

The Magnitude of Genetic and Environmental Influences on Parental and Observational Measures of Behavioral Inhibition and Shyness in Toddlerhood

Ashley K. Smith · Soo H. Rhee · Robin P. Corley ·
Naomi P. Friedman · John K. Hewitt ·
JoAnn L. Robinson

Received: 25 February 2012 / Accepted: 5 June 2012 / Published online: 18 July 2012
© Springer Science+Business Media, LLC 2012

Abstract Behavioral inhibition is a temperamental trait that refers to slow approach to novel items, shyness towards new people, and fearfulness in new situations, and individuals may develop inhibited response styles by as early as 2 years of age. There are important methodological considerations in the assessment of early temperament, with parental report and observational measures providing both corroborative and unique data. The present study examined behavioral inhibition measured by parental report and observational measures in a genetically informative sample to delineate the agreement between the methods and the uniqueness of each method, and to estimate the magnitude of genetic and environmental influences on the common and unique variance. The biometric, psychometric, and rater bias models were conducted to study the covariance between measurement modalities. Overall, the results suggested a common phenotype was assessed by both parents and observers. The latent phenotype underlying parental and observational measures of behavioral inhibition was moderately to substantially heritable.

Keywords Temperament · Behavioral inhibition · Rater bias · Toddler · Shyness · Twin study methods · Parent report · Observational measures

Introduction

Temperamental traits are thought to be relatively stable behavioral characteristics that are observable in early childhood and at least partially affected by genetic influences (Kagan and Saudino 2001). Behavioral inhibition is a temperamental trait that reflects inhibition or withdrawal to novelty. Individuals develop inhibited (or disinhibited) response styles in early toddlerhood, and individual differences in behavioral inhibition seem somewhat stable across the life course (e.g., Kagan and Saudino 2001). Behavioral inhibition is often studied in the context of how one relates to one's environment, and is a general term encompassing slow approach to novel items, shyness towards new people, and fearfulness in new situations (e.g., Kagan et al. 1984; Robinson et al. 1992). Different terms are used in the literature to describe constructs similar to behavioral inhibition, such as inhibition (Caspi et al. 1996), fearfulness (e.g., Goldsmith and Lemery 2000) and shyness (specific to situations with new people; Emde et al. 1992). The term behavioral inhibition is used in the present study as a general term to encompass the range of constructs associated with inhibited response styles; however, when referring to specific studies, the term used in that study is used.

Several longitudinal studies have found associations between early behavioral inhibition and later negative adjustment and psychopathology, suggesting the importance of studying its development and etiology. For example, Sanson et al. (1996) found moderate stability in

Edited by Yoon-Mi Hur.

A. K. Smith (✉) · S. H. Rhee · J. K. Hewitt
Department of Psychology and Neuroscience,
University of Colorado, Boulder, CO, USA
e-mail: ashksmith@gmail.com

A. K. Smith · S. H. Rhee · R. P. Corley ·
N. P. Friedman · J. K. Hewitt
Institute for Behavioral Genetics, University of Colorado,
Boulder, CO, USA

J. L. Robinson
Department of Human Development and Family Studies,
University of Connecticut, Storrs, CT, USA

shyness from infancy to early childhood, and that shyness was associated with “difficultness” in infancy and internalizing problems at age 6. Another study found associations between inhibition and shyness in toddlerhood and internalizing symptoms during adolescence; those with stable high or increasing levels of shyness over the course of the study had the most internalizing symptoms, impaired academic performance, and fewer positive interactions with peers and parents (Letcher et al. 2009). Caspi et al. (1996) found longitudinal associations between inhibition at 3 years and psychological disorders at age 21; those who were inhibited at age 3 were at an increased risk of having one or more psychological disorders, including depression and anxiety; unexpectedly, alcohol problems and violent crimes were also associated with early inhibition, but only in males. Additionally, childhood behavioral inhibition was also found to be associated with negative emotionality, decreased pleasure from social encounters, and delayed life events in early adulthood (Gest 1997). Conversely, some researchers suggest that behavioral inhibition may be a protective factor against externalizing behaviors (e.g., harm avoidance, or avoiding novel/dangerous activities, is hypothesized to predict fewer antisocial behaviors, Lahey and Waldman 2003) or associated with more positive outcomes, such as early conscience development (Kochanska et al. 1994) and decreased impulsivity and better self restraint (Askan and Kochanska 2004).

Through twin studies, researchers are able to estimate the magnitude of genetic influences (i.e., heritability) and environmental influences by assessing differences in the similarity between monozygotic (i.e., genetically identical) and dizygotic twins. These studies indicate a moderate to high heritability for behavioral inhibition. For example, Eley et al. (2003) found high heritability for shyness and inhibition ($h^2 = .66$ for females and $h^2 = .76$ for males) in a twin study of preschoolers. Emde et al. (1992) found moderate heritability estimates for observed shyness ($h^2 = .49$) and a behavioral inhibition paradigm ($h^2 = .56$). Another study reported significant genetic, but not shared environmental, influences on parent-rated shyness in adolescent same-sex sibling-pairs (i.e., MZ twins, DZ twins, and full siblings raised in intact families; full siblings, half siblings and unrelated siblings raised in stepfamilies; Saudino et al. 1995).

Differences in heritability estimates of behavioral inhibition may be partially due to differences in the method of assessment across studies. Parental questionnaires are commonly used in studies of early temperament. Observational measures are used much less often, but provide valuable information in multi-method studies of temperament. Together, the measures may provide both corroborative and unique data in the estimation of heritability estimates (Hewitt et al. 1992; Emde et al. 1992; Kagan

2001; Kagan and Saudino 2001). However, both parental questionnaires and observational data have methodological limitations that make interpretation of studies using these assessment methods difficult.

Parental questionnaires are cost- and time-efficient, and gather information from generalized perspectives of behaviors across situations. However, parental questionnaires may be affected by rater biases, parental phenotypes, and contrast effects. Rater bias is the tendency of a rater to over- or underestimate scores consistently (e.g., Neale and Cardon 1992). This tendency may reflect phenotypes of the parents, with parents having unique interpretations of questions regarding their children’s behaviors depending on their own experiences, personality traits, and other variables (e.g., van der Valk et al. 2001; Rothbart and Goldsmith 1985). Another potential methodological limitation of parent report is the contrast effect, which refers to the tendency for parents to overestimate differences between dizygotic twins compared to parents of monozygotic twins, and leads to inflated genetic estimates in some cases (Saudino et al. 2000; Emde et al. 1992).

Observational measures allow unbiased individuals to assess temperament. Usually, two observers rate the individual twin behaviors independently, so rater biases do not contribute to the correlation between the ratings of the twins. However, observational measures only provide data for certain contexts. This can be problematic for a construct like behavioral inhibition, as individuals may exhibit high levels of inhibited responses in some situations (e.g., in a laboratory setting) but not others. Observational measures only capture a snapshot of behavior (e.g., what is observed during the specific protocol); reliability/repeatability of behaviors is an important issue to consider. Because observational measures are generally limited to small slices of an individual’s behavioral repertoire, they may reflect state, rather than trait, behaviors (Wachs 1992, cited in Sanson et al. 1996). Thus, observational measures are potentially susceptible to random error based on the generally short segments of observational data collected (e.g., Kendler and Baker 2007). Additionally, because raters are typically following a protocol and aiming for high inter-rater reliability, consistency between ratings may appear as a bias in ratings contributing to instrument variance (e.g., Cronbach et al. 1963). The protocol bias will appear as systematic variance in behaviors, especially when assessed across time, which would not be shared in common with parent ratings.

