Published in final edited form as:

Infancy. 2015 ; 20(1): 98–114. doi:10.1111/infa.12063.

The relation between reactivity at 4 months and Behavioral Inhibition in the second year: Replication Across Three Independent Samples

Nathan A. Fox¹, Nancy Snidman², Sara A. Haas¹, Kathryn A. Degnan¹, and Jerome Kagan³

¹Department of Human Development and Quantitative Methodology, University of Maryland, College Park, MD 20742

²Department of Psychology, University of Massachusetts Boston, MA 02125

³Department of Psychology, Harvard University, Cambridge MA 02138

Abstract

This paper examines the predictive relations between two infant temperamental biases assessed at 4 months and inhibited behavior during the first two years of life in three independent samples from two research laboratories. Although each sample used slightly different criteria for classifying infants, the results across samples were consistent. Infants of both genders who displayed high levels of motor activity and distress to unfamiliar events were more inhibited at 14 months of age. By 24 months there were significant sex differences: boys identified as high reactive were more inhibited than high reactive girls.

Keywords

temperament; behavioral inhibition; infant reactivity

INTRODUCTION

Young infants vary in their patterns of responsiveness to unfamiliar events. Some of this variation has an origin in one or more temperamental biases. Previous work suggested that variation in motor activity and fretting/crying in response to unfamiliar events reflects the two temperamental biases called high and low reactivity, which, in turn, predict inhibited or uninhibited behavior in the second year (Kagan, Reznick, Snidman, Gibbons, & Johnson, 1988; Kagan & Snidman, 1991). Four-month-old infants who display very high levels of limb/trunk movement and distress (called high reactive) are biased to show avoidance of and withdrawal to unfamiliar stimuli or contexts during early childhood, a pattern called behavioral inhibition. Those who show very low levels of motor activity and distress at four months (called low reactive) are biased to become sociable and bold and are called uninhibited (Fox, Henderson, Rubin, Calkins, & Schmidt, 2001; Kagan and Snidman, 1991).

Although the results from different laboratories find similar biological correlates among the high and low reactive infant groups, one node of disagreement concerns the predictive power of the 4-month classifications of high or low reactivity. One set of studies found that the predictive relations between behavioral inhibition at 1 and 2 years and behavior at later years were significantly higher than the relations between high or low reactivity at 4-months and behavioral inhibition among toddlers (Fox et al., 2001). A second set of studies however, reported that the four-month temperamental biases were better predictors of the behavior and physiology of older children than indexes of behavioral inhibition in the second year. Variations across the samples in recruitment and methodology in selecting the high and low reactive groups could be a source of the differences between labs.

The current paper compares methods and outcomes across the two laboratories and three independent samples. In addition, we examined the possibility of sex differences in the expression of temperament and prediction to behavioral inhibition across the samples. Previous research has been mixed with regard to the role of sex differences in temperament. One set of findings suggests that high reactivity predicted social wariness for boys at 4 years of age, but not for girls (Henderson, Fox, and Rubin, 2001). Similarly, other researchers have shown that the continuity of behavioral inhibition may be stronger in boys than in girls (Crockenberg and Smith, 1982; Fagan, 1990). On the other hand, Kagan (1994) reported that the girls in his sample were, on average, more inhibited at 14 and 21 months. We therefore examined relations in three samples between infant reactivity and behavioral inhibition at 14 and 24 months in both males and females.

METHODS

Participants

Sample 1—Subjects were infants from a large metropolitan area in the Northeast United States who were Caucasian and middle-class. All infants were born within 2 weeks of due date and reported no significant medical difficulties. At 4 month olds, (+/- 1 week) more than 500 children were screened for high or low reactivity using the battery of unfamiliar stimuli designed by Kagan and colleagues (Kagan & Snidman, 1991; Kagan, 1994). High reactive infants were defined as those who displayed high frequencies of both motor movement (arms, trunk, and legs) and fretting or crying. Of these infants, 60% were selected for longitudinal study because they displayed extremely high or low scores for motor activity and/or distress. The group showing high levels of both variables was called high reactive (20% of the total sample) and the group showing low levels on both measures were called low reactive (40% of the total sample). For the analyses presented in this paper, we included only the 128, four-month-old infants who were observed at both 14 and 21 months as they encountered a series of unfamiliar objects or people in unfamiliar laboratory settings (High Reactive Group: $n = 55$, 28 males; Low Reactive Group: $n = 73$, 36 males). Refer to Kagan (1994) for reasons for attrition. The frequencies of avoidant behaviors and/or crying were used to construct an index of behavioral inhibition (Kagan & Snidman, 1991).

