Supplemental Methods

Measurements (see Supplemental Table 1 for overview of the study design).

14- & 24-month Observed Behavioral Inhibition (BI) – Infants' reactions to unfamiliar stimuli in the laboratory were coded to provide an index of BI [see (1-4) for a complete description of these procedures]. At 14 months, the unfamiliar stimuli consisted of: (1) an unfamiliar room/environment, (2) an adult stranger, and (3) a novel toy/object. At 24 months, the infant was presented with identical stimuli, and in addition, the infant's reactions to an adult stranger dressed in a clown costume were recorded, as was his/her willingness to crawl through an inflatable tunnel when encouraged to do so by the experimenter.

At 14 and 24 months, a standardized composite index of BI was computed based on the infant's reactions to these unfamiliar stimuli (i.e., latency to touch a toy, vocalize, or approach a stranger, and proximity to mother). Scores on the index of BI ranged from -1.85 to 3.00 at 14 months, and from -2.30-2.56 at 24 months. Inter-coder reliability was computed for 15% of the sample at 14 months and 24% of the sample at 24 months. Pearson correlations ranged from .85-1.00 at 14 months and .77-.97 at 24 months.

4- & 7-year Observation of Social Reticence – In early childhood, inhibition was assessed during children's play with unfamiliar peers. Specifically, children were observed in a same-age, same-sex quartet playgroup at 4 and 7 years of age to assess their reactions to and interactions with unfamiliar peers. Each playgroup consisted of one child high in social reticence (SR)/BI in the laboratory at the previous visit (one-half standard deviation or more above the mean), one child who exhibited very low SR/BI in the

previous laboratory visit (one-half standard deviation or more below the mean), and two average children (within one standard deviation of the mean). Data from the free-play session were included in this study. For this task children were left alone in the playroom for 15 minutes with age-appropriate toys, while their mothers remained in a waiting area.

The Play Observation Scale (POS; 5) was used to code behaviors in the play sessions. Reticent behavior composites at both 4 and 7 years of age were created by summing two categories: *Onlooking behavior,* defined as "the child observes the other children's activities without attempting to play," and *Unoccupied behavior*, defined as "the child demonstrates an absence of focus or intent" (see 2 and 5 for additional details). This sum was divided by the total number of observed segments minus the number of segments that could not be coded to create the proportion of time spent displaying *Reticence*. Three independent coders observed the tapes in approximately 90, 10-second intervals. Cohen's kappas ranged from .71-.86 at age 4 and .84-.88 at age 7. Scores ranged from 0-.85 (M = .18, SD = .16) at age 4 and 0-1.0 (M = .14, SD = .16) at age 7, with higher scores indicating greater reticence.

15-year Response Monitoring – Participants were administered a letters version of the flanker task (6). Stimuli were presented on a 15 inch computer monitor situated approximately 70 cm from the subjects. A trial began with a centrally presented cue in the shape of an asterisk for 200 ms followed by a blank screen for 300 ms. The target display of letters then appeared for 250 ms. Participants were required to respond within 1100 ms of the target display and the inter-trial interval randomly varied in length from 0 to 400 ms

Participants completed a practice block of 20 trials in order to become

accustomed to the button box and computer stimuli**.** Both reaction time and performance accuracy were recorded for each trial. If the response window elapsed, or if the participant