Child Development, September/October 2002, Volume 73, Number 5, Pages 1474–1485

Continuity and Change in Inhibited and Uninhibited Children

Marcie Pfeifer, H. H. Goldsmith, Richard J. Davidson, and Maureen Rickman

After screening 368 toddlers and selecting 77 into extremely inhibited, extremely uninhibited, and intermediate groups, 63 children (82%) were followed up at 4 and 7 years. Minority subgroups of both the inhibited and uninhibited children showed continuity on outcomes consisting of questionnaire measures of shyness, inhibitory control, and impulsivity, as well as multiepisode observational measures of behavioral inhibition and exuberance. Change from both inhibited and uninhibited status from the toddler age was more common than remaining extremely inhibited or uninhibited, but that change was largely constrained to the middle of the distribution.

INTRODUCTION

Research on behavioral inhibition, fearfulness, shyness, and related behavioral constructs in the first few years of life has advanced to the point where one can consider an initial set of questions to be at least tentatively answered. One of the first questions concerned the identity of the component behaviors of inhibition, and how these components change as the child matures. For example, inhibited 21-month-old children show long latencies to interact with an unfamiliar female examiner and retreat from unfamiliar people or objects rather than play or explore (Kagan, Reznick, & Snidman, 1987). At age 4 years, inhibited children are more likely to isolate themselves in a kindergarten classroom rather than interact with a teacher or peer (Kagan Reznick, Clarke, Snidman, & Garcia Coll, 1984; Kagan et al., 1987; Kagan, Snidman, & Arcus, 1993); they also withdraw in structured laboratory-based situations with peers (Rubin, Stewart, & Coplan, 1995) and, in general, show reduced social competence (Fox et al., 1995). These behaviors are outside infants' repertoires, but 4-month-olds who react to stimulation with crying and high activity tend to become inhibited later (Calkins, Fox, & Marshall, 1996; Kagan & Snidman, 1991).

Another question in this line of research concerns stability, in the sense of maintenance of rank order from infancy and toddlerhood through the early school years. For example, 4-year-old children who exhibited longer latencies to first unsolicited utterances toward a female stranger in a play situation tended, at age 8, to show longer latencies to first unsolicited utterance in a conversation about a circus. Behavioral inhibition was more stable than IQ during this age range (Asendorpf, 1992, 1993). Stability from infancy and toddlerhood into childhood has been much harder to demonstrate. A series of longitudinal studies from the laboratories of Kagan and Fox have been the most informative. A detailed review of this work is beyond the present article's scope, but the most prominent studies consider only extreme groups. The significant stability in these studies is usually emphasized, but is coupled with clear evidence that many inhibited and uninhibited infants and toddlers move into the normal range during early childhood (Kagan, Snidman, & Arcus, 1998). Perhaps because of the search for stability, the fear and shyness-related measures that do not show stability are sometimes de-emphasized. However, as an empirical point of departure, it is useful to recall that one of the first findings with regard to stability was that only 68% of infants were consistently reclassified by their behavior in the laboratory 1 month after initial testing (Garcia Coll, Kagan, & Reznick, 1984). These and other issues concerning continuity are treated in detail by Kagan (1994) and Reznick (1989). Among the longitudinal studies that examined the full range of fearfulness (not fully synonymous with behavioral inhibition), Lemery, Goldsmith, Klinnert, and Mrazek (1999) used three different maternal report questionnaires to measure fearfulness at six ages from 3 to 48 months. An autoregressive simplex model best fit the data, suggesting that only intermediate forms of fearfulness (with no additional effect of initial fearfulness status) mediated stability of these maternal reports. There was also evidence of some discontinuity between 12 and 18 months, a transition period too early to affect the interpretation of the present study.

Although less directly relevant to this study, the genetics and physiology of behavioral inhibition have received considerable research attention. Some rhesus macaque sires regularly produce anxious offspring (Suomi, 1987), suggesting transmission of a heritable trait. Several human twin studies support an inference of moderate heritability, including studies that

 0009-3920/2002/7305-0010

employed behavioral observations (DiLalla, Kagan, & Reznick, 1994; Emde & Hewitt, 2001; Goldsmith & Campos, 1986; Matheny, 1983, 1989; Plomin & Rowe, 1979), as well as parental report temperament questionnaires (e.g., Emde & Hewitt, 2001; Goldsmith, Buss, & Lemery, 1997; Goldsmith, Lemery, Buss, & Campos, 1999). In human research on physiological correlates of inhibition, one focus has been on arousal of the sympathetic branch of the autonomic nervous system (Kagan et al., 1987; Kagan et al., 1993), with some emphasis on the parasympathetic nervous system as well (Porges & Doussard-Roosevelt, 1998). The physiological literature has focused on salivary cortisol levels, heart rate and heart rate variability (Snidman, 1989), blood pressure, pupil dilation, temperature asymmetry (Kagan, Snidman, & Peterson, 2000), and asymmetry of power in the α band of the electroencephalogram (EEG) in frontal cortical regions (for a review, see Kagan, 1994). These data are not reviewed in this article because the current study did not include physiological measures. However, research that has shown greater relative right frontal EEG activation in early development among behaviorally inhibited children (Davidson & Fox, 1982; Fox, 1991; Fox & Davidson, 1988) is particularly relevant because the participants in the current study were drawn from a sample in which inhibited children also showed relatively greater right frontal activation (for more detail, see Davidson, 1993).

Work on delineation of the construct, its stability, its genetics, and its physiological underpinnings has advanced to the point where systematic reviews are available (Kagan, 1998; Rothbart & Bates, 1998). Indeed, the sense that behavioral inhibition is serving as the prototype for future investigations of other temperamental traits provides a strong incentive to pursue a detailed understanding of the inhibition concept itself and the children who are classified by it. The current study of 7-year-old children, who were classified during toddlerhood as inhibited, intermediate, or uninhibited, pursued this understanding.

As a first goal, this study sought to replicate the behavioral findings regarding the stability of inhibition. This goal was accomplished by exposing toddlers to inhibition-eliciting situations similar—and in some cases nearly identical—to those used by Kagan and colleagues. The present study also examined longitudinal temperament questionnaire data using both variable-centered and person-centered approaches, and considered observational and questionnaire data jointly. In contrast to most previous research, this research incorporated an intermediate group (neither inhibited nor uninhibited). Although this might appear to be a purely methodological issue, the study of an intermediate group allows for the determination of whether the uninhibited or inhibited group, or both, deviated from a normative group, and this distinction can be practically important (Kalin, Larson, Shelton, & Davidson, 1998). The implications are quite different if, on an outcome measure, the inhibited group differs significantly from uninhibited and intermediate groups (with the latter two being indistinguishable) versus the uninhibited group differing significantly from the inhibited and intermediate groups (with the latter two being indistinguishable).