Sample 2—Infants were recruited in two waves. Wave 1 consisted of infants from a large metropolitan area in the Mideast United States who were predominately Caucasian (95%),

and middle- class. All infants were born within 2 weeks of due date and reported no significant medical difficulties. At 4 months (+/- 1 week), 208 children were screened for high or low reactivity using the battery of unfamiliar stimuli designed by Kagan and colleagues (Kagan & Snidman, 1991). The average scores for motor and negative affect for the first 25 percent of the sample were used to determine the criteria for classification. Infants who were above the mean on both motor reactivity and negative affect were classified as high reactive. Infants who were below the mean for both motor reactivity and negative affect were classified as low reactive. Of these infants, 81 were selected for longitudinal study because they displayed extremely high or low scores for motor activity and/or negative affect. The group showing high levels of both variables were called high reactive (representing 14% of the total sample); the group showing low levels on both measures were low reactive (representing approximately 14% of the total sample) (Calkins, Fox, & Marshall, 1996). For the analyses presented in this paper, we included only infants ($n = 53$) who were observed at both 14 and 24 months of age when measures of behavioral inhibition were collected (High Reactive Group: $n = 25$, 12 males; Low Reactive Group: $n = 28$, 12 males) (Fox et al., 2001). Reasons for attrition were unrelated to reactivity grouping, and are further explained in Fox et al., 2001).

A second wave of 225, four-month-old infants from the same metropolitan area in Maryland was observed with the same procedures. The motor activity and negative affect were coded on a Likert scale from 1 to 7 (1 = low reactivity, 7 = extreme reactivity). Subjects whose scores were above 4 on both variables were categorized as high reactive. Subjects whose scores were less than 3 on motor reactivity and negative reactivity were called low reactive. Of these infants, 49 infants were selected for longitudinal study because they displayed extremely high or low scores for motor activity and/or negative affect. The group showing high levels of both variables were called high reactive (representing 12% of the total sample); the group showing low levels on both measures were low reactive (representing approximately 10% of the total sample). For the analyses presented in this paper, we included only infants ($n = 42$) who were observed at both 14 and 24 months of age when measures of behavioral inhibition were collected (High Reactive Group: $n = 23$, 10 males; Low Reactive Group: $n = 19$, 9 males) (Fox et al., 2001). Reasons for attrition were unrelated to reactivity grouping, and are further explained in Fox et al., 2001).

In order for subjects from both Wave 1 and Wave 2 to be coded with the same metric, the 208 subjects from Wave 1 were also coded with the Likert system used to score infants in Wave 2. Correlations between the Likert scales and discrete coded variables were $r = .51$ for motor reactivity: $p < .001$ and $r = .71$ for negative affect: $p < .001$.

The final make up of Sample 2 analyzed in this paper combined Wave 1 and Wave 2, resulting in 95 infants, 48 (22 males) of which were in the Higher Reactive, and 47 (21 males) of which were Low Reactive.

Sample 3—A third sample was recruited from the same large metropolitan area in the Mideast United States. All infants were born within 2 weeks of due date and reported no significant medical difficulties. At 4 months (+/- 1 week) 779 infants were screened for high or low reactivity with the same battery of unfamiliar stimuli used by Kagan and

colleagues (e.g., Kagan & Snidman, 1991). Mean scores for motor activity, negative affect and positive affect of the first 100 infants were used to determine category membership. Subjects whose scores were above the mean for both motor reactivity and negative affect were classified as high reactive (approximately 11% of the total sample). Subjects whose scores were above the mean for both motor reactivity and positive affect were classified as high positive reactive (approximately 9% of the total sample). Infants who did not meet the criteria for either group, and whose scores on both motor and negative reactivity were close to the mean of the entire sample, were selected as a “control” group (representing approximately 7% of the total sample). Of these, 291 subjects were selected for longitudinal study. A total of 187 infants (64.3%) were Caucasian, 41 (14.1%) were African American, and 63 (21.6%) were of another ethnicity. (Hane, Fox, Henderson, & Marshall, 2008). For the analyses presented in this paper, we included only infants ($n = 140$) who were observed at 24 months of age when measures of behavioral inhibition were collected (High Reactive Group: $n = 89$, 34 males; Low Reactive Group: $n = 51$, 23 males). Infants in this sample were not seen at 14 months of age (White, McDermott, Degnan, Henderson, & Fox, 2011). Reasons for attrition were unrelated to reactivity grouping, and are further explained in White et al. (2011).

