equations 15 and 16, optimum family sizes were obtained numerically using

TABLE 7

Precision of predicted SEs for groups composed of two families

Scheme	Error%					
	$\hat{\sigma}_{Ab}^2$	$\hat{\sigma}_{Abs}$	$\hat{\sigma}_{As}^2$	$\hat{\sigma}_{TBV}^2$	$\hat{T}^2$	$\hat{r}_g$
Basic	3	2	-2	-3	0	2
Alt. 1	1	-5	-2	-1	5	147
Alt. 2	4	-1	1	-3	-1	140
Alt. 3	1	2	-1	-1	5	101
Alt. 4	5	7	1	-3	-2	95
Alt. 5	2	-2	-2	-1	5	121
Alt. 6	3	1	-2	-3	1	1
Alt. 7	4	3	1	-5	-3	117
Alt. 8	3	2	-1	-3	1	2
Alt. 9	-1	2	4	0	4	-1
Alt. 10	0	1	6	3	6	1
Alt. 11	2	3	-1	-2	-1	-1
Alt. 12	4	1	-2	-3	3	3
Alt. 13	2	2	1	-2	1	3
Alt. 14	5	1	-3	-4	1	1
Alt. 15	3	7	-1	0	4	4
Alt. 16	2	2	3	2	5	2
Alt. 17	0	1	4	7	10	-3
Alt. 18	0	6	-1	3	6	7
Alt. 19	8	-1	1	-4	-5	-44
Alt. 20	8	6	4	2	1	-36

See Table 5 for description of schemes. Error% = 100% × (predicted – empirical)/empirical. Empirical SEs were based on stochastic simulations with 1000 replicates. With 1000 replicates, the relative SE of the empirical SE equals SE( $\hat{\sigma}$ )/ $\sigma \approx 1/\sqrt{2000} \approx 0.022$  (STUART and ORD 1994). Therefore, when predictions equals the true values, E[error %] = 0, and the SE of error% given the predicted value equals ~2.2%. Then the expected absolute error equals E[|error %|]  $\approx$  1.8%, and |error %| > 4.4% indicates significant bias ( $P < 0.05$ ; two-sided, not accounting for multiple testing).

Equation 7 by varying  $n$  from 1 through  $T$ , and storing the family size yielding the lowest predicted SE encountered in this range. Results showed that Equations 15 and 16 are very accurate. For all schemes in Table 5, the relative difference between the numerically obtained minimum SE and the result from Equations 15 and 16 was smaller than 1% (results not shown).

An impression of optimum family sizes can be obtained by using estimates available in the literature. Using estimates of BERGSMA *et al.* (2008) for growth rate in fattening pigs with  $n_w \approx 8$ , optimum half-sib family sizes for groups composed at random are  $n_{opt,\sigma_{Ab}^2} = 19$  for DGE, and  $n_{opt,\sigma_{As}^2} = 200$  for IGE. For groups composed of two families, those values are  $n_{opt,\sigma_{Ab}^2} = 29$  and  $n_{opt,\sigma_{As}^2} = 69$ . Using estimates of ELLEN *et al.* (2008; line WB) for survival time in cannibalistic laying hens with  $n_w = 4$ , optimum half-sib family sizes are  $n_{opt,\sigma_{Ab}^2} = 42$  and  $n_{opt,\sigma_{As}^2} = 113$  for groups composed at random, and  $n_{opt,\sigma_{Ab}^2} = 52$  and  $n_{opt,\sigma_{As}^2} = 82$  for groups composed of two families. These values suggest that estimation of IGEs using groups composed at random may require rather large families. Moreover, the differ-

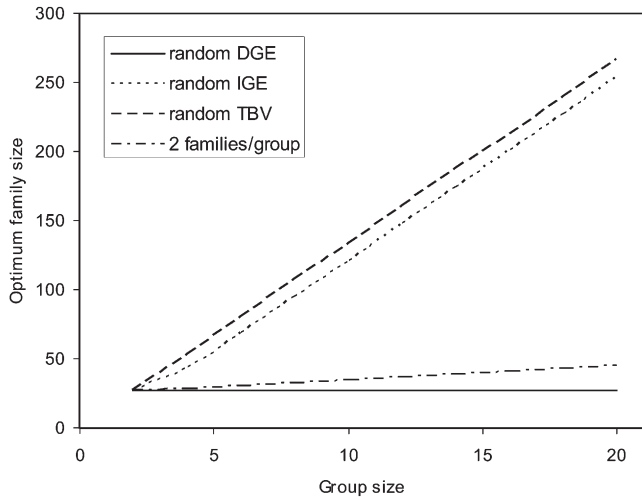


FIGURE 1.—Optimum half-sib family size for schemes with groups either composed at random or consisting of two families. A value of  $d = 0.5$  was used for the entire figure, and therefore, the phenotypic variance is the same for all group sizes. (See BIJMA 2010, the accompanying article in this issue, and section “Dilution of IGEs”.) With  $h_D^2 = h_S^2 = 0.3$ ,  $\sigma_{P_D}^2 = \sigma_{P_S, n_w=2}^2 = 1$ ,  $r_g = r_E = 0$ ,  $r = 0.25$ . For schemes with two families per group, optimum family size was identical for  $\sigma_{A_D}^2$ ,  $\sigma_{A_S}^2$ , and  $\sigma_{TBV}^2$ .

ence between optimum family size for DGEs *vs.* IGEs is much larger with groups composed at random than with groups composed of two families. This pattern is typical, as illustrated in Figure 1. Hence, when interest is in both DGEs and IGEs, the use of groups composed of two families is attractive, because optimum family sizes are more similar for both parameters.

OPTIMUM GROUP SIZE

The above equations shown that the SEs of estimated genetic parameters depend on group size. Relative performance of experimental designs may, therefore, depend on group size as well, and statistical power can be improved by optimizing group size.

**Dilution of IGEs:** The effect of group size on the SE of estimated genetic parameters is complicated by a potential effect of group size on the magnitude of IGEs (ARANGO *et al.* 2005; HADFIELD and WILSON 2007; BIJMA 2010). An individual’s IGE on a single recipient may become smaller in larger groups, because its total IGE is divided over more group mates. The degree of dilution, meaning the decrease of indirect effects in larger groups, will depend on the trait of interest. With competition for a finite amount of feed per group, for example, an individual consuming 1 kg has an average indirect effect of  $P_{S,i} = -1/(n_w - 1)$  on each of its group mates. Hence, the indirect effect is inversely proportional to the number of group mates, indicating full dilution. The other extreme of no dilution may be illustrated by a highly contagious infectious disease, where an individual may infect all its group mates, irrespective of group size. Here the indirect effect each

group mate receives is independent of group size, indicating no dilution.

Following BIJMA (2010), dilution of indirect effects was modeled as

$$P_{S,i,n_w} = \frac{1}{(n_w - 1)^d} P_{S,i,2}, \tag{18}$$

where  $P_{S,i,n_w}$  is the indirect effect of individual  $i$  in a group of  $n_w$  members,  $P_{S,i,2}$  is the indirect effect of  $i$  when it has a single group member (*i.e.*, when  $n_w = 2$ ), and  $d$  is the degree of dilution. With no dilution,  $d = 0$ , indirect effects do not depend on group size,  $P_{S,i,n_w} = P_{S,i,2}$ , as with the infectious disease. With full dilution,  $d = 1$ , indirect effects are inversely proportional to the number of group members,  $P_{S,i,n_w} = P_{S,i,2}/(n_w - 1)$ , as with the finite amount of feed. The magnitude of  $d$  can be estimated empirically (ARANGO *et al.* 2005; BIJMA 2010; CANARIO *et al.* 2010). Dilution affects the relationship of phenotypic variance with group size; for  $d < 0.5$  phenotypic variance increases with group size, whereas for  $d > 0.5$  phenotypic variance decreases with group size (BIJMA 2010). It is assumed here that IGEs and nongenetic indirect effects are diluted in the same manner as phenotypic indirect effects (*i.e.*, Equation 18 applies also to  $A_S$  and  $E_S$ ).

The following results illustrate the effect of group size on the SEs of estimated variance components, which will depend on the degree of dilution. Results were very similar for fixed and optimum family sizes and are presented for optimum family sizes only, first for groups composed at random, and subsequently for groups composed of two families.

**Groups composed at random:** Depending on the situation, the size of an experiment may be limited either by the total number of individuals or by the number of groups. Results are presented first for a fixed number of individuals and subsequently for a fixed number of groups.

*Fixed number of individuals (T):* With fixed  $T$ , the number of groups decreases as group size increases,  $n_g = T/n_w$ . The effect of group size on SEs differs between DGEs and IGEs. For DGEs, SEs increased with group size when  $d < 0.5$  and decreased with group size when  $d > 0.5$  (Figure 2A). This relationship originates from the impact of dilution on phenotypic variance, which increases with group size for  $d < 0.5$ .