The limitations of assessment methods may have an impact on the estimate of heritability and environmental influences on phenotypes for many aspects of behavior. For example, a meta-analysis examining the putatively environmental measure life events found that parent reports yield higher heritability estimates than observational

measures (Kendler and Baker 2007). Also, assessment method was a significant moderator of heritability in a meta-analysis examining antisocial behaviors (Rhee and Waldman 2002).

There have been mixed findings of low to moderate correlations between the two methods of assessing behavioral inhibition (e.g., Emde et al. 1992; Kagan 2001; Kagan and Saudino 2001). These correlations suggest that it may be possible to derive a more valid phenotype by examining both methods simultaneously. To evaluate this possibility, it is necessary to understand whether behavioral inhibition assessed via two different methods is merely correlated, or represents a common view of the phenotype. One way to assess the degree to which parent report and observational measures are assessing the same latent phenotypic construct is to fit theory-driven models to the data. Hewitt et al. (1992) proposed several steps in analyzing rater agreement in twin data. First, the biometric model (Fig. 1a) takes an agnostic approach to the data, and makes no assumptions regarding the source of the correlation between measured phenotypes. The psychometric model (Fig. 1b) assumes that a common phenotype is assessed by both raters, and reflects a trait of the individual. This model estimates the genetic and environmental influences affecting the common phenotype, as well as the genetic and environmental influences associated with each variable's unique variance. The rater bias model (Fig. 1c) also assumes that a common phenotype is assessed by both raters while accounting for the tendency of individual raters to consistently over- or underestimate scores. In the rater bias model, genetic and environmental influences are estimated for the common phenotype, and the rater bias and residual variance is estimated for each variable's unique variance. In the case of observations, rater biases do not contribute to the correlation between the ratings of the two twins, as two different individuals rate the two twins. Therefore, the "rater bias" is simply the source of covariance between the two twins over and above the covariance due to the latent common phenotype (e.g., protocol/instrument bias).

The present study investigated the magnitude of genetic and environmental influences on behavioral inhibition in toddlerhood utilizing a longitudinal twin study approach with observational and parent-reported questionnaire data. The first study aim was to assess whether these measures reflect a common view of the toddler's behaviorally inhibited phenotype. The second aim was to assess the extent to which the common variance shared by parental and observational measures is influenced by genetic and environmental influences. We were also able to assess the extent to which the unique variance of each measure is influenced by genetic and environmental factors (in the psychometric model), or by rater bias (in the rater bias model).

Methods

Sample

Data were collected as part of the MacArthur Longitudinal Twin Study (MALTS). Same-sex twin pairs were recruited from the Colorado Twin Registry of twins born from 1984 to 1990 (Rhea et al. 2006). Twins were selected based on high birth weight (greater than 3.86 lb) and general good health. Zygosity was determined via a ten-item assessment of twin similarity completed by examiners working with the twins. For cases in which the examiners were unsure, a panel of nine or more highly polymorphic simple tandem repeat markers was used to confirm zygosity. For the present study, behavioral inhibition data were collected as part of a larger research protocol during home and lab visits (occurring within 2 weeks of each other) at 14, 20, 24, and 36 months. All analyses were conducted at each time point. Table 1 provides the sample sizes of dizygotic and monozygotic individuals within twin pairs with behavioral inhibition data at each time point.

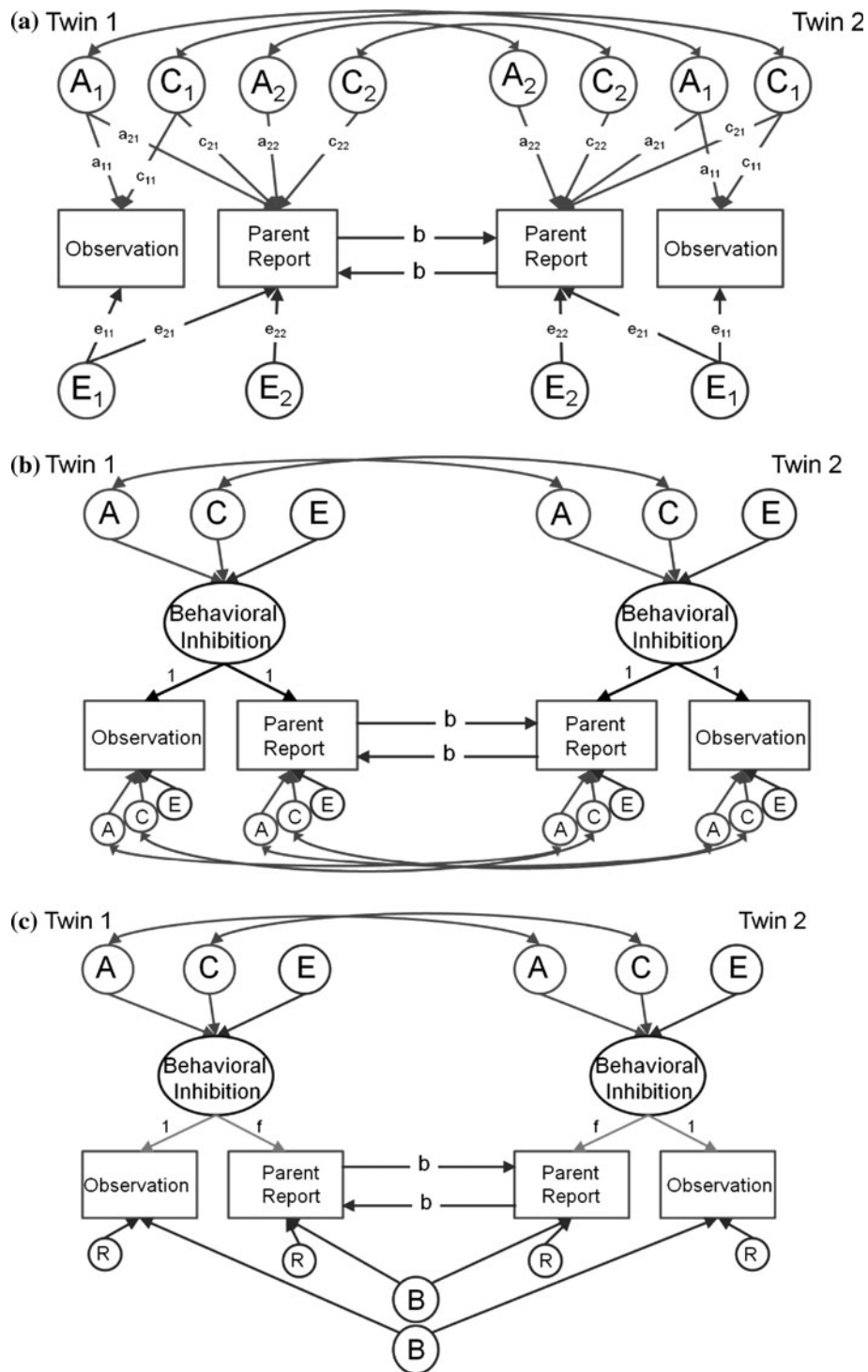
Assessment methods

Observational measures

Observational measures of behavioral inhibition were collected during home and lab visits, which were videotaped and later rated by trained research assistants. The home and lab video recordings were rated by separate examiner assistants to ensure both twins were not assessed by the same individual.