Behavioral Inhibition—The frequency of avoidant behaviors or crying to a series of unfamiliar events defines behavioral inhibition. For Sample 1, crying to any episode or failure to approach a small number of events designated a priori as novel were coded as “fear” responses at 14 and 21 months (See Table 2 for more details). The definition of behavioral inhibition in Sample 1 was the total number of episodes during which the child displayed a fear response (range of 0 to 13; a majority had 2 or fewer fears).

For sample 2, seen at 14 and 24 months, and sample 3 at 24 months, the index of inhibition was comparable to the one used with Sample 1. The behavioral inhibition scores for samples 2 and 3 included latency to approach an unfamiliar object or person, as well as time spent in close proximity to the caregiver. The final values for behavioral inhibition for Samples 2 and 3 were standardized scores. (Table 2 presents the details of the different coding systems. See also Kagan, 1994; Fox et al. 2001; White et al., 2011).

Although results from each of the three samples have been published, a comparison across samples had not been examined.

RESULTS

4-month Reactivity Groupings

To examine whether motor reactivity¹ at 4 months differed between the low and high reactivity groups in samples 1, 2, & 3 and whether observed differences varied by sample and sex, a $3 \times 2 \times 2$ (sample $\times$ reactivity group $\times$ sex) factorial analysis of variance (ANOVA) was computed. A significant main effect of reactivity group ($F_{1, 351} = 209.210$, $p < .001$, $\eta_p^2 = .373$) revealed that the high reactive infants ($M = .54$, $SD = .95$) were

¹Motor Reactivity and Behavioral Inhibition scores of each of the three samples were z-scored in order to account for differences in measurement scales.

significantly more active and distressed at 4 months than the low reactive infants ($M = -.71$, $SD = .68$). There were no main effects of sample ($\eta_p^2 = .001$), sex ($\eta_p^2 = .002$), or interactions of reactivity group and sex ($\eta_p^2 = .001$), reactivity group and sample ($\eta_p^2 = .001$), sex and sample ($\eta_p^2 = .002$), or group, sex, and sample ($\eta_p^2 = .001$).

14 month Behavioral Inhibition

To examine whether behavioral inhibition¹ at 14 months differed as a function of reactivity, sample, and gender, a $2 \times 2 \times 2$ (sample x reactivity group x sex) factorial analysis of variance (ANOVA) was computed. Sample 3 was not included in this analysis because 14-month behavioral inhibition data were not collected for this sample. A significant main effect of reactivity group ($F_{1, 206} = 56.253$, $p < .001$, $\eta_p^2 = .214$) revealed that high reactive infants ($M = .50$, $SD = 1.08$) were significantly more inhibited at 14 months than low reactive infants ($M = -.41$, $SD = .68$). (See Figure 1). There was no significant main effect of sample ($\eta_p^2 = .002$), or interaction of sample and reactivity group ($\eta_p^2 = .012$). However, there were two 2-way interactions involving sex and reactivity ($F_{1, 206} = 4.86$, $p < .05$, $\eta_p^2 = .02$) and sex and sample ($F_{1, 206} = 6.52$, $p < .05$, $\eta_p^2 = .03$). Furthermore, these were superseded by a significant three-way interaction of sex x reactivity x sample ($F_{2, 206} = 4.76$, $p < .05$, $\eta_p^2 = .02$).

Post hoc analyses using the Bonferroni post hoc criterion to examine significant differences between reactivity groups for each sex within each sample confirmed that, in sample 1, the high reactive infants were significantly more inhibited at 14 months than the low reactive group for both males ($F_{1, 206} = 25.30$, $p < .001$, $\eta_p^2 = .11$) and females ($F_{1, 206} = 25.13$, $p < .001$, $\eta_p^2 = .11$). In sample 2, the high reactive group was significantly more inhibited at 14 months than the low reactive group for males ($F_{1, 206} = 19.82$, $p < .001$, $\eta_p^2 = .09$), but not for females ($\eta_p^2 = .002$) (See Figure 1).

