Published in final edited form as:

Dev Psychol. 2016 March ; 52(3): 391–399. doi:10.1037/dev0000090.

The development of inhibitory control in early childhood: A twin study from 2-3 years

Jeffrey R. Gagne¹ and Kimberly J. Saudino²

¹University of Texas at Arlington

²Boston University

Abstract

Parent and lab-based observer ratings were employed to examine genetic and environmental influences on continuity and change in inhibitory control (IC) in over 300 twin-pairs assessed longitudinally at 2 and 3 years of age. Genetic influences accounted for approximately 60% of the variance in parent-rated IC at both ages. Although many of the same genetic effects on parent-rated IC were stable across age, there were also novel genetic effects that emerged at age 3 (i.e., genetic factors contributed to both continuity and change in parent ratings of IC). Observed IC displayed a different developmental pattern. Genetic influences were moderate at age 2 (38%) and nonsignificant at age 3 (6%). Change in observed IC across early childhood was due to shared and nonshared environmental factors. Findings indicate that it is important to consider the measurement of IC when interpreting developmental and etiological findings.

Keywords

inhibitory control; continuity and change; etiology; multi-method assessment; early childhood

Introduction

The temperament dimension of inhibitory control (IC) involves the regulation and inhibition of pre-potent behavior, typically in response to instructions or expectations (Kochanska, Murray, Jacques, Koenig & Vandegeest, 1996; Rothbart, 1989). Common examples of IC in early childhood include refraining from obtaining a forbidden snack or running into a busy street in the context of a parent or adult instructing the child to avoid these acts. Children with high IC are able to regulate their behavioral responses in these situations, while those with lower IC tend to behave much more impulsively. In Rothbart's theory of temperament, IC is a component of the broader factor of effortful control (EC), which includes the ability to inhibit responses (i.e., IC), but also involves the *activation* of responses and the process of executive attention (Rothbart & Ahadi, 1994; Rothbart & Bates, 2006). Emotional and motor reactivity emerge in infancy, and as the child enters toddlerhood, more self-regulatory systems of temperament influence this reactivity, exemplified by the emergence of EC (Rothbart & Bates, 2006). Relatedly, the development of executive attention between the

Address correspondence to: Jeffrey R. Gagne, Department of Psychology, University of Texas at Arlington, Box 19528, Arlington, TX 76019, jgagne@uta.edu, Phone: (817) 272-0922, Fax: (817) 272-2364.

ages of 2 to 7 is characterized by developmental changes in important skills associated with negotiating conflict/inhibition, error detection and the slowing of behavioral responses (Rothbart & Bates, 2006). According to Rothbart & Bates (2006), the development of EC and IC is strongly related to executive attention focusing skill. Therefore, for most children, IC emerges late in the second year of life, and develops throughout the toddlerhood and preschool periods (Gagne & Goldsmith, 2011; Gagne & Saudino, 2010; Rothbart, 1989). Understanding IC in early childhood is particularly important because individuals with low IC have more cognitive and socio-emotional development issues such as externalizing behavior problems and associated Attention Deficit Hyperactivity Disorder (ADHD) psychopathology (Eisenberg et al., 2001; 2004; Gagne, Saudino & Asherson, 2011; Schachar, Tannock, Marriott, & Logan, 1995).

The early development of problem behavior typically occurs in the toddlerhood and preschool periods (Campbell, 1995; Keenan & Wakschlag, 2000), and children with behavior problems at this early stage are at increased risk for several poor developmental outcomes (Saudino, Carter, Purper-Ouakil, & Gorwood, 2008). Therefore, it is important to consider the normative development of temperament traits that influence behavioral maladjustment earlier in childhood. Research on IC as an aspect of early emerging temperament (vs. IC as a complex executive function in later childhood) has yielded important initial findings on behavioral maladjustment and etiology. Consistent with EF findings in middle childhood, low IC and EC temperaments have been linked to higher levels of attention and externalizing behavior problems from early childhood through adolescence (Eisenberg et al., 2001; 2003; 2005; Gagne et al., 2011; Lemery-Chalfant, Doelger & Goldsmith, 2008; Murray & Kochanska, 2002; Nigg, Quamma, Greenberg, & Kusche 1999; Olson, Schilling, & Bates, 1999; Polderman et al., 2009; Valiente et al., 2003).

Although IC emerges in early childhood, it is most often studied in middle childhood from the executive functioning (EF) perspective, using standard cognitively-based EF tasks to assess IC often in the context of diagnostic assessments of ADHD. Investigations of earlier-emerging IC are rare because standard EF tasks are too difficult for younger children to complete, and children are not often diagnosed with ADHD as early as toddlerhood. However, basic inhibition and delay aversion IC tasks can be used with children as young as 2 years of age within the framework of a laboratory assessment of temperament. For example, the preschool version of the Laboratory Temperament Assessment Battery (Lab-TAB; Goldsmith, Reilly, Lemery, Longley & Prescott, 1995), a comprehensive temperament assessment battery that is designed to multiple dimensions of temperament, includes several behavioral episodes designed to elicit IC in young children. Despite this, most early temperament research exploring IC and related dimensions in typically developing samples employs only parent-ratings which can be susceptible to various rating biases, including contrast effects (Saudino, 2003). One noteworthy exception, a longitudinal study of EC in early childhood used composited mother report and laboratory measures of delay ability, slowing motor activity, effortful attention, voice lowering, and suppression/initiation of responses to a signal at 22, 33 and 45 months (Kochanska & Knaack, 2003; Kochanska, Murray & Harlan, 2000). These measures of EC were increasingly stable and coherent across development, such that developmental stability was commensurate with the stability

of IQ from 33 to 45 months (Kochanska et al., 2000; Rothbart & Bates, 2006). However, the focus of this study was the broader domain of EC, not IC specifically.

A small number of studies have also found significant genetic variance for IC (Gagne & Goldsmith, 2011; Gagne & Saudino, 2010) and EC temperament dimensions (Goldsmith, Buss, & Lemery, 1997; Lemery-Chalfant et al., 2008; Yamagata, Takahashi, Kijima, Maekawa, Ono, & Ando, 2005). Only two of these studies have examined early IC in the context of a multi-method approach employing parent ratings and lab-based assessments (e.g., Gagne & Goldsmith, 2011; Gagne & Saudino, 2010). Both used Lab-TAB IC episodes as behavioral assessments, and standard parent rating questionnaires in moderately-sized twin samples. In the Boston University Twin Project (BUTP), heritability estimates were .38 and .58 for lab-based observed and parent-rated IC, respectively, at age 2 (Gagne & Saudino, 2010). In a study of 3-year-old twins in the Genetics of Emotional Ontogeny (GEO) project, parent ratings of IC also showed significant genetic variance (.63), but there was no evidence of significant heritability for Lab-TAB IC (Gagne & Goldsmith, 2011). Familial resemblance for this measure of IC was due to shared environmental influences. Across the two studies, therefore, it appears that genetic influences are no longer significant at age 3 when IC is assessed using laboratory ratings, and suggests change in the etiology of IC. In contrast to the lab-based measures of IC, parent ratings showed significant genetic variance in both studies and hint at a stable etiology. Although both the BUTP and GEO samples reflect racial/ethnic distributions of their respective geographic areas (Boston metro-area; greater Madison, WI area), the two samples are very similar in terms of race (88.2-93.2% Caucasian) and socioeconomic status (largely middle class with a range from low to high). Thus, demographic differences are unlikely to explain the differences in outcomes.

These cross-sectional findings are intriguing and suggest that conclusions about developmental change in IC might differ across measures, but longitudinal studies in which both lab-based and parent-rated measures of IC are obtained in the *same sample* are needed to confirm stability and change in the factors that influence individual differences in IC across the transition from infancy to early childhood. In addition, twin studies of early IC focusing on one time point do not inform about the genetic and environmental contributions to continuity and change in IC. Genes switch on and off throughout development and there may be changes in the quantity and quality of genetic effects across age (Plomin & Nesselroade, 1990). Similarly, the environments that children experience differ across age and the role of the environment on individual differences in IC may change as a result. This dynamic nature of genes and environments means that there may be developmental differences in genetic and environmental influences on IC across age (Saudino & Wang, 2012). Consequently, behavioral genetics findings about the etiology of IC at one age may not apply to another.