During the home visit, the first 5 min of the examiners' visit to the family home was recorded by one examiner while the other examiner interacted with the mother and twins. Examiners introduced themselves, presented two toys (one for each twin), and placed identifying vests on the twins. This segment was later rated to assess each toddler's level of shyness to the examiners and inhibition to the research equipment. The outcome factor scores of the segment refer to the tendency for each toddler to approach novelty (e.g., the toy and/or the examiner; "Approach") and the tendency for the toddler to cling to their mother ("Cling").¹ Each twin in the dyad was assigned to work with a separate examiner during the home visit. The toddlers were in separate rooms during most of the home procedures. Examiners assessed toddlers' behaviors,

¹ Confirmatory and exploratory factor analyses (EFA) were conducted in Mplus to determine the factor structure of the home shyness variables. The variables loading on the "Approach" factor were average shyness, proximity to the researcher, and playing with the novel toy. The variables loading on the "Cling" factor were proximity to mother, touching mother, and clinging to mother.



including level of fearfulness, using the Infant Behavior Record (IBR; Matheny 1980) during the Bayley Scales of Infant Development (Bayley 1969; collected at 14, 20, 24 months) or during the Stanford–Binet (Terman and Merrill 1973; collected at 36 months) and over the course of all other home procedures. The “IBR Fear” item used in the present study was the average of the toddler’s “reaction

to the new or strange; e.g., strangers, strange surroundings, test materials” on a nine-point scale across the two ratings (i.e., during the Bayley/Stanford–Binet and the aggregate across all other procedures).

During the lab visit, toddlers participated in a behavioral inhibition paradigm (based on research by Kagan and colleagues (e.g., Kagan et al. 1984), which was conducted

Fig. 1 a Biometric model. $A1$ genetic influences shared in common by observation and parent report and $A2$ genetic influences specific to parent report. Correlation between Twin 1 genetic influences and Twin 2 genetic influences is 1.0 for MZ twins and .5 for DZ twins. a_{11} genetic influences on observations, a_{21} common genetic influences on parent reports, a_{22} unique genetic influences on parent reports, $C1$ shared environmental influences shared in common by observation and parent report, and $C2$ shared environmental influences specific to parent report. Correlation between Twin 1 shared environmental influences and Twin 2 shared environmental influences is 1.0 for MZ twins and DZ twins. c_{11} shared environmental influences on observations, c_{21} common shared environmental influences on parent reports, c_{22} unique shared environmental influences on parent reports, $E1$ nonshared environmental influences shared in common observation and parent report, $E2$ nonshared environmental influences specific to parent report. e_{11} nonshared environmental influences on observations, e_{21} common nonshared environmental influences on parent reports, e_{22} unique nonshared environmental influences on parent reports, and b contrast effects. **b** Psychometric model. Correlation between Twin 1 genetic influences and Twin 2 genetic influences is 1.0 for MZ twins and .5 for DZ twins. Correlation between Twin 1 shared environmental influences and Twin 2 shared environmental influences is 1.0 for MZ twins and DZ twins. b contrast effects. **c** Rater bias model. Correlation between Twin 1 genetic influences and Twin 2 genetic influences is 1.0 for MZ twins and .5 for DZ twins. Correlation between Twin 1 shared environmental influences and Twin 2 shared environmental influences is 1.0 for MZ twins and DZ twins. b contrast effects, f freed latent factor loading on parent report, B rater bias parameter, and R residual variance

in a playroom to assess toddler reactions to novel items and strangers. The measures included latency to leave parent upon entering the playroom, latency to approach toys, latency to approach stranger, latency to approach a novel

object (e.g., monster mask), and time spent close to mother during the phases (i.e., toys, stranger, and novel object). The paradigm changed to a risk room setting at the 36-month visit, in which the latency for the toddler to approach different items (i.e., balance beam, box, mattress, and scary mask) was recorded. A composite score (“BI Task”) was created from the standardized scores of the variables. At the end of the lab visits at 14, 20, and 24 months, examiners rated the overall shyness of each twin during the lab assessments on a four-point ordinal scale (i.e., not shy to very shy; “Lab Visit Shyness”). This variable was not assessed during the 36-month lab visit.

Parent report measures

Parental ratings of toddler behavioral inhibition were collected using questionnaires. Shyness was ascertained using an average of the mother and father ratings on the Colorado Childhood Temperament Inventory (“CCTI Shy,” CCTI; Rowe and Plomin 1977). Fearfulness was measured via mother ratings of fearfulness on the Differential Emotions Scale (“DES Fear,” DES; Izard et al. 1980); father ratings were not collected for this scale. Approach was assessed using the average of mother and father ratings on the Toddler Temperament Scale (“TTS Approach,” TTS; Carey and McDevitt 1978). When available, averages of parental ratings were used to reduce measurement error.

Table 1 Sample size and means by sex across time points

	Method	Descriptives								Regression analyses ⁺		
		MZ				DZ				Gender	Zygosity	G*Z int
		Male, n	Female, n	M: mean (SD)	F: mean (SD)	Male, n	Female, n	M: mean (SD)	F: mean (SD)	β	β	β
14 months	Observation	212	229	-.10 (.69)	.11 (.72)	188	158	-.03 (.67)	.00 (.66)	.09*	-.08	-.09
	Parent report	194	204	-.18 (.88)	.14 (.90)	172	153	.01 (.97)	.02 (1.06)	.09*	-.06	-.12*
20 months	Observation	193	200	-.10 (.77)	.18 (.84)	172	148	-.08 (.78)	-.01 (.75)	.11**	-.12*	-.10
	Parent report	165	176	-.06 (.88)	.11 (.90)	161	142	-.07 (.98)	.01 (1.01)	.07	-.06	-.04
24 months	Observation	186	200	-.13 (.80)	.15 (.80)	174	152	.08 (.72)	-.09 (.79)	.04	-.16**	-.20**
	Parent report	176	187	-.09 (.95)	.17 (.89)	165	148	-.02 (.98)	-.08 (.94)	.05	-.14*	-.12*
36 months	Observation	176	200	-.04 (.71)	.20 (.81)	158	154	-.12 (.71)	.04 (.82)	.13**	-.10*	-.04
	Parent report	161	174	-.10 (.94)	.26 (.90)	147	141	-.09 (.95)	-.10 (.98)	.09*	-.19**	-.14*

MZ monozygotic twins, DZ dizygotic twins, G*Z int gender \times zygosity interaction term, n number of individual twins, M male, F female, β standardized beta coefficient

* Significant p value of $<.05$; ** significant p value of $<.01$. ⁺ Full model regression analyses were significant at $p < .01$, with the exception of parent report at 24 months, which was significant at $p < .05$

Statistical analysis

Data preparation

All variables included in the analyses were assessed for normality of their distributions. In cases where the skewness/kurtosis values were greater than 1.0, the variables were log transformed or square root transformed to achieve normal distributions. Variables violating the assumptions of normality even after transformation were ordinalized while ensuring adequate sample sizes in each category to avoid the issue of zero or small cells.

Factor analysis

Factor analyses were conducted in Mplus (v6.1; Muthén and Muthén 1998–2010), which allows the simultaneous analysis of categorical and continuous variables. EFA and confirmatory factor analyses (CFA) were conducted at each time point to examine the underlying factor structure of the data. The cluster option was used when calculating standard errors and χ^2 tests in Mplus to account for non-independence of the data within twin pairs (Muthén and Muthén 1998–2010). Given that the χ^2 is sensitive to sample size, the comparative fit index (CFI; Bentler 1990), the Tucker–Lewis index (TLI; Bentler 1990), and the root mean square error of approximation (RMSEA; Browne and Cudeck 1993) were also examined. A CFI and TLI greater than .95 and RMSEA less than .06 indicate good model fit (Hu and Bentler 1998). EFA and CFA of all measures yielded two factors, a parent-report factor and an observational factor, across all ages. Based on these findings, factor scores computed by Mplus were used for the following analyses.