The second goal was to shift the research focus toward uninhibited children. Our assumption was that being uninhibited might not simply be "the other side of the coin" of being inhibited, even if the groups are initially defined as occupying opposite poles of behavioral rating scales. Thus, we sought to characterize more fully the later temperament of previously uninhibited toddlers. To do this, a set of measures for a construct labeled exuberance was developed. Exuberance is an intense, approach-oriented facet of positive emotionality (as opposed to, e.g., contentment). At high levels, it includes connotations of risk taking and lack of restraint. In contrast to most studies, the present study's exuberance measures were operationally and conceptually distinct from the follow-up inhibition measures. Thus, it was operationally possible given our measures—although conceptually unlikely—for an individual to be both inhibited and exuberant on follow-up.

METHOD

Participants

Participants were 63 seven- to eight-year-old children (M = 7.64 years, SD = .20; 38 males, 25 females). These participants were selected because they had been classified as either extremely inhibited, extremely uninhibited, or intermediate at age 32 months, based on behavioral reactions to several novel situations within a play session. A total of 368 children (208 males,160 females), recruited from published birth announcements, participated in these earlier play sessions. Initially, 77 highly selected children were recruited for follow-up (24 inhibited, 24 intermediate, 29 uninhibited), and studied by Davidson (1993), Finman (1992), and Rickman and Davidson (1994). Approximately 5 years later, 63 of the 77 families (82%) again agreed to come to the laboratory. Of these 63 children, 23 had been classified as inhibited as toddlers, 16 had been classified as intermediate, and 24 had been classified as uninhibited. Each of these three groups was approximately 60% males, selected so that gender would not differ across the groups (Finman, 1992). This intentional control over the gender ratio meant that investigations of gender differences were inappropriate for this sample, and the sample size was chosen on the assumption that males and females would be pooled. The analyses (presented in the Results section) checked for gender effects. Given the sample size of the present study and the possibility of chance findings, there was simply no good evidence for gender differences; however, the least inhibited 7-year-olds tended to be boys. Because this issue was not a main focus of the present study, it is not further addressed.

Behavioral Screening at Age 32 Months

At age 32 months, each 25-min play session included two previously unacquainted same-sex children. Classification was based on several discrete variables: time proximal to mother, latency to speak, latency to touch first toy, latency to enter a toy tunnel, latency to approach a robot that had entered the room without warning, and latency to approach a stranger and take a toy from this stranger. Time proximal to the mother while not playing and total time talking were also scored. Specific criteria on each of these variables were designed to yield groups of inhibited and uninhibited children (for details, see Finman, 1992). For example, to be classified as inhibited, children must have spent >9 min within arm's length of the mother while not playing, and they must not have approached the stranger. The inhibited children also met relevant criteria on the other variables, and the application of all the criteria resulted in only 7% of the children being classified as inhibited. Approximately 10% of the 368 children met the criteria for being uninhibited, although not all of these could be recruited for the follow-up study. The single variable that best differentiated the two extreme groups was time proximal to the mother; thus, being within one half of a standard deviation of the mean on this variable was a criterion for inclusion in the intermediate group. After this screen for intermediate children, the criterion of latency to touch the first toy being less than one third of a standard deviation above the mean was imposed. Then, in further selecting the intermediate group, some variety in the other behaviors (listed above) was sought and gender was approximately balanced. Thus, the inhibited and uninhibited groups were more extreme in this study than is typical in the literature, and the intermediate group was more heterogeneous than the two extreme groups. Variation within these highly selected groups was not further analyzed.

Behavioral Assessment at Age 7 Years

Each session lasted approximately 2 hours, during which 23 episodes selected from the Preschool Laboratory Temperament Assessment Battery (PS-Lab-TAB; Goldsmith, Reilly, Lemery, Longley, & Prescott, 1993) were administered. Anticipating the results, variables from 6 of these 23 episodes contributed to two composite measures that significantly differentiated the inhibited, intermediate, and uninhibited groups. (Space constraints preclude description in this article of the other 17 PS-Lab-TAB episodes that were used. A manual is available on request from the second author.)

These two composite measures were designated inhibition and exuberance. The inhibition composite was derived from episodes called Stranger Approach, Scary Mask, and Risk Room. In the Stranger Approach episode (similar to many in the literature), a stranger approached the child while alone in a testing room. Following a detailed script, the stranger entered the room and engaged the child in a conversation for a short time before leaving. In the Scary Mask episode, a friendly female stranger wearing a snarling wolf mask confronted the child. First, the friendly stranger talked with the child while wearing the mask, and then removed the mask and invited the child to touch and wear it. In a third episode, modeled directly from Kagan's work (Reznick, Kagan, Snidman, Gersten, Baak, & Rosenberg, 1986), the child was escorted into a room filled with slightly risky play objects including a 7-ft-long cloth tunnel, a balance beam, a gorilla head mask mounted on a black stand, a large black box with a jagged opening, and a mattress with a set of stairs beside it (to allow jumping onto the mattress). After playing freely with these objects for 5 min, the child was asked by the experimenter to engage in the risky behavior afforded by each object. To varying degrees, these three episodes incorporated elements of novelty and mild threat.

The exuberance composite was derived from episodes called Make That Car Go; Surprise! It's a Pop-Up Snake; and Popping Bubbles. The Make That Car Go episode allowed the child to race a remotely controlled car on a track. The child and experimenter raced against one another. The Surprise! It's a Pop-Up Snake episode began with the experimenter surprising the child with a fake can of candy actually containing two coiled, paper-covered springs, referred to as "pop-up snakes." Then, the child surprised his/her parent with the pop-up snakes, with the child's anticipation of surprising the parent being of key interest. In the Popping Bubbles episode, the child was allowed to blow and pop bubbles with a bubble toy. The experimenter prompted the child to pop bubbles with different body parts including hands, elbows, and legs. These games have the formal characteristics theoretically specified as relevant to joy (Campos, Barrett, Lamb, Goldsmith, & Stenberg, 1983). The child was given a significant goal (such as surprising the parent or popping many bubbles) that was attained during the episode (i.e., the parent acted surprised or bubbles were popped). Furthermore, each game invited high-intensity, exuberant responses from the child.