The present study addresses this gap in the literature by following the BUTP sample when the twins were 3 years of age. Phenotypic research examining the development of IC in early childhood finds moderate stability with a cross-age correlation of .44 between 22 and 33 months of age (Kochanska et al., 2000). Therefore, from an individual differences perspective there is evidence of both stability and change in IC. Little is known, however, about the genetic and environmental contributions to this stability and change. Different

genes and different environmental influences can be relevant to individual differences in IC across age (Saudino, 2012). In addition, different methods of assessing IC may have varying patterns of genetic and environmental variance and contributions to stability and change, an issue that remains largely unexplored. Understanding the factors that influence the development of IC in early childhood is particularly important because this is a period when IC ability is emerging. Genetic influences on developmental change in IC during early childhood may be related to the overlapping development of neural systems underlying self-regulation and executive functioning in early childhood (Berger, 2011). Thus, clarifying the factors that are involved in IC development from a multi-method perspective could have significant implications for the broader field of child development and potentially inform related intervention strategies.

We used longitudinal multivariate model-fitting analyses to estimate the relative contributions of genetic and environmental influences to individual differences in IC at ages 2 and 3, and to evaluate genetic and environmental contributions to continuity and change in IC across age. Our multi-method approach within the same sample allowed us to explore whether parent ratings and lab-based measures yield similar findings regarding the etiology of developmental change in IC in early childhood. If conclusions based on the cross-sectional findings reviewed above are correct, we predicted that the two methods would tell a different developmental story. Specifically, that parent ratings of IC would show significant heritability and common genetic variance across age; whereas for laboratory-assessed IC, shared environmental factors would explain most of the variance at age 3 (representing developmental change). This differential pattern of findings in the same individuals would suggest that the two methods of assessing IC tap different aspects of IC, which in turn, have different underlying developmental processes.

Methods

Participants

The sample comprised 145 monozygotic or MZ twin pairs, and 168 same-sex dizygotic (DZ) twin pairs participating in the BUTP, a longitudinal multi-method study of temperament in early childhood (see Gagne & Saudino, 2010 for details on recruitment procedures). Twins were assessed within 4 weeks of their second and third birthdays. Zygosity was determined through DNA cheek swab samples. Twins with low birth weight (< 1750 g), early gestational age (< 34 weeks) or significant health problems were confidentially identified by the Massachusetts Registry of Vital Statistics and were not recruited into the study. There were approximately equal distributions of male and female (52.8%: 47.2%) twins in the total sample. The racial distribution of the sample was 87.0% White, 3.24% Black, 1.94% Asian, 0.32% Native American, and 7.46% mixed which is generally representative of the state of Massachusetts at the time of data collection (2006 Census of Population and Housing). 4.89% of the sample was classified as Hispanic/Latino. The mean socioeconomic status of the participants was middle class according to the Hollingshead index ($M = 51.2$, $SD = 10.87$), although it ranged from 22-66.

Procedure

The procedure involved two visits, 48 hours apart, to the BUTP laboratory at each age. With few exceptions, noted below, the lab procedures and administration of the assessments were similar at both ages. During the first visit at each age, one twin was assessed in a standardized cognitive test situation while the other twin participated in laboratory situations designed to assess temperament. At the second visit, the situations were reversed. The present analyses focus on behaviors assessed during the temperament episodes and therefore the cognitive testing will not be discussed further. All assessments were administered by research technicians and advanced graduate students who received extensive training on child assessment and Lab-TAB protocols. Within a twin pair, each twin was assessed by a different tester who assessed the same child at both visits. Temperament was assessed via episodes from Lab-TAB (Goldsmith et al., 1995). Because of the specific focus of the BUTP, only the activity level and inhibitory control episodes of the Lab-TAB were administered. Within each visit, the assessments took approximately one hour to complete. The assignment of first- and second-born twins to the cognitive and temperament situations was counterbalanced across the study such that half of all first-born twins participated in the cognitive situation at the first lab visit and half in the second lab visit.

Laboratory measures of inhibitory control

Laboratory-based assessments of IC were conducted with the Lab-TAB episodes “Dinky Toys,” “Snack Delay,” and “Gift.” These episodes employ delay ability tasks that were adapted from Kochanska’s work on IC and EC (Kochanska et al., 1996; Kochanska et al., 2000). Episodes were administered following the procedures outlined in the Lab-TAB manual. Although the episodes are similar across age, there are some subtle differences to take into account the development in cognitive abilities across age. All Lab-TAB IC data was coded from episode videos using the standard Lab-TAB coding protocol (Goldsmith et al., 1995). Coders were trained to a criterion of 90% inter-rater reliability before they were permitted to code episodes independently. Ten percent of the sample was rated by a second observer and inter-rater agreement for the Lab-TAB IC composite variable was high ($r = .89$, $p < .01$).

In the *Dinky Toys* episode, children were given the choice to select one out of an array of colorful and attractive toys, assessing the ability to inhibit the impulse to select multiple toys. At age 2, children were presented with 6 toys from which to choose, and at age 3 they were presented a small container of approximately 20 toys. There were two separate trials during the episode at both ages. Both trials were coded for child’s initial approach to the stimuli, latency to touch the first toy (at age 3 latency to touch the container was also included), latency to choose a toy, style of touching (does not touch any toy, picks first toy, touches two or more toys, shuffles through toys quickly), frequency of touch (age 2 only), number of toys touched, level of distress, following directions, comprehension and interest, and a global rating of impulsivity. For the *Snack Delay* episode, children were offered a snack (a candy or a cracker), and were instructed to wait for a signal before eating it. Experimenters placed the snack under a clear plastic cup and children were told that they needed to wait until the experimenter rang a bell before they could retrieve and eat the snack. There was one practice trial with no wait time before the experimenter rang the bell,

and six test trials with different pause lengths (5s, 10s, 0s, 20s, 0s, 30s at age 2; 20s, 30s, 0s, 40s, 10s, 60s at age 3). The variables coded for each trial included global IC during the instruction phase and the trial, whether the child waited until the signal before eating the snack, the presence of fidgeting and self-distracting behaviors, the latency to fidget and self-distract, and the latency to eat the snack. During the *Gift* episode, children were presented with a wrapped gift and were required to wait two minutes before opening it. Variables for this episode included child IC during instruction phase, fidgeting and distraction behavior, latency to fidget, latency to self-distract, latency to open the gift, frequency of self-corrections, distress, and whether the child opened the gift before the end of the 2-minute interval.

A composite IC score was formed following Goldsmith et al. (1995). The variables used in the summary scores for each episode were selected on the basis of principal component analyses (loadings ranged from .56 to .97). All item-level data were converted to z-scores and averaged across trials. The Dinky Toys summary score at age 2 was created by averaging style of touching, frequency of touches, number of toys touched, latency to choose, following directions, and impulsivity scores. At age 3, the same variables were used to form the summary score with the exception of the latency to choose and frequency of touches. The Snack Delay summary scores were based on the mean of global IC, whether the child waited to eat the snack, and latency to eat at age 2; and the mean of global IC and instruction phase IC, whether the child waited to eat the snack, and average fidgeting behavior at age 3. The age 2 summary score for Gift comprised opening the gift early, frequency and latency to self-distract, and latency to open the gift. At age 3 the summary score also included the frequency of fidgeting. All Lab-TAB episode summary scores were significantly correlated (r s = .10 - .17), and an overall composite of observed IC was computed by averaging the standardized summary scores for each episode at each age.

Parent ratings of inhibitory control

Parent ratings of IC were obtained at both ages using the Inhibitory Control subscale of the Toddler Behavior Assessment Questionnaire-Revised (TBAQ-R; Goldsmith, 1996). The majority of questionnaires were completed by mothers (94%). The TBAQ-R requires the parent to make judgments of IC behaviors in specific situations that occurred over the last month (e.g., “When asked to wait for something (like a toy or a snack), how often did your child find something to distract her/himself until it was time?”) and is rated on a Likert scale from 1 (*never*) to 7 (*always*). This measure is well-established, reliable, valid, and is appropriate for temperament assessment in early childhood (Goldsmith, 1996). Published estimates of internal consistencies for the TBAQ-R range from .86 to .89 (Goldsmith, 1996), and in the current study, Cronbach’s alpha was .82 for the IC subscale.