Model fitting

Model fitting analyses were conducted in Mx (Neale et al. 2003) using the factor scores from the CFA described above. Saturated models, which estimate the means, variances, and covariances without constraints, were fit to the data at each time point to estimate the fit of the other models tested. Null models, which estimate the means and variances only, were fit to the data at each time point to allow the estimation of TLI and RMSEA for each estimated model (Neale et al. 2003; null model results available upon request). To determine whether observation and parent reports are measuring the same underlying behavioral inhibition phenotype, we fit the biometric, psychometric, and the rater bias models (Hewitt et al. 1992; Neale and Maes 2004) to the data. We also assessed the degree to which the latent phenotype underlying the parent report and observation measures are influenced by genetic and

environmental influences, and whether the unique variance of each method is better explained by genetic/environmental influences (psychometric model) or rater bias and residual variance (rater bias model). Figure 1 illustrates each model.

All models in the present study follow the basic twin analysis assumption for estimating genetic, shared environmental, and unique environmental influences (with the exception of the saturated models). The correlations between genetic influences on the twins are set to be equal to 1.0 for monozygotic twins and 0.5 for dizygotic twins because monozygotic twins are genetically identical, whereas dizygotic twins share an average of 50 % of their alleles identical by descent. The correlations between shared environmental influences are set to 1.0 for both monozygotic and dizygotic twins, following the equal environments assumption (e.g., Kendler et al. 1993; Eaves et al. 2003; Derks et al. 2006), which postulates that twins are raised in similar environments, regardless of the degree to which they share genes. Unique environmental influences do not correlate between twin pairs because these influences are unique to each individual (e.g., measurement error, unique environmental stressors). Additionally, an estimate of contrast effects for parent reports was included in each model because it has been demonstrated that parents of dizygotic twins tend to rate their twins as less similar (thus leading to larger variance) than parents of monozygotic twins on measures of temperament (e.g., Saudino et al. 2000).

The biometric model (Fig. 1a) estimates the common genetic (a^2), shared environmental (c^2), and nonshared environmental (e^2) influences on the observations and decomposes the variance of parent reports between those that are shared in common with observations and those that are unique to parent reports. The a_{11} path refers to the genetic influences on observations, the a_{21} path refers to the common genetic influences on parent reports, and the a_{22} path refers to the unique genetic influences on parent reports. By squaring the standardized parameter estimates for each path, an estimate of heritability (a^2) is obtained. The remaining estimates (c^2 and e^2) follow the same pattern.

The psychometric model (Fig. 1b) takes a latent variable approach to fitting the data. It is assumed that there is a latent variable of behavioral inhibition explaining the covariance between observation and parent report. The a^2 , c^2 , and e^2 are estimated for the latent variable. The latent variable factor loadings are fixed to 1 in order for the model to be identified (Hewitt et al. 1992). Separate a^2 , c^2 , and e^2 are estimated for the unique variance of observations and parent reports.

The rater bias model (Fig. 1c; Hewitt et al. 1992) is similar to the psychometric model, in that there is an

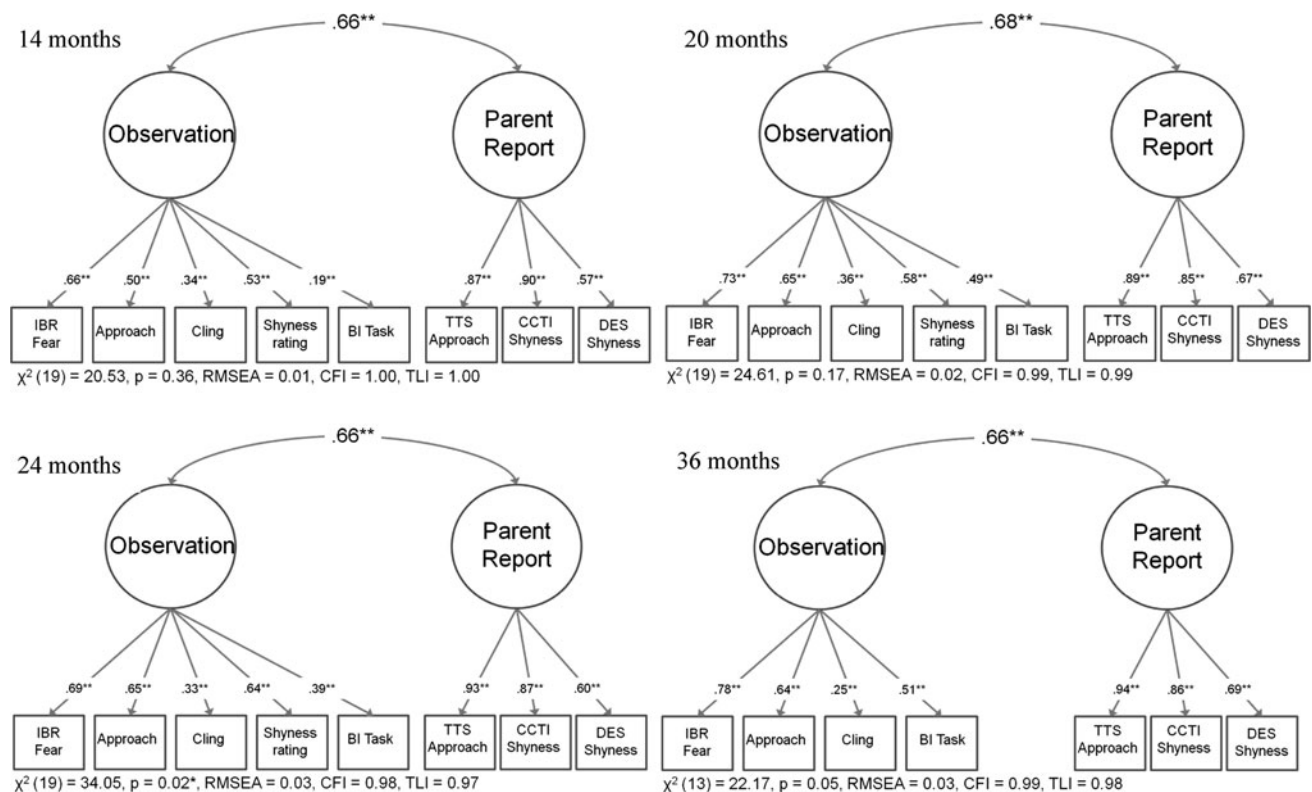


Fig. 2 CFA results. *Significant p value of $<.05$; **significant p value of $<.01$. *IBR Fear* Infant Behavior Record fearfulness ratings, *Approach* Toddler's tendency to approach novel items/strangers during home visit, *Cling* Toddler's tendency to cling to mother during home visit, *shyness rating* shyness during lab visit, *BI task* behavioral

inhibition task, TTS Approach Toddler Temperament Scale approach rating, *CCTI Shyness* Colorado Childhood Temperament Inventory shyness rating, and *DES Shyness* Differential Emotions Scale fearfulness rating

underlying latent common phenotype. However, the latent variable loading on observational measures was fixed to 1 and the loading on parent report was allowed to be free. One loading had to be fixed to 1 for the model to be identified (Hewitt et al. 1992). Additionally, the unique variance of observation and parent report measures are assessed for rater bias (i.e., the tendency for individuals to consistently rater behavior higher/lower compared to other raters) and nonshared environmental influences.

As noted above, the parameterization of the latent factor differs between the psychometric and rater bias models. In the rater bias model, the loading on parent report was allowed to be free whereas in the psychometric model, both loadings were fixed to one. As Hewitt et al. (1992) noted, fixing the factor loadings to be one in the psychometric model allows the model to be identified and unique parameterization of the variances of each measure; otherwise, shared environmental influences in the psychometric model would be equal to the bias parameters in the rater bias model and nonshared environmental influences in the psychometric model would be equal to the residual variance parameter in the rater bias model.