Coding and Data Reduction for Behavioral Measures

All sessions were videotaped and scored later by trained coders. The raw variables within each episode were scored within temporal intervals judged most appropriate for each episode—usually 5 to 10 s; these variables took the form of latencies, means, intensities, or frequencies. The variables were checked for skewness and transformed to a more normal distribution as needed, then converted to z scores. Behavioral scores were further aggregated by unit weighting, based on their intercorrelations, to obtain more robust measures.

Table 1 summarizes three types of psychometric data from each of the six episodes (described above) that were components of the inhibition and exuberance composites. The center section of Table 1 shows measures of scorer agreement, κs in the case of all variables besides latencies, for which percentage agreement was tallied within 1 s. Within each episode, two to four measures contributed to the composite score representing the episode (the Risk Room episode yielded two composites). As shown in Table 1, the median κ was .81 (*range* = .60–1.00), and the median percentage agreement for latencies was 90% (*range* = 88%–90%). The right column shows the mean correlation among the measures within each episode; these correlations support the formation of each within-episode composite. For example, in the first entry in Table 1, the three Stranger Approach variables (verbal hesitancy, activity decrease, and sad facial expression) had a mean correlation of .33, *range* = .26–.45, all $p < .05$.

The mean correlations among the narrower, within-episode composites for the two broad domains of inhibition (four composites) and exuberance (three composites) are also provided in Table 1 (footnotes b and d, respectively). These correlations justify formation of the broad inhibition and exuberance composites used in the remainder of the analyses. These broader composites were formed by averaging the four (for inhibition) or three (for exuberance) narrower composites.

Parental Report Assessment of Temperament

The behavioral assessment was supplemented with questionnaire data. At age 4 and 7 years, the Children's Behavior Questionnaire (CBQ; Rothbart, Ahadi,

Table 1 Constituent Measures of Final Behavioral Composites

Temperament Composite	Constituent Measure[a]	Mean Intervariable r of Measures within Episode
Inhibition[b]	Stranger Approach episode: Composite of activity decrease, κ = .76, verbal hesitancy, κ = .92, and sad facial expression, κ = .72.	.33
	Scary Mask episode: Composite of fearful facial expression[c] and avoidance behavior, κ = .90, scores.	.33
	Risk Room episode #1: Composite of number of objects touched, κ = 1.00, and latency to touch first object (90% agreement within 1 s), plus approach toward parent, κ = .81.	.37
	Risk Room episode #2: composite of tentativeness of play, κ = .96, and latency to vocalizations (90% agreement within 1 s).	.43
Exuberance[d]	Make That Car Go episode: Composite of approach behaviors, κ = .75, smiling, κ = .73, laughter, κ = .96, and latency to smile (90% agreement within 1 s).	.46
	Surprise! It's a Pop-Up Snake episode: Composite of laughter, κ = .82, approach, κ = .67, and smiling, κ = .60.	.47
	Popping Bubbles episode: Composite of smiling, κ = .66, latency to smile (88% agreement within 1 s), and laughter, κ = .95.	.63

[a] Information presented is the episode, followed by measures contributing to the composite, with κ for scoring agreement for each measure.
[b] Mean r of constituent measures across episodes = .23.
[c] There were too few instances of fearful facial expressions to compute κ.
[d] Mean r of constituent measures across episodes = .51.

& Hershey, 1994) was administered to mothers. Results of the 4-year administration, along with parental self-report data on their own personalities, were the focus of an earlier report (Rickman & Davidson, 1994). The CBQ was designed to assess temperament in children ages 3 to 8 years. Alpha coefficients for the scales range from .67 to .94 for the 15 scales. Parents decided whether each item was true or untrue (rated on a 7-point scale) of the child within the past 6 months. Items from the three scales used in the present analyses included: (1) Shyness: "Sometimes seems nervous when talking to adults s/he has just met"; (2) Inhibitory Control: "Is usually able to resist temptation when told he/she is not supposed to do something"; (3) Impulsivity: "When eager to go outside, sometimes rushes out without putting on the right clothes." For some analyses, these latter two scales were composited, as explained below.

At age 7 years, the Colorado Childhood Temperament Inventory (CCTI; Rowe & Plomin, 1977)—a 30-item, six-scale parental rating instrument—was also employed. Although this questionnaire has well-documented, satisfactory, psychometric characteristics, its scales yielded no significant differences or trends based on children's earlier classification as inhibited, intermediate, or uninhibited. Thus, the data from this questionnaire are not presented in the Results section.

RESULTS

Stability of Inhibition

The first goal was to replicate and extend prior stability research on inhibition. This goal was addressed in three phases: analyzing whether prior classification predicted inhibited behavior 4 years later in the laboratory; analyzing whether prior classification predicted mothers' ratings of shyness at ages 4 and 7 years; and considering whether stability in shyness from 4 to 7 years was itself associated with prior classification.

Laboratory follow-up with the inhibition composite. An ANOVA with toddler classification as inhibited, intermediate, or uninhibited as predictors, yielded an $F(2, 60) = 3.39$, $p < .05$, with a Scheffé post hoc test also significant for the difference between the inhibited ($M = .22$, $SE = .14$) and uninhibited ($M = -.17$, $SE = .07$) groups. As expected, the previously inhibited group scored highest on the inhibition composite at 7 years, with the previously intermediate group's mean ($M = .00$, $SE = .12$) about halfway between the other two, but not significantly different from either of them.

Questionnaire follow-up with the CBQ Shyness Scale. An ANOVA comparable to that for the laboratory measure was performed for age 4 CBQ Shyness, the scale with content most similar to inhibition as measured during toddlerhood. (The alternative to the Shyness Scale, CBQ Inhibitory Control, does not include novelty- or threat-related content, nor does it tap withdrawal or expressive fearful responses; rather CBQ Inhibitory Control taps the ability to refrain from prepotent, approach-related responses when asked to do so, or when the situation demands doing so.) The inhibited group ($M = 4.74$, $SE = .29$) showed the most shyness, and the intermediate group ($M = 3.77$, $SE = .28$) was also more shy than the previously uninhibited group ($M = 2.62$, $SE = .26$), $F(2, 55) = 15.87$, $p < .05$. CBQ Shyness at age 4 was correlated only .15 (*ns*) with the age 7 inhibition composite from the laboratory, suggesting that these two follow-up measures tap somewhat different aspects of inhibited behavior, but that both are related to prior classification.