Statistical approach

Descriptive statistics—Tests of mean sex and zygosity differences, and phenotypic correlational analyses were conducted for parent- and observer-rated IC within and across each time point. These tests were corrected to account for the nested nature of twin data. Generalized Estimating Equation models for correlated data were employed to test for mean differences (sex and zygosity) in IC (Liang & Zeger, 1986; Zeger & Liang, 1986), and dyad-

level correlations were computed with Griffin and Gonzalez' GGEXCH programs for data wherein partners are exchangeable, such as twin data (Griffin & Gonzalez, 1995; O'Connor, 2004).

Twin correlations—Twin intraclass correlations were computed using a double entry procedure in order to provide an index of twin similarity for both parent-rated and lab-based IC. When MZ twin correlations exceed DZ twin correlations for a trait, it suggests that genetic factors contribute to individual differences for that trait. Cross-age, cross-twin correlations where twin A's score for IC at age 2 is correlated with twin B's score for IC at age 3, and vice versa, were calculated for both measures of IC. These cross-correlations are used as an initial attempt to explore multivariate behavior genetic analyses. When cross-age, cross-twin correlations for MZ twins exceed those of DZ twins genetic contributions to the covariance across age are suggested. This is further tested using more powerful model-fitting analyses. Because twin covariances can be inflated by the variance due to gender, scores for all IC variables in the genetic analyses were residualized for sex effects (McGue & Bouchard, 1984).

Model-fitting—To examine genetic and environmental sources of covariance between IC scores across 2 and 3 years of age, bivariate Cholesky decomposition ACE models were fit to raw data for parent-rated and observed IC, separately, using Mx maximum-likelihood model-fitting procedures (Neale, 2003). A path diagram of this model is depicted in Figure 1. The observed phenotypic variances of the IC variables are represented by the rectangles, and the circles represent latent genetic and environmental variables. A1, C1, and E1 signify the genetic, shared environmental, and nonshared environmental factors that influence IC at both age 2 and 3; and A2, C2, and E2 are genetic, shared environmental, and nonshared environmental factors unique to IC at age 3. These models estimate genetic and environmental variance components for IC at each age, and genetic and environmental correlations (i.e., r_g , r_c , r_e) between assessments of IC at age 2 and 3 (i.e., the degree to which genetic or environmental factors for IC at age 2 overlap with those at age 3, *independent of the heritability* at each age). Under this model, "continuity" reflects effects that persist across age and "change" refers to novel effects that emerge at age 3 independent of those operating at age 2.

The overall fit of each ACE model was assessed by calculating twice the difference between the negative log-likelihood (-2LL) of the model and that of a saturated model (i.e., a model in which the variance/covariance structure is not estimated and all variances and covariances for MZ and DZ twins are estimated). The difference in -2LL is asymptotically distributed as χ^2 with degrees of freedom equal to the difference in the number of parameters in the full model and that in the saturated model.

Results

Descriptive statistics and phenotypic correlations

Table 1 presents the means and standard deviations of observer- and parent-rated IC at 2 and 3 years of age for the full sample, males and females, and across twin zygosity. Effect sizes estimated as Cohen's d expressing group differences in standard deviation units between

males and females, and MZ and DZ twins are also reported. Females had higher levels of IC than males across all measures at both time points, and there were no significant differences between MZ and DZ twins. The effect sizes showed that males and females differed by 21-25% of a standard deviation on observed IC, and 33-44% of a standard deviation on parent ratings of IC (gender differences were greater at age 2 for both types of IC assessments). Observer ratings of IC did not significantly correlate across age ($r = .09, p > .05$). In contrast, parents' perceptions of child IC were stable from 2 to 3 years of age ($r = .51, p < .01$). At both ages, observer and parent ratings were only modestly associated (age 2: $r = .19, p < .05$; age 3: $r = .14, p < .05$).

Twin correlations

Twin correlations are presented in Table 2. MZ twin intraclass correlations exceeded DZ intraclass correlations in all cases except with observer ratings of IC at age 3. This pattern of findings indicates the presence of genetic influences on parent- and observer-rated IC at age 2 and parent-rated IC at age 3, but not on observed IC at age 3. Cross-age, cross-twin correlations within each measure of IC for MZ twins exceeded those for the DZ twins, suggesting that the stability in IC across 2 and 3 years of age is partially due to overlapping genetic factors.

Model-fitting

The fit statistics for the bivariate Cholesky decomposition models for both parent-rated and observed IC are presented in Table 3. For both methods of assessing IC, the χ^2 for the ACE model was nonsignificant indicating that the model provided a good fit to the data. Estimates of genetic and environmental variances at each age and covariances across age indexing genetic and environmental overlap between ages 2 and 3 years are presented in Table 4. The variance components for IC at age 3 are decomposed into variances that persist from age 2 ("Continuity") and variances that are unique to age 3 ("Change"). "Continuity" reflects the genetic or environmental effects that are common across both ages and "change" refers to novel effects that emerge at age 3 independent of previous effects operating at age 2. For each variance component (i.e., a^2 , c^2 , and e^2) at age 3, summing the values for continuity and change provides an estimate of the total variance for that component (roughly equivalent to 1.0). For example, the total genetic variance for parent-rated IC at age 3 is .23 (continuity) + .39 (change) = .62, the total shared environmental variance is .13 (continuity) + .00 (change) = .13, and the total nonshared environmental variance is .01 (continuity) + .25 (change) = .26.

For parent-rated IC genetic factors accounted for approximately 60% of the variance at both ages. Shared environmental influences were modest and significant at only 2 years of age. Nonshared environmental influences, which include measurement error, accounted for the remaining variance at both ages. There was significant genetic ($r_g = .61$) and nonshared environmental ($r_e = .23$) covariance between parent ratings of IC at age 2 and 3 indicating that genetic and nonshared environmental factors contribute to continuity in parent ratings of IC across age. Phenotypic change is also due to genetic and nonshared environmental influences. Sixty-three percent of the genetic influences ($.39 / .62 = .63$), and 96% of the

nonshared environmental influences ($.25 / .26 = .96$) on parent-rated IC at age 3 were independent of prior effects at age 2 (i.e., were unique to age 3).

For observer-rated IC, a different pattern emerged. At 2 years of age, genetic influences were significant, accounting for 38% of the variance in observed IC, however, at age 3 genetic influences were nonsignificant and accounted for only 6% of the variance. In contrast, shared environmental variance was estimated as 0% at age 2, but was moderate (24%) and significant at age 3. Thus, familial resemblance in observed IC in the lab was due to genetic factors at 2 years and shared environmental factors at 3 years. Neither the genetic, shared environmental or nonshared environmental correlations across age were significant. This was not unexpected given that the age-to-age phenotypic correlation for observed IC was not significant indicating little stability. There were some modest genetic and nonshared environmental effects that persist across age, although those effects accounted for only 7% of the variance in observed IC at age 3. Change in observed IC is due to novel shared and nonshared environmental influences at age 3.

Discussion

This study examined genetic and environmental influences on individual differences in IC from 2 to 3 years of age using parent and lab-based measures. Previous research has either been confined to one form of assessment (e.g., parent ratings) or to only one age in early childhood (age 2 or 3 in independent samples). This multi-method investigation is the first longitudinal twin study to explore factors that influence continuity and change in IC in early childhood. Employing both types of IC ratings at age 2 and 3 in the same sample provides important information about the early development and etiology of this temperament trait. As is typical of much temperament research (e.g., Gagne et al., 2011; Goldsmith, Rieser-Danner & Briggs, 1991; Saudino, Wertz, Gagne & Chawla, 2004), the agreement between the two methods of IC was low, suggesting important differences between parent and lab ratings of IC in early childhood. This is further supported by different patterns of continuity and change for each method of assessment. These findings indicate that parent and lab-based temperament measures assess different aspects of IC.

As expected, genetic and nonshared environmental influences were present for parent ratings of IC at age 3. Although significant at age 2, shared environmental influences were not significant one year later. The phenotypic association between parent-assessed IC across age was primarily due to genetic covariance and to a lesser extent, nonshared environmental covariance. At age 3, effects of the nonshared environment were almost entirely new, whereas approximately half of the genetic effects were new. Therefore, continuity from age-to-age is largely due to genetic factors, and change is due to both genetic and nonshared environmental factors. The appearance of new nonshared environmental influences at 3 years is consistent with previous studies of other temperament dimensions (McGue, Bacon & Lykken, 1993; Saudino, 2012), and likely represents developmental increases in a child's exposure to unique environmental experiences as they age. Nonshared environmental variance also includes measurement error, however, given that our measures were highly reliable, it is unlikely that measurement error is responsible for all nonshared effects and allows for more interesting interpretations. Relevant sources of nonshared environment may

include differential treatment by parents and other experiences that are specific to the child (Plomin, DeFries, McClearn & Rutter, 1997). For example, one twin may receive more negativity from a parent, or experience an accident or illness that their co-twin did not. Between the ages of 2 and 3, parents may increasingly structure activities (Saudino, 2012) or develop higher expectations of their children related to IC and to the extent to which these differ across co-twins, they would be nonshared environmental effects.