For IGEs, SEs increased with group size, particularly with strong dilution (Figure 2C). When  $d = 0$ , SEs were identical for groups of 2 and 3 individuals. This relationship originates from two mechanisms. First, group size affects the variance of the record of interest,  $\sigma_z^2$  (Table 3), higher  $\sigma_z^2$  yielding higher SEs. With  $d = 0$ ,  $\sigma_z^2$  was minimal for groups of  $\sim 3$  individuals, explaining lowest SEs for those group sizes. Second, with dilution,  $\sigma_{A_k}^2$  decreases with group size, which decreases the intraclass correlation (Equation 25b) and therefore increases SEs. Hence, with strong dilution, smaller groups are considerably better. Results for total genetic variance were similar to

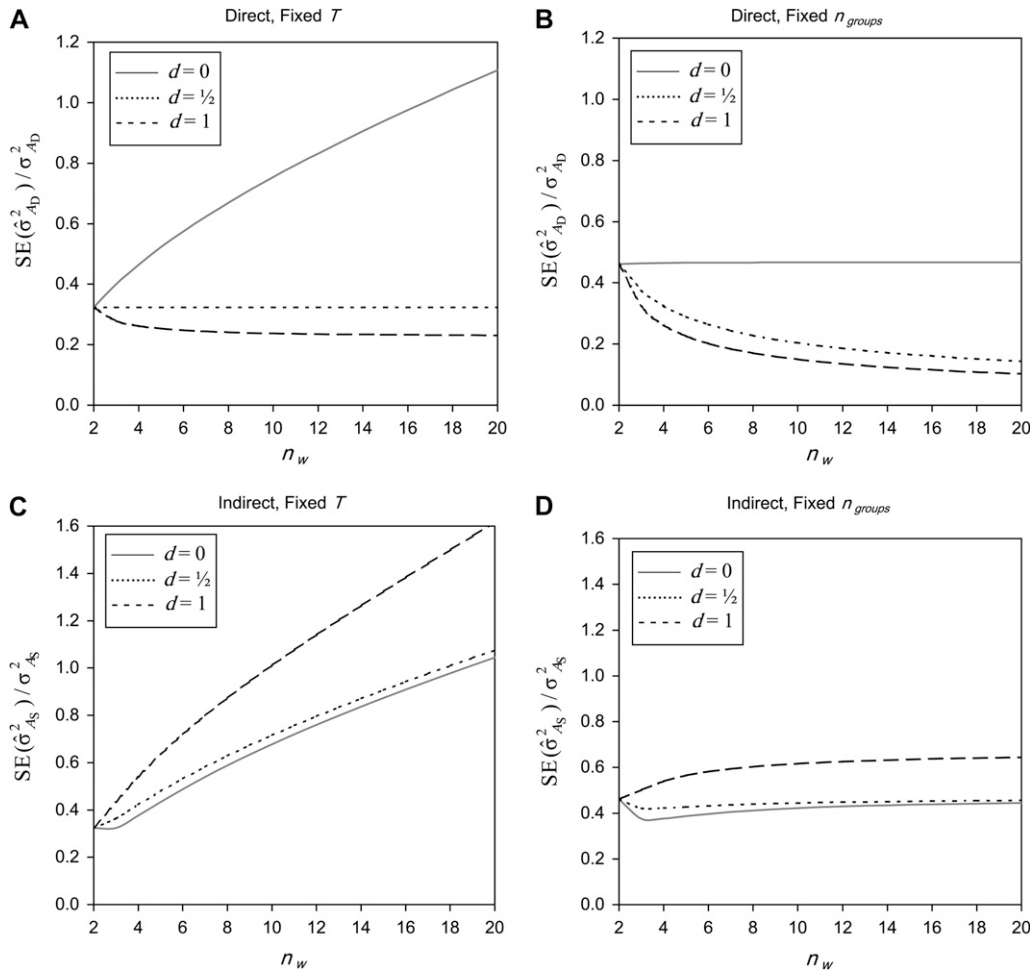


FIGURE 2.—Relationship between SEs and group size, for groups composed at random. (A) Direct genetic variance,  $T = 2000$ . (B) Direct genetic variance,  $n_{\text{groups}} = 500$ . (C) Indirect genetic variance,  $T = 2000$ . (D) Indirect genetic variance,  $n_{\text{groups}} = 500$ . (A–D)  $h_D^2 = h_S^2 = 0.3$ ,  $\sigma_{Pb}^2 = \sigma_{Pb,2}^2 = 1$ .  $r_g = r_E = 0$ ,  $r = 0.25$  (half-sibs). Family size is optimum (Equation 15). Because  $\sigma_{A_s}^2$  varies with group size when  $d \neq 0$ , results focus on the relative SE,  $\text{SE}(\hat{\sigma}_{A_s}^2)/\sigma_{A_s}^2$ .

those for indirect genetic variance, smaller groups yielding lower SEs (results not shown).

**Fixed number of groups ( $n_g$ ):** With a fixed number of groups, the total number of individuals increases as group size increases,  $T = n_g n_w$ . The effect of group size on SEs differs between DGEs and IGEs. For direct effects, relative SEs decreased with group size, except for  $d = 0$ , in which case SEs were nearly independent of group size (Figure 2B). This relationship originates from the increase of  $T$  with group size, which is counterbalanced by an increase of  $\sigma_p^2$  when  $d = 0$ . Hence, for  $d > 0$ ,  $\text{SE}(\hat{\sigma}_{Ab}^2)$  can be decreased substantially by using more individuals.

For IGEs, SEs were lowest for groups of two or three individuals (Figure 2D), and increased slightly with larger group sizes. For  $d = 0$ , groups of three individuals were clearly better than groups of two individuals. This relationship originates from an increase in phenotypic variance and dilution of IGEs, which counterbalances the benefit of greater  $T$ . Results for total genetic variance were similar to those for indirect genetic variance.

In conclusion, to estimate the variance due to IGEs while using groups composed at random, it is optimal to have small groups of two or three individuals. This result applies irrespective of whether the number of individ-

uals or the number of groups is the limiting factor in the experimental design.

**Groups composed of two families:** Results are presented first for a fixed number of individuals and subsequently for a fixed number of groups.

**Fixed number of individuals ( $T$ ):** For direct effects, SEs mostly increased with group size, particularly for  $d = 0$  (Figure 3A). This result originates from an increase in  $\sigma_z^2$  with group size, which reduces the intraclass correlation and thus increases the SE of  $\hat{\sigma}_{Ab}^2$  in larger groups. Also for IGEs, SEs increased with group size, particularly for  $d = 1$  (Figure 3C). This relationship originates from two mechanisms. First, with dilution ( $d > 0$ ),  $\sigma_{A_s}^2$  decreases with group size, which increases the relative SE. Hence, stronger dilution favors smaller groups. Second, when groups are composed of two families, an increase in group size tends to increase the intraclass correlation, particularly when  $\sigma_{A_s}^2$  is small, which favors larger groups. Compared to groups composed at random, the second mechanism limits the increase of SEs with group size (Figure 3C *vs.* Figure 2C). Results for total genetic variance were similar to those for indirect genetic variance (not shown).

**Fixed number of groups ( $n_g$ ):** For direct effects, relative SEs decreased with group size, particularly for  $d = 1$