This finding was replicated with the age 7 CBQ Shyness data. Previously inhibited children showed the most shyness ($M = 4.41$, $SE = .29$), with the intermediate ($M = 3.09$, $SE = .28$) and uninhibited ($M = 2.64$, $SE = .24$) groups showing less, $F(2, 55) = 12.87$, $p < .05$. Scheffé post hoc tests confirmed that the significant difference was between the inhibited and uninhibited groups. Replicating the findings at age 4, CBQ Shyness at age 7 was correlated only .24 (*ns*) with the concurrent laboratory inhibition composite.

Stability of CBQ Shyness from 4 to 7 years. The CBQ Shyness Scale showed high stability, $r = .77$, $p < .05$. Of course, both the children's actual behavior and the mothers' stable but possibly inaccurate perceptions might be sources of this observed stability.

Examined was whether this stability was associated with prior classification. The scatterplot of the .77 correlation in Figure 1 shows quite clearly that the children rated as stably shy by their mothers had earlier been inhibited in the laboratory. For instance, an optimal raw cutoff score (for which no empirical guidelines exist) of 4.9 or above at both ages for classification as shy on the 1 to 7 response scale yielded 8 stably shy children, all of whom were formerly inhibited. Change from extreme toddler status was almost always to the intermediate range rather than the other extreme (an exception was that 1 previously inhibited child was later reported as moderately bold [nonshy] at 4 years and quite bold at 7 years). About half of the previously inhibited children were not stably shy, with the exact number obviously being dependent on the cutoff score for defining shy.

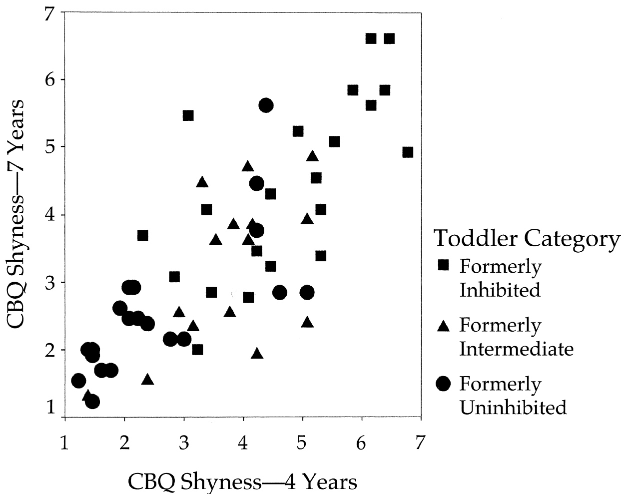

Figure 1 Stability of Child Behavior Checklist (CBQ) Shyness, based on prior classification.

Stability of Lack of Inhibition

As indicated in the Introduction, we did not view the uninhibited behavioral pattern as simply the opposite pole of inhibition. Thus, the second goal of this study was to characterize more fully the later temperament of previously uninhibited toddlers. Continuing with the examination of Figure 1, the conclusions just reached (above) for prior inhibition and later shyness were paralleled by findings for the previously uninhibited group. Children who were rated as stably bold by their mothers had earlier been uninhibited in the laboratory, with a few exceptions depending on the cutoff score for a bold classification. For example, an optimal raw cutoff score of 3 at both ages on the 1 to 7 response scale for classification as bold yielded 18 stably bold children, of whom 15 were uninhibited toddlers and the other 3 were formerly intermediate. These descriptive findings highlight the considerable individual developmental change that coexists with continuity demonstrated at the group level.

Laboratory measure of exuberance related to earlier lack of inhibition. For exuberance, the ANOVA with toddler classification as the predictor variable yielded $F(2, 60) = 4.66$, $p < .05$. The formerly intermediate group ($M = -.33$, $SE = .15$) scored lower than did the inhibited group ($M = -.04$, $SE = .14$), which showed less exuberance than did the formerly uninhibited group ($M = .26$, $SE = .11$). The Scheffé test on the difference between the intermediate and uninhibited groups was significant, $p < .05$, but tests on the other two comparisons were not significant. The laboratory measure of exuberance was correlated $-.09$ with laboratory inhibition, measured during the same 2-hr period, thus confirming the independence of the two constructs.

Temperament questionnaire scales related to earlier lack of inhibition. The analyses described above (see Figure 1 and related text) established that CBQ Shyness was associated with earlier status as uninhibited. Two other CBQ scales, Impulsivity and Inhibitory Control, were conceptually related to a lack of inhibition and perhaps to its less social aspects. These two scales correlated $-.55$ at 4 years and $-.69$ at 7 years, which suggested that they could be combined to yield more parsimonious results. Thus the Inhibitory Control Scale was subtracted from the Impulsivity Scale. The resulting composite, called Impulsive/Low Control (computed so that higher positive scores meant greater impulsivity and lesser inhibitory control), was correlated $-.49$ with CBQ Shyness at age 4 years. Results of the ANOVA with toddler status predicting Impulsive/Low Control were significant, $F(2, 55) = 5.63$, $p < .05$. The group means were in the expected order (uninhibited, .40; intermediate, .02; inhibited, $-.44$; with *SE*s of .19, .17, and .18, respectively), with a significant Scheffé post hoc test on the contrasts showing that the previously uninhibited group was more impulsive/less controlled than was the inhibited group. The intermediate group was not significantly different from either extreme group.

A very close replication of these results was obtained at age 7 years. The age 7 Impulsive/Low Control composite was also correlated $-.49$ with concurrent CBQ Shyness. Results of the ANOVA with toddler status predicting age 7 Impulsive/Low Control were significant and of similar strength as the age 4 findings, $F(2, 55) = 5.05$, $p < .05$. The group means were again in the expected order (uninhibited, .42; intermediate, $-.06$; inhibited, $-.40$; with *SE*s of .20, .16, and .20, respectively), with the only significant Scheffé post hoc test on the contrasts showing that the previously uninhibited group was more impulsive/less controlled than was the inhibited group.

The age 7 CBQ Impulsive/Low Control composite correlated .36 with the concurrent laboratory-based exuberance composite and $-.43$ with laboratory-based inhibition, both ps $< .05$; corresponding correlations for the age 4 CBQ Impulsive/Low Control composite with later laboratory-based composites were in the predicted direction, but nonsignificant. Thus, the concurrent—but not the longitudinal–questionnaire and Lab-TAB results converged, and both age 7 and age 4 Impulsive/Low Control measures were predicted by earlier toddler status in the same way.