Although the heritabilities for parent-rated IC were the same at ages 2 and 3, this does not mean that the same genetic factors operate across age. The genetic correlation across age was moderate and significant new genetic effects emerged at age 3. A similar developmental pattern has been found for other temperament dimensions across the same ages. For instance, there is evidence of novel genetic effects emerging at age 3 for activity level (Saudino & Cherny, 2001; Saudino, 2012). Self-regulation increases in early childhood typically in the third year (Kochanska et al., 2000), and IC is considered an aspect of self-regulation. Therefore, new genetic effects on IC could reflect emerging self-regulation in the third year of life. This developmental change may be associated with ongoing neurological maturation, including the frontal lobes, underlying the ability to focus attention and inhibit cognitions and motor behavior (Berger, 2011). Shared environmental influences were present for parent-rated IC at age 2 but not at age 3. However, although shared environmental variance was not significant at 3 years, overlapping confidence intervals suggest that estimates do not differ across age. A cautious interpretation is that there may be some modest shared environmental influences on parent-rated IC that persists across age.

The pattern of results was markedly different for our laboratory-based measure of IC. Consistent with previous temperament research in early childhood (e.g., Gagne & Goldsmith, 2011), the stability of observer ratings of IC from age 2 to 3 was much lower than that of our parent ratings indicating considerable developmental change. There is also considerable developmental change at the level of etiology, as the moderate genetic influences on Lab-TAB IC at age 2 did not replicate at age 3. Genetic effects were very modest and nonsignificant when assessed one year later. Shared environmental variance showed a reverse pattern. At age 2, there were no significant shared environmental influences on observed IC in the lab, but at 3 years novel shared environmental variance emerged as a significant source of variation in this measure. Thus, in addition to the change in genetic effects across age, developmental change in IC within the context of the lab was also due to new shared and nonshared environmental influences at age 3. This change reflects exposure to both new family-wide experiences as well as experiences that are specific to each child. As children become more physically and cognitively autonomous, parents may begin teaching them to learn to control their impulses and as a result, constrain IC behavior at age 3. Some families are introducing their children to daycare at this age, where they are expected to increasingly regulate and inhibit their behavior in a more structured setting. These new sources of environmental experiences at age 3 would most likely be reflected in familial similarity (shared environment) if parent behavior and care environments do not differ across twins. The presence of slight methodological alterations in the assessment of laboratory IC across age may also contribute to developmental change, but is likely minimal given that the procedures were highly similar across age.

These results are consistent with prior research in several ways. First, our finding that parent ratings of IC show higher age-to-age stability than observer ratings is common in the literature (Saudino, 2003). In fact, the GEO study found a similar pattern of high stability for parent ratings and no stability for Lab-TAB ratings of anger across 12 and 36 months (Gagne & Goldsmith, 2011). Although it is possible that the higher stability for parent ratings is due to stability in the parent's rating behaviors or their expectancies, this is unlikely given that it is genetic factors that largely explain the phenotypic age-to-age correlation for IC in the present study. That is, approximately two-thirds of the stability correlation for parent ratings was due to overlapping genetic effects across age. The fact that there was no age-to-age genetic covariance for observed IC in the lab likely explains the lower stability. Similar results have emerged for parent and actigraph-assessed activity level in the BUTP which found that the stability of parent ratings was entirely due to genetic effects (Saudino, 2012), suggesting that the higher stability in parent-rated temperament is not simply due to rater biases (i.e., biases would contribute to environmental covariance across age).

Lower genetic variance in lab-based assessments is also more common than with parent ratings. In GEO, Lab-TAB ratings of IC at 36 months of age also showed relatively low genetic influences and significant shared environmental influences (Gagne & Goldsmith, 2011). The difficult question is what is driving these phenotypic and etiological differences between parent and lab-based measures of temperament? The explanation may not be as simple as broad differences between parent and observer ratings. BUTP observer ratings of activity level using the Infant Behavior Record (IBR) showed substantial stability from age 2 to 3, significant genetic variance at both ages, and substantial genetic covariance across age (Saudino, 2012), indicating that the overall lack of stability and genetic effects on IC at age 3 are not simply a function of observational methods but may reflect specific aspects of the Lab-TAB IC assessment.

The findings of significant genetic influences on parent-rated IC at both ages and Lab-TAB IC at age 2 are consistent with a lab-based study of IC in twins aged 9-18 years that used a Stroop task to assess IC (Polderman et al., 2009). The Stroop task assesses a more complex set of executive functions, however, both the Stroop and the simpler delay aversion/response inhibition tasks employed in the Lab-TAB tap similar underlying constructs. This makes the lack of genetic effects on the Lab-TAB at age 3 somewhat more puzzling. Although genes and environments are dynamic in nature, it seems very unlikely that IC is influenced by genetic factors at age 2, by only the environment at age 3, and then genetic factors again in later childhood. Moreover, parent ratings of IC show substantial heritability across ages 2 and 3 in the same children. The fact that the Lab-TAB at age 3 in both the current study and GEO study does not yield findings that are consistent with IC assessed with other methods, or even with itself across age, suggests that while the Lab-TAB may be sensitive to individual differences in IC at age 2, it may not be at age 3. This explanation is supported by a slight, although non-significant reduction in variance across age. In addition, the DZ co-twins are becoming more similar at age 3. The DZ correlation for observed IC at age 3 was significantly higher than that at age 2 indicating that the differences between co-twins is narrowing. Although the MZ correlation is lower at age 3 it was not significantly different from the age 2 correlation. Evidence of low stability and a significant increase in the DZ

correlation is consistent with our speculation that the measure may be less discriminating at age 3. However, variance did not significantly change across age for the Lab-TAB assessments, and unlike DZ twins, MZ twins did not become significantly more correlated. Additional research with other lab-based measures of IC at age 3 is needed to provide further resolution on this.

Another important consideration is that the different Lab-TAB IC assessments may be more or less appropriate for use at age 2 vs. age 3. Following Goldsmith et al. (1995), we composited multiple Lab-TAB episodes in our scoring for the observational measures of IC at both ages in a fairly consistent manner. Perhaps one episode is more sensitive to individual differences at age 2, and some episodes show more stability across ages while others do not. There is some evidence from the phenotypic literature showing that Snack Delay performance improves from 18 to 30 months, ultimately reaching a ceiling effect by late preschool (Spinrad et al., 2007) or early elementary school age (Murray & Kochanska, 2002). Conversely, other investigations suggest that Snack Delay is developmentally appropriate for IC assessment in 3-4 year olds (Bassett et al., 2012; Wiebe et al., 2011). Although used in the current study at 2 and 3 years of age, Dinky Toys was originally designed for preschool age children (Kochanska et al., 1996) and has been found to be difficult for children aged 30 months (Spinrad et al., 2007). Therefore, it could be that Snack Delay may be approaching a ceiling effect in 3-year-olds while Dinky Toys is more appropriate for the 3-year-olds (vs. 2-year-olds). In our data, Dinky Toys and Snack Delay were slightly more skewed at age 3, but distributions of the 3 Lab-TAB IC episodes at age 3 were fairly similar to those at age two and fairly normal. In addition, within- and across-episode stability from age 2 to 3 was fairly low (correlations ranged from .01-.17). Although we can't definitively say why there is a difference in genetic and environmental influences in the LAB-TAB IC across age 2 and 3, the present study shows that this measure is differentially influenced across age. Researchers should not assume that a measure of IC that is genetically influenced at one age will be genetically influenced at another.

A limitation of the current study is the modest sample size, which prevents certain genetic model-fitting analyses (e.g., sex-limitation analyses) due to lack of sufficient power. However, longitudinal temperament investigations with larger samples typically focus on questionnaires rather than comprehensive laboratory assessments, and there are no large longitudinal twin samples with lab-based assessments of IC. The BUTP includes extensive lab-based assessments of IC at 2 and 3 years of age as well as standard parent-rating questionnaires. This comprehensive multi-method approach to assessment across age allows for more comprehensive conclusions about the etiology and development of early IC than studies that rely on parent ratings only. Our findings support stability and change in early IC, and suggest that the phenotypes measured in parent and lab-based assessments are somewhat different. Although this pattern of results varies somewhat with the results of projects that use only questionnaires and find both phenotypic stability and consistent etiological results across age, we replicated previous findings in an independent sample using the same methodologies (Gagne & Goldsmith, 2011). Many questions remain, but the complexity of early IC development is clear.