Additional post hoc analyses examining sample differences for each sex within each reactivity group revealed that within the high reactivity group, sample 1 males were less inhibited than sample 2 males ($F_{1, 206} = 4.12$, $p < .05$, $\eta_p^2 = .02$), while sample 1 females were more inhibited than sample 2 females ($F_{1, 206} = 6.93$, $p < .01$, $\eta_p^2 = .03$). There were no sample differences within the low reactivity group for males ($\eta_p^2 = .01$) or females ($\eta_p^2 = .007$).

21 or 24-Month Behavioral Inhibition

To examine whether behavioral inhibition¹ at 21/24 months differed between the low and high reactivity groups in samples 1, 2 & 3, and whether these differences varied by sample and sex, a $3 \times 2 \times 2$ (sample x reactivity group x sex) factorial analysis of variance (ANOVA) was computed. A significant main effect of reactivity group ($F_{1, 335} = 11.845$, $p < .001$, $\eta_p^2 = .034$), revealed that high reactive infants ($M = .17$, $SD = 1.04$) were more behaviorally inhibited at 21/24 months than low reactive infants ($M = -.18$, $SD = .91$) (See Figure 2a). There was no significant effect of sample ($\eta_p^2 = .001$), nor an interaction of sample and reactivity group ($\eta_p^2 = .004$), or three-way interaction ($\eta_p^2 = .004$). However, there was a significant two-way interaction of sex and reactivity group ($F_{1, 335} = 4.33$, $p < .$

05, $\eta_p^2 = .012$) and a significant two-way interaction of sex and sample ($F_{2, 335} = 3.19, p < .05, \eta_p^2 = .019$).

Post hoc analyses using the Bonferroni post hoc criterion to examine significant differences between reactivity groups for each sex confirmed that the high reactive group was significantly more inhibited at 21/24 months than the low reactive group for males ($F_{1,335} = 14.33, p < .001, \eta_p^2 = .04$), but not for females ($\eta_p^2 = .003$) (See Figure 2b).

Additional post hoc analyses examining sex differences for each sample revealed that in sample 1, males were less inhibited than females ($F_{1,206} = 4.12, p < .05, \eta_p^2 = .02$). However, there were no differences in inhibition scores between males and females for sample 2 ($\eta_p^2 = .001$) or sample 3 ($\eta_p^2 = .001$).

DISCUSSION

The hypothesis that four-month-old infants who displayed high or low levels of motor activity and distress to unfamiliar events would differ in wariness and avoidance in the second year was affirmed for the assessment at 14 months in Samples 1 and 2, and at 21/24 months in Samples 1, 2 and 3. These data suggest that 4-month reactivity reflects early temperamental biases predictive of differences in behavioral inhibition at two years of age.

Although the two laboratories used slightly different coding methods and paradigms, an infant's classification as high or low reactive predicted behavioral inhibition at 14 months, and at 21 or 24 months for all three samples. These findings support previously published research that high and low reactivities retain some features of their temperament through adolescence. (Kagan and Snidman 2004; Kagan Snidman, Kahn & Towsley, 2007; Schwartz, Wright, Shin, Kagan, & Rauch, 2003; Fox, Henderson, Marshall, Nichols, & Ghera, 2005).

The current data suggest that there is greater preservation of behavioral inhibition in males during the first two years of life, despite a lack of sex differences in reactivity in at four months. One possible post hoc explanation for this observation is that these sex differences are a result of variation in experiences during the first two years of life. Parents may exhibit an intrusive caregiving style with timid sons because they violate the sex role stereotype for males but shyness is less inconsistent with the stereotype for girls (Park, Belsky, Putnam, & Crnic, 1997). Sex differences in rate of maturation of frontal lobe processes might also explain these data. The incentives for an inhibited response were unfamiliar events that were unexpected and none were serious threats to harm or loss of a parent's comfort. Because frontal lobe functions are able to mute a "fear" response to the unfamiliar, partly through projections to the amygdala, and these projections are likely to be precocious in girls, it is reasonable to suggest that this biological difference could explain the sex difference (Hepper, Shannon, & Dorman 1997).

