Accommodation of genotype-environment covariance in a longitudinal twin design

In the classical twin study, genetic and environmental influences on a phenotype are usually estimated under the assumption that genotype-environment covariance (GE covariance) is absent. We explore possibilities to accommodate GE covariance in longitudinal data using the genetic simplex model. First, the genetic simplex model is presented, accompanied by a brief summary of results found in cognitive developmental studies. Second, GE covariance is specified via niche picking and sibling effects. Third, numerical and analytical identification is established, and the statistical power to detect GE covariance is examined. In a simplex model comprising four time points, GE covariance can be accommodated by introducing phenotype to environment cross-lagged pathways, either within or between twins. By using different parameter constraints within the genetic simplex, the extended models are numerically and analytically identified. The power to detect GE covariance is relatively low and therefore large sample sizes are needed.

Authors: Johanna M. de Kort* , Conor V. Dolan*,** and Dorret I. Boomsma**

*Department of Psychology, University of Amsterdam
**Department of Biological Psychology, Free University Amsterdam

Correspondence to:
Johanna M. de Kort, University of Amsterdam, Department of Psychology, Weesperplein 4, 1018 XA Amsterdam, the Netherlands, e-mail: jjmdekort@gmail.com

Quantitative genetics is concerned with determining the genetic and environmental influences on behavioural variance within a well-defined population. To determine which portion of the phenotypic variance is due to genetic and environmental influences, researchers often use the classical twin design (Figure 1), i.e., the comparison of monozygotic (MZ) and dizygotic (DZ) twins growing up together (Eaves, Last, Martin, & Jinks, 1977; Van Dongen, Slagboom, Draisma, Martin, & Boomsma, 2012). Using this design, quantitative genetic studies have produced a wealth of results concerning the genetic and environmental influences to the observed multivariate and longitudinal covariance structure of different complex traits, such as cognitive abilities (see Plomin, DeFries, McClearn, & McGuffin, 2008). The application of the classical twin design to longitudinal data has shown that genetic and environmental influences are present throughout the lifespan. For a range of traits, the contribution of genetic influences to the phenotypic variance tends to increase, while the contribution of the environmental influences decreases from childhood into adulthood. For traits such as IQ, common environmental influences (i.e., environmental influences shared by twins that contribute to their similarity) are present prior to adolescence, but decrease in importance later in adolescence, while unique environmental influences (i.e. environmental effects unique to each twin which contribute to the dissimilarity of twins) are present throughout (e.g., Bartels, Rietveld, Van Baal, & Boomsma, 2002; Boomsma et al., 2002; Cardon, Fulker, & DeFries, 1992; Hoekstra, Bartels, & Boomsma, 2007; Petrill et al., 2004; Rietveld, Dolan, Van Baal, & Boomsma, 2000).
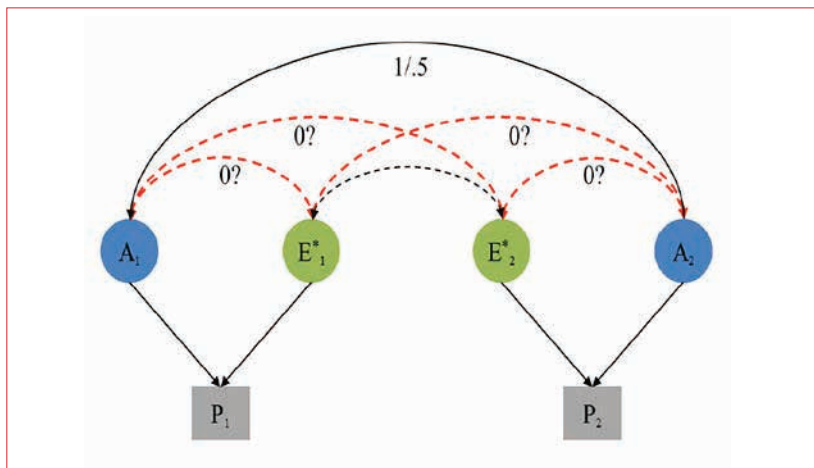
Figure 1 Classical twin-design model. This is the simplified representation of the classical twin model. In this simplified representation, (A) represents the additive genetic influences, (E*) represents the total environmental influences, and (P) the phenotype. The correlation between the environmental effects (E1* and E2*) is due to common environmental effects (C). Correlation between A1 and A2 equals 1 in monozygotic, and .5 in dizygotic twins, due to genetic resemblance. In this model GE covariance (dotted arrows) and GxE interaction are assumed to be absent

Most applications of the classical twin design come with well-known model assumptions, including absence of genotype by environment interaction (GxE interaction), zero genotype-environment covariance (GE covariance) (see dotted arrows in Figure 1), and random mating (i.e., zero spousal correlation; Eaves et al., 1977). These assumptions are known, or are suspected, to be violated to some degree in different complex traits. For instance, it is well known that assortative mating plays a role in intelligence (i.e., a positive correlation between IQ test scores of spouses; Eaves, 1973). For example, if data on parents of twins are available this information can be accommodated in the twin model (e.g., Martin, Eaves, Heath, Jardine, Feingoldt, & Eysenck, 1986). GxE interaction, i.e., moderation of genetic effects by environmental variables, or a dependence of environmental exposures on genotype, has been assessed thanks to advances in statistical modelling, enabling researchers to incorporate measured moderators, such as SES, into the twin model (Purcell, 2002; Harden, Turkheimer, & Loehlin, 2006; Boomsma & Martin, 2002). GE covariance has generally received less attention, although theoretically GE covariance is probably important, and has been hypothesised to explain increased heritability with age (Kan, Wicherts, Dolan, & Van der Maas, under revision).

The absence of GE covariance is certainly a strong assumption for many complex traits (Plomin et al., 2008). This assumption is often made pragmatically; in a design that includes only MZ and DZ twins applied to univariate data obtained at a single occasion, the covariance between genetic and environmental influences is not identified, and therefore cannot always be estimated. Here we explore whether GE covariance can be estimated from longitudinal data in the classical twin design using the genetic simplex (Boomsma & Molenaar, 1987). The genetic simplex provides a decomposition of phenotypic variance into genetic and environmental components at each measurement occasion (Figure 2). In addition, the genetic simplex expresses the phenotypic stability, i.e., the phenotypic correlation of a trait over time, in terms of genetic and environmental stability. In the standard genetic simplex, GE covariance is assumed to be absent as there is no direct or indirect pathway between genotypic and environmental components (Figure 2).