This investigation not only confirms previous findings, but also recommends a variety of issues for future study. In particular, multi-method investigations that can address questions of genetic and environmental contributions to developmental stability and change, as well as the development of more sensitive lab-based protocols for IC assessment in early childhood will be important. Toddlerhood is a brief period of development, and further longitudinal study that extends into school age is needed. Larger sample sizes in future studies will also allow for a more in-depth analysis of gender differences in IC and how gender may affect etiology. Investigations that harness the power of molecular genetics would also be beneficial. Although linkage and association studies of psychological phenotypes have been increasing in recent years, we have very little information on the specific genes associated with IC. Finally, studies that extend our findings of genetic covariance between IC and behavior problems at age 2 (Gagne et al., 2011) to preschool and school-age children will help to clarify the importance of IC in the development of externalizing psychopathology. Further study in this area has important public health implications as it is possible that training and behavioral interventions that focus on IC may be beneficial to the treatment of ADHD and other externalizing disorders. However, the feasibility and validation of these types of interventions is uncertain until we conduct clinical research examining this issue. The BUTP sample includes behavior problems measures at age 3 as well as molecular genetic data. Therefore, we intend to explore some of these research questions in future analyses, remaining mindful that genetic and environmental contributions to development in early IC are complicated and should not be evaluated without consideration of assessment methodology and age.

Appendix

A. TBAQ-R 2003 Version

Each of the 11 scales has 10 items.

Activity Level (AL) Limb, trunk, or locomotor movement during a variety of daily situations, including free play, confinement, or quiet activities.

Anger (AN) Crying, protesting, hitting, pouting, or other signs of anger in situations involving conflict with the respondent or another child.

Attention (AT) Concentration on an object of attention, resisting distraction. The ability to transfer attentional focus from one activity/task to another.

Inhibitory Control (IC) The capacity to restrict or alter behavior under instruction.

Interest (IN) Duration of task engagement in ongoing solitary play.

Object Fear (OF) Distress, withdrawal, and hedonically fearful vocalizations when exposed to a variety of objects and nonsocial situations.

Pleasure (PL) Smiling, laughter, and other hedonically positive vocalizations or playful activity in a variety of non-threatening or mildly novel situations.

Sadness (SAD) Negative affect and lowered mood and energy related to suffering, disappointment, and object loss.

Sensory Defensiveness (SD) Negative reaction to auditory or tactile sensory stimuli (has two 5-item subscales)

Social Fear (SF) Inhibition, distress, withdrawal (vs approach), or signs of shyness in novel or uncertainty provoking situations.

Soothability (SOO) Recovery from distress

Toddler Behavior Assessment Questionnaire-R (TBAQ-R; Goldsmith, 2003)

INSTRUCTIONS: Please read carefully before starting. As you read each description of the child's behavior below, please indicate how often your child did this during the **last month** by circling one of the numbers. These numbers indicate how often you observed your child in the behavior described during the **last month**.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Never	Very rarely	Less than half the time	About half half the time	More than half the time	Almost always	Always	Not applicable

The "Not Applicable" option is used when you did not see your child in the situation described during the last month. For example, if the situation mentions your child going to the doctor and there was no time during the last month when your child went to the doctor, circle the (NA) option. "Not Applicable" is different from "Never;" "Never" is used when you saw your child in the situation described but s/he never engaged in the behavior mentioned during the **last month**.

Please be sure to circle a number or NA for every item.

1. When playing inside the house or apartment (for example, because of bad weather), how often did **TWIN A** run through the house?

1 2 3 4 5 6 7 NA

2. When playing on a movable toy, such as a tricycle, how often did **TWIN A** attempt to go as fast as s/he could?

1 2 3 4 5 6 7 NA

3. While playing alone in a sandbox or playing with dolls, how often did **TWIN A** remain interested for 10 minutes or longer?

1 2 3 4 5 6 7 NA

4. When in a high place (for example, on a balcony), how often did s/he seem afraid?

1 2 3 4 5 6 7 NA

5. How often did **TWIN A** react noticeably when a low-pitched sound started suddenly (such as an air conditioner, a heating system, a refrigerator, or a vacuum from another room)?
1 2 3 4 5 6 7 NA
6. When you removed something **TWIN A** should not have been playing with, how often did s/he follow your request without signs of anger?
1 2 3 4 5 6 7 NA
7. When you removed something **TWIN A** should not have been playing with, how often did s/he protest (scream or grab objects back)?
1 2 3 4 5 6 7 NA
8. While going through a tunnel or over a bridge while in a car or other vehicle, how often did s/he seem scared?
1 2 3 4 5 6 7 NA
9. When engaged in play with her/his favorite toy, how often did **TWIN A** play for 5 minutes or less?
1 2 3 4 5 6 7 NA
10. When engaged in play with her/his favorite toy, how often did **TWIN A** play for more than 10 minutes?
1 2 3 4 5 6 7 NA
11. When at the doctor's office or a clinic, how often did **TWIN A** cling or hold on to you and not want to let go?
1 2 3 4 5 6 7 NA
12. When you disapprove of **TWIN A**'s behavior, how often did **TWIN A** have hurt feelings?
1 2 3 4 5 6 7 NA
13. While coloring by her/himself, how often did **TWIN A** continue to color alone for 10-20 minutes?
1 2 3 4 5 6 7 NA
14. When a dog or other large animal approached **TWIN A**, how often did s/he cling to you fearfully?
1 2 3 4 5 6 7 NA
15. When playing quietly with one of her/his favorite toys, how often did **TWIN A** smile?
1 2 3 4 5 6 7 NA

16. When **TWIN A** wanted to play outside but you said "no," how often did s/he protest by crying loudly, pouting, frowning, sulking, or looking mad?
1 2 3 4 5 6 7 NA
17. When looking at picture books by her/himself, how often did **TWIN A** lose interest or get bored quickly?
1 2 3 4 5 6 7 NA
18. When asked to wait for something (like a toy or a snack), how often did **TWIN A** wait patiently?
1 2 3 4 5 6 7 NA
19. When asked to wait for something (like a toy or a snack), how often did **TWIN A** find something to distract her/himself until it was time?
1 2 3 4 5 6 7 NA
20. How often did **TWIN A** seem to be alarmed when s/he heard sirens (such as a police, fire, or an ambulance siren) in the distance?
1 2 3 4 5 6 7 NA
21. How often did **TWIN A** ask or gesture for the volume of loud music, radio, or TV to be lowered?
1 2 3 4 5 6 7 NA
22. If you or someone else in your family was tickling, wrestling, or playfully chasing **TWIN A**, how often did s/he laugh?
1 2 3 4 5 6 7 NA
23. When a favorite toy was lost or broken, how often did **TWIN A** cry sadly?
1 2 3 4 5 6 7 NA
24. When a favorite toy was lost or broken, how often did **TWIN A** show no sign of sadness?
1 2 3 4 5 6 7 NA
25. After s/he got a bump or scrape, how often did **TWIN A** forget about it after a couple minutes?
1 2 3 4 5 6 7 NA
26. After s/he got a bump or scrape, how often did **TWIN A** remain upset for a long time?
1 2 3 4 5 6 7 NA
27. How often during the past month did **TWIN A** play games which involved running around, banging, or dumping out toys?
1 2 3 4 5 6 7 NA