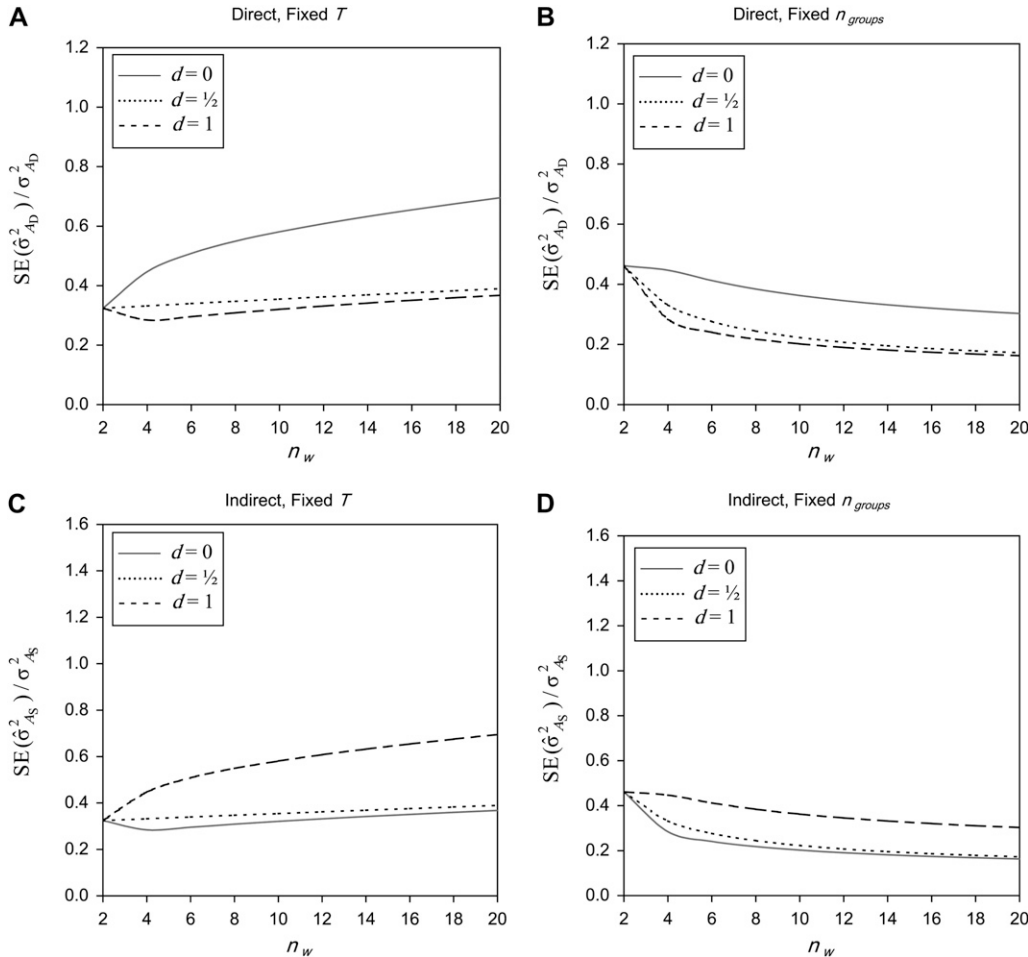


FIGURE 3.—Relationship between SEs and group size, for groups composed of two families. (A) Direct genetic variance,  $T = 2000$ . (B) Direct genetic variance,  $n_{groups} = 500$ . (C) Indirect genetic variance,  $T = 2000$ . (D) Indirect genetic variance,  $n_{groups} = 500$ . (A–D)  $h_D^2 = h_S^2 = 0.3$ ,  $\sigma_{P_D}^2 = \sigma_{P_{S,2}}^2 = 1$ .  $r_G = r_E = 0$ ,  $r = 0.25$  (half-sibs). Family size is optimum (Equation 16). Because  $\sigma_{A_S}^2$  varies with group size when  $d \neq 0$ , results focus on the relative SE,  $SE(\hat{\sigma}_{A_S}^2) / \sigma_{A_S}^2$ .

(Figure 3B). This relationship originates mainly from the increase of  $T$  with group size. Also for IGEs, relative SEs decreased with group size (Figure 3D). Benefits of increasing group size were largest for small  $d$  and small  $\sigma_{A_S}^2$ . For example, with  $d = 0$  and  $\sigma_{P_{S,2}}^2 = 0.2$ , the  $SE(\hat{\sigma}_{A_S}^2) / \sigma_{A_S}^2$  decreased from 0.51 for  $n_w = 2$  to 0.10 for  $n_w = 20$  (results not shown in Figure 3). Hence, when IGEs are expected to be small and the number of groups is the limiting factor in the experimental design, then the use of large groups, each composed of two families, is a powerful scheme.

Figures 2 and 3 show that the impact of dilution is opposite for DGEs *vs.* IGEs. For direct effects, SEs are lowest with full dilution, whereas for IGEs, SEs are lowest with no dilution. Moreover, Figure 2, A and B, *vs.* Figure 2, C and D, shows that, for random groups and  $d > 0.5$ , there is a conflict between optimum group sizes for direct genetic variance, which are large, and optimum group sizes for indirect genetic variance, which are small. This conflict is largely absent when groups are composed of two families, particularly when the number of groups is limiting (Figure 3, A and B, *vs.* Figure 3, C and D).

In conclusion, when groups are composed of two families, optimum group size to estimate the variance due to IGEs depends on the factor limiting the size of the experiment; small groups tend to be optimal when the

number of individuals is limiting, whereas large groups tend to be optimal when the number of groups is limiting.

**Group composition:** To compare both experimental designs, predicted SEs were compared between schemes with groups composed at random and schemes with groups composed of two families, for all schemes listed in Table 3 (results not shown). Differences between both designs were small for DGEs. For IGEs, schemes with groups composed of two families were superior, particularly when groups were large (Table 5, Alt. 10). Predicted SEs were identical for both designs when groups consisted of two individuals.

Figure 4, A and B, shows that schemes with groups composed of two families become increasingly superior when group size increases, superiority being largest for  $d = 0$ . Figure 4, C and D, shows the relative performance of both schemes when compared at optimum group sizes, as a function of  $d$ . When the total number of individuals is limiting, superiority of two-family schemes is limited and only present for  $d < 0.5$  (Figure 4C). When the number of groups is limiting, superiority of two-family schemes is substantial (Figure 4D). In that case, optimum group size for two-family schemes is large ( $n_w$  was limited to a maximum of 20 in Figure 4), whereas optimum group size for random schemes is small, ranging from two to four. In conclusion, if number of

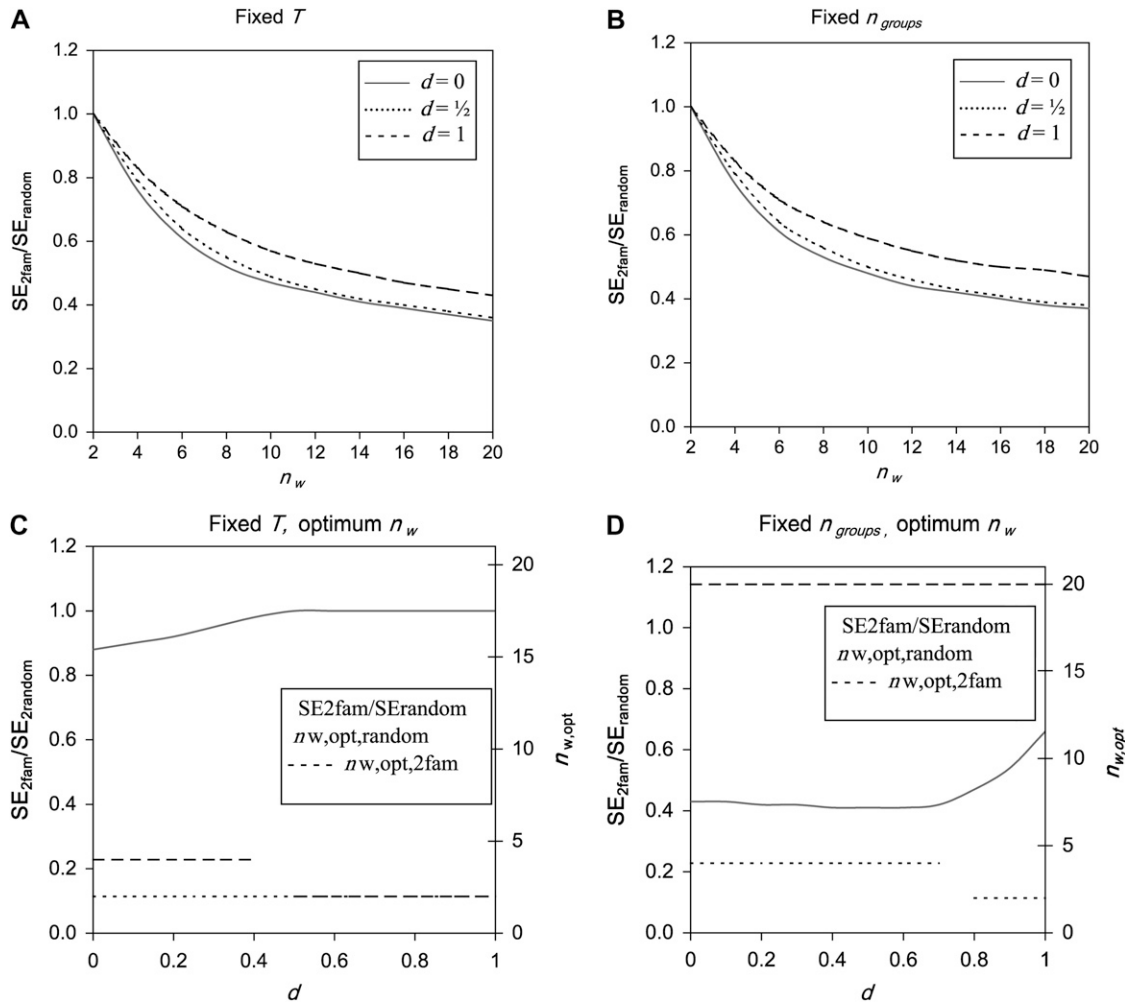


FIGURE 4.—Ratio of  $SE(\hat{\sigma}_A^2)$  for schemes with two families per group over that for schemes with random groups. (A) For fixed  $T = 2000$ . (B) For fixed  $n_{groups} = 500$ . (C) For fixed  $T = 2000$  and optimum group size. (D) For fixed  $n_{groups} = 500$  and optimum group size. (A–D)  $h_D^2 = h_S^2 = 0.3$ ,  $\sigma_{P_D}^2 = \sigma_{P_{S,2}}^2 = 1$ ,  $r_g = r_E = 0$ ,  $r = 0.25$  (half-sibs). Family size is optimum (Equations 15 and 16).

groups is the limiting factor in the experiment and it is possible to use groups of more than two members, then there is substantial benefit of putting two families in each group rather than using groups composed at random.