Developmental psychologists and behaviour geneticists, however, have long recognised definite processes giving rise to GE covariance (Carey, 1986; Eaves et al., 1977; Loehlin & DeFries, 1987; Plomin, DeFries, & Loehlin, 1977; Scarr, 1992; Scarr & McCartney, 1983). An important theoretical distinction is made between *passive*, *reactive*, and *active* GE covariance (Loehlin & DeFries, 1987; Plomin et al., 1977; Scarr, 1992; Scarr & McCartney, 1983): passive GE covariance arises when parents supply both genes and environment during the development of their offspring (i.e., smart parents transmit 'smart' genes and provide a 'smart' environment); reactive GE covariance arises when certain genotypes evoke certain reactions in the environment (e.g. 'smart' individuals evoke 'smart' reactions from their environment); and active GE covariance arises when individuals actively seek out environments consistent with their phenotype (i.e., 'smart' children seeking out a 'smart' environment). Provided that individual differences in the phenotype are at least partially due to genetic factors, these processes give rise to GE covariance.

Two conceptualisations of GE covariance are *niche picking* and *sibling effects*. Niche picking gives rise to within-individual GE covariance, as it involves an individual's choice or preference for certain environments, based on personal interest, talent, and personality (Scarr, 1992, Scarr & McCartney, 1983). This process thus implies a pathway between the individual's genotype and his or her environment, possibly mediated via the phenotype. Sibling effects give rise to between-individual GE covariance, as one sibling might directly or indirectly influence the other sibling's environment (Eaves, 1976; Carey, 1986), creating a pathway from one individual's genotype toward another individual's environment.

The aim of the present paper is to consider the specification of GE covariance processes in the genetic simplex and to explore the possibility to incorporate GE covariance in longitudinal twin models. We limit ourselves to the processes of niche picking and sibling effects in the simplex model including additive genetic (A) and unique environmental effect (E), and in a special case of a model with A, E, and common environmental effects (C) (see below). The setup of this paper is as follows: First, we represent the genetic simplex model, which we use as our starting model in which we incorporate GE covariance. Second, we consider the specification of GE covariance as arising through the processes of niche picking and sibling effects in three models: the within twin member's *niche picking model*, the between twin members' *sibling effects model*, and the combination of these two in the *combined model*. Third, we investigate the identification and resolution of these extended models, and compute the power to detect the parameters which give rise to GE covariance. We conclude with a brief discussion.

The genetic simplex

The genetic simplex model (Boomsma & Molenaar, 1987) has been used extensively to model longitudinal data in the classical twin design (e.g., Bartels et al., 2002; Bishop et al., 2003; Cardon et al., 1992; Petrill et al., 2004; Rietveld et al., 2000). The genetic simplex involves the regression of the phenotype measure at time point t, P_{tij}, on the additive genetic (A_{tij}), common (C_{tij}), and unique environmental variables (E_{tij}):
$$P_{tij} = A_{tij} + C_{tij} + E_{tij} + e_{tij}$$
(1)
where t denotes the measurement occasion (t=1...T), i denotes the twin pair, and *j* denotes the twin member. The term e_{tij} represents an occasion-specific residual, which may include genetic and environmental influences along with measurement error. Assuming the variables A, C, and E are uncorrelated, and given a correction for the occasion specific variance var(e_t) (e.g., if var(e_t) is a pure measurement error, this would be a correction for attenuation), the implied decomposition of variance at occasion t is
$$\text{var}(P_t) = \text{var}(A_t) + \text{var}(C_t) + \text{var}(E_t),$$
(2)
and the narrow sense heritability is $h^2_t = \text{var}(A_t)/[\text{var}(A_t)+\text{var}(C_t)+\text{var}(E_t)]$. The phenotypic stability is modelled by specifying autoregressive processes for A_t, C_t, and E_t. Limiting the equations to the additive genetic process, this entails the regression of A_{t+1} on A_t:
$$A_{t+1ij} = \beta_{At+1}A_{tij} + \zeta_{At+1},$$
(3)

where β_{At+1} is the autoregressive coefficient and ζ_{At+1} is the residual, or innovation term. The implied variance decomposition is var(A_{t+1}) = β_{At+1}^2var(A_t) + var(ζ_{At+1}), where var(ζ_{At+1}) is the residual or innovation variance. The covariance between A at t and $t+1$ equals cov($A_t A_{t+1}$) = β_{At+1}var(A_t). We may also consider the percentage of explained variance in this regression, i.e., $R_{At+1}^2 = \beta A_{t+1}^2$var($A_t$) / [$\beta A_{t+1}^2$var(At) + var($\zeta A_{t+1}$)]. Note that this percentage depends on the relative magnitudes of the autoregressive coefficient, β_{At+1}, and the residual variance, var(ζ_{At+1}). The regression model applies to C_{tij} and E_{tij} as well, so that the phenotypic covariance of the phenotype at t and t+1 is decomposed as follows:
$$\text{cov}(P_t P_{t+1}) = \beta_{At+1}\text{var}(A_t) + \beta_{Ct+1}\text{var}(C_t) + \beta_{Et+1}\text{var}(E_t).$$
(4)

The genetic simplex provides an informative decomposition of the phenotypic variance at each occasion and of the contribution of genetic and environmental effect to the stability and change over time. Note that in the simplex (i.e., excluding the parameters giving rise to GE covariance), the first and the last occasion specific variances (var(e_1) & var(e_2)) are not identified. Identification can be achieved by setting these terms to zero, or by the imposition of the constraints var(e_1) = var(e_2) and var(e_{T-1}) = var(e_T). During our model evaluation, we imposed these latter equality constraints. Also note that the model includes several special cases. For instance, if the parameters of β_A approach zero this implies that genetic effects do not contribute to stability. If var(ζ_A) approaches zero (given β_A are not equal to zero), the genetic stability is perfect (i.e., R_A^2 approach 1). If this is the case throughout the time period considered, the (genetic part of the) autoregressive model tends towards a single common factor model (Bishop, et al., 2003).