28. While playing with a detailed or complicated toy (such as a big doll house or toy garage), how often did **TWIN A** become easily bored or restless?
1 2 3 4 5 6 7 NA
29. When in the bathtub, how often did **TWIN A** babble or talk happily?
1 2 3 4 5 6 7 NA
30. When being dressed or undressed, how often did **TWIN A** lie or sit quietly long enough for you to get her/him ready?
1 2 3 4 5 6 7 NA
31. When s/he saw other children while in the park or playground, how often did **TWIN A** join in the laughing and giggling?
1 2 3 4 5 6 7 NA
32. When being gently rocked or hugged, how often did **TWIN A** smile?
1 2 3 4 5 6 7 NA
33. When it was time for bed or a nap and **TWIN A** did not want to go, how often did s/he protest by crying loudly?
1 2 3 4 5 6 7 NA
34. When given a wrapped package or a new toy in a bag, how often did **TWIN A** laugh or squeal with joy?
1 2 3 4 5 6 7 NA
35. While reading a story of average length to **TWIN A**, how often did s/he pay attention to your reading during the entire story?
1 2 3 4 5 6 7 NA
36. When **TWIN A** needed to sit still, as in a waiting room or a restaurant, how often did s/he play quietly?
1 2 3 4 5 6 7 NA
37. When **TWIN A** needed to sit still, as in a waiting room or a restaurant, how often did s/he try to climb all over other chairs?
1 2 3 4 5 6 7 NA
38. When looking at picture books on her/his own, how often did **TWIN A** stay interested in the book for more than 10 minutes?
1 2 3 4 5 6 7 NA
39. If a stranger came to your house or your apartment, how often did **TWIN A** "warm up" to the stranger within 10 minutes?
1 2 3 4 5 6 7 NA

40. When you told **TWIN A** that s/he would have to play alone for a short time, how often did just one activity or object keep her/him busy?
1 2 3 4 5 6 7 NA
41. When you removed something **TWIN A** should not have been playing with and did not want to give up, how often did s/he stay upset for 10 minutes or longer?
1 2 3 4 5 6 7 NA
42. When you removed something **TWIN A** should not have been playing with and did not want to give up, how often did s/he calm down within 5 minutes?
1 2 3 4 5 6 7 NA
43. If a room suddenly became dark, how often did **TWIN A** cry?
1 2 3 4 5 6 7 NA
44. In a restaurant, at the movies, or other similar situations where there is a need to sit in a chair, how often did **TWIN A** have trouble sitting still?
1 2 3 4 5 6 7 NA
45. In a restaurant, at the movies, or other similar situations where there is a need to sit in a chair, how often did **TWIN A** sit patiently and quietly?
1 2 3 4 5 6 7 NA
46. When one of the parents' friends who did not have daily contact with **TWIN A** visited the home, how often did **TWIN A** talk much less than usual?
1 2 3 4 5 6 7 NA
47. When one of the parents' friends who did not have daily contact with **TWIN A** visited the home, how often did **TWIN A** enthusiastically greet them?
1 2 3 4 5 6 7 NA
48. When one of the parents' friends who did not have daily contact with **TWIN A** visited the home, how often did **TWIN A** squeal with joy?
1 2 3 4 5 6 7 NA
49. When one of the parents' friends who did not have daily contact with **TWIN A** visited the home, how often did **TWIN A** smile?
1 2 3 4 5 6 7 NA
50. While shopping, if you did not agree to buy **TWIN A** a toy that s/he wanted, how often did s/he protest by whining or physically struggling when you tried to separate her/him from the toy?
1 2 3 4 5 6 7 NA

51. If you were not able to give immediate attention to **TWIN A** because you were busy (for example, you were cooking dinner or talking on the phone), how often did s/he cry loudly?
1 2 3 4 5 6 7 NA
52. When **TWIN A** saw an insect up close, how often did s/he run away?
1 2 3 4 5 6 7 NA
53. When first visiting a babysitting co-op, daycare center, or nursery, how often did **TWIN A** cry when not being held by her/his parent and resist being put down?
1 2 3 4 5 6 7 NA
54. When first visiting a babysitting co-op, daycare center, or nursery, how often did **TWIN A** immediately begin to explore?
1 2 3 4 5 6 7 NA
55. When you turned off the television set (because it was bedtime, dinnertime, or time to leave), how often did **TWIN A** throw a tantrum or get really mad?
1 2 3 4 5 6 7 NA
56. How often did **TWIN A** seem overly sensitive to, or irritated by, certain sounds, voices or music?
1 2 3 4 5 6 7 NA
57. During a loud storm, how often did s/he look afraid?
1 2 3 4 5 6 7 NA
58. While watching a favorite children's television program such as *Sesame Street*, how often did **TWIN A** pay attention to the whole show?
1 2 3 4 5 6 7 NA
59. While watching a favorite children's television program such as *Sesame Street*, how often did **TWIN A** watch only the first few minutes of the show before showing signs of restlessness?
1 2 3 4 5 6 7 NA
60. When you removed something **TWIN A** should not have been playing with, how often did s/he become sad?
1 2 3 4 5 6 7 NA
61. How often was **TWIN A** easily able to stop activities when asked to do so?
1 2 3 4 5 6 7 NA
62. How often was your child able to keep from running around the house when asked not to?
1 2 3 4 5 6 7 NA

63. How often was **TWIN A** distracted by background sounds that do not bother most other people?
1 2 3 4 5 6 7 NA
64. When **TWIN A** saw blood or cuts on her/himself or others, how often did s/he look really scared?
1 2 3 4 5 6 7 NA
65. How often was **TWIN A** able to complete a task that required a series of separate steps (such as a simple jigsaw puzzle)?
1 2 3 4 5 6 7 NA
66. How often did **TWIN A** have difficulty waiting for her/his turn to play?
1 2 3 4 5 6 7 NA
67. When first arriving at a favorite place to play, how often did **TWIN A** get really excited?
1 2 3 4 5 6 7 NA
68. When touching a new object, how often did **TWIN A** seem concerned by how smooth or rough the texture was?
1 2 3 4 5 6 7 NA
69. When in the bathtub, how often did **TWIN A** splash or kick?
1 2 3 4 5 6 7 NA
70. When being dressed or undressed, how often did **TWIN A** squirm or try to get away?
1 2 3 4 5 6 7 NA
71. When in a big crowd, how often did s/he become afraid?
1 2 3 4 5 6 7 NA
72. If plans didn't work out that **TWIN A** was looking forward to, how often did **TWIN A** become sad?
1 2 3 4 5 6 7 NA
73. In unfamiliar places, how often did s/he cling to you fearfully?
1 2 3 4 5 6 7 NA
74. When unable to do some task, how often did **TWIN A** feel depressed?
1 2 3 4 5 6 7 NA
75. When **TWIN A** knew her/his parents were about to leave her/him at home, how often did **TWIN A** cling to her/his parents?
1 2 3 4 5 6 7 NA

76. When **TWIN A** knew her/his parents were about to leave her/him at home, how often did **TWIN A** show no sign of being upset?
1 2 3 4 5 6 7 NA
77. How often did **TWIN A** object to scratchy or stiff clothing fabrics?
1 2 3 4 5 6 7 NA
78. How often did **TWIN A** object to changes in articles of clothing that fit snugly or tightly (for example, putting on a hat, wearing gloves, getting new shoes)?
1 2 3 4 5 6 7 NA
79. When tired, how often did **TWIN A** become tearful?
1 2 3 4 5 6 7 NA
80. When it was time for bed or a nap and **TWIN A** did not want to go, how often did **TWIN A** have difficulty settling down?
1 2 3 4 5 6 7 NA
81. When it was time for bed or a nap and **TWIN A** did not want to go, how often did s/he fall asleep within 10 minutes?
1 2 3 4 5 6 7 NA
82. When it was time for bed or a nap and **TWIN A** did not want to go, how often did s/he whimper or sob?
1 2 3 4 5 6 7 NA
83. How often did **TWIN A** play alone with her/his favorite toy for 30 minutes or longer?
1 2 3 4 5 6 7 NA
84. How often did **TWIN A** play alone with her/his favorite toy for 10 minutes or longer?
1 2 3 4 5 6 7 NA
85. When another child took away a favorite toy that **TWIN A** was playing with, how often did s/he object?
1 2 3 4 5 6 7 NA
86. When another child took away a favorite toy that **TWIN A** was playing with, how often did s/he try to hit, kick, or bite the other child?
1 2 3 4 5 6 7 NA
87. When playing alone, how often did **TWIN A** become easily distracted?
1 2 3 4 5 6 7 NA

88. When asked to wait for something (like a toy or a snack), how often did **TWIN A** go after it anyway?
1 2 3 4 5 6 7 NA
89. When placed in a car seat or stroller, how often did **TWIN A** squirm?
1 2 3 4 5 6 7 NA
90. When placed in a car seat or stroller, how often did **TWIN A** sit still?
1 2 3 4 5 6 7 NA
91. How often did **TWIN A** refuse to touch a sticky or gooey substance (for example, shaving cream, mayonnaise, toothpaste, or mud)?
1 2 3 4 5 6 7 NA
92. How often did **TWIN A** move from one task or activity to another without completing any?
1 2 3 4 5 6 7 NA
93. When told to do something s/he did not want to do, how often did **TWIN A** become tearful?
1 2 3 4 5 6 7 NA
94. How often did **TWIN A** have trouble focusing on a task without guidance?
1 2 3 4 5 6 7 NA
95. How often did **TWIN A** object to the feeling of a comb moving through her/his hair or a toothbrush touching her/his gums?
1 2 3 4 5 6 7 NA
96. When s/he heard a sad story, how often did **TWIN A** whimper or sob?
1 2 3 4 5 6 7 NA
97. When you were comforting your upset child, how often did s/he cheer up within 5 minutes?
1 2 3 4 5 6 7 NA
98. When you were comforting your upset child, how often did s/he calm down quickly?
1 2 3 4 5 6 7 NA
99. When **TWIN A** watched a favorite television program, how often did s/he watch for 5 minutes or less?
1 2 3 4 5 6 7 NA
100. When **TWIN A** watched a favorite television program, how often did s/he watch for 10 minutes or longer?