There is increasing evidence suggesting sex differences in the later outcomes of BI, especially social anxiety. The higher prevalence in males may be due in part to the fact that extreme levels of shyness and timidity are less acceptable in boys than in girls, (Doey, Coplan, & Kingsbury, 2013). For example, stronger associations between toddler shyness

and a multitude of less adaptive outcomes, including peer exclusion, depression, anxiety, and low self-esteem, have been reported for boys relative to girls (e.g., Caspi, Elder, & Bem, 1988; Coplan, Closson, & Arbeau, 2007; Spangler & Gazelle, 2009).

However, not all studies report that inhibited boys are at a greater risk for internalizing problems. Schwartz, Snidman, and Kagan (1999) found a stronger association between BI in the second year and social anxiety in adolescent girls. The results in this paper suggest that sex differences begin early. Future work examining caregiving differences in high reactive infants as a function of the child's sex will illuminate this puzzle.

Acknowledgments

We would like to thank all of the families who participated in this research. Funding for the research from Fox's lab is from the National Institutes of Health (R37HD17899 and U01 MH093349-04).

References

- Calkins SD, Fox NA, Marshall TR. Behavioral and physiological antecedents of inhibited and uninhibited behavior. *Child Development*. 1996; 67:523–540. [PubMed: 8625726]
- Caspi A, Elder GH, Bem DJ. Moving away from the world: Life-course patterns of shy children. *Developmental psychology*. 1988; 24(6):824.
- Coplan RJ, Closson L, Arbeau K. Gender differences in the behavioral associates of loneliness and social dissatisfaction in kindergarten. *Journal of Child Psychology and Psychiatry*. 2007; 48:988–995. [PubMed: 17914999]
- Crockenberg SB, Smith P. Antecedents of mother-infant interaction and infant irritability in the first three months of life. *Infant Behavior and Development*. 1982; 5(2):105–119.
- Doey L, Coplan RJ, Kingsbury M. Bashful Boys and Coy Girls: A Review of Gender Differences in Childhood Shyness. *Sex Roles*. 2013:1–12.
- Fagan J. The interaction between child sex and temperament in predicting behavior problems of preschool age children in day care. *Early Child Development and Care*. 1990; 59(1):1–9.
- Fox NA, Henderson HA, Rubin KH, Calkins SD, Schmidt LA. Continuity and discontinuity of behavioral inhibition and exuberance: Psychophysiological and behavioral influences across the first four years of life. *Child Development*. 2001; 72:1–21. [PubMed: 11280472]
- Fox NA, Henderson HA, Marshall PJ, Nichols KE, Ghera MM. Behavioral inhibition: Linking biology and behavior within a developmental framework. *Annu Rev Psychol*. 2005; 56:235–262. [PubMed: 15709935]
- Hane AA, Fox NA, Henderson HA, Marshall PJ. Behavioral reactivity and approach-withdrawal bias in infancy. *Developmental Psychology*. 2008; 44(5):1491. [PubMed: 18793079]
- Henderson HA, Fox NA, Rubin KH. Temperamental contributions to social behavior: The moderating roles of frontal EEG asymmetry and gender. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2001; 40(1):68–74. [PubMed: 11195566]
- Hepper PG, Shannon EA, Dornan JC. Sex differences in fetal mouth movements. *Lancet*. 1997; 350:1820. [PubMed: 9428256]
- Kagan, J. *Galen's Prophecy*. New York: Basic Books; 1994.
- Kagan J, Reznick JS, Snidman N, Gibbons J, Johnson MO. Childhood derivatives of inhibition and lack of inhibition to the unfamiliar. *Child Development*. 1988:1580–1589. [PubMed: 3208569]
- Kagan J, Snidman N. Infant predictors of inhibited and uninhibited profiles. *Psychological Science*. 1991; 2(1):40–44.
- Kagan, J.; Snidman, N. *The Long Shadow of Temperament*. Belknap Press/Harvard University Press; 2004.