Also examined was the stability of CBQ Impulsive/Low Control scores from age 4 to 7 years. The

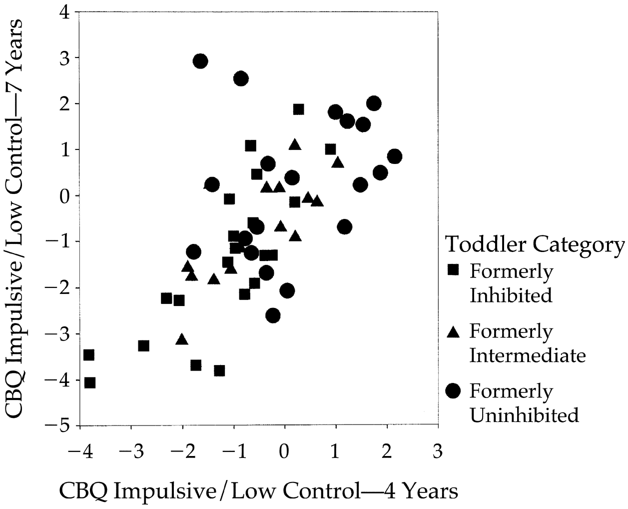

Figure 2 Stability of Child Behavior Checklist (CBQ) Impulsive/low control, based on prior classification.

correlation, $r = .67$, $p < .05$, was examined with a scatterplot (see Figure 2) analogous to that shown in Figure 1 for CBQ Shyness. Interpretation of this scatterplot was also similar to that offered for Figure 1 (see above). A substantial majority of the children who were rated as impulsive and low in control at both 4 and 7 years by their mothers (upper right corner of figure) had been uninhibited toddlers (i.e., 7 of the 10 most extreme on the bivariate distribution of the age 4 and 7 CBQ composites). The lower left corner of Figure 2 shows that 8 of the 10 most extreme cases on the bivariate distribution (stably rated as low impulsive and high control) had been inhibited as toddlers. As in Figure 1, the majority of both previously inhibited and previously uninhibited children were not rated stably toward the corresponding pole on CBQ Impulsive/Low Control, but change to a position near the opposite extreme was rare. One might question whether Figures 1 and 2 supply nonredundant information. That is, are the extreme cases in each figure all the same individuals? In short, the answer is "no," with the exception that the three cases in the lower left corner of Figure 2 (stably non-impulsive and highly controlled) were all stably shy as well. Otherwise, the extreme cases in each figure are distributed through the corresponding extreme and the middle range in the other figure.

Person-Centered Analyses of Change

Other than Figures 1 and 2, the analyses presented above were largely variable centered. However, person-centered analyses lend a complementary perspective. Looking back at Figures 1 and 2, the normative developmental pattern for extreme toddlers was to change into the broad middle of the distribution. Change to the opposite pole of either inhibition or lack of inhibition was rare for lack of inhibition and nonexistent for inhibition, at least from the perspective of maternal report.

Do the laboratory-based outcomes at age 7 years support this conclusion from the questionnaire data? Before this question could be answered from a person-centered perspective, it was necessary to derive a cutoff point (or points) for the distributions of laboratory inhibition and exuberance scores that would classify the children into groups. There was no obvious a priori method for determining this cutoff score, but, for the purposes of the present study, it was reasonable to set a cutoff score that maximized the differentiation of the toddler groups at 7 years, and then use this maximally discriminative cross-classification to study change. That is, it would have been inappropriate to view the resulting significance levels for the χ^2 as accurate, because the cutoff point was placed at the point that best discriminated the extreme group at follow-up and thus maximized chance factors. However, this procedure did yield an appropriate baseline for examining change.

When all possible 2×2 cross-classifications were examined, two widely separated cuttoff scores that yielded significant associations were apparent. The results for the inhibited toddler group are shown in the top portion of Table 2. That group was contrasted to the combined intermediate and uninhibited groups. All possible cutoff scores were examined to select regions of the distribution that yielded the highest χ^2. For the inhibition domain, one cutoff score (see left-hand side of Table 2) split the 7 most inhibited 7-year-olds from all of the other children. According to this analysis, 17 of the 23 children (74%) who had been inhibited as toddlers fell below the cutoff score on the 7-year laboratory inhibition composite. A second cutoff score was closer to the middle of the distribution, as shown in Table 2. This second cross-classification captured about half (11 out of 23) of the inhibited toddlers as inhibited at age 7.

Next performed was a parallel cross-classification for uninhibited toddlers followed up with the laboratory-based exuberance composite at age 7 years. The results (bottom portion of Table 2) were similar to the inhibition results, with one exception. The first high value of χ^2 (4.044) was obtained for a cutoff score that designated only 5 children as highly exuberant (see bottom left of Table 2). Twenty formerly uninhibited toddlers (83%) were not in this extreme group. A second cross-classification (see bottom right of Table 2)

Table 2 Cross-Classifications of 7-Year Laboratory Composites by Toddler Status, with Cutoff Scores that Maximize χ^2

	Optimal Cut toward the High Pole of the Distribution: Toddler Classification		Optimal Cut toward the Low Pole of the Distribution: Toddler Classification	
Age-7 inhibition	Not Inhibited	Inhibited	Not Inhibited	Inhibited
Not inhibited	39	17	32	12
Inhibited	1	6	8	11
	$\chi^2(1, N = 63) = 8.226, p = .0041$		$\chi^2(1, N = 63) = 5.368, p = .0205$	
Age-7 exuberance	Not Uninhibited	Uninhibited	Not Uninhibited	Uninhibited
Not exuberant	38	20	16	1
Exuberant	1	4	23	23
	$\chi^2(1, N = 63) = 4.044, p = .0443$		$\chi^2(1, N = 63) = 10.24, p = .0014$	

yielded a stronger discrimination by setting the cutoff score nearer to the low pole (nonexuberant) of the distribution (in contrast to the middle of the distribution, as for inhibition in the top right of Table 2). The results of this second cross-classification showed that 23 of the 24 previously uninhibited toddlers were not lower in exuberance at age 7 than the cutoff score.

The distribution of cases that led to the cross-classifications in Table 2 are depicted graphically in Figure 3. The prior classification of each child is shown by the three types of shading in the histogram. Individuals who scored within narrow intervals on the age 7 variables are grouped together within single columns in Figure 3.