Twin studies based on the genetic simplex have provided detailed information on the contributions of genetic and environmental factors to the longitudinal covariance structure of complex traits, such as cognitive abilities. During the development of cognitive abilities during early childhood, the influences of additive genetic components (A) follow a simplex pattern (i.e., both β_A and var(ζ_C) greater than zero; Bishop et al., 2003; Cardon et al., 1992; Rietveld et al., 2000, Petrill et al., 2004). As such, additive genetic influences are both a source of stability and change. Unique environmental influences (E) mostly contribute to instability, as β_E are relatively low and var(ζ_E) are non-zero (Bartels et al., 2002; Cardon et al., 1992; Petrill et al., 2004; Rietveld et al., 2000). Common environmental influences (C) mostly contribute to stability during early development, as β_C tends to approach unity and var(ζ_C) tend to zero (Bartels et
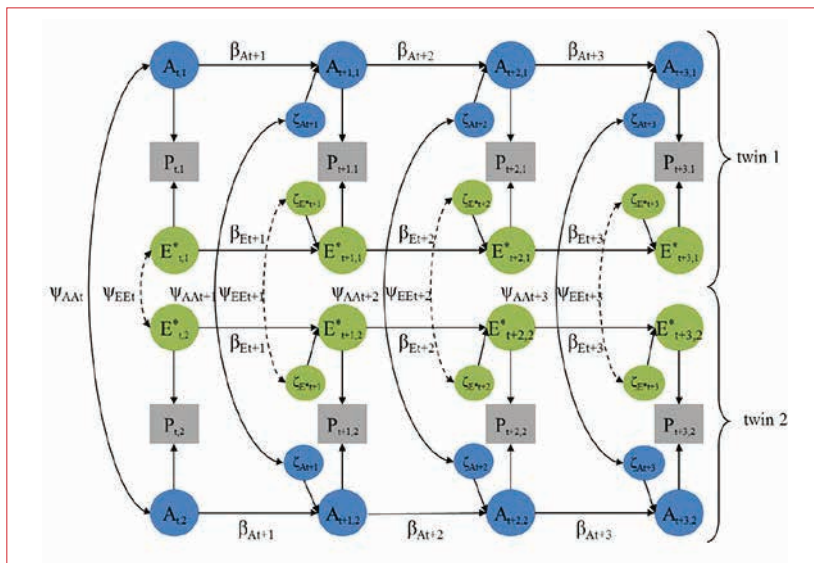
Figure 2 Simplified basic genetic simplex model, depicted with the minimum set of time points necessary to be able to identify the model. Note that E* represents total environmental influences, which are correlated between twins due to common environmental influences (C) depicted by the dotted pathways (AE *model). If this pathway is dropped, the model reduces to the AE model, in which only additive genetic variance (A) and unique environment (E) influence the phenotype (P)
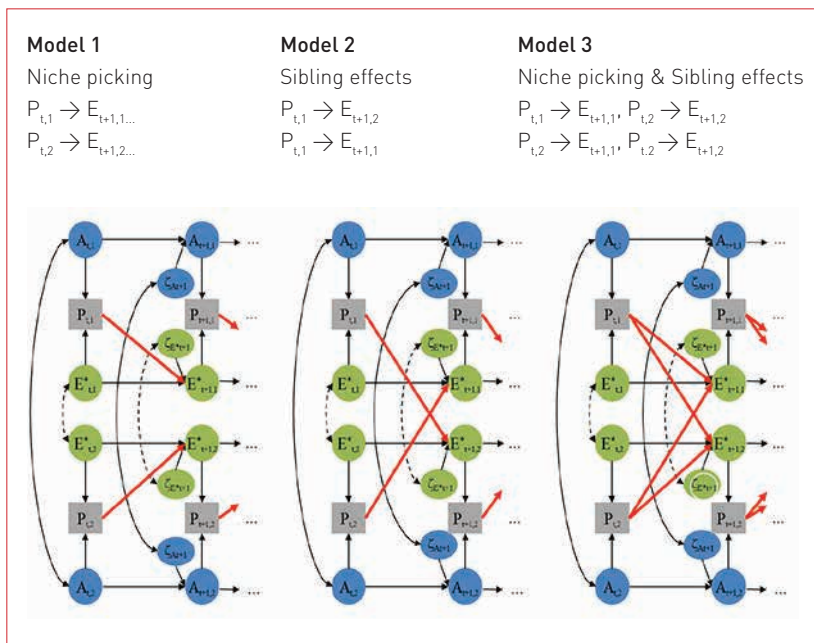

Figure 3 Path diagrams of three extensions of the basic genetic simplex model. To avoid clutter, only two time points (t=t, t+1) are depicted. The correlations between twins for var(A) (at t=t) and var(ζA) (at t=t+1,...) equals 1 and .5 in MZ and DZ twins, respectively. The covariance between the total environmental effects var(E) (at t=t) and var(ζE) (at t=t+1,...) are estimated, to accommodate common environmental effects

al., 2002; Bishop et al., 2003; Cardon et al., 1992; Petrill et al., 2004; Rietveld et al., 2000). Common environmental influences decrease in magnitude later in life, disappearing altogether in late adolescence. In addition, it has been established that the relative contribution of *A* increases, and that of *E* decreases over time (i.e. heritability increases over time, e.g., Bartels et al., 2002; Bishop et al., 2003; Boomsma et al., 2002; Haworth et al., 2010; Petrill et al., 2004,). Although the contributions of heritability and environment are robust and well established, the role of GE covariance has not been taken into account in these longitudinal studies.

Methods

Introducing GE covariance processes

We took the genetic simplex (Figure 2) as our starting model to introduce parameters giving rise to GE covariance. The simplex, as shown, accommodates common environmental influences (*C*) by the specification of correlated environmental influences (dotted arrows) rather than by the specification of a separate simplex process for C. By assessing the total environmental effects ($E^*=C+E$), instead of estimating each component separately, the specification and investigation of GE covariance originating in sibling effects and niche picking is greatly simplified[1]. So we considered two different models namely; 1) the A*E* model in which only additive genetic variance and unique environmental variance influence the phenotypic variance (i.e. the pathway between $E_{,,1}$, and $E_{,,2}$ is not included; 2) the A*E** model in which the unique environmental effects and the common environmental effects are captured in one term namely E^*.