## DISCUSSION

IGEs attract increasing attention, as illustrated by the increasing number of publications (*e.g.*, BLEAKLEY and BRODIE 2009; WILSON *et al.* 2009; CHENOWETH *et al.* 2010; HSU *et al.* 2010; MCGLOTHLIN *et al.* 2010; MUIR *et al.* 2010). This development is warranted, given the large impact of IGEs on both the direction and magnitude of response to genetic selection and the increasing empirical evidence of their existence (see Introduction). Thus IGEs exist and cannot be ignored. Accurate knowledge of the genetic parameters of IGEs and of their relationship to DGEs is required to quantify the consequences for response to selection, in natural, agricultural, and laboratory populations.

This work has provided the tools to quantify the required size of experiments aiming to identify IGEs and to optimize experimental designs. The expressions for SEs of estimated genetic parameters are accurate for the variance of DGEs and IGEs, for their covariance, and for the total genetic variance. The expression for the SE of the estimated genetic correlation between DGEs and IGEs requires further improvement. An expression for optimum family size was provided as well, and optimum group size was investigated numerically. For IGEs, schemes with groups composed of two families were clearly better than schemes with groups composed at random. The R-package SE.IGE was developed, which predicts SEs and identifies optimum family and group sizes for each parameter.

Results suggest that, when optimum designs are implemented, presence of IGEs can be demonstrated successfully when the number of individuals is at least  $\sim 1000$ – $2000$ , or the number of groups is at least  $\sim 250$ – $500$  (Figures 2 and 3; obviously, this will depend on the true parameters, which are unknown *a priori*). For

example, when using the estimates of BERGSMAN *et al.* (2008), the *P*-value for the variance of IGEs is just below 5% when  $T = 2000$  in a half-sib design with two families per group and groups of eight members. Using the estimates of ELLEN *et al.* (2008; line WB),  $T = 2000$ , and in groups of four members, the corresponding *P*-value is below 5% for a half-sib design and below 1% for a full-sib design. Those experimental sizes are within the feasible range for many species in agriculture and aquaculture and for the smaller laboratory species. This work has considered a simple population structure, where relatedness is the same for all families and members of different families are unrelated. Extension to more complex family structures, such as full-sib families within half-sib families, would be useful.

This work has compared two experimental designs, showing that schemes with groups composed of two families yield lower SEs for the variance of IGEs than schemes with groups composed at random. The mechanism is as follows. With  $\frac{1}{2}n_w$  family members in a group, the phenotypes of the group mates belonging to the other family contain  $\frac{1}{2}n_w$  IGEs from the focal family. With a random group composition, in contrast, the phenotypes of the group mates contain only a single IGE from the focal family. Hence, the two-family scheme has the effect of increasing the family component in the phenotype, which increases the intraclass correlation. This increase in intraclass correlation can be observed in the expression for  $\sigma_z^2$ , which has a factor  $n_w^2$  in the denominator (Table 4). Since larger effects can be estimated more precisely, this reduces the  $SE(\hat{\sigma}_{A_k}^2)$ . This effect is counteracted by a difference in the effective number of records per family, which equals  $m = n$  for random schemes but  $m = n/(\frac{1}{2}n_w)$  for two-family schemes (Tables 3 and 4). Hence, moving from random schemes to two-family schemes increases the relative size of the effect by a factor of  $\frac{1}{2}n_w$ , but reduces the effective number of records by the same factor. Investigation of Equation 7, however, reveals a positive net effect.

Although a formal proof of optimality of the two-family design is difficult, the above suggests that this design is probably optimal or near optimal. Consider the full range of schemes with  $n_f$  families per group, each family contributing  $n_m$  members, so that  $n_w = n_f n_m$ . The extremes of this range are given by taking all members from different families,  $n_f = n_w$  and  $n_m = 1$ , which is near identical to the random scheme, and by the two-family scheme,  $n_f = 2$  and  $n_m = \frac{1}{2}n_w$ . (With a single family per group,  $n_f = 1$  and  $n_m = n_w$ , genetic parameters of DGEs and IGEs are statistically nonidentifiable.) Within this range, the size of the family effect in the phenotypes of the group mates of that family equals  $n_m A_{S,k}$ , and the number of effective records per family equals  $m = n/n_m$ . Investigation of Equation 7 shows that increasing the size of the effect while reducing the effective number of records by the same factor reduces the SE of  $\hat{\sigma}_f^2$  (unless  $m$  is very small). Hence, this suggests that

using the highest possible number of family members per group is optimal, which is the two-family scheme.

In any power calculation, predicted SEs depend on the true values of the parameters to be estimated. Hence, application of prediction equations presented here requires prior assumptions on those true values. Such prior assumptions may be obtained from knowledge of traits underlying the indirect effects, *e.g.*, from the intensity of behavioral interactions among individuals and the heritability of such behaviors or from existing estimates of genetic parameters taken from the literature. The expressions for SEs presented here are formulated in terms of the true parameters for total direct and indirect effect,  $\sigma_{P_D}^2$ ,  $\sigma_{P_{DS}}^2$ , and  $\sigma_{P_S}^2$ , and the corresponding genetic parameters,  $\sigma_{A_D}^2$ ,  $\sigma_{A_{DS}}^2$ , and  $\sigma_{A_S}^2$ . However, while the genetic parameters are statistically identifiable, the parameters for total direct and indirect effect are not. The reason is that the nongenetic parameters  $\sigma_{E_D}^2$ ,  $\sigma_{E_{DS}}^2$ , and  $\sigma_{E_S}^2$  are nonidentifiable, because there exist multiple combinations of those parameters yielding the same nongenetic variances between and within groups. The between-group variance equals  $\sigma_g^2 = 2\sigma_{E_{DS}}^2 + (n - 2)\sigma_{E_S}^2$ , whereas the within-group variance equals  $\sigma_e^2 = \sigma_{E_D}^2 - 2\sigma_{E_{DS}}^2 + \sigma_{E_S}^2$ . Hence, the literature will include estimates of  $\sigma_g^2$  and  $\sigma_e^2$ , rather than  $\sigma_{P_D}^2$ ,  $\sigma_{P_{DS}}^2$ , and  $\sigma_{P_S}^2$  (*e.g.*, Appendix C). This issue is resolved by reformulating the expressions in Tables 3 and 4 in terms of  $\sigma_g^2$  and  $\sigma_e^2$ , using

$$\sigma_{P_D}^2 + (n_w - 1)\sigma_{P_S}^2 = \sigma_{A_D}^2 + (n_w - 1)\sigma_{A_S}^2 + \sigma_g^2 + \sigma_e^2, \tag{19}$$

and

$$2\sigma_{P_{DS}}^2 + (n_w - 2)\sigma_{P_S}^2 = 2\sigma_{A_{DS}}^2 + (n_w - 2)\sigma_{A_S}^2 + \sigma_g^2. \tag{20}$$

Hence, expression presented here can be implemented using prior values taken from the literature, on the condition that authors present not only the estimated genetic parameters, but also the estimated between-group and residual variance.

I thank Seteng Alemu and Jørgen Ødegård for helpful discussion on this topic, and Albart Coster for assistance with R. The R-package SE.IGE can be downloaded from the repository of R-packages, CRAN <http://cran.r-project.org/package=SE.IGE>, following the usual method to install R packages. This research was financially supported by the Dutch science council (Nederlandse organisatie voor Wetenschappelijk Onderzoek) and was coordinated by the Netherlands Technology Foundation (Stichting Technische Wetenschappen).

*Note added in proof:* See P. Bijma (pp. 1029–1031) in this issue, for a related work.

#### LITERATURE CITED

- AGRAWAL, A. F., E. D. BRODIE, III and M. J. WADE, 2001 On indirect genetic effects in structured populations. *Am. Nat.* **158**: 308–323.  
 ARANGO, J., I. MISZTAL, S. TSURUTA, M. CUBERTSON and W. HERRING, 2005 Estimation of variance components including competitive effects of Large White growing gilts. *J. Anim. Sci.* **83**: 1241–1246.