In summary, Table 2 and the associated raw data suggest three conclusions from a person-centered perspective: (1) a majority of many extreme toddlers were no longer extreme at age 7 years—regardless of their classification in Table 2—when followed up with conceptually and empirically related laboratory measures; (2) a small group of extreme toddlers (6 inhibited, or 26%; 4 uninhibited, or 17%) continued to be extreme at age 7, even when cutoff scores were set conservatively; and (3) extreme toddlers tended not to change to the other pole of the distribution, and this tendency was particularly salient for uninhibited toddlers in this sample.

A next step in the person-centered analyses began with the recognition that there was no special reason that the toddler age assessment should be considered the true baseline for change. For instance, the analyses might have been structured differently by considering

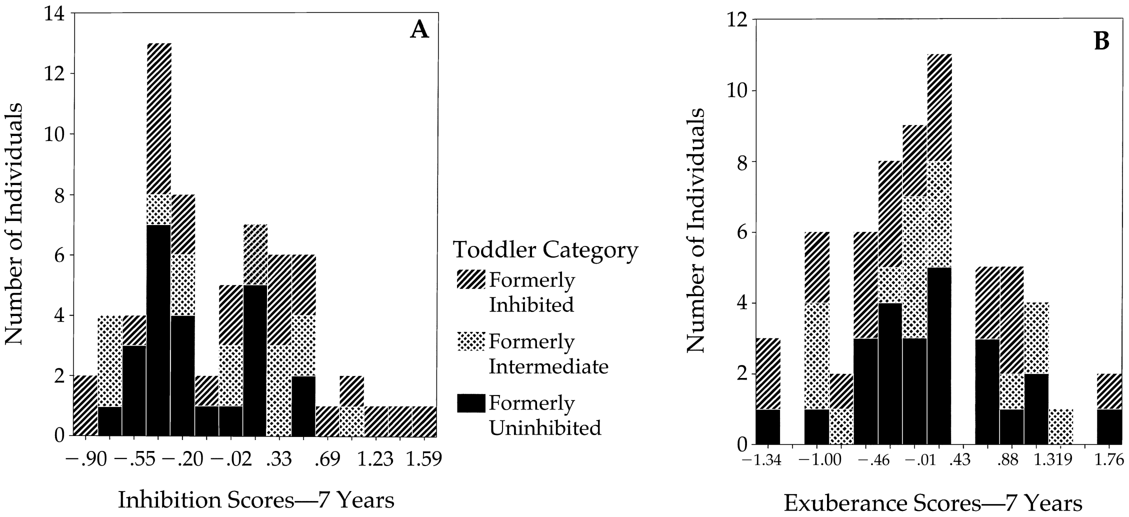

Figure 3 Histogram illustrating the position of previously classified toddlers on the distribution of laboratory-based (A) inhibition scores and (B) exuberance scores at age 7 years.

both the toddler behavioral observation and 4-year maternal-rated CBQ Shyness as selection criteria, to ensure that the sample was stably shy during this early period and that the inhibited/shy behavioral pattern was perceived by mothers as well as apparent in the laboratory. If the inhibited toddlers had been required to also be rated above the median on CBQ Shyness at age 4 years, 15 out of 21 participants (who also had 7-year CBQ data) would have been retained as the inhibited group. In the context of discovery, it is interesting to note that 14 of these 15 were still above the median on CBQ Shyness ratings at age 7, and the 15th participant was only slightly below the median. However, of particular cautionary note, only 7 of these 15 formerly shy children were clearly above the median on the laboratory-based inhibition composite at age 7, with 4 essentially at the median and the other 4 scoring toward the noninhibited pole of the laboratory-based measure. Also relevant to these person-centered analyses is the fact that 3 individuals, all classified as inhibited as toddlers, fell at the extremely inhibited pole of the age-7 inhibition laboratory composite. The mothers of only 2 of these 3 children completed CBQs, and both mothers rated their children as average in CBQ Shyness at both 4 and 7 years. Thus, the laboratory results and the maternal ratings diverged in these cases.

DISCUSSION

Continuity was demonstrated for both inhibition and lack of inhibition. Minority subgroups of both previously inhibited and uninhibited children remained extreme. The evidence for a type of limited discontinuity was also striking. A majority or near-majority subgroup of extreme children did not remain extreme, but also did not change to the other pole of the distribution. That is, at least half, and by most measures a majority, of previously inhibited children changed so that they lay in a middle range of inhibition-related measures by age 7. Previously uninhibited children tended to change as well, although to a lesser degree. These findings of limited change have been interpreted as temperament setting constraints on developmental outcome (Kagan, 1998).

One virtue of the present study's findings is that they held for both maternal report questionnaire measures and for extensive laboratory-based measures. Moreover, separate follow-up questionnaire and laboratory measures were used to distinguish the inhibited and uninhibited groups. This use of separate follow-up measures advanced our goal of highlighting the uninhibited group. Most prior studies have used one set of measures to define both groups initially, as was done in this study, but have also followed up with only one set of measures. This common strategy encourages a view of uninhibited status as the "other side of coin" from inhibited status. The strategy used in the current research yielded evidence for the "polar opposite" view in some instances, but not in others.

One salient feature of this study is that a very large sample was screened so that quite extreme groups could be selected. Only 7% of the sample was selected as inhibited, whereas other prominent studies (see Kagan, 1998) have suggested that about 15% of a base population satisfy criteria for inhibition. The straightforward expectation was that selection of even more extreme children would exaggerate differences between the inhibited and uninhibited groups, but it is conceivable that this was not the case. In any case, the substantial amount of change observed in this study should not be attributed to the participants being less inhibited or uninhibited than those in other studies.

Another feature of this study was the use of an intermediate group, which has proven to be valuable in other studies. For instance, Kalin et al. (1998) found differences in plasma cortisol levels in young rhesus monkeys that had left versus right frontal asymmetry in EEG activation. The use of an intermediate group (no strong EEG asymmetry) allowed for the inference that the left frontally activated monkeys were low in cortisol level, rather than the right frontally activated group being high in cortisol. For the present study, the intermediate group also allowed us to infer whether the formerly inhibited or the formerly uninhibited group was "deviant" in cases of a difference between the two extreme groups. An example occurred with the age-7 CBQ Shyness data. Previously inhibited children showed the most shyness, with the intermediate and uninhibited groups showing less—and not significantly different—shyness. The value of including the intermediate group was also apparent in results for the age 7 laboratory-based exuberance composite; the intermediate rather than the inhibited group exhibited the fewest exuberant behaviors. Perhaps children with a history of normal-range behavior are better able to dampen extreme externalizing behaviors, even when the context encourages high energy and activity, as it did for the exuberant behaviors in the laboratory. For the laboratory-based inhibition composite as well as all of the CBQ scales reported as differentiating the groups, the intermediate group mean always lay between the means of the two extreme groups. This "intermediate outcome" of the formerly intermediate group is probably the implicit interpretation when only two extreme groups are compared, but empirical confirmation of the interpretation is useful.