By introducing crossed lagged phenotype to environment pathways within the two longitudinal models, we accommodated GE covariance within (i.e., niche picking) and between twins (i.e., sibling effects). Specifically, we viewed niche picking as the influence of phenotypic variable at occasion *t* on the environmental variable at time point *t+1* within each individual (Figure 3, Model 1). We accommodated sibling effects by introducing a cross lagged pathway from the phenotypic variable of one twin member at occasion *t* on the environment of the other twin member at time point *t+1* (Figure 3, Model 2)[2]. Finally, these two models can be combined (Figure 3,

---

[1] Note that our AE* simplex model is nested under the standard ACE simplex model, i.e., the standard ACE simplex will fit data generated with our AE* simplex model. The AE* simplex implies that in the standard ACE simplex the autoregressive coefficients of the E simplex equal those of the C simplex. This nesting is amenable to statistical testing.

[2] Note that this parameterisation of sibling effects deviates from previous methods such as that of Carey (1986) in which a polynomial was used to estimate GE covariance.
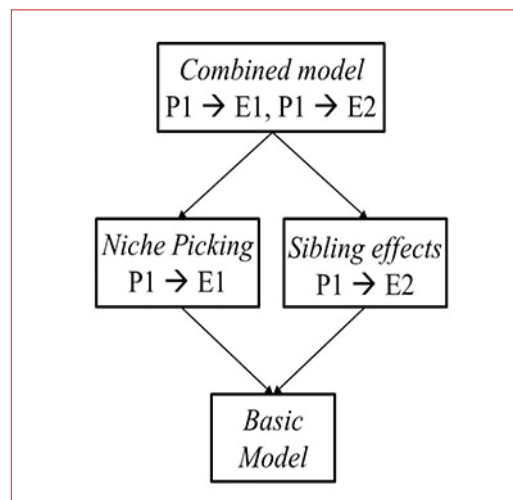
Figure 4 Nesting of the models of interest

Model 3), incorporating niche picking and sibling effects simultaneously. Note that we did not consider the direct path from A to E. Therefore we worked under the assumption that any effect of A on E must be mediated by the phenotype P. Still the pathway from P to E does imply GE covariance, as with this path in place, A and E are connected indirectly. For instance, in model 1 of Figure 3, the covariance between $A_{t,1}$ and $E_{t+1},1$ is due to the path from $A_{t,1}$ to $P_{t,1}$, and from $P_{t,1}$ to $E_{t,1}$.

Model evaluation

To establish whether extending the simplex model (i.e., the proposed cross lagged pathways) is practically feasible, we evaluated the extended models with respect to local identification, resolution, and power. First, we established model identification, which concerns the question whether the unknown parameters in the model can be estimated uniquely given appropriate longitudinal twin data. We distinguished between numerical and analytical identification. We considered both, because analytical identification does not rule out empirical under-identification. Empirical identification implies that fitting the true model to exact population MZ and DZ matrices produces a zero χ^2 value and perfect recovery of the parameter estimates regardless of variation in the starting values. A model is analytically identified if the Jacobian matrix of the model is of full column rank (Bekker, Merckens, & Wansbeek, 1993). The elements in the Jacobian matrix are the derivatives of each element in the population (MZ and DZ) covariance matrices to the unknown parameters (Derks, Dolan, & Boomsma, 2006; see also Bollen & Bauldry, 2010). We established analytical identification using the Maple program (Heck, 1993). To specify GE covariance, we added the cross lagged parameters (i.e., the additional parameters in the models depicted in Figure 3) to the basic

simplex, without giving any additional constraints. If this model is not identified, we proceeded by imposing constraints on the parameters underlying GE covariance or on the other parameters in the model (Maple input is available on request). Second, we determined the resolution to see how well the competing models can distinguish between different effects. It is necessary to establish that two models (say model 1 and 2, as depicted in Figure 3), while both being identified, are not equivalent (i.e., they should not fit the data containing different effects equally well). To establish this, we fitted data generated according to one model and fitted all other models, which should result in misfit expressed in χ^2 values greater than zero. Third, we computed the power of each model to detect the parameters underlying the GE covariance, given an α of .05. To calculate the power, we first constructed MZ and DZ population covariance matrices according to the model of interest, i.e., giving the parameters underlying the GE covariance a certain value. Fitting the true model will then produce a χ^2 statistic of zero. Dropping the parameter of interest, i.e., those associated with niche picking and/or sibling effects, will result in a positive χ^2 statistic. This statistic can be used to calculate the power to detect the parameters underlying the GE covariance (Satorra & Saris, 1985). We computed the power for all nested models (Figure 4) using sample sizes up to 3000 twins and a fixed α of .05 (R scripts are available on request).

Calculation of covariance matrices

The numerical population MZ and DZ covariance matrices are calculated in four different scenarios (no GE covariance; GE covariance in the form of niche picking, GE covariance in the form of sibling effect, GE covariance in the form of a combined effect; see Figures 2 and 3), the two different models (AE and AE*), four time points and 1000 MZ and 1000 DZ twin pairs (see Table 1 for parameter values). In the AE* models, we included common environmental variance as the covariance between the environmental variables. To accommodate increasing heritability, the genetic innovations terms $var(\zeta_A)$ and the autoregressive coefficients β_A increase with time, while the values for the environmental innovations terms $var(\zeta_E)$ and the autoregressive coefficients β_E decrease. The strength of the niche picking effect (β_{PE} i.e. the GE covariance due to paths from Pt,1 to Et+1,1, and from $P_{t,2}$ to $E_{t+1,2}$, see Figure 3) is set to equal .1 for the first time point t, adding a value of .01 for each additional time point. The strength of the sibling effects (β_{PE*} i.e. GE covariance due the path from $P_{t,1}$ to $E_{t+1,2}$ and $P_{t,2}$ to $E_{t+1,1}$, see Figure 3) is set to .05 at time point one, again adding a value of .01 for each additional time point (* indicates these parameters concern the sibling effects). We chose the

Table 1 Overview of the parameter values used to calculate the MZ and DZ covariance matrices