- BERGSMAS, R., E. KANIS, E. F. KNOL and P. BIJMA, 2008 The contribution of social effects to heritable variation in finishing traits of domestic pigs (*Sus scrofa*). *Genetics* **178**: 1559–1570.
- BIJMA, P., 2010 Multilevel selection 4: modeling the relationship of indirect genetic effects and group size. *Genetics* **186**: 1029–1031.
- BIJMA, P., and M. J. WADE, 2008 The joint effects of kin, multilevel selection and indirect genetic effects on response to genetic selection. *J. Evol. Biol.* **21**: 1175–1188.
- BIJMA, P., W. M. MUIR and J. A. M. VAN ARENDONK, 2007a Multilevel selection 1: quantitative genetics of inheritance and response to selection. *Genetics* **175**: 277–288.
- BIJMA, P., W. M. MUIR, E. D. ELLEN, J. B. WOLF and J. A. M. VAN ARENDONK, 2007b Multilevel selection 2: estimating the genetic parameters determining inheritance and response to selection. *Genetics* **175**: 289–299.
- BLEAKLEY, B. H., and E. D. BRODIE, III, 2009 Indirect genetic effects influence antipredator behavior in guppies: estimates of the coefficient of interaction  $\psi$  and the inheritance of reciprocity. *Evolution* **63**: 1796–1806.
- CANARIO, L., N. LUNDEHEIM and P. BIJMA, 2010 Pig growth is affected by social genetic effects and social litter effects that depend on group size. Proceedings of the 9th World Congress on Genetics Applied to Livestock Production. German Society of Animal Science, Leipzig, Germany, p. 87.
- CANTET, R. J. C., and E. P. CAPPAS, 2008 On identifiability of (co)variance components in animal models with competition effects. *J. Anim. Breed. Genet.* **125**: 371–381.
- CAPPAS, E. P., and R. J. C. CANTET, 2008 Direct and competition additive effects in tree breeding: Bayesian estimation from an individual mixed tree model. *Silvae Genet.* **57**: 45–57.
- CHEN, C. Y., S. D. KACHMAN, R. K. JOHNSON, S. NEWMAN and L. D. VAN VLECK, 2008 Estimation of genetic parameters for average daily gain using models with competition effects. *J. Anim. Sci.* **86**: 2525–2530.
- CHENOWETH, S. F., H. D. RUNDLE and M. W. BLOWS, 2010 Experimental evidence for the evolution of indirect genetic effects: changes in the interaction effect coefficient,  $\Psi$ , due to sexual selection. *Evolution* **64**: 1849–1856.
- CHEVERUD, J. M., 1984 Evolution by kin selection: a quantitative genetic model illustrated by maternal performance in mice. *Evolution* **38**: 766–777.
- CHEVERUD, J. M., and A. J. MOORE, 1994 Quantitative genetics and the role of the environment provided by relatives in the evolution of behavior, pp. 60–100 in *Quantitative Genetic Studies of Behavioral Evolution*, edited by C. R. B. BOAKE. University of Chicago Press, Chicago.
- CLUTTON-BROCK, T. H., 2002 Breeding together: kin selection, reciprocity and mutualism in cooperative animal societies. *Nature* **373**: 209–216.
- DARWIN, C., 1859 *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, Ed. 1. John Murray, London.
- DICKERSON, G. E., 1947 Composition of hog carcasses as influenced by heritable differences in rate and economy of gain. *Iowa Agric. Exp. Stn. Res. Bull.* **354**: 492–524.
- ELLEN, E. D., J. VISSCHER, J. A. M. VAN ARENDONK and P. BIJMA, 2008 Survival of laying hens: genetic parameters for direct and associative effects in three purebred layer lines. *Poultry Sci.* **87**: 233–239.
- EMIK, L. O., and C. E. TERRIL, 1949 Systematic procedures for calculating inbreeding coefficients. *J. Hered.* **40**: 51–55.
- FALCONER, D. S., 1965 Maternal effects and selection response, pp. 763–774 in *Genetics Today, Proceedings of the XIth International Congress on Genetics*, Vol. 3, edited by S. J. GEERTS. Pergamon, Oxford.
- FALCONER, D. S., and T. F. C. MACKAY, 1996 *Introduction to Quantitative Genetics*. Pearson Education, Harlow, UK.
- FRANK, S. A., 1998 *Foundations of Social Evolution*. Princeton University Press, Princeton, NJ.
- GILMOUR, A. R., B. J. GOGEL, B. R. CULLIS and R. THOMPSON, 2006 *ASReml User Guide*, Release 2.0, VSN International, Hemel Hempstead, UK.
- 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**: 127–142.
- GRIFFING, B., 1976 Selection in reference to biological groups. VI. Analysis of full sib groups. *Genetics* **82**: 723–731.
- GRIFFING, B., 1981a A theory of natural-selection incorporating interaction among individuals. 1. The modeling process. *J. Theor. Biol.* **89**: 635–658.
- GRIFFING, B., 1981b A theory of natural-selection incorporating interaction among individuals. 2. Use of related groups. *J. Theor. Biol.* **89**: 659–677.
- HADFIELD, J. D., and A. J. WILSON, 2007 Multilevel selection 3: modeling the effects of interacting individuals as a function of group size. *Genetics* **177**: 667–668.
- HAMILTON, W. D., 1964 The genetical evolution of social behavior. *J. Theor. Biol.* **7**: 1–52.
- HENDERSON, C. R., 1976 A simple method for the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* **32**: 69–83.
- HSU, W. L., R. K. JOHNSON and L. D. VANVLECK, 2010 Effect of pen mates on growth, backfat depth, and longissimus muscle area of swine. *J. Anim. Sci.* **88**: 895–902.
- KELLER, L., 1999 *Levels of Selection in Evolution*. Princeton University Press, Princeton, NJ.
- KIRKPATRICK, M., and R. LANDE, 1989 The evolution of maternal characters. *Evolution* **43**: 485–503.
- KRUUK, L. E. B., 2004 Estimating genetic parameters in natural populations using the ‘animal model.’ *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**: 873–890.
- LYNCH, M., and B. WALSH, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.
- MCGLOTHLIN, J. W., and E. D. BRODIE, III, 2009 How to measure indirect genetic effects: the congruence of trait-based and variance-partitioning approaches. *Evolution* **63**: 1785–1795.
- MCGLOTHLIN, J. W., A. J. MOORE, J. B. WOLF and E. D. BRODIE, III, 2010 Interacting phenotypes and the evolutionary process. III. Social evolution. *Evolution* **64**: 2558–2574.
- MOORE, A. J., E. D. BRODIE, III and J. B. WOLF, 1997 Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. *Evolution* **51**: 1352–1362.
- MOUSSEAU, T. A., and C. W. FOX, eds., 1998 *Maternal effects as adaptations*. Oxford University Press, New York.
- MUIR, W. M., 1996 Group selection for adaptation to multiple-hen cages: selection program and direct responses. *Poult. Sci.* **75**: 447–458.
- MUIR, W. M., 2005 Incorporation of competitive effects in forest tree or animal breeding programs. *Genetics* **170**: 1247–1259.
- MUIR, W. M., and A. SCHINCKEL, 2002 Incorporation of competitive effects in breeding programs to improve productivity and animal well being. Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, communication no. 14–07.
- MUIR, W. M., P. BIJMA, C. Y. CHEN, I. MISZTAL and M. J. WADE, 2010 Advances in selection theory including experimental demonstrations: the theory of selection with social competition. Proceedings of the 9th World Congress on Genetics Applied to Livestock Production. German Society of Animal Science, Leipzig, Germany, p. 141.
- MUTIC, J. J., and J. B. WOLF, 2007 Indirect genetic effects from ecological interactions in *Arabidopsis thaliana*. *Mol. Ecol.* **16**: 2371–2381.
- PATTERSON, H. D., and R. THOMPSON, 1971 Recovery of interblock information when block sizes are unequal. *Biometrika* **58**: 545–554.
- PETFIELD, D., S. F. CHENOWETH, H. D. RUNDLE and M. W. BLOWS, 2005 Genetic variance in female condition predicts indirect genetic variance in male sexual display traits. *Proc. Natl. Acad. Sci. USA* **102**: 6045–6050.
- ROBERTSON, A., 1959 Experimental design in the evaluation of genetic parameters. *Biometrics* **15**: 219–226.
- STUART, A., and J. K. ORD, 1994 *Kendall's Advanced Theory of Statistics, Vol. 1, Distribution Theory*. Hodder Education, London.
- STUART, A., J. K. ORD and S. ARNOLD, 1999 *Kendall's Advanced Theory of Statistics, Vol. 2A, Classical Inference and the Linear Model*, Ed. 6. Arnold, London.
- VAN VLECK, L. D., and J. P. CASSADY, 2004 Modification of MTDFREML to estimate variance due to genetic competition effects. *J. Anim. Sci.* **82**(Suppl. 2): 38 (Abstr.).



VAN VLECK, L. D., and J. P. CASSADY, 2005 Unexpected estimates of variance components with a true model containing genetic competition effects. *J. Anim. Sci.* **83**: 68–74.  
 WILLHAM, R. L., 1963 The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics* **19**: 18–27.  
 WILSON, A. J., U. GELIN, M-C. PERRON and D. RÉALE, 2009 Indirect genetic effects and the evolution of aggression in a vertebrate system. *Proc. R. Soc. B* **276**: 533–541.