The present study also had some clear limitations. Perhaps the most important is the single-day behavioral assessment at both the initial toddler assessment and at the age 7 follow-up. Most researchers would agree that even a single occasion of laboratory-based assessment of temperament is an advance over—or at least a useful supplement to—questionnaires (Kagan, 1998). However, laboratory-based behavioral assessment has its own set of limitations (Rothbart & Goldsmith, 1985), including the effects of novelty of the laboratory, other mood effects, and idiosyncratic background factors such as unrecognized, impending illnesses. Among others, Seifer, Sameroff, Barrett, and Krafchuk (1994) have demonstrated that the reliability of behavioral observations increases notably over multiple observational sessions. One implication of such demonstrations is that the size of the association between observed temperament and a reliable criterion would increase substantially if temperament were observed on a sufficient number of occasions to allow for reliability to approach a ceiling. In the present research, this principle applies to the interpretation of the convergence between questionnaire- and laboratory-based results at age 7 years, where laboratory-based exuberance and inhibition correlated only moderately with conceptually related CBQ scales (Impulsivity and Inhibitory Control, and Shyness, respectively). These findings contribute to the construct validation of both the CBQ and laboratory measures, although the moderate size of the cross-method correlations—although consistent with other studies—cautions against equating laboratory and questionnaire measures with similar names.

Another limitation was sample size. Although the initial sample consisted of 368 toddlers, the small subgroup who were extreme and the subgroups of those who showed distinctive patterns of continuity or change were small. Thus, the descriptive observation that uninhibited toddlers changed less than did inhibited toddlers is only tentative. Assuming that this difference is real, its cause is unknown. Two possibilities are that parents exert successful socialization pressure on some shy children (Kagan, 1994) or that experiences with formal schooling render the inhibition-eliciting contexts less novel and threatening for many previously shy children.

A question that will concern temperament researchers is why the results from two commonly used temperament questionnaires, the CCTI and the CBQ, differed so dramatically. None of the CCTI scales differentiated the formerly inhibited versus uninhibited children. This might be explained by the scale construction of the CCTI and resulting item content of the scales. The CCTI includes fear, anger, and sadness with the category of negative emotionality, which might have insufficient specificity to distinguish inhibited from uninhibited children in the present research design. Fear and shyness are, of course, separable constructs, and scales with shyness content rather than fear content are perhaps more likely to capture prior differences in laboratory-based inhibition. The behavioral inhibition literature has generally not relied on easy-to-use parental report, and although parental reports can certainly be overused, this report does demonstrate their value, or at least the value of a questionnaire such as the CBQ with its highly differentiated content scales and relatively high number of items per scale.

Finally, the person-centered analyses provided a more clinical view of continuity and change. When stability was built into either the outcome measure or the classification, more clearcut cases of continuity over time were apparent, and this held for both questionnaire and laboratory measures. On the other hand, given the increasing clinical attention to childhood shyness and social anxiety (Albano, Chorpita, & Barlow, 1996), the existence of cases in which single-occasion laboratory-based observation and maternal report of outcome diverged substantially should be a warning flag that age and method of assessment can drastically affect the predictive power of these measures.

ACKNOWLEDGMENTS

This research was supported by National Institutes of Health grants R01 MH50560 (to H. H. G.,) and P50-MH52354 (to Wisconsin Center for Affective Science, R. J. D., Director). The authors appreciate contributions to this project by Rona Finman and Caroline Bollinger.

ADDRESSES AND AFFILIATIONS

Corresponding author: H. H. Goldsmith, Department of Psychology, 1202 West Johnson Street, University of Wisconsin, Madison, WI 53706; e-mail: hhgoldsm@facstaff.wisc.edu. Marcie Pfeifer, Richard J. Davidson, and Maureen Rickman are also at University of Wisconsin, Madison.

REFERENCES

Albano, A. M., Chorpita, B. F., & Barlow, D. H. (1996). Childhood anxiety disorders. In E. J. Mash & R. A. Barkley (Eds.), *Child psychopathology* (pp. 196–241). New York: Guilford Press.

Asendorpf, J. B. (1992). A Brunswikean approach to trait continuity: Application to shyness. *Journal of Personality, 60*, 53–77.

Asendorpf, J. B. (1993). Beyond temperament: A two-factorial coping model of the development of inhibition during childhood. In K. H. Rubin & J. B. Asendorpf (Eds.), *Social withdrawal, inhibition, and shyness in childhood* (pp. 265–289). Hillsdale, NJ: Erlbaum.

Calkins, S. D., Fox, N. A., & Marshall, T. R. (1996). Behavioral and physiological antecedents of inhibited and uninhibited behavior. *Child Development, 67*, 523–540.

Campos, J. J., Barrett, K. C., Lamb, M. E., Goldsmith, H. H., & Stenberg, C. (1983). Socioemotional development. In M. M. Haith & J. J. Campos (Eds.), *Handbook of child psychology: Vol. 2. Infancy and developmental psychobiology* (pp. 783–916). New York: Wiley.

Davidson, R. J. (1993). Childhood temperament and cerebral asymmetry: A neurobiological substrate of behavioral inhibition. In K. H. Rubin & J. B. Asendorpf (Eds.), *Social withdrawal, inhibition, and shyness in childhood* (pp. 31–48). Hillsdale, NJ: Erlbaum.

Davidson, R. J., & Fox, N. A. (1982). Asymmetrical brain activity discriminates between positive versus negative affective stimuli in human infants. *Science, 218*, 1235–1237.

DiLalla, L. F., Kagan, J. J., & Reznick, J. S. (1994). Genetic etiology of behavioral inhibition among 2-year-old children. *Infant Behavior and Development, 17*, 405–412.

Emde, R. N., & Hewitt, J. K. (Eds.). (2001). I*nfancy to early childhood: Genetic and environmental influences on developmental change*. New York: Oxford University Press.