Parameter	Value given at time point			
	t	t+1	t+2	t+3
Ψ_A	10	2	3	4
$\Psi_{AA(MZ/DZ)}$	10/5	2/1	3/1.5	4/2
Ψ_E	10	3	2.5	2
Ψ_{EE}	2	1	1	1
var(e)	3	3	2	2
var(z_A)		2	3	4
var(z_E)		3	2.5	2
β_A		0.6	0.7	0.8
β_E		0.2	0.25	0.3
β_{PE}		0.1	0.11	0.12
β_{PE^*}		0.05	0.06	0.07

Table 2 Overview of constraints, the number of parameters used to estimate cross lagged GE covariance, and analytical identification. The same results are found for both the AE and the AE*models

Identifying constraint	Max # of parameters for GE covariance	Is the model identified?		
		Niche picking	Sibling effects	Combined Model
$\beta_{Pet+1} = \beta_{PEt+2} = \beta_{Pet+3}$	1	Yes	-	No
$\beta_{Pet+1}^* = \beta_{PEt+2}^* = \beta_{Pet+3}^*$	1	-	Yes	No
$\beta_{Pet+1} = \beta_{PEt+2} = \beta_{PEt+3}$ & $\beta_{Pet+1}^* = \beta_{PEt+2}^* = \beta_{PEt+3}^*$	1	-	-	Yes
$\beta_{PE} = \delta_{00} + \delta_{01}(t-2)$	2	Yes	-	No
$\beta_{PE}^* = \delta_{00}^* + \delta_{01}^*(t-2)$	2	-	Yes	No
$\beta_{PE} = \delta_{00} + \delta_{01}(t-2)$ & $\beta_{Pet}^* = \delta_{00}^* + \delta_{01}^*(t-2)$	2	-	-	Yes
$\beta_{At+1} = \beta_{At+2} = \beta_{At+3}$	3	Yes	Yes	Yes
$\beta_{Et+1} = \beta_{Et+2} = \beta_{Et+3}$	3	Yes	Yes	Yes
$\beta_{At+1} = \beta_{At+2} = \beta_{At+3}$ & $\beta_{Et+1} = \beta_{Et+2} = \beta_{Et+3}$	3	Yes	Yes	Yes
var(z_{At+1}) = var(z_{At+2}) = var(z_{At+3})	3	No	Yes	No
var(z_{Et+1}) = var(z_{Et+2}) = var(z_{Et+3})	3	No	Yes	No
var(z_{At+1}) = var(z_{At+2}) = var(z_{At+3}) & var(z_{Et+1}) = var(z_{Et+2}) = var(z_{Et+3})	3	No	Yes	No
$\Psi_{Et+1} = \Psi_{Et+2} = \Psi_{Et+3}$	3	No	Yes	No
var(e_{t+1}) = var(e_{t+2}) = var(e_{t+3}) = var(e_{t+3})	3	No	Yes	No

occasion specific variance (var(e_t)) to approach 20% of the phenotypic variance. While the parameter values chosen here are somewhat arbitrary, the parameters do give rise to summary statistics that resemble those reported in the literature on cognitive abilities. That is, given the present parameter values, heritability increases over time (h^2 = .50, .622, .679, & .774). We performed numerical analyses using R (R Development Core Team, 2012) and LISREL 8.80 (Jöreskog & Sörbom, 2006).

Results

Model identification

We first established analytic identification of the three GE covariance extensions (Figure 3) in both the AE and the AE* models. To establish which constraints are identifying, we started with the most unconstraint model, a model without any equality constraints on the parameters, except for the standard equality constraints on the occasion specific variance mentioned above (i.e. var(e_t) = var(e_{t+1}) and var(e_{t+2}) = var(e_{t+3})), and worked through different constraints to see if the models were identified (see Table 2). Note that the basic model is identified if the parameters underlying GE covariance are fixed to zero.

The analytical identification procedures indicated that none of the extended models are identified without additional constraints. This was true in both the AE and the AE* models. For each extension (i.e., either for niche picking, sibling effects, or these effects combined), we determined which restrictions rendered the models identified. To this end, we first explored the possibilities within the parameters used to model GE covariance. One way to restrict the GE covariance parameters is by constraining the GE covariance parameters to be equal over time (i.e., for niche picking model: $\beta_{PEt+1} = \beta_{PEt+2} = \beta_{PEt+3}$, for sibling effects model: $\beta_{PEt+1}^* = \beta_{PEt+2}^* = \beta_{PEt+3}^*$, and for the combined model: $\beta_{PEt+1} = \beta_{PEt+2} = \beta_{PEt+3}$ & $\beta_{PEt+1}^* = \beta_{PEt+2}^* = \beta_{PEt+3}^*$). These equality constraints resulted in identification of the models in both the AE and the AE* models. A less restrictive identifying constraint is the use of a two parameter model (i.e., $\beta_{PE} = \delta_{00} + \delta_{01}(t-2)$), in which δ_{00} resembles the intercept of the regression (i.e. the initial influence of GE covariance) and δ_{01} the direction coefficient of the regression slope coefficient (i.e. the change in the influence of GE covariance with time). By using the two parameter model we allowed linear changes in the GE covariance estimates over time. We used the following parameters for the niche picking model: $\beta_{PE} = \delta_{00} + \delta_{01}(t-2)$, the sibling effects model $\beta_{PE}^* = \delta_{00}^* + \delta_{01}^*(t-2)$, and for the combined model $\beta_{PE} = \delta_{00} + \delta_{01}(t-2)$ & $\beta_{PEt}^* = \delta_{00}^* + \delta_{01}^*(t-2)$). Again this identifying constraint resulted in model identification. By using different constraints for the GE covariance parameter, the extended models are thus identified.