Model fit comparisons were conducted with a χ^2 difference test of the -2 log likelihood (-2ll) scores when the models being compared were nested (i.e., the comparisons between the saturated model and biometric, psychometric, and rater bias models, and the comparison between the biometric and rater bias models; Hewitt et al. (1992) provide substantive information regarding model comparisons). In comparing the fit of the biometric and psychometric models, which are not nested and have the same degrees of freedom due to the latent factor loading constraint in the psychometric model, the -2ll and Akaike's information criterion (AIC) values were compared, with lower -2ll and AIC values indicating better fit. To determine the overall fit of each model, saturated models (which estimate the means, variances, and covariances without constraints) were fit at each time point, and a χ^2 difference test was calculated between the saturated model and estimated models. Fit indices (i.e., RMSEA and TLI) were estimated using information from the saturated, null, and estimated models (Neale et al. 2003). To assess the statistical significance of parameter estimates, individual parameter paths were dropped and a one degree of

freedom χ^2 difference test was computed to determine if dropping the parameter resulted in a significant decrement in fit. When a parameter could not be dropped from a model (i.e., specific nonshared environmental influences, the variance unique to each twin, include measurement error, and are thus necessary in the biometric models), 95 % confidence intervals were calculated to determine statistical significance.

Results

Factor analysis

Eight variables were included in the EFA of the behavioral inhibition data. Initial EFA results suggested that there were two factors underlying the behavioral inhibition measures, given that there were two eigenvalues greater than one, and the two-factor model fit the data well at each age (results available upon request). The five observational measures significantly loaded on the first factor and the three parental report measures significantly loaded on the second factor.

The factor structures suggested by the EFA were then tested using CFA. Figure 2 presents the CFA results. At each age, the two-factor model suggested by the EFA fit the data well according to model fit statistics (as described above). At each age, all factor loadings were significant at $p < .01$, and the correlations between the factors were significant at $p < .01$. Based on these findings, factor scores computed by Mplus were used for the following analyses.

Descriptive statistics and correlations

Table 1 provides descriptive statistics (means and standard deviations) for boys and girls by zygosity groups (MZ and DZ) at each time point. Regressions were conducted to determine whether the behavioral inhibition scores were significantly different between gender groups, zygosity groups, and whether there was an interaction between gender and zygosity (Table 1). Generally, girls had higher behavioral inhibition scores than boys, and there appeared to be an interaction between zygosity and gender at some ages, with monozygotic females generally having higher behavioral inhibition ratings than dizygotic females.

Table 2 provides the phenotypic, cross-twin within-trait, and cross-twin cross-trait correlations. The phenotypic correlations were fixed to be equal across zygosity because of the assumption that phenotypic correlations will not vary as a function of zygosity, but were free to vary across sex. The phenotypic correlations indicated that observational and parent report measures were significantly correlated

within individuals. The cross-twin within-trait correlations provide evidence for genetic influences on behavioral inhibition because the monozygotic twin correlations were consistently larger than the dizygotic twin correlations. The cross-twin within-trait correlations provided evidence of contrast effects present in the parental ratings (Neale and Maes 2004), as the ratings from parents of dizygotic twin were weakly and/or negatively correlated, but the observer ratings of dizygotic twins were statistically significant and positive. Further, MZ and DZ variances could be equated for observation measures, but not parent report, for which DZ variance was larger than MZ variance, a predicted consequence of contrast effects or sibling competition (Neale and Cardon 1992). Because of these findings, a contrast effect for parent report was modeled in the following analyses.

Model fitting

Sex differences

A sex difference model was tested in Mplus at each time point in models estimating the phenotypic, cross-twin within-trait, and cross-twin cross-trait correlations to determine whether there were significant sex differences. A model allowing the correlations to vary across sex provided a better fit to the data only at 36 months ($\Delta\chi^2(7) = 14.36$, $p = .05$; data available upon request). Cross-twin, cross-trait correlations were larger in the female group at 36 months. Also, we compared the fit of the model where the parameter estimates were constrained to be equal across sex and where the parameter estimates were free to vary across sex when testing the biometric, psychometric, and rater bias models in Mx. Similarly, there was only evidence of sex differences in the parameter estimates at 36 months (all $\Delta\chi^2$ p values $< .01$; data available upon request). Based on these results, sex differences were described at 36 months only.

Contrast effect estimate

As noted above, the negative DZ correlations and the greater DZ variances than MZ variances suggested the parent report measures were affected by a contrast parameter (Neale and Maes 2004). Therefore, a contrast effect was included for parent reports in each model tested in Mx (Fig. 1). For all models (i.e., biometric, psychometric, and rater bias) at each time point, the contrast parameter was negative, indicating that a higher inhibition rating for one twin by the parent was contributing to a lower rating for the second twin and vice versa (14 months: $b = -.29$ to $-.40$; 20 months: $b = -.42$ to $-.45$; 24 months: $b = -.22$ to $-.25$; 36 months: $b = -.17$ to $-.48$).

Table 2 Phenotypic, cross-twin within-trait, and cross-twin cross-trait correlations

	14 months				20 months				24 months				36 months			
	Obs-t1	Parent-t1	Obs-t2	Parent-t2	Obs-t1	Parent-t1	Obs-t2	Parent-t2	Obs-t1	Parent-t1	Obs-t2	Parent-t2	Obs-t1	Parent-t1	Obs-t2	Parent-t2
MZF																
Obs-t1	1.0				1.0				1.0				1.0			
Parent-t1	0.41**	1.0			0.49**	1.0			0.49**	1.0			0.50**	1.0		
Obs-t2	0.52**	0.35**	1.0		0.65**	0.37**	1.0		0.69**	0.45**	1.0		0.70**	0.46**	1.0	
Parent-t2	0.35**	0.48**	0.41**	1.0	0.37**	0.58**	0.49**	1.0	0.45**	0.59**	0.49**	1.0	0.46**	0.60**	0.50**	1.0
DZF																
Obs-t1	1.0				1.0				1.0				1.0			
Parent-t1	0.41**	1.0			0.49**	1.0			0.49**	1.0			0.50**	1.0		
Obs-t2	0.40**	0.07	1.0		0.50**	0.15**	1.0		0.46**	0.19**	1.0		0.56**	0.15	1.0	
Parent-t2	0.07	-0.13*	0.41**	1.0	0.15**	-0.06	0.49**	1.0	0.19**	0.07	0.49**	1.0	0.15	-0.02	0.50**	1.0
MZM																
Obs-t1	1.0				1.0				1.0				1.0			
Parent-t1	0.44**	1.0			0.47**	1.0			0.46**	1.0			0.41**	1.0		
Obs-t2	0.55**	0.38**	1.0		0.61**	0.36**	1.0		0.71**	0.42**	1.0		0.49**	0.30**	1.0	
Parent-t2	0.38**	0.52**	0.44**	1.0	0.36**	0.58**	0.47**	1.0	0.42**	0.51**	0.46**	1.0	0.30**	0.61**	0.41**	1.0
DZM																
Obs-t1	1.0				1.0				1.0				1.0			
Parent-t1	0.44**	1.0			0.47**	1.0			0.46**	1.0			0.41**	1.0		
Obs-t2	0.43**	0.08	1.0		0.47**	0.15**	1.0		0.48**	0.18**	1.0		0.21*	0.06	1.0	
Parent-t2	0.08	-0.14*	0.44**	1.0	0.15**	-0.06	0.47**	1.0	0.18**	0.06	0.46**	1.0	0.06	-0.02	0.41**	1.0