1 2 3 4 5 6 7 NA

101.When hearing very loud noises (for example, a siren, a train whistle, or machinery operating), how often did s/he become distressed?

1 2 3 4 5 6 7 NA

102.When **TWIN A** wanted to eat something sweet before dinner was finished but did not get it, how often did s/he protest by crying loudly?

1 2 3 4 5 6 7 NA

103.When asked not to play with something, how often did **TWIN A** follow your request?

1 2 3 4 5 6 7 NA

104.When making a discovery (such as two Lego pieces fitting together, learning to stack blocks, or learning to turn a light switch on and off), how often did **TWIN A** smile?

1 2 3 4 5 6 7 NA

105.When **TWIN A** was approached by a stranger when you and s/he were out (for example, shopping), how often did **TWIN A** babble or talk?

1 2 3 4 5 6 7 NA

106.When **TWIN A** was approached by a stranger when you and s/he were out (for example, shopping), how often did **TWIN A** show distress or cry?

1 2 3 4 5 6 7 NA

107.When asked to do so, how often did **TWIN A** lower her/his voice immediately?

1 2 3 4 5 6 7 NA

108.When asked to do so, how often did **TWIN A** wait before beginning a new activity?

1 2 3 4 5 6 7 NA

109.When **TWIN A** was left to calm her/himself down after being upset, how often was s/he able to do so within a couple of minutes?

1 2 3 4 5 6 7 NA

110.When **TWIN A** was disappointed about having to leave a place where s/he was having fun, how often did s/he calm down quickly?

1 2 3 4 5 6 7 NA

B. Correlation table that provides the correlations amongst all predictor variables

Phenotypic Correlations: 24- and 36-month Parent-rated and Observed Inhibitory Control.

	Observed IC 36 months	Parent IC 24 months	Parent IC 36 months
Observed IC 24 months	.09	.19 **	.11 *
Observed IC 36 months		.13 *	.14 **
Parent IC 24 months			.51 **

Note:

*
 $p < .05$,
**
 $p < .01$.

References

- Berger, A. Self-regulation: Brain, cognition and development. American Psychological Association; Washington, DC: 2011.
- Bassett H, Denham S, Wyatt T, Warren-Khot H. Refining the preschool self-regulation assessment for use in preschool classrooms. *Infant and Child Development*. 2012; 21:596–616.
- Campbell SB. Behavior problems in preschool children: A recent review of the research. *Journal of Child Psychology*. 1995; 36:113–149.
- Eisenberg N, Cumberland A, Spinrad TL, Fabes RA, Shepard SA, Reiser M, Valiente C, Murphy SH, Losoya SH, Guthrie IK. The relations of regulation and emotionality to children's externalizing and internalizing problem behavior. *Child Development*. 2001; 72:1112–1134. [PubMed: 11480937]
- Eisenberg N, Sadovsky A, Spinrad TL, Fabes RA, Losoya SH, Valiente C, Reiser M, Cumberland A, Shepard SA. The relations of problem behavior status to children's negative emotionality, effortful control, and impulsivity; Concurrent relations and prediction of change. *Developmental Psychology*. 2005; 41:193–211. [PubMed: 15656749]
- Eisenberg N, Spinrad TL, Fabes RA, Reiser M, Cumberland A, Shepard SA, Valiente C, Losoya SH, Guthrie IK, Thompson M. The relations of effortful control and impulsivity to children's resiliency and adjustment. *Child Development*. 2004; 75:25–46. [PubMed: 15015673]
- Eisenberg N, Valiente C, Fabes RA, Smith CL, Reiser M, Shepard SA, Losoya SH, Guthrie IK, Murphy BC, Cumberland A. The relations of effortful control and ego control to children's resiliency and social functioning. *Developmental Psychology*. 2003; 39:761–776. [PubMed: 12859128]
- Gagne JR, Goldsmith HH. A longitudinal analysis of anger and inhibitory control in twins from 12 to 36 months of age. *Developmental Science*. 2011; 14:112–124. [PubMed: 21159093]
- Gagne JR, Saudino KJ. Wait for it! The etiology of inhibitory control in early childhood. *Behavior Genetics*. 2010; 40:327–337. [PubMed: 19936910]
- Gagne JR, Saudino KJ, Asherson P. The genetic etiology of inhibitory control and behavior problems at 24 months of age. *Journal of Child Psychology and Psychiatry*. 2011; 52:1155–1163. [PubMed: 21627653]
- Goldsmith HH. Studying temperament via construction of the Toddler Behavior Assessment Questionnaire. *Child Development*. 1996; 67:218–235. [PubMed: 8605830]
- Goldsmith HH, Buss KA, Lemery KS. Toddler and childhood temperament: expanded content, stronger genetic evidence, new evidence for the importance of environment. *Developmental Psychology*. 1997; 33:891–905. [PubMed: 9383612]

- Goldsmith HH, Reilly J, Lemery KS, Longley S, Prescott A. The laboratory temperament assessment battery-preschool version: Description of procedures. 1995 Unpublished manuscript.
- Goldsmith HH, Rieser-Danner LA, Briggs S. Evaluating convergent and discriminant validity of temperament questionnaires for preschoolers, toddlers, and infants. *Developmental Psychology*. 1991; 27:566–579.
- Griffin D, Gonzalez R. Correlational analysis of dyad-level data in the exchangeable case. *Psychological Bulletin*. 1995; 118:430–439.
- Keenan K, Wakschlag L. More than the terrible twos: The nature and severity of behavior problems in clinic-referred preschool children. *Journal of Abnormal Child Psychology*. 2000; 28:33–46. [PubMed: 10772348]
- Kochanska G, Knaack A. Effortful control as a personality characteristic of young children: Antecedents, correlates and consequences. *Journal of Personality*. 2003; 71:1087–1112. [PubMed: 14633059]
- Kochanska G, Murray KT, Harlan ET. Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology*. 2000; 36:220–232. [PubMed: 10749079]
- Kochanska G, Murray K, Jacques TY, Koenig AL, Vandegeest KA. IC in young children and its role in emerging internalization. *Child Development*. 1996; 67:490–507. [PubMed: 8625724]
- Lemery-Chalfant K, Doelger L, Goldsmith HH. Genetic relations between effortful and attentional control and symptoms of psychopathology in middle childhood. *Infant and Child Development*. 2008; 17:365–385. [PubMed: 27076792]
- Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73:13–22.
- McGue M, Bacon S, Lykken DT. Personality stability and change in early adulthood: A behavioural genetic analysis. *Developmental Psychology*. 1993; 29:96–109.
- McGue M, Bouchard TJ Jr. Adjustment of twin data for the effects of age and sex. *Behavior Genetics*. 1984; 14:325–343. [PubMed: 6542356]
- Murray KT, Kochanska G. Effortful control: Factor structure and relation to externalizing and internalizing behaviors. *Journal of Abnormal Child Psychology*. 2002; 30:503–514. [PubMed: 12403153]
- Neale, MC. Mx: Statistical modeling. 6th. Department of Psychiatry; Medical College of Virginia: 2003.
- Nigg JT, Quamma JP, Greenberg MT, Kusche CA. A two-year longitudinal study of neuropsychological and cognitive performance in relation to behavioral problems and competencies in elementary school children. *Journal of Abnormal Child Psychology*. 1999; 27:51–63. [PubMed: 10197406]
- O'Connor BP. SPSS and SAS programs for addressing interdependence and basic levels-of-analysis issues in psychological data. *Behavior Research Methods, Instruments, & Computers*. 2004; 36:17–28.
- Olson SL, Schilling EM, Bates JE. Measurement of impulsivity: Construct coherence, longitudinal stability, and relationship with externalizing problems in middle childhood and adolescence. *Journal of Abnormal Child Psychology*. 1999; 27:151–165. [PubMed: 10400061]
- Plomin R, Nesselroade JR. Behavioural genetics and personality changes. *Journal of Personality*. 1990; 58:191–220. [PubMed: 2198339]
- Plomin, R.; DeFries, JC.; McClearn, GE.; Rutter, M. *Behavioral Genetics*. 3rd. W. J. Freeman and Company; New York: 1997.
- Polderman TJC, de Geus EJC, Hoekstra RA, Bartels M, van Leeuwen M, Verhulst FC, Posthuma D, Boomsma DI. Attention problems, inhibitory control, and intelligence index overlapping genetic factors: A study in 9-, 12-, and 18-year-old twins. *Neuropsychology*. 2009; 23:381–391. [PubMed: 19413451]
- Rothbart, MK. Temperament and development. In: Kohnstamm, GA.; Bates, JA.; Rothbart, MK., editors. *Temperament in childhood*. Wiley; New York: 1989. p. 187–247.
- Rothbart MK, Ahadi SA. Temperament and the development of personality. *Journal of Abnormal Psychology*. 1994; 103:55–66. [PubMed: 8040481]