Second, we explored constraints on other parameters in the model to determine if these constraints rendered the parameters underlying GE covariance identified (without imposing any constraints on these parameters). As can be seen in Table 2, in the sibling effects model, many different constraints render the sibling effect parameters (i.e., model 2 in Figure 3)

Table 3 Overview of χ² values obtained when fitting different models to different data sets

χ² values obtained when fitting different models in AE model

Fitted model	Data generating model			
	Basic	*Niche picking*	*Sibling effects*	*Combined*
Basic	-	1.14	17.02[+]	22.62[+]
Niche picking	Perfect	-13.8[+]	16.75	
Sibling effects	Perfect	.77	- 1.71	
Combined model	Perfect	Perfect	Perfect	-

χ² values obtained when fitting different models in AE* model

Fitted model	Data generating model			
	Basic	*Niche picking*	*Sibling effects*	*Combined*
Basic	-	.67	21.74	28.14
Niche picking	Perfect	-16.16	20.33	
Sibling effects	Perfect	.55	- .47	
Combined model	Perfect	Perfect	Perfect	-

+ Models that experience computational problems when certain parameter values are used to calculate the MZ and DZ covariance matrices

identified. Within the niche picking model and combined model (models 1 and 3 in Figure 3), only constraints on the autoregressive coefficients (i.e., either $\beta_{At+1} = \beta_{At+2} = \beta_{At+3}$ or $\beta_{Et+1} = \beta_{Et+2} = \beta_{Et+3}$ or $\beta A_{t+1} = \beta_{At+2} = \beta_{At+3}$ & $\beta_{Et+1} = \beta_{Et+2} = \beta_{Et+3}$) rendered the GE covariance parameters identified.

Lastly, we established numerical identification for each of the three GE covariance extensions (see Figure 3) in both the AE and the AE* models using the two parameter model to estimate GE covariance. To do so, we first calculated the population MZ and DZ covariance matrices, to which we fitted the data generating model, i.e., the true model under which the covariance matrix is calculated, in LISREL 8.80 (Jöreskog & Sörbom, 2006). Although our results are limited to the parameter values chosen, we had no trouble fitting these models in LISREL. This suggests that, given the chosen parameter values, empirical under-identification was not a problem.

Resolution of the models

To determine whether the models were distinguishable, we generated MZ and DZ covariance matrices for all different effects (no effect, niche picking effect, sibling effect, combined effect) for both AE and the AE* models, and fitted various competing models to these covariance matrices (see Table 3). For instance, we fitted the niche picking model to covariance matrices generated with the sibling effects.

Our analyses led to several noteworthy observations (Table 3). First, in both AE and AE* models, fitting the basic model (i.e., no GE covariance) to the covariance matrices including GE covariance parameter leads to deviations from the zero χ² value. This shows the possibility to distinguish our proposed GE covariance models from the basic genetic simplex model. The low χ² value obtained when fitting the basic model to niche picking data, indicates low power given any reasonable α. Thus given the chosen parameters values, niche picking (i.e., within individual GE covariance) has a relatively weak effect on the phenotypic covariance structure. The higher χ² values, obtained when fitting the basic model to the sibling effects (i.e. between twin GE covariance) and combined model, indicate greater power, and thus a stronger effect on the phenotypic covariance structure. Second, when fitting the different GE covariance models to covariance matrices generated under the basic genetic simplex (i.e., fitting models with GE covariance parameters to data where GE covariance is absent) led to perfect model fit, as expected. This shows that the GE covariance parameters are correctly estimated to be zero when a GE covariance effect is absent. Third, when fitting the sibling effects model to niche picking or combined data, the model fit is almost perfect. This again indicates that the niche picking effect is hard to detect and to distinguish from the sibling effect. Fourth, fitting the niche picking model to the sibling effects and combined data led to large χ² values, which indicates that when the sibling effect is present, the niche picking model will not fit well.

Statistical power

The statistical power to detect different forms of GE covariance depends on the sample size and on α. Table 4 and Figure 5 give an overview of the number of twins needed to attain certain power given an α of .05. It can be concluded that, in terms of power, detecting niche picking is more difficult than detecting sibling effects. This conclusion is in line with the results presented earlier, where we found that the χ² values were lower when fitting the basic simplex to data including the niche picking effect than to data including the sibling effects. The greatest power is found for the detection of the combined effects. It should be noted that this is an omnibus test, in which the power to detect sibling effects and niche picking effects are combined. When computing the power to detect these effects separately, it is clear that the sibling effects are easier to detect (Figure 5).

Table 4 The power, non-centrality parameter (*italic*), and degrees of freedom, given an α of .05, for different sample sizes for both the AE models and the AE* models

Data generating model	Fitted model	df	AE models 2x 500	AE models 2x 1000	AE models 2x 1500	AE* models 2x 500	AE* models 2x 1000	AE* models 2x 1500
Niche picking	*Basic model*	2	.10	.15	.20	.08	.10	.13
			.57	*1.14*	*1.71*	*.34*	*.67*	*1.01*
Sibling effects	*Basic model*	2	.75	.97	1.00	.85	.99	1.00
			8.51	*17.02*	*25.53*	*10.87*	*21.74*	*32.61*
Combined model	*Basic model*	4	.78	.98	1.00	.87	1.00	1.00
			11.31	*22.62*	*33.93*	*14.07*	*28.14*	*42.21*
Combined model	*Niche picking*	2	.74	.96	1.00	.82	.99	1.00
			8.38	*16.75*	*25.12*	*10.16*	*20.33*	*30.49*
Combined model	*Sibling effects*	2	.12	.20	.28	.07	.09	.11
			.86	*1.71*	*2.56*	*.24*	*.47*	*.70*

Figure 5 Graphical representation of ratio between sample size and power, given an α of .05, for the different models

Discussion

The aim of this paper was to specify processes giving rise to GE covariance within the genetic simplex model. To model GE covariance in the genetic simplex, we introduced phenotype to environment cross lagged relationships, representing niche picking effects, sibling effects, and the combined effects. We considered two models: one model with additive genetics and unique environmental influences (AE), and one model in which we accommodated the common environmental influences by covariance between E of each twin (AE*). First, we demonstrated the possibility to accommodate GE covariance in both the AE and AE* simplex models. The additional GE covariance parameters are identified under various identifying constraints. Identifying constraints may be imposed on the parameters accounting for the GE covariance. For instance, equality constraints and the use of a two parameter model (constraining the change in the parameters to be linear) rendered the model identified. Identification can also be achieved by imposing constraints on the standard parameters in the genetic simplex (e.g., the autoregressive coefficients). Given such constraints the parameter used to model GE covariance can be estimated freely at each time point. Second, we showed that it is possible, in principle, to determine whether an effect of GE covariance is present or not, as fitting a different model than the data generating model leads to non-zero χ^2 values. Third, we showed that relatively large sample sizes are needed to reach sufficient power to detect GE covariance effects, given our present parameter values. It turns out that the power to detect GE covariance depends on the type of effect. Larger sample sizes are needed to detect the niche picking effects than the sibling effects or combined effects. As power depends on the number of observations, we expect that adding time points to the models will lead to greater power in addition to simply increasing the sample size.