Obs-t1 observational factor, twin 1; *Parent-t1* parent rating factor, twin 1; *Obs-t2* observational factor, twin 2; *Parent-t2* parent rating factor, twin 2; *MZF* monozygotic female group; *DZF* dizygotic female group; *MZM* monozygotic male group; *DZM* dizygotic male group

* Significant p value of <0.05; ** significant p value of <0.01

Table 3 Proportion of variance explained by the parameter estimates from the biometric model

	Method	A1	C1	E1	A2	C2	E2
14 months	Observation	0.25**	0.29*	0.47*			
	Parent report	0.68**	0.00	0.00	0.11**	0.00	0.20*
20 months	Observation	0.28**	0.35**	0.37**			
	Parent report	0.18**	0.14	0.01	0.27	0.34	0.08
24 months	Observation	0.47**	0.23	0.30*			
	Parent report	0.29**	0.04	0.00	0.31*	0.16	0.19
36 months—female	Observation	0.32**	0.39*	0.29*			
	Parent report	0.46**	0.02	0.00	0.17	0.23	0.12
36 months—male	Observation	0.49**	0.00	0.51*			
	Parent report	0.23**	0.00	0.01*	0.54	0.00	0.22*

Parameter estimates were individually dropped from the full model to test statistical significance. A series of one degree of freedom χ^2 difference tests were computed comparing the full model to each model with the dropped parameter. Statistical significance is indicated when dropping the parameter from the model results in a significant decrement in fit. The specific nonshared environmental parameters could not be dropped from the model, thus 95 % confidence intervals were obtained for the unstandardized estimates to determine statistical significance (not shown; available upon request)

A1 genetic influences shared in common by observation and parent report, *A2* genetic influences specific to parent report, *C1* shared environmental influences shared in common by observation and parent report, *C2* shared environmental influences specific to parent report, *E1* nonshared environmental influences shared in common by observation and parent report, *E2* nonshared environmental influences specific to parent report

* Significant *p* value of <.05; ** significant *p* value of <.01

Table 4 Standardized parameter estimates from the psychometric model

	Method	Behavioral inhibition latent factor			Proportion of method variance explained			
		A	C	E	Latent	A _u	C _u	E _u
14 months	Observation	0.66**	0.25	0.09**	0.62	0.00	0.00	0.38**
	Parent report				0.29	0.31**	0.30	0.11*
20 months	Observation	0.43**	0.46	0.11	0.66	0.00	0.04	0.30**
	Parent report				0.37	0.25**	0.33	0.04*
24 months	Observation	0.75**	0.20	0.05	0.63	0.00	0.10	0.26**
	Parent report				0.38	0.31**	0.13	0.17**
36 months—female	Observation	0.53**	0.43	0.04	0.65	0.00	0.10	0.26**
	Parent report				0.36	0.13	0.46	0.05*
36 months—male	Observation	0.81*	0.00	0.19	0.56	0.04	0.00	0.40**
	Parent report				0.30	0.52**	0.00	0.18**

The subscript *u* refers to unique A, C, and E influences on each method's variance after accounting for proportion of variance explained by the latent factor. Parameter estimates were individually dropped from the full model to test statistical significance. A series of one degree of freedom χ^2 difference tests were computed comparing the full model to each model with the dropped parameter. Statistical significance is indicated when dropping the parameter from the model results in a significant decrement in fit. Statistical significance of the variance explained by the latent variable was not assessed because the factor loadings are fixed in the psychometric model

* Significant *p* value of <.05; ** significant *p* value of <.01

Biometric model

The standardized variance components (i.e., proportion of variance explained) from the biometric model are presented in Table 3. The biometric model provided evidence of common genetic influences between parent report and observational measures at each time point. The parameter estimates for observational measures suggest small to

moderate genetic and environmental influences on the behavioral inhibition phenotype. The parent report variance was decomposed into that shared in common with observational measures (*A1*, *C1*, and *E1*) and that unique to parent report (*A2*, *C2*, and *E2*). Genetic influences shared in common by parent report and observation measures (*A1*) had moderate influences on parent report, whereas the *C1* and *E1* estimates indicate little to no common

environmental influences between parent report and observational measures. A2, which influences only parent report, indicated moderate to substantial genetic effects; C2 and E2 indicated small to moderate environmental influences on parent report.

Psychometric model

The standardized variance components from the psychometric model are presented in Table 4. The results from the psychometric model suggest that there are moderate to substantial genetic influences and small to moderate environmental influences on the latent behavioral inhibition factor underlying observations and parent reports. At 36 months, there were substantial genetic influences on the latent behavioral inhibition factor and no shared environmental influences in males, in contrast to moderate genetic and shared environmental influences in females.

The latent phenotype accounted for a moderate to substantial proportion of the observational measure variance. Unique nonshared environmental influences accounted for a moderate proportion of the remaining variance of observational measures, whereas little to no variance was explained by unique genetic or shared environmental influences. The latent phenotype accounted for a moderate proportion of the parent report variance. There were modest to moderate unique genetic and environmental influences on parent report. At 36 months, there were moderate unique genetic and no unique shared environmental influences on parent report in males, whereas there

were modest unique genetic and moderate unique shared environmental influences on parent report in females.

Rater bias model

Standardized variance components (i.e., proportion of variance explained) from the rater bias model are provided in Table 5. The rater bias model indicated that genetic influences on the latent common phenotype were moderate to substantial. Shared environmental influences accounted for moderate variance of the latent variable at 20 months, and a small proportion of the variances in females at 36 months.

Observational measures were moderately influenced by the latent variable, rater bias, and residual variance. Parental report measures had a substantial proportion of variance explained by the latent variable, a small proportion of variance explained by the residual variance, and little to no variance explained by rater bias.

Model fit comparisons

Table 6 provides the model fit statistics. No model fit significantly worse than the saturated model, and all fit indices indicate that the estimated models fit the data well (i.e., RMSEA smaller than .05 and TLI greater than .95). Psychometric and biometric models have the same degrees of freedom, so a test of χ^2 difference could not be calculated. At 14, 20, and 36 months, the rater bias model did not fit significantly worse than the biometric model. At 20,

Table 5 Standardized parameter estimates from the rater bias model

	Method	Latent variable estimates			Proportion of method variance explained		
		A	C	E	Latent variable	Rater bias	Residual variance
14 months	Observation	0.91**	0.00	0.09*	0.22	0.33	0.45*
	Parent report				0.87	0.01	0.12
20 months	Observation	0.53**	0.37	0.10**	0.27	0.36	0.37*
	Parent report				0.92	0.08	0.00
24 months	Observation	0.92**	0.06	0.02	0.36	0.32	0.32*
	Parent report				0.68	0.12	0.21*
36 months—female	Observation	0.83**	0.13	0.05	0.30	0.40	0.29*
	Parent report				0.78	0.13	0.09
36 months—male	Observation	0.81**	0.00	0.19	0.18	0.26	0.56*
	Parent report				0.94	0.00	0.06

Parameter estimates were individually dropped from the full model to test statistical significance. A series of one degree of freedom χ^2 difference tests were computed comparing the full model to each model with the dropped parameter. Statistical significance is indicated when dropping the parameter from the model results in a significant decrement in fit. The residual variance parameters could not be dropped from the model; 95 % confidence intervals were obtained for the unstandardized estimates to determine statistical significance (not shown; available upon request). Statistical significance of the variance explained by the latent variable was not assessed because the factor loadings were fixed for the first variable in the rater bias model