Finman, R. (1992). *Behavioral and psychophysiological correlates of inhibition to the unfamiliar in children*. Unpublished doctoral dissertation, University of Wisconsin, Madison.

Fox, N. A. (1991). It it's not left, it's right. *American Psychologist, 46*, 863–872.

Fox, N. A., & Davidson, R. J. (1988). Patterns of brain electrical activity during facial signs of emotion in ten month old infants. *Developmental Psychology, 24*, 230–236.

Fox, N. A., Rubin, K. H., Calkins, S. D., Marshall, T. R., Coplan, R. J., Porges, S. W., Long, J. M., & Stewart, S. (1995). Frontal activation asymmetry and social competence at four years of age. *Child Development, 66*, 1770–1784.

Garcia Coll, C. T., Kagan, J., & Reznick, J. S. (1984). Behavioral inhibition in young children. *Child Development, 55*, 1005–1019.

Goldsmith, H. H., Buss, K. A., & Lemery, K. S. (1997). Toddler and childhood temperament: Expanded content, stronger genetic evidence, new evidence for the importance of environment. *Developmental Psychology, 33*, 891–905.

Goldsmith, H. H., & Campos, J. J. (1986). Fundamental issues in the study of early temperament: The Denver Twin Temperament Study. *Advances in Developmental Psychology, 4*, 231–283.

Goldsmith, H. H., Lemery, K. S., Buss, K. A., & Campos, J. J. (1999). Genetic analyses of focal aspects of infant temperament. *Developmental Psychology, 35*, 972–985.

Goldsmith, H. H., Reilly, J., Lemery, K. S., Longley, S., & Prescott, A. (1993). *Preliminary manual for the Preschool Temperament Assessment Battery* (version 1.0, Technical Report). Department of Psychology, University of Wisconsin, Madison.

Kagan, J. (1994). *Galen's prophecy: Temperament in human nature*. New York: Basic Books.

Kagan, J. (1998). Biology and the child. In N. Eisenberg (Ed.), W. Damon (Series Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (pp. 177–235). New York: Wiley.

Kagan, J., Reznick, J. S., Clarke, C., Snidman, N., & Garcia Coll, C. (1984). Behavioral inhibition to the unfamiliar. *Child Development, 55*, 2212–2225.

Kagan, J., Reznick, J. S., & Snidman, N. (1987). The physiology and psychology of behavioral inhibition in children. *Child Development, 58*, 1459–1473.

Kagan, J., & Snidman, N. (1991). Infant predictors of inhibited and uninhibited profiles. *Psychological Science, 2*, 40–44.

Kagan, J., Snidman, N., & Arcus, D. (1993). On the temperamental categories of inhibited and uninhibited children. In K. H. Rubin & J. B. Asendorpf (Eds.), *Social withdrawal, inhibition, and shyness in childhood* (pp. 19–28). Hillsdale, NJ: Erlbaum.

Kagan, J., Snidman, N., & Arcus, D. (1998). Childhood derivatives of high and low reactivity in infancy. *Child Development, 69*, 1483–1493.

Kagan, J., Snidman, N., & Peterson, E. (2000). Temperature asymmetry and behavior. *Developmental Psychobiology, 37*, 186–193.

Kalin, N. H., Larson, C., Shelton, S. E., & Davidson, R. J. (1998). Asymmetric frontal brain activity, cortisol, and behavior associated with fearful temperaments in Rhesus monkeys. *Behavioral Neuroscience, 112*, 286–292.

Lemery, K. S., Goldsmith, H. H., Klinnert, M. D., & Mrazek, D. A. (1999). Developmental models of infant and childhood temperament. *Developmental Psychology, 35*, 189–204.

Matheny, A. P. (1983). A longitudinal twin study of stability of components from Bayley's Infant Behavior Record. *Child Development, 54*, 356–360.

Matheny, A. P. (1989). Children's behavioral inhibition over age and across situations: Genetic similarity for a trait during change. *Journal of Personality, 57*, 215–226.

Plomin, R., & Rowe, D. C. (1979). Genetic and environmental etiology of social behavior in infancy. *Developmental Psychology, 15*, 62–72.

Porges, S. W., & Doussard-Roosevelt, J. A. (1998). The psychophysiology of temperament. In J. D. Noshpitz (Ed.), *Handbook of child and adolescent psychiatry*. New York: Wiley.

Reznick, J. S. (Ed.). (1989). *Perspectives on behavioral inhibition*. Chicago: The University of Chicago Press.

Reznick, J. S., Kagan, J., Snidman, N., Gersten, M., Baak, K., & Rosenberg, A. (1986). Inhibited and uninhibited children: A follow-up study. *Child Development, 57*, 660–680.

Rickman, M. D., & Davidson, R. J. (1994). Personality and behavior in parents of temperamentally inhibited and uninhibited children. *Developmental Psychology, 30*, 346–354.

Rothbart, M. K., Ahadi, S. A., & Hershey, K. L. (1994). Temperament and social behavior in childhood. *Merrill Palmer Quarterly, 40*, 21–39.

Rothbart, M. K., & Bates, J. E. (1998). Temperament. In N. Eisenberg (Ed.), W. Damon (Series Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (pp. 105–176). New York: Wiley.

Rothbart, M. K., & Goldsmith, H. H. (1985). Three approaches to the study of infant temperament. *Developmental Review, 5*, 237–250.

Rowe, D. C., & Plomin, R. (1977). Temperament in early childhood. *Journal of Personality Assessment, 41*, 151–156.

Rubin, K. H., Stewart, S. L., & Coplan, R. J. (1995). Social withdrawal in childhood: Conceptual and empirical perspectives. *Advances in Clinical Child Psychology, 17*, 157–196.

Seifer, R., Sameroff, A. J., Barrett, L. C., & Krafchuk, E. (1994). Infant temperament measured by multiple observations and mother report. *Child Development, 65*, 1478–1490.

Snidman, N. (1989). Behavioral inhibition and sympathetic influence on the cardiovascular system. In J. S. Reznick (Ed.), *Perspectives on behavioral inhibition* (pp. 51–70). Chicago: The University of Chicago Press.

Suomi, S. J. (1987). Genetic and maternal contributions to individual differences in rhesus monkey biobehavioral development. In N. A. Krasnegor, E. M. Blass, M. A. Hofer, & W. P. Smotherman (Eds.), *Perinatal development: A psychobiological perspective* (pp. 397–420). New York: Academic Press.