We emphasise that the present study is a first step towards establishing viable twin models including processes giving rise to GE covariance. Our present results are limited in the following respect. First, our results are limited to the scenarios considered, both in terms of measurement occasions (T=4) and of our choice of parameter values in our numerical results. Increasing the number of occasions is not likely to given rise to problems of identification. However, fewer occasions (say, 2 or 3) requires further study.

Second, our results concerning power and resolution depend wholly on our choice of parameter values, and are limited accordingly. More extensive power analyses were beyond the present scope, but we
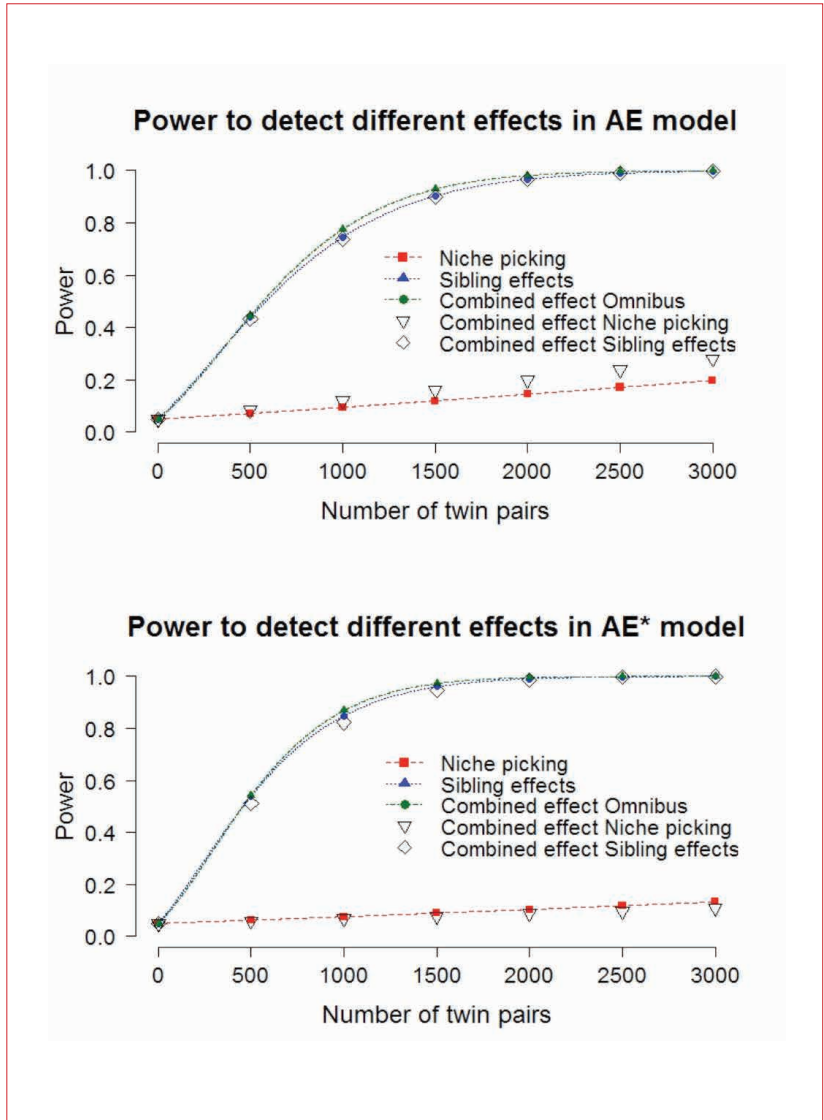
note that such analyses pose no great problem to carry out, and can be tailored to the researcher's specific expectations. Our explorations of other parameter values showed that identification did not depend on the exact values (as expected). However, we did find that certain choices of parameters resulted in computational problems in fitting the basic (i.e., excluding parameters giving rise to GE covariance) genetic simplex. Notably, low values of the environmental autoregressive coefficient (e.g., $\beta_{Et+1} = .1$, $\beta_{Et+2} = .15$, $\beta_{Et+3} = .2$) in the sibling effects and the combined model rendered the basic simplex model computationally hard to fit as the occasion specific residual variances assumed negative values. This problem can be resolved by fixing these variances to zero.

Finally, we have only considered the AE model and the AE* model. The AE model is standard in the absence of common environmental influences (C). The AE* model treats common and unique environmental influences as 'total environmental effects', rather than explicitly modelling separate E and C processes. The AE* model is nested under the ACE model (as the ACE model with equal autoregressive C and E parameters implies the AE* model). In our current exploration of GE covariance, we only considered processes giving rise to AE covariance or AE* covariance. We have not addressed other sources of covariance, such as AC covariance, which are distinct from AE* and AE covariance, as these forms were beyond the scope of this article. We hope to extend our present results to the ACE model in the near future.

We conclude that sibling interaction and niche picking, conceptualised as the regression of environmental influences (E or E*) on the phenotypic variable, can be accommodated in the genetic simplex models considered here. While these models are identifiable given appropriate constraints, the issue of power requires attention, as does the generalisation to the standard ACE model. The application of these models, given adequate sample sizes, will ultimately allow one to establish whether these sources of GE covariance play any role in complex phenotypes, as is often suggested (e.g., in discussions of cognitive abilities).

References

Bartels, M., Rietveld, M. J. H., Van Baal, G. C. M., & Boomsma, D. I. (2002). Genetic and environmental influences on development of intelligence. *Behavior Genetics, 32*, 237-249.

Bekker, P. A., Merkens, A., & Wansbeek, T. J. (1993). *Identification, equivalent models, and computer algebra*. Boston, MA: Academic Press.