WILSON, E. O., 1975 *Sociobiology: The New Synthesis*. Belknap Press of Harvard University Press, Cambridge, MA.  
 WOLF, J. B., E. D. BRODIE, III, A. J. CHEVERUD, A. J. MOORE and M. J. WADE, 1998 Evolutionary consequences of indirect genetic effects. *Trends Ecol. Evol.* **13**: 64–69.

Communicating editor: M. W. FELDMAN

APPENDIX A

This appendix shows the derivations of Equations 7, 9, 12, and 13.

**Equation 7:** With one-way ANOVA, expected mean squares are (STUART *et al.* 1999)

$$EMS_f = m\sigma_f^2 + \sigma_c^2 \tag{A1a}$$

$$EMS_c = \sigma_c^2, \tag{A1b}$$

so that the estimated family variance equals

$$\hat{\sigma}_f^2 = (MS_f - MS_c)/m. \tag{A2}$$

The variance of  $\hat{\sigma}_f^2$  follows from the variances of the mean squares, using the general result that (STUART *et al.* 1999)

$$\text{Var}(MS) = \frac{2EMS^2}{df}. \tag{A3}$$

Combining Equations A1–A3 yields

$$\text{Var}(\hat{\sigma}_f^2) \approx \frac{2}{N-1} \left[ \sigma_f^4 + \frac{2\sigma_f^2\sigma_c^2}{m} + \frac{\sigma_c^4}{m(m-1)} \right]. \tag{A4}$$

The last term in this expression uses  $(T-1)/T \approx 1$ . Finally, using  $\hat{\sigma}_A^2 = \hat{\sigma}_f^2/r$ ,  $r$  denoting additive genetic relatedness among family members, yields Equation 7. The  $r$  equals 1/2 for full-sib families and 1/4 for half-sib families (FALCONER and MACKAY 1996).

**Equation 9:** The derivation of  $SE(\hat{\sigma}_{Abs})$  is based on an approximation of the variance of an estimated covariance. Consider  $df$  paired observations on two random variables, say  $x$  and  $y$ , following a bivariate normal distribution with zero mean,  $\text{Cov}(x_i, x_{j \neq i}) = \text{Cov}(y_i, y_{j \neq i}) = \text{Cov}(x_i, y_{j \neq i}) = 0$ , and  $\text{Cov}(x_i, y_i) = \text{Cov}(x, y) = \rho\sigma_x\sigma_y$ , where  $\rho$  denotes the correlation between  $x$  and  $y$  within a pair. (With respect to a covariance, the assumption of zero mean is equivalent to assuming the mean known. Hence, this derivation ignores that the mean has to be estimated from the sample.) Interest is in  $\text{Var}(\hat{\sigma}_{xy})$ . With

$$E(x) = E(y) = 0, \quad \hat{\sigma}_{xy} = \frac{1}{df} \sum_{df} xy,$$

so that  $\text{Var}(\hat{\sigma}_{xy}) = (1/df)\text{Var}(xy)$ . Next, we use  $\text{Var}(xy) = \text{Var}[E(xy|y)] + E[\text{Var}(xy|y)]$  (STUART and ORD 1994). Under normality,  $E(xy|y) = yE(x|y) = \rho(\sigma_x/\sigma_y)y^2$ . From the moments of the normal distribution (STUART and ORD 1994),  $\text{Var}(y^2) = 2\sigma_y^4$ , so that  $\text{Var}[E(xy|y)] = 2\rho^2\sigma_x^2\sigma_y^2$ . Moreover, under normality,  $\text{Var}(xy|y) = y^2\text{Var}(x|y) = (1 - \rho^2)\sigma_x^2y^2$ , so that  $E[\text{Var}(xy|y)] = (1 - \rho^2)\sigma_x^2\sigma_y^2$ . Combining both terms and substituting  $\rho^2\sigma_x^2\sigma_y^2 = \sigma_{xy}^2$  yields

$$\text{Var}(\hat{\sigma}_{xy}) \approx \frac{\sigma_x^2\sigma_y^2 + \sigma_{xy}^2}{df}. \tag{A5}$$

Application of this result to the direct-social genetic covariance requires knowledge of  $\sigma_x^2$ ,  $\sigma_y^2$ , and  $\sigma_{xy}$ . Because DGE and IGEs are not directly observed, one cannot simply substitute those parameters by the true genetic parameters. Rather, the  $\sigma_x^2$  and  $\sigma_y^2$  represent the “effective variance” of the “record” providing information on DGE and IGE, respectively. Those effective variances can be backcalculated from the SEs of the estimated variances of  $x$  and  $y$ . Using the general result that  $\text{Var}(\hat{\sigma}_x^2) = 2\sigma_x^4/df$ , the effective variances are  $\sigma_x^2 = SE(\hat{\sigma}_x^2)\sqrt{df/2}$ , and  $\sigma_y^2 = SE(\hat{\sigma}_y^2)\sqrt{df/2}$ . Moreover, the  $\sigma_{xy}$  in Equation A5 represents the covariance between the records providing information on the DGE and IGE, respectively. Because covariances are not affected by taking averages, the true value of  $\sigma_{xy}$  can be used. Substituting the above results into Equation A5 yields

$$\text{Var}(\hat{\sigma}_{xy}) \approx \frac{SE(\hat{\sigma}_x^2)SE(\hat{\sigma}_y^2)}{2} + \frac{\sigma_{xy}^2}{df}. \tag{A6}$$

Applying this result to DGEs and IGEs, substituting  $df = N-1$ , and taking the square root yields Equation 9.

**Equations 12:** The estimated genetic correlation equals

$$\hat{r}_g = \frac{\hat{\sigma}_{Abs}}{\hat{\sigma}_{Ab} \hat{\sigma}_{As}}.$$

This derivation ignores covariances among  $\hat{\sigma}_{Ab}^2$ ,  $\hat{\sigma}_{As}^2$ , and  $\hat{\sigma}_{Abs}$ . Then  $\text{Var}(\hat{r}_g)$  can be predicted using a first-

order Taylor-series approximation of a ratio, say  $x/y$ , giving  $x/y \approx \bar{x}/\bar{y} + \bar{y}^{-1}(x - \bar{x}) - \bar{x}\bar{y}^{-2}(y - \bar{y})$ . Assuming  $\text{Cov}(x, y) \approx 0$ ,

$$\text{Var}(x/y) \approx \sigma_x^2/\bar{y}^2 + \bar{x}^2\sigma_y^2/\bar{y}^4. \quad (\text{A7})$$

Substituting  $\bar{x} = \sigma_{A_{\text{obs}}}$ ,  $\sigma_x^2 = \text{Var}[\hat{\sigma}_{A_{\text{obs}}}]$ ,  $\bar{y} = \sigma_{A_b}\sigma_{A_s}$ , and  $\sigma_y^2 = \text{Var}[\hat{\sigma}_{A_b}\hat{\sigma}_{A_s}]$  and taking the square root yield Equation 12.

**Equation 13:** Application of Equation 12 requires  $\text{Var}[\hat{\sigma}_{A_b}\hat{\sigma}_{A_s}]$ , the variance of a product. Consider the product  $\sqrt{x}\sqrt{y}$ , where  $x$  corresponds to  $\hat{\sigma}_{A_b}^2$ , and  $y$  corresponds to  $\hat{\sigma}_{A_s}^2$ . Using  $\text{Var}(ab) = \text{Var}[E(ab|b)] + E[\text{Var}(ab|b)]$ , and assuming  $\text{Cov}(\sqrt{x}, \sqrt{y}) \approx 0$  and  $\sqrt{x} \approx \sqrt{\bar{x}}$ , yields  $\text{Var}[\sqrt{x}\sqrt{y}] \approx [\text{Var}(\sqrt{x}) + \bar{x}] [\text{Var}(\sqrt{y}) + \bar{y}] - \bar{x}\bar{y}$ . Next, a first-order Taylor-series approximation for the square root of a variable yields  $\sqrt{x} \approx \sqrt{\bar{x}} + (1/2\sqrt{\bar{x}})(x - \bar{x})$ , so that  $\text{Var}(\sqrt{x}) \approx \sigma_x^2/4\bar{x}$ . Analogously  $\text{Var}(\sqrt{y}) \approx \sigma_y^2/4\bar{y}$ . Substituting those results yields

$$\text{Var}[\sqrt{x}\sqrt{y}] \approx \left[ \frac{\sigma_x^2}{4\bar{x}} + \bar{x} \right] \left[ \frac{\sigma_y^2}{4\bar{y}} + \bar{y} \right] - \bar{x}\bar{y}. \quad (\text{A8})$$

Substituting  $\bar{x} = \sigma_{A_b}^2$ ,  $\sigma_x^2 = \text{Var}(\hat{\sigma}_{A_b}^2)$ ,  $\bar{y} = \sigma_{A_s}^2$ , and  $\sigma_y^2 = \text{Var}(\hat{\sigma}_{A_s}^2)$  yields Equation 13.