- Kagan J, Snidman N, Kahn V, Towsley S, Steinberg L, Fox NA. The preservation of two infant temperaments into adolescence. *Monographs of the Society for Research in Child Development*. 2007;i-95.
- Park SY, Belsky J, Putnam S, Crnic K. Infant emotionality, parenting, and 3-year inhibition: Exploring stability and lawful discontinuity in a male sample. *Developmental Psychology*. 1997; 33(2):218. [PubMed: 9147831]
- Schwartz CE, Snidman N, Kagan J. Adolescent social anxiety as an outcome of inhibited temperament in childhood. *Journal of the American Academy of Child & Adolescent Psychiatry*. 1999; 38(8): 1008–1015. [PubMed: 10434493]
- Schwartz CE, Wright CI, Shin LM, Kagan J, Rauch SL. Inhibited and uninhibited infants “grown up”: Adult amygdalar response to novelty. *Science*. 2003; 300(5627):1952–1953. [PubMed: 12817151]
- Spangler T, Gazelle H. Anxious solitude, unsociability, and peer exclusion in middle childhood: A multitrait–multimethod matrix. *Social Development*. 2009; 18(4):833–856.
- White LK, McDermott JM, Degnan KA, Henderson HA, Fox NA. Behavioral inhibition and anxiety: the moderating roles of inhibitory control and attention shifting. *Journal of abnormal child psychology*. 2011; 39(5):735–747. [PubMed: 21301953]

Figure 1.
14-Month Behavioral Inhibition Scores

Figure 2a

Figure 2.

Figure 2a: 21/24-Month Behavioral Inhibition Scores

Figure 2b: 24-Month Inhibition by Sex and Reactivity Group

Table 1

Description of Testing Procedures

Sample	4 month Visit	14 month Visit	21 or 24 Month Visit
Sample 1	<p>1 Mobiles: A series of 3 mobiles back and forth in front of the infant's face (mobiles had 1, 2, and 7 colorful objects). Each mobile was moved back and forth for 20 seconds and each mobile was shown 3 times (for a total of 9 trials).</p> <p>2 Nonsense Syllables: Infants heard a female voice speaking three different syllables (ma, pa, ga) at three different increasing volumes, for a total of nine trials.</p> <p>3 Smiling Mother: Mother stood in front of her child once smiling but not talking, for one minute.</p>	<p>Episode 1: Placement of electrodes, Episode 2: Placement of a blood pressure cuff, Episode 3: Facial disapproval from an examiner or the mother, Episode 4: A noisy rotating wheel, Episode 5: Request to taste liquid from a dropper, Episode 6: Invitation to approach an unfamiliar object (a robot) or unfamiliar adults, despite a friendly invitation to do so.</p>	<p>Episode 1: An unfamiliar room containing novel objects Episode 2: Application of heart rate electrodes Episode 3: Blood pressure cuff, Episode 4: Request to accept liquid through an eyedropper Episode 5: An unfamiliar woman, Episode 6: A person dressed as a clown, and Episode 7: A radio-controlled robot. (Seen at 21 months)</p>
Sample 2	<p>1 Mobiles: The visual stimuli were two sets of three mobiles differing in the number of hanging characters (1, 3, or 6 characters; 1 set of brightly colored Winnie the Pooh and a 2nd set of Mickey Mouse mobiles) that were displayed at the infant's eye level, 12 inches from the face for nine trials of 20 sec each, with an interstimulus interval of 10 sec</p> <p>2 Nonsense Syllables: The first set of auditory stimuli were a set of taped nonsense syllables (ma, ga, pa) presented in three consecutive 10-s trials, with 5-s interstimulus intervals. The second set of auditory stimuli was sentences (8 short sentences, 6 s in duration, followed by a 2-s interstimulus interval) presented in pairs. The pairs differed in the number of voices speaking and, as a result, the volume of the presentation.</p>	<p>Episode 1: 5 minutes of free play Episode 2: Stranger play with truck (1 min quiet play before invite, 2-3 minutes total) Episode 3: Stranger brings in robot (2 minutes) Episode 4: Reaction to stranger dressed in clown costume Episode 5: Willingness to climb through inflatable tunnel. (Seen at 24 months)</p>	
Sample 3	<p>1 Mobiles: The visual stimuli were two sets of three mobiles differing in the number of hanging characters (1, 3, or 6 characters; 1 set of brightly colored Winnie the Pooh and a 2nd set of Mickey Mouse mobiles) that were displayed at the infant's eye level, 12 inches from the face for nine trials of 20 sec each, with an interstimulus interval of 10 sec</p> <p>2 Nonsense Syllables: The first set of auditory stimuli were a set of taped nonsense syllables (ma, ga, pa) presented in three consecutive 10-s trials, with 5-s interstimulus intervals. The second set of auditory stimuli was sentences (8 short sentences, 6 s in duration, followed by a 2-s interstimulus interval) presented in pairs. The pairs differed in the number of voices speaking and, as a result, the volume of the presentation.</p>	<p>N/A</p>	<p>Episode 1: 5 minutes of free play Episode 2: Stranger play with truck (1 min quiet play before invite-2-3 minutes total) Episode 3: Stranger brings in robot (2 minutes) Episode 4: Reaction to stranger dressed in clown costume Episode 5: Willingness to climb through inflatable tunnel. (Seen at 24 months)</p>