Bishop, E. G., Cherny, S. S., Corley, R., Plomin, R., DeFries, J. C., & Hewitt, J. K. (2003). Development genetic analysis of general cognitive ability from 1 to 12 years in a sample of adoptees, biological siblings and twins. *Intelligence, 31*, 31-49.

Bollen, K. A., & Bauldry, S. (2010). Model identification and computer algebra. *Sociological Methods & Research, 39*, 127-156.

Boomsma, D. I., Martin, N. G. (2002). Gene-environment interactions. In H. D'haenen, J. A. den Boer, P. Wilner (Eds), *Biological Psychiatry* (181-187). John Wiley & Sons Ltd.

Boomsma, D. I., & Molenaar, P. C. M. (1987). The genetic analysis of repeated measures. I. Simplex models. *Behavior Genetics, 17*, 111-123.

Boomsma, D. I., Vink, J. M., van Beijsterveldt, T. C., de Geus, E. J., Beem, A. L., Mulder, E. J., Derks, E. M., Riese, H., Willemsen, G. A., Bartels, M., van den Berg, M., Kupper, N. H., Polderman, T. J., Posthuma, D., Rietveld, M. J., Stubbe, J. H., Knol, L. I., Stroet, T., & Van Baal G. C. (2002). Netherlands twin register: A focus on longitudinal research. *Twin Research, 5,* 401-406.

Cardon, L. R., Fulker, D. W., & DeFries, J. C. (1992). Continuity and change in general cognitive ability from 1 to 7 years of age. *Developmental Psychology, 28*, 64-73.

Carey, G. (1986). Sibling imitation and contras effects. *Behavior Genetics, 16*, 319-341.

Derks, E. M., Dolan, C. V., & Boomsma D. I. (2006). A test of the equal environment assumption (EEA) in multivariate twin studies. *Twin Research and Human Genetics, 9*, 403-11.

Van Dongen, J., Slagboom, P. E., Draisma, H. H., Martin, N. G., Boomsma, D. I. (2012). The continuing value of twin studies in the omics era. *Nature Reviews Genetics, 13*, 640-653.

Eaves, L. J. (1973). Assortative mating and intelligence: An analysis of pedigree data. *Heredity, 30*, 199-210.

Eaves, L. J. (1976). A model for sibling effects in man. *Heredity, 36*, 205-214.

Eaves, L. J., Last, K., Martin N. G., & Jinks, J. L. (1977). A progressive approach to non-additivity and genotype-environmental covariance in the analysis of human differences. *British Journal of Mathematical Statistical Psychology, 30*, 1-42.

Harden K. P., Turkheimer, E., & Loehlin, J. C. (2006). Genotype by environment interaction in adolescents' cognitive aptitude, *Behavior Genetics*, DOI 10.1007/s10519-006-9113-4.

Haworth C. M. A., Wright, M. J., Luciano, M., Martin, N. G., de Geus, E. J. C., van Beijsterveldt, C. E. M., Bartels, M., Posthuma, D., Boomsma, D. I., Davis, O. S. P., Kovas, Y., Corley, R. P., DeFries, J. C., Hewitt, J. K., Olson, R. K., Rhea, S-A., Wadsworth, S. J., Iacono, W.G., McGue, M., Thompson, L. A., Hart. S. A., Petrill, S. A., Lubinski, D., & Plomin, R. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood, *Molecular Psychiatry, 15*, 1112–1120.

Heck, A. (1993). *Introduction to Maple, a computer algebra system*. New York; Springer-Verlag.

Hoekstra, R. A., Bartels, M., Boomsma, D. I. (2007). Longitudinal genetic study of verbal and nonverbal IQ from early childhood to young adulthood. *Learning and Individual Differences, 17*, 97-114.

Jöreskog, K. G., & Sörbom, D. (2006). *LISREL 8.80 for Windows [Computer Software]*. Lincolnwood, IL: Scientific Software International, Inc.

Kan, K-J, Wicherts, J. M., Dolan, C. V., & Van der Maas, H. L. J. (2012). On the nature and nurture of intelligence and specific cognitive abilities: the more heritable, the more culture dependent. [Under revision].

Loehlin, J. C., & DeFries, J. C. (1987). Genotype-environment correlation and IQ. *Behavior Genetics, 17*, 263-277.

Martin, N. G., Eaves, L. J., Heath, A. C., Jardine, R., Feingoldt, L. M., & Eysenck, H. J. (1986). Transmission of social attitudes. *Proceedings of the National Academy of Sciences of the United States of America, 83*, 4364-4368.

Petrill, S. A., Hewitt, J. K., Cherny, S. S., Lipton, P. A., Plomin, R., Corley, R., & DeFries, J. C. (2004). Genetic and environmental contributions to general cognitive ability through the first 16 years of life. *Developmental Psychology, 40*, 805-812.

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin, 84*, 309-322.

Plomin, R., DeFries J. C., McClearn, G. E., & McGuffin, P. (2008). *Behavioral Genetics*. New York; Worth Publishers

Plomin, R., Loehlin, J. C., & DeFries, J. C. (1985). Genetic and environmental components of 'environmental' influences. *Developmental Psychology, 21*, 391-402.

Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research, 5*, 554–771.

Rietveld, M. J. H., van Baal, G. C. M., Dolan, C. V., & Boomsma, D. I. (2000). Genetic factor analyses of specific cognitive abilities in 5-year-old Dutch children. *Behavioral Genetics, 30*, 29-40.

R Development Core Team (2012). 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/.

Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika, 50*, 83-90.

Scarr, S. (1992). Developmental theories for the 1990s: Development and Individual differences. *Child Development, 63*, 1-19.

Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype à environment effects. *Child Development, 54*, 424-435.

JOHANNA M. DE KORT
Following the Research Master Psychology at the UvA. She is interested in structural equation modelling of longitudinal data, with applications in the fields of psychometrics, quantitative genetics, developmental psychopathology, and clinical psychology.

CONOR V. DOLAN
Interested in the theory of structural equation modelling, with applications in the areas of psychometrics, quantitative genetics, and cognitive abilities.

DORRET I. BOOMSMA
She established the Netherlands Twin Register which is used as a resource for research projects that aim to quantify the influence of genes on phenotypic differences between individuals and identify the responsible variants.