## APPENDIX B

This appendix shows the derivation of results presented in Tables 3 and 4.

**Table 3,  $\sigma_z^2$ , and  $\sigma_f^2$  for groups composed at random:** When groups are composed at random with respect to family, group mates are genetically independent. The effective number of records in this design equals the number of individuals per family (see derivations below), so that

$$m = n. \quad (\text{B1})$$

From Equation 1, phenotypic variance equals

$$\sigma_P^2 = \sigma_{P_b}^2 + (n_w - 1)\sigma_{P_s}^2, \quad (\text{B2})$$

and the covariance between phenotypes of group mates equals

$$\text{Cov}_w = 2\sigma_{P_{\text{DS}}} + (n_w - 2)\sigma_{P_s}^2. \quad (\text{B3})$$

*Direct genetic variance:* The DGE of a family is expressed in the phenotypes of the family members themselves. Hence, in terms of Equation 6, the record of interest for DGEs is the phenotypic value of the  $l$ th member of family  $k$ ,

$$z_{kl} = P_{kl}. \quad (\text{B4})$$

Thus the full- and between-family variances are

$$\sigma_z^2 = \sigma_P^2 \quad (\text{B5a})$$

$$\sigma_f^2 = r\sigma_{A_b}^2. \quad (\text{B5b})$$

*Indirect genetic variance:* The IGE of an individual is expressed in the phenotypes of its group mates. Hence, the group mates of the members of a family provide information on the IGE of that family. Thus the record of interest for IGEs is the mean phenotypic value of the  $n_w - 1$  group mates of the  $l$ th member of family  $k$ ,

$$z_{kl} = \frac{1}{n_w - 1} \sum_{j=1}^{n_w-1} P_j, \quad (\text{B6})$$

$j$  denoting the group mates of  $l$ . From Equation B6, the full variance equals

$$\sigma_z^2 = \frac{\sigma_P^2 + (n_w - 2)\text{Cov}_w}{n_w - 1}. \quad (\text{B7a})$$

The expected value of  $z_{kl}$ , given family  $k$ , equals the IGE of family  $k$ ,  $f_k = A_{S,k}$ . Thus the between-family variance equals

$$\sigma_f^2 = r\sigma_{A_s}^2. \quad (\text{B7b})$$

*Total heritable variance:* An individual's total heritable effect equals  $\text{TBV}_i = A_{D,i} + (n_w - 1)A_{S,i}$ , of which the  $A_{D,i}$  is expressed in its own phenotype, and the  $A_{S,i}$  in the phenotypes of each of its  $n_w - 1$  group mates. Hence, the summed phenotypes of the focal individual and its  $n_w - 1$  group mates expresses the TBV of the focal individual. Thus the record of interest for TBVs is the sum of phenotypic values of all  $n_w$  individuals belonging to the group of the  $l$ th member of family  $k$ ,

$$z_{kl} = \sum_{j=1}^{n_w} P_j, \quad (\text{B8})$$

$j$  denoting group members of  $l$ , including  $l$  itself. From Equation B8, the full variance equals

$$\sigma_z^2 = n_w[\sigma_P^2 + (n_w - 1)\text{Cov}_w]. \quad (\text{B9a})$$

The expected value of  $z_{kl}$ , given family  $k$ , equals the TBV of family  $k$ ,  $f_k = \text{TBV}_k$ . Thus the between-family variance equals

$$\sigma_f^2 = r\sigma_{\text{TBV}}^2. \quad (\text{B9b})$$

**Table 4,  $\sigma_z^2$ , and  $\sigma_f^2$  for groups composed of two families:** In this design, each group consists of mem-

bers of two families, each contributing  $\frac{1}{2}n_w$  individuals (Table 2). The effective number of records in this design equals the number of groups to which a family contributes (see derivations below), so that

$$m = 2n/n_w. \tag{B10}$$

Phenotypic variance depends on relatedness. Relatedness within group equals  $r$  between members of the same family and zero between members of different families. From

$$P_i = P_{D,i} + \sum_{j=1}^{n_w-1} P_{S,j},$$

phenotypic variance equals

$$\begin{aligned} \sigma_P^2 &= \sigma_{P_D}^2 + (n_w - 1)\sigma_{P_S}^2 + 2 \sum_{j=1}^{n_w-1} \text{Cov}(P_{D,i}, P_{S,j}) \\ &+ \sum_{j=1}^{n_w-1} \sum_{j'=1}^{n_w-2} \text{Cov}(P_{S,j}, P_{S,j'})_{j \neq j'}. \end{aligned}$$

Only covariances between family member are non-zero. Since  $i$  has  $\frac{1}{2}n_w - 1$  family members in the same group,

$$2 \sum_{j=1}^{n_w-1} \text{Cov}(P_{D,i}, P_{S,j}) = 2r\left(\frac{1}{2}n_w - 1\right)\sigma_{A_{DS}}.$$

In

$$\sum_{j=1}^{n_w-1} \sum_{j'=1}^{n_w-2} \text{Cov}(P_{S,j}, P_{S,j'})_{j \neq j'}$$

there are  $(\frac{1}{2}n_w - 1)$  members of the family of  $i$  together contributing  $(\frac{1}{2}n_w - 1)(\frac{1}{2}n_w - 2)r\sigma_{A_S}^2$ , and  $\frac{1}{2}n_w$  members of the other family together contributing  $\frac{1}{2}n_w(\frac{1}{2}n_w - 1)r\sigma_{A_S}^2$ , yielding a total of  $2r(\frac{1}{2}n_w - 1)^2\sigma_{A_S}^2$ . Collecting terms yields

$$\begin{aligned} \sigma_P^2 &= \sigma_{P_D}^2 + (n_w - 1)\sigma_{P_S}^2 + 2r\left(\frac{1}{2}n_w - 1\right) \\ &\left[\sigma_{A_{DS}} + \left(\frac{1}{2}n_w - 1\right)\sigma_{A_S}^2\right]. \end{aligned} \tag{B11}$$

The covariance between phenotypes of group mates belonging to the same family follows from

$$\text{Cov}\left(P_{D,i} + \sum_{j=1}^{n_w-1} P_{S,j}, P_{D,i'} + \sum_{j'=1}^{n_w-1} P_{S,j'}\right),$$

where  $i$  and  $i'$  belong to the same family, and  $j$  and  $j'$  denote the group mates of  $i$  and  $i'$ , respectively. This covariance can be split into a term due to unrelated individuals, which is the same as for groups composed at

random and equals  $2\sigma_{P_{DS}} + (n_w - 2)\sigma_{P_S}^2$ , and a remaining term due to relatedness. Carefully collecting terms shows that components due to relatedness are: (1)  $r\sigma_{A_D}^2$  from  $\text{Cov}(P_{D,i}, P_{D,i'})$ , (2)  $2r(\frac{1}{2}n_w - 2)\sigma_{A_{DS}}$  from the covariance between DGEs and IGEs among members of the same family, (3)  $(\frac{1}{4}n_w^2 - 1.5n_w + 3)r\sigma_{A_S}^2$  from IGEs of the focal family, and (4)  $(\frac{1}{4}n_w^2 - \frac{1}{2}n_w)r\sigma_{A_S}^2$  from IGEs of the other family. Collecting terms yields

$$\begin{aligned} \text{Cov}_{w,\text{fam}} &= 2\sigma_{P_{DS}} + (n_w - 2)\sigma_{P_S}^2 \\ &+ r\left[\sigma_{A_D}^2 + 2\left(\frac{1}{2}n_w - 2\right)\sigma_{A_{DS}} + \left(\frac{1}{2}n_w^2 - 2n_w + 3\right)\sigma_{A_S}^2\right]. \end{aligned} \tag{B12}$$

The covariance between phenotypes of group mates belonging to different families can be derived analogously. Components due to unrelated individuals contribute  $2\sigma_{P_{DS}} + (n_w - 2)\sigma_{P_S}^2$ . Components due to relatedness are  $2(\frac{1}{2}n_w - 1)r\sigma_{A_{DS}}$  from the covariances between DGEs and IGEs of family members and  $2(\frac{1}{2}n_w - 1)^2r\sigma_{A_S}^2$  from the covariances between IGEs of family members. Collecting terms yields

$$\begin{aligned} \text{Cov}_{w,\text{nonfam}} &= 2\sigma_{P_{DS}} + (n_w - 2)\sigma_{P_S}^2 \\ &+ 2r\left(\frac{1}{2}n_w - 1\right)\left[\sigma_{A_{DS}} + \left(\frac{1}{2}n_w - 1\right)\sigma_{A_S}^2\right]. \end{aligned} \tag{B13}$$