Table 2

Descriptions of Scoring Measures

Sample	4 month Visit	14 month Visit	21 or 24 Month Visit
Sample 1 Thistle	<p>Motor Reactivity: Total # of movements of both arms, both legs, and hyperextension of the arms or legs; (1 pt. each); Bursts: Total # of occurrence of bursts of both arms or legs (two or more movements in quick succession) (2pts each) Negative Reactivity: Crying: Total # of seconds the child cried across the battery Fretting/Fussing: Total # of trials on which these two occurred Other Coded Reactivity: Arching: Total occurrences of arching of the back (2pts each); Fret/Fuss: # of trials on which these two responses occurred. Positive Vocalizations: # of trials on which response occurred; Smiling # of trials on which response occurred. Final Scores: Sum of frequencies for both Motor Reactivity and Negative Reactivity.</p>	<p>Total number of fears (including fretting or crying, or reluctance to approach) to an unfamiliar event or procedure. (Seen at 21 months)</p>	<p>Total number of fears (including fretting or crying, or reluctance to approach) to an unfamiliar event or procedure. (Seen at 21 months)</p>
Sample 2 Cohort 1: Thistle	<p>Motor Reactivity: Total # of movements of both arms, both legs, and hyperextension of the arms or legs; Can be one or two in succession; Bursts: Total # of occurrence of bursts of both arms or legs (three or more movements in quick succession) (1 pt. each) Arching: Total occurrences of arching of the back (1pts each) Negative Reactivity: Crying: Total # of seconds the child cried across the battery; Fret/Fuss: # of trials on which these two responses occurred. Positive Reactivity: Positive Vocalizations: # of trials on which response occurred; Smiling # of trials on which response occurred. Final Scores: Sum of frequencies for Motor Reactivity, Negative Reactivity, and Positive Reactivity.</p>	<p>Composite score of the following measures: Latency to touch Toy (free play), Latency to vocalize (free play), Time spent in proximity to mother (free play), Latency to vocalize to stranger, Latency to approach stranger, Latency to vocalize to approach robot, Latency to approach robot, Time spent in proximity to mother (robot), Time spent in proximity to mother (truck).</p>	<p>Composite score of the following measures: Time spent in proximity to mother (free play), Time spent in proximity to mother (truck), Time spent in proximity to mother (robot), Time spent in proximity to mother (tunnel), Latency to approach/touch truck, Latency to approach/touch robot, Latency to pass through tunnel. (Seen at 24 months)</p>
Cohort 1 Likert	<p>Motor Reactivity: movements of both arms, both legs, and hyperextension of the arms or legs; Can be one or two in succession Bursts: of occurrence of bursts of both arms or legs Arching: Total occurrences of arching of the back (three or more movements in quick succession) (all scored on a Likert Scale from 1 to 7) Negative Reactivity: Crying: seconds the child cried across the battery; Fret/Fuss:(scored on a Likert Scale from 1 to 7). (all scored on a Likert Scale from 1 to 7) Positive Reactivity: Positive Vocalizations; Smiling (all scored on a Likert Scale from 1 to 7). Final Scores: Composite score of Likert scores for Motor Reactivity, Negative Reactivity, and Positive Reactivity.</p>	<p>Composite score of the following measures: Latency to touch Toy (free play), Latency to vocalize (free play), Time spent in proximity to mother (free play), Latency to vocalize to stranger, Latency to approach stranger, Latency to vocalize to approach robot, Latency to approach robot, Time spent in proximity to mother (robot), Time spent in proximity to mother (truck).</p>	<p>Composite score of the following measures: Time spent in proximity to mother (free play), Time spent in proximity to mother (truck), Time spent in proximity to mother (robot), Time spent in proximity to mother (tunnel), Latency to approach/touch truck, Latency to approach/touch robot, Latency to pass through tunnel. (Seen at 24 months)</p>
Cohort 2 Likert	<p>Motor Reactivity: movements of both arms, both legs, and hyperextension of the arms or legs; Can be one or two in succession Bursts: of occurrence of bursts of both arms or legs Arching: Total occurrences of arching of the back (three or more movements in quick succession) (all scored on a Likert Scale from 1 to 7) Negative Reactivity: Crying: seconds the child cried across the battery; Fret/Fuss:(scored on a Likert Scale from 1 to 7). (all scored on a Likert Scale from 1 to 7) Positive Reactivity: Positive Vocalizations; Smiling (all scored on a Likert Scale from 1 to 7). Final Scores: Composite score of Likert scores for Motor Reactivity, Negative Reactivity, and Positive Reactivity.</p>	<p>Composite score of the following measures: Latency to touch Toy (free play), Latency to vocalize (free play), Time spent in proximity to mother (free play), Latency to vocalize to stranger, Latency to approach stranger, Latency to vocalize to approach robot, Latency to approach robot, Time spent in proximity to mother (robot), Time spent in proximity to mother (truck).</p>	<p>Composite score of the following measures: Time spent in proximity to mother (free play), Time spent in proximity to mother (truck), Time spent in proximity to mother (robot), Time spent in proximity to mother (tunnel), Latency to approach/touch truck, Latency to approach/touch robot, Latency to pass through tunnel. (Seen at 24 months)</p>
Sample 3	<p>Motor Reactivity: movements of both arms, both legs, and hyperextension of the arms or legs; Can be one or two in succession Bursts: of occurrence of bursts of both arms or legs Arching: Total occurrences of arching of the back (three or more movements in quick succession) (all scored on a Likert Scale from 1 to 7) Negative Reactivity: Crying: seconds the child cried across the battery; Fret/Fuss:(scored on a Likert Scale from 1 to 7). (all scored on a Likert Scale from 1 to 7) Positive Reactivity: Positive Vocalizations; Smiling (all scored on a Likert Scale from 1 to 7). Final Scores: Composite score of Likert scores for Motor Reactivity, Negative Reactivity, and Positive Reactivity.</p>	<p>Composite score of the following measures: Time spent in proximity to mother (free play), Time spent in proximity to mother (truck), Time spent in proximity to mother (robot), Time spent in proximity to mother (tunnel), Latency to approach/touch truck, Latency to approach/touch robot, Latency to pass through tunnel. (Seen at 24 months)</p>	<p>Composite score of the following measures: Time spent in proximity to mother (free play), Time spent in proximity to mother (truck), Time spent in proximity to mother (robot), Time spent in proximity to mother (tunnel), Latency to approach/touch truck, Latency to approach/touch robot, Latency to pass through tunnel. (Seen at 24 months)</p>