- Rothbart, MK.; Bates, JE. Handbook of child psychology. In: Eisenberg, N., editor. Social, emotional, and personality development. 6th. Vol. 3. John Wiley & Sons, Inc; Hoboken, NJ: 2006. p. 99-166. Temperament
- Saudino KJ. Sources of continuity and change in activity level in early childhood. *Child Development*. 2012; 83:266–281. [PubMed: 22103336]
- Saudino KJ. The need to consider contrast effects in parent-rated temperament. *Infant Behavior & Development*. 2003; 26:118–120.
- Saudino, KJ.; Wang, M. Quantitative and Molecular Genetics of Temperament. In: Zentner, M.; Shiner, R., editors. *The Handbook of Temperament*. Guilford; New York, NY: 2012. p. 315-346.
- Saudino KJ, Carter AS, Purper-Ouakil D, Gorwood P. The etiology of behavioral problems and competencies in very young twins. *Journal of Abnormal Psychology*. 2008; 117:48–62. [PubMed: 18266485]
- Saudino, KJ.; Cherny, SS. Parental ratings of temperament in twins. *Infancy and early childhood: Genetic and environmental influences on developmental change*. Emde, RN.; Hewitt, JK., editors. Oxford University Press; New York, NY: 2001. p. 73-88.
- Saudino KJ, Wertz AE, Gagne JR, Chawla S. Night and day: are siblings as different in temperament as parents say they are? *Journal of Personality and Social Psychology*. 2004; 87:698–706. [PubMed: 15535780]
- Schachar R, Tannock R, Marriott M, Logan G. Deficient IC in attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology*. 1995; 23:411–437. [PubMed: 7560554]
- Spinrad T, Eisenberg N, Gaertner B. Measures of effortful regulation for young children. *Infant Mental Health Journal*. 2007; 28(6):606–626. [PubMed: 18066395]
- Valiente C, Eisenberg N, Smith CL, Reiser M, Fabes RA, Losoya S, Guthrie IK, Murphy BC. The relations of effortful control and reactive control to children's externalizing problems: A longitudinal assessment. *Journal of Personality*. 2003; 71:1171–1196. [PubMed: 14633062]
- Wiebe S, Sheffield T, Nelson J, Clark C, Chevalier N, Espy K. The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*. 2011; 108:436–452. [PubMed: 20884004]
- Yamagata S, Takahashi Y, Kijima N, Maekawa H, Ono Y, Ando J. Genetic and environmental etiology of effortful control. *Twin Research and Human Genetics*. 2005; 8:300–306. [PubMed: 16176712]
- Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*. 1986; 42:121–130. [PubMed: 3719049]

Figure 1. Bivariate Cholesky Model

Table 1

Sample Sizes, Means (and Standard Deviations) by Sex and Zygosity, Effect Sizes of Sex and Zygosity Differences for Observer- and Parent-rated Inhibitory Control at age 2 and 3.

	<i>n</i>	Overall Mean (SD)	Male Mean (SD)	Female Mean (SD)	MZ Mean (SD)	DZ Mean (SD)	Effect Size Gender	Effect Size Zygosity
Observed IC (2 years)	<i>n</i>	0 (.65)	−0.07 (.66)	0.09 (.64)	0.02 (.64)	−.01 (.67)	−.25 **	.05
Observed IC (3 years)	606	−.01 (.61)	−0.07 (.63)	0.06 (.58)	−0.03 (.64)	0.01 (.58)	−.21 *	−.07
Parent IC (2 years)	613	38.85 (8.80)	37.09 (8.88)	40.84 (8.30)	38.91 (8.81)	38.81 (8.81)	−.44 **	.01
Parent IC (3 years)	588	41.28 (9.0)	39.89 (8.83)	42.86 (8.94)	41.31 (8.95)	41.25 (9.06)	−.33 **	.01

Note:

Effect size estimated as Cohen's *d* express group differences in standard deviation units. Sample sizes for all sub-samples are available.

*
 $p < .05$,

**
 $p < .01$.

Table 2

Twin Intraclass Correlations, Cross-age Cross-twin Correlations: Observed and Parent-rated Inhibitory Control (within each measure) at age 2 and 3.

Twin Intraclass Correlations							
Observed IC 2 years		Parent IC 2 years		Observed IC 3 years		Parent IC 3 years	
MZ	DZ	MZ	DZ	MZ	DZ	MZ	DZ
.38**	.12*	.87**	.55**	.26**	.35**	.73**	.42**
Cross-age Cross-twin Correlations							
Observed IC				Parent IC			
MZ		DZ		MZ		DZ	
.17**		.12*		.58**		.47**	

Note.

MZ=monozygotic twins, DZ=dizygotic twins.

* $p < .05$,

** $p < .01$.

Table 3

Fit Statistics for Bivariate Cholesky Models of Parent-rated and Observed Inhibitory Control across 2 and 3 years of age.

Variables and Model	Overall Fit of Model					
	-2LL	df	χ^2	df	p	AIC
Parent-Rated IC						
Saturated	8066.91	1173				
ACE Cholesky	8074.22	1184	7.31	11	.77	-14.69
Observed IC						
Saturated	2198.95	1168				
ACE Cholesky	2206.61	1179	7.66	11	.74	-14.34

Note. -2LL= Likelihood Statistic; *df*= degrees of freedom; *AIC*= Akaike's Information Criterion.

Table 4

Multivariate Estimates of Genetic and Environmental Variance (Standardized Parameter Estimates), and Genetic and Environmental Correlations (and 95% Confidence Intervals) for Parent-rated and Observed Inhibitory Control at age 2 and 3 using the best-fitting Bivariate Cholesky Models.

IC Measure	Variance Components						Covariances					
	Age 2 Years			Age 3 Years								
	a^2	c^2	e^2	Continuity			Change					
				a^2	c^2	e^2	a^2	c^2	e^2	r_g	r_c	r_e
Parent Rating	.61 (.78;.87)	.23 (.01;.41)	.17 (.13;.22)	.23 (.08;.45)	.13 (.00;.33)	.01 (.00;.04)	.39 (.19;.48)	.00 (-.17;.17)	.25 (.19;.32)	.61 (.43;.80)	1.0 (-1.0;1.0)	.23 (.06;.39)
Observer Rating	.38 (.14;.51)	.00 (.00;.18)	.62 (.49;.76)	.06 (.00;.25)	.00 (-.36;.36)	.01 (-.04;.00)	.00 (-.20;.20)	.24 (.03;.33)	.70 (.59;.81)	1.0 (-1.0;1.0)	1.0 (-1.0;1.0)	-.11 (-.25;.03)

Note. IC=Inhibitory Control, a^2 =genetic variance, c^2 =shared environmental variance, e^2 =nonshared environmental variance, r_g =genetic correlation, r_c =shared environmental correlation, r_e =nonshared environmental correlation. Unstandardized parameter estimates are available from the first author by request.