* Significant p value of <.05; ** significant p value of <.01

Table 6 Model fit statistics/comparisons

	Model	–2ll	df	AIC	TLI	RMSEA	Δ	$\Delta - 211$	Δdf	p
14 months	(1) Saturated	3449.02	1,576	297.02						
	(2) Biometric	3493.27	1,620	253.27	1.00	0.01	2 vs. 1	44.25	44	.46
	(3) Psychometric	3497.48	1,620	257.48	0.99	0.02	3 vs. 1	48.45	44	.30
	(4) Rater bias	3493.42	1,621	251.42	1.00	0.00	4 vs. 1	44.39	45	.50
							4 vs. 2	0.14	1	.70
20 months	(5) Saturated	3341.16	1,576	189.16						
	(6) Biometric	3396.38	1,620	156.38	0.99	0.04	6 vs. 5	55.22	44	.12
	(7) Psychometric	3396.40	1,620	156.40	0.99	0.04	7 vs. 5	55.23	44	.12
	(8) Rater bias	3399.67	1,621	157.67	0.98	0.04	8 vs. 5	58.51	45	.09
							8 vs. 6	3.29	1	.07
24 months	(9) Saturated	3366.66	1,576	214.66						
	(10) Biometric	3410.34	1,620	170.34	1.00	0.00	10 vs. 9	43.68	44	.49
	(11) Psychometric	3410.34	1,620	170.34	1.00	0.00	11 vs. 9	43.68	44	.49
	(12) Rater bias	3415.33	1,621	173.33	1.00	0.02	12 vs. 9	48.67	45	.33
							12 vs. 10	4.99	1	.03
36 months	(13) Saturated	3287.62	1,576	135.62						
	(14) Biometric	3327.00	1,608	111.00	1.01	0.00	14 vs. 13	39.38	32	.17
	(15) Psychometric	3327.60	1,608	111.60	1.01	0.00	15 vs. 13	39.98	32	.16
	(16) Rater bias	3331.05	1,610	111.05	1.00	0.00	16 vs. 13	43.43	34	.13
							16 vs. 14	4.06	2	.13

–2ll negative two log-likelihood, *df* degrees of freedom, *AIC* Akaike's information criterion, *TLI* Tucker–Lewis index, *RMSEA* root mean square error of approximation

24, and 36 months, the psychometric model had an almost identical model fit as the biometric model. Generally, these findings indicate that a behavioral inhibition latent construct fits the data well and support a common phenotype approach.

Discussion

The present study was conducted to assess the extent to which observational and parent report measures assess a common behavioral inhibition phenotype, and the degree to which the phenotype is influenced by genetic and environmental factors. The aims were addressed by fitting theory-driven models to the data (Hewitt et al. 1992). We found that the covariance between observational and parent report measures was significant, and either rater bias or psychometric models fit the data well at each age. Overall, these results provide support for a latent common phenotype underlying observations and parent report at each age. Additionally, all analyses indicated moderate to substantial genetic influences on the common latent phenotype of behavioral inhibition. These findings are consistent with earlier reports on the heritability of shyness and other inhibited behaviors (e.g., Emde et al. 1992; Eley et al. 2003). It also suggests that the common phenotype

underlying both parent report and observations is more heritable than the unique variance of these measures, which are also influenced by measurement error.

The proportion of variance of each method explained by the behavioral inhibition latent variable was reversed in size between the psychometric and rater bias models. When interpreting this result, it is important to note that the model constraints affected the individual parameters within each model. In the psychometric model, the latent variable factor loadings are both fixed to one, whereas in the rater bias model, the latent variable factor loading is freed for parent report and fixed to one for observations. Thus, the parameter estimates cannot be compared directly across models.

Additionally, it is important to note that the heritability of the latent variable is also not comparable across the psychometric and rater bias models because there is only one source of genetic influences in the rater bias model, whereas there are genetic influences on both the latent variable and method-specific variances in the psychometric model. It is possible that the limited genetic/environmental structure of the model of the rater bias model have influenced the results, with the latent variable having a greater influence on the parent report than the observations.

In the present study, we found that girls have a higher level of behavioral inhibition than boys, with this gender

difference being significant at most ages. The differences observed in the present study (see beta coefficients for gender, Table 1) are consistent with the conclusion of a meta-analysis indicating that girls have slightly higher rates of shyness than boys (weighted mean effect size $d = -0.10$; Else-Quest et al. 2006). We also found that the magnitude of genetic and environmental influences could be constrained across gender from 14 to 24 months, but significant sex differences in the parameter estimates emerged at 36 months, with suggestion of greater shared environmental influences on behavioral inhibition in girls than in boys. One possible explanation for the appearance of sex differences in the parameters for the first time at age 36 months is the emergence of differential social rules, in which males face higher social consequences (e.g., peer exclusion, negative parent response) for inhibited behaviors compared to females (e.g., Rubin and Coplan 2004; Karevold et al. 2011). For example, one review found rates and types of problematic behaviors diverge between the sexes around age 4, and the authors posit that this may be due to socialization factors and the development of adaptive skills (Keenan and Shaw 1997). However, it may also be possible that parents and/or examiners have different thresholds for rating inhibited behaviors in boys versus girls, although it is unclear why that difference would emerge at 36 months. Another possibility is that we did not have statistical power to detect small sex differences from 14 to 24 months.

Twin studies examining behavioral inhibition have reported slightly mixed findings of gender effect on the heritability of behavioral inhibition. For example, a study that examined parent reported shyness in toddlers (in the same sample examined here) reported no significant decrement in the fit of a model constraining parameters to be equal across gender compared to that allowing parameters to be free across gender (Rhee et al. 2007). Eley et al. (2003) found significant sex differences in a twin study of shyness/inhibition 4-year-olds, with results indicating that the magnitude of additive genetic effects and contrast effects were greater in males, whereas the magnitude of nonshared environmental influences was larger in females. It is possible that differences in the measurement of behavioral inhibition and age at which behavioral inhibition was assessed may have led to inconsistent results in the literature.

One of the strengths of the present study was the availability of several observational and parental report measures. Creating factor scores from a variety of measures likely reduced the measurement error inherent in any singular measure. Additionally, the availability of longitudinal data allowed us to examine the variables across different time points in toddlerhood. Overall, the results were consistent across time points, which provide higher

confidence in our findings. However, a limitation of the study was that the assessment of some of the variables changed over time, which limited our ability to examine these data longitudinally. Additionally, the use of the same sample over time may contribute to consistency within the study, and these results need to be replicated with other study samples. Also, some models' fit could not be compared statistically because they were not nested (i.e., psychometric and biometric models); however, the examination of $-2\ln$ and AIC values allowed us to examine whether the psychometric model fit as well as the biometric models (i.e., smaller values indicates better fit). Another potential limitation is that the sample size may be small for finding small effects, especially in complex multivariate models.

In sum, the present study found that observational and parental report measures of toddler behavioral inhibition share significant variance and reflect a common view of toddler behavioral inhibition (Aim 1). Additionally, there is a moderately to substantially heritable latent common phenotype underlying observational and parent report measures of behavioral inhibition (Aim 2). By combining across the two assessment methods, we can derive a more valid phenotype. By examining the more valid latent common phenotype, studies examining behavioral inhibition as a predictor of later outcomes or those examining the genetics of behavioral inhibition may increase their power while being less affected by the methodological limitations inherent in observations or parent reports.

Acknowledgments This research was supported by grants from the MacArthur Foundation, and grants NIH HD010333, HD050346, HD007289, and MH063207. The authors thank the participants and research assistants for their participation and assistance with this project.