Sample	4 month Visit	14 month Visit	21 or 24 Month Visit
	Scores: Composite score of Likert scores for Motor Reactivity, Negative Reactivity, and Positive Reactivity. Composite score of Likert scores for Motor Reactivity, Negative Reactivity, and Positive Reactivity.		
	truck, Latency to approach/touch robot, Latency to pass through tunnel. (Seen at 24 months)		

Table 3

Descriptives

Time Point	Reactivity Group	Sex	Mean	Std. Error	Lower 95% CI	Upper 95% CI
4-month Motor Reactivity	Low Reactive	Male	−0.68	0.096	−0.868	−0.492
		Female	−0.757	0.088	−0.93	−0.583
	High Reactive	Male	0.616	0.092	0.434	0.797
		Female	0.56	0.085	0.393	0.727
14-month Behavioral Inhibition	Low Reactive	Male	−0.494	0.125	−0.74	−0.248
		Female	−0.308	0.112	−0.529	−0.088
	High Reactive	Male	0.684	0.128	0.431	0.937
		Female	0.34	0.121	0.101	0.579
21/24-month Behavioral Inhibition	Low Reactive	Male	−0.416	0.115	−0.612	−0.161
		Female	0.015	0.107	−0.194	0.225
	High Reactive	Male	0.192	0.109	−0.002	0.426
		Female	0.163	0.103	−0.039	0.364