Moreover, the variance of the mean phenotype of the  $\frac{1}{2}n_w$  group mates belonging to the same family equals

$$\text{Var}(\bar{P}_{\text{fam}}) = \frac{\sigma_P^2 + (\frac{1}{2}n_w - 1)\text{Cov}_{w,\text{fam}}}{\frac{1}{2}n_w}. \tag{B14}$$

*Direct genetic variance:* With two families per group, the mean phenotype of the  $\frac{1}{2}n_w$  group mates of a family depends not only on the DGE of that family, but also on its IGE. Given family  $k$ , the expected mean of the  $\frac{1}{2}n_w$  members of family  $k$  in group  $l$  equals  $E[\bar{P}_{kl}|k] = A_{D,k} + (\frac{1}{2}n_w - 1)A_{S,k}$ . The  $(\frac{1}{2}n_w - 1)A_{S,k}$  arises because each family member has  $(\frac{1}{2}n_w - 1)$  group mates belonging to its own family. Moreover, given family  $k$ , the expected mean phenotype of the other family in the group, say  $k'$ , equals  $E[\bar{P}_{k'l}|k] = \frac{1}{2}n_w A_{S,k}$ . A linear combination of both expected means yields an estimate of the DGE of family  $k$ ,  $E[\bar{P}_{kl} - \phi \bar{P}_{k'l}] = A_{D,k}$ , where

$$\phi = \frac{\frac{1}{2}n_w - 1}{\frac{1}{2}n_w}. \tag{B15}$$

Thus the record of interest for DGEs is

$$z_{kl} = \bar{P}_{kl} - \phi \bar{P}_{k'l}, \tag{B16}$$

where  $\bar{P}_{kl}$  is the mean phenotypic value of the  $\frac{1}{2}n_w$  member of family  $k$  in the  $l$ th group to which family  $k$  contributes, and  $\bar{P}_{k'l}$  is the mean phenotypic value of the  $\frac{1}{2}n_w$  members of the other family in that group. Hence,  $l$  indexes the 1 through  $2n/n_w$  groups to which a family

contributes, so that the number of records equals  $m = 2n/n_w$ . From Equation B16, the full variance equals

$$\sigma_z^2 = (1 + \phi^2) \text{Var}(\bar{P}_{\text{fam}}) - 2\phi \text{Cov}_{w,\text{nonfam}}, \quad (\text{B17a})$$

and the between-family variance equals

$$\sigma_f^2 = r\sigma_{A_D}^2. \quad (\text{B17b})$$

*Indirect genetic variance:* As shown above, given family  $k$ , the expected mean phenotype of the other family in group  $l$  equals  $E[\bar{P}_{k'l}|k] = \frac{1}{2}n_w A_{S,k}$ , so that  $E[\bar{P}_{k'l}|k]/(\frac{1}{2}n_w) = A_{S,k}$ . Thus the record of interest for IGEs is

$$z_{kl} = \frac{\bar{P}_{k'l}}{\frac{1}{2}n_w}. \quad (\text{B18})$$

The full variance, therefore, equals

$$\sigma_z^2 = \frac{\text{Var}(\bar{P}_{\text{fam}})}{\frac{1}{4}n_w^2}, \quad (\text{B19a})$$

and the between-family variance equals

$$\sigma_f^2 = r\sigma_{A_S}^2. \quad (\text{B19b})$$

*Total heritable variance:* In any group, each of the  $\frac{1}{2}n_w$  members of family  $k$  receives a contribution  $A_{D,k} + (\frac{1}{2}n_w - 1)A_{S,k}$  to its phenotype. Moreover, each of the  $\frac{1}{2}n_w$  members of the other family,  $k'$ , receives a contribution  $\frac{1}{2}n_w A_{S,k}$ . Hence, given family  $k$ , the expected value of the sum of phenotypes of all  $n_w$  members of group  $l$  equals  $E[\sum_{j=1}^{n_w} P_{kl,j}|k] = \frac{1}{2}n_w[A_{D,k} + (n_w - 1)A_{S,k}] = \frac{1}{2}n_w \text{TBV}_k$ , where  $j$  denotes members of the  $l$ th group to which family  $k$  contributes. Thus the record of interest for the TBVs is

$$z_{kl} = \frac{1}{\frac{1}{2}n_w} \sum_{j=1}^{n_w} P_{kl,j}. \quad (\text{B20})$$

The full variance, therefore, equals

$$\sigma_z^2 = \frac{4}{n_w} \left[ \sigma_P^2 + \left( \frac{1}{2}n_w - 1 \right) \text{Cov}_{w,\text{fam}} + \frac{1}{2}n_w \text{Cov}_{w,\text{nonfam}} \right], \quad (\text{B21a})$$

and the between-family variance equals

$$\sigma_f^2 = r\sigma_{\text{TBV}}^2. \quad (\text{B21b})$$

### APPENDIX C

**Stochastic simulation and data analysis:** Simulated population structures are described in the main text. DGEs and IGEs were sampled from a bivariate normal distribution,

$$\begin{bmatrix} \mathbf{a}_D \\ \mathbf{a}_S \end{bmatrix} \sim N(\mathbf{0}, \mathbf{C} \otimes \mathbf{A}),$$

where  $\mathbf{a}_D$  and  $\mathbf{a}_S$  are vectors of DGEs and IGEs for all individuals, respectively,

$$\mathbf{C} = \begin{bmatrix} \sigma_{A_D}^2 & \sigma_{A_{DS}} \\ \sigma_{A_{DS}} & \sigma_{A_S}^2 \end{bmatrix},$$

and  $\mathbf{A}$  is the so-called numerator relationship matrix, containing coefficients of additive genetic relatedness between all individuals (EMIK and TERRIL 1949; HENDERSON 1976). Elements  $A_{ij}$  of  $\mathbf{A}$  equaled  $r$  when  $i$  and  $j$  were from the same family, and zero otherwise.

Variance components were estimated from simulated data using restricted maximum likelihood (REML) as implemented in ASReml (GILMOUR *et al.* 2006), with the model (ARANGO *et al.* 2005; MUIR 2005; BIJMA *et al.* 2007b)

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Z}_D \mathbf{a}_D + \mathbf{Z}_S \mathbf{a}_S + \mathbf{Z}_g \mathbf{g} + \mathbf{e}, \quad (\text{C1})$$

where  $\mathbf{y}$  is a vector of phenotypes of all individuals,  $\boldsymbol{\mu}$  the overall mean,  $\mathbf{Z}_D$  an incidence matrix linking phenotypes of individuals to their own DGE,  $\mathbf{Z}_S$  an incidence matrix linking phenotypes of individuals to the IGEs of their group mates, and  $\mathbf{Z}_g$  an incidence matrix linking phenotypes of individuals to their group. The  $\mathbf{g}$  is a vector of random group effects with  $\text{Var}(\mathbf{g}) = \mathbf{I}\sigma_g^2$ , and  $\mathbf{e}$  is a vector of residuals with  $\text{Var}(\mathbf{e}) = \mathbf{I}\sigma_e^2$ ,  $\mathbf{I}$  denoting an identity matrix. Results showed that variance components estimated using Equation C1 were unbiased. Fitting model C1 yields estimates of the genetic parameters of interest,  $\sigma_{A_D}^2$ ,  $\sigma_{A_{DS}}$ , and  $\sigma_{A_S}^2$ , and of the nongenetic parameters  $\sigma_g^2$  and  $\sigma_e^2$ . Variance components for nongenetic direct and indirect effects,  $\sigma_{E_D}^2$ ,  $\sigma_{E_{DS}}$ , and  $\sigma_{E_S}^2$  are not statistically identifiable, and therefore not fitted (see DISCUSSION).

The  $\mathbf{g}$  accounts for the covariance between group mates arising from nonheritable indirect effects,  $\sigma_g^2 = 2\sigma_{E_{DS}} + (n_w - 2)\sigma_{E_S}^2$  (BERGSMÄ *et al.* 2008). If  $\sigma_{E_{DS}}$  is negative and sufficiently large, the  $2\sigma_{E_{DS}} + (n_w - 2)\sigma_{E_S}^2$  can be a negative value, so that the nongenetic covariance between group mates is negative, and fitting a random group effect is inappropriate. In those cases, therefore, the random group effect was omitted from the model, and residuals were allowed to be correlated within group, using  $\text{Var}(\mathbf{e}) = \mathbf{R}\sigma_e^2$ , where  $R_{ii} = 1$ ,  $R_{ij} = \rho$  when  $i$  and  $j$  are in the same group, and  $R_{ij} = 0$  when  $i$  and  $j$  are in different groups. The  $\rho$  represents the correlation between residuals of group mates, which is estimated from the data (BIJMA *et al.* 2007b). Both models are equivalent when  $\rho > 0$  (BERGSMÄ *et al.* 2008). When  $\rho > 0$ , we fitted random group effects rather than correlated residuals, because it was computationally less demanding and showed better convergence.