References

- Askan N, Kochanska G (2004) Links between systems of inhibition from infancy to preschool years. *Child Dev* 75(5):1477–1490
- Bayley N (1969) Manual for the Bayley scales of infant development. Psychological Corporation, New York
- Bentler PM (1990) Comparative fit indexes in structural models. *Psychol Bull* 107:238–246
- Browne MW, Cudeck R (1993) Alternative ways of assessing model fit. In: Bollen KA, Long JS (eds) *Testing structural equation models*. Sage, Newbury Park, pp 136–162
- Carey WB, McDevitt SC (1978) Stability and change in individual temperament diagnoses from infancy to early childhood. *Am Acad Child Psychiatry* 17:331–337
- Caspi A, Moffitt TE, Newman DL, Silva PA (1996) Behavioral observations at age 3 years predict adult psychiatric disorders: longitudinal evidence from a birth cohort. *Arch Gen Psychiatry* 53:1033–1039
- Cronbach LJ, Rajaratnam K, Gleser GC (1963) Theory of generalizability: a liberalization of reliability theory. *Br J Math Stat Psychol* 16:137–163

- Derks EM, Dolan CV, Boomsma DI (2006) A test of the equal environment assumption (EEA) in multivariate twin studies. *Twin Res Hum Genet* 9(3):403–411
- Eaves L, Foley D, Silberg J (2003) Has the ‘equal environments’ assumption been tested in twin studies? *Twin Res* 6:486–489
- Eley TC, Bolton D, O’Connor TG, Perrin S, Smith P, Plomin R (2003) A twin study of anxiety-related behaviors in pre-school children. *J Child Psychol Psychiatry* 44(7):945–960
- Else-Quest NM, Shibley Hyde J, Goldsmith HH, Van Hulle CA (2006) Gender differences in temperament: a meta-analysis. *Psychol Bull* 132(1):33–72
- Emde RN, Plomin R, Robinson J, Corley R, DeFries J, Fulker DW, Reznick JS, Campos J, Kagan J, Zahn-Waxler C (1992) Temperament, emotion, and cognition at fourteen months: the MacArthur longitudinal twin study. *Child Dev* 63:1437–1455
- Gest SD (1997) Behavioral inhibition: stability and associations with adaptation from childhood to early adulthood. *J Pers Soc Psychol* 72(2):467–475
- Goldsmith HH, Lemery KS (2000) Linking temperamental fearfulness and anxiety symptoms: a behavior-genetic perspective. *Biol Psychiatry* 48:1199–1209
- Hewitt JK, Silberg JL, Neale MC, Eaves LJ, Erickson M (1992) The analysis of parental ratings of children’s behavior using LISREL. *Behav Genet* 22(3):293–317
- Hu L, Bentler PM (1998) Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychol Methods* 3:424–453
- Izard CE, Huebner R, Risser D, McGinnes GC, Dougherty L (1980) The young infant’s ability to produce discrete emotional expressions. *Dev Psychol* 16:132–140
- Kagan J (2001) The structure of temperament. In: Emde RN, Hewitt JK (eds) *Infancy to early childhood: genetic and environmental influences on developmental change*. Oxford University Press, Oxford, pp 45–51
- Kagan J, Saudino KJ (2001) Behavioral inhibition and related temperaments. In: Emde RN, Hewitt JK (eds) *Infancy to early childhood: genetic and environmental influences on developmental change*. Oxford University Press, Oxford, pp 111–119
- Kagan J, Reznick JS, Clarke C, Snidman N, Garcia-Coll C (1984) Behavioral inhibition to the unfamiliar. *Child Dev* 55:2212–2225
- Karevold E, Coplan R, Stoolmiller M, Mathiesen KS (2011) A longitudinal study of the links between temperamental shyness, activity, and trajectories of internalising problems from infancy to middle childhood. *Aust J Psychol* 63:36–43
- Keenan K, Shaw D (1997) Developmental and social influences on young girls’ early problem behavior. *Psychol Bull* 121(1):95–113
- Kendler KS, Baker JH (2007) Genetic influences on measures of the environment: a systematic review. *Psychol Med* 37:615–626
- Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ (1993) A test of the equal-environment assumption in twin studies of psychiatric illness. *Behav Genet* 23(1):21–27
- Kochanska G, DeVet K, Goldman M, Murray K, Putnam SP (1994) Maternal reports of conscience development and temperament in young children. *Child Dev* 65(3):852–868
- Lahey BB, Waldman ID (2003) A developmental propensity model of the origins of conduct problems during childhood and adolescence. In: Lahey BB, Moffitt TE, Caspi A (eds) *Causes of conduct disorder and juvenile delinquency*. Guilford, New York, pp 76–117
- Letcher P, Smart D, Sanson A, Toumbourou JW (2009) Psychosocial precursors and correlates of differing internalizing trajectories from 3 to 15 years. *Soc Dev* 18(3):618–646
- Matheny AP (1980) Bayley’s infant behavior record: behavioral components and twin analyses. *Child Dev* 51(4):1157–1167
- Muthén LK, Muthén BO (1998–2010) *Mplus user’s guide*, 6th edn. Muthén & Muthén, Los Angeles
- Neale MC, Cardon LR (1992) *Methodology for genetic studies of twins and families*. Kluwer, Dordrecht
- Neale MC, Maes HHM (2004) *Methodology for genetic studies of twins and families*. Kluwer, Dordrecht
- Neale MC, Boker SM, Xie G, Maes HH (2003) *Mx: statistical modeling*, 6th edn. Department of Psychiatry, VCU, Richmond
- Rhea S, Gross AA, Haberstick BC, Corley RP (2006) Colorado twin registry. *Twin Res Hum Genet* 9(6):941–949
- Rhee SH, Waldman ID (2002) Genetic and environmental influences on antisocial behavior: a meta-analysis of twin and adoption studies. *Psychol Bull* 128(3):490–529
- Rhee SH, Cosgrove VE, Schmitz S, Haberstick BC, Corley RC, Hewitt JK (2007) Early childhood temperament and the covariation between internalizing and externalizing behavior in school-aged children. *Twin Res Hum Genet* 10(1):33–44
- Robinson JL, Kagan J, Reznick JS, Corley R (1992) The heritability of inhibited and uninhibited behavior: a twin study. *Dev Psychol* 28(6):1030–1037
- Rothbart MK, Goldsmith HH (1985) Three approaches to the study of infant temperament. *Dev Rev* 5:237–260
- Rowe DC, Plomin R (1977) Temperament in early childhood. *J Pers Assess* 41:150–156
- Rubin KH, Coplan RJ (2004) Paying attention to and not neglecting social withdrawal and social isolation. *Merrill-Palmer Q* 50(4):506–534
- Sanson A, Pedlow R, Cann W, Prior M, Oberklaid F (1996) Shyness ratings: stability and correlates in early childhood. *Int J Behav Dev* 19(4):705–724
- Saudino KJ, McGuire S, Reiss D, Hetherington EM, Plomin R (1995) Parent ratings of EAS temperaments in twins, full siblings, half siblings, and step siblings. *J Pers Soc Psychol* 68(4):723–733
- Saudino KJ, Cherny SS, Plomin R (2000) Parent ratings of temperament in twins: explaining the ‘too low’ DZ correlations. *Twin Res* 3:224–233
- Terman LH, Merrill MA (1973) *Stanford–Binet intelligence scale: 1972 norms edition*. Houghton-Mifflin, Boston
- van der Valk JC, van den Oord EJCG, Verhulst FC, Boomsma DI (2001) Using parental ratings to study the etiology of 3-year-old twins’ problem behaviors: different views or rater bias? *J Child Psychol Psychiatry* 42(7):921–931
- Wachs TD (1992) Issues in the measurement of biological and behavioral indices of temperament. In: Paper presented at the ninth occasional temperament conference, Bloomington, Oct 29–31 1992 (cited in Sanson et al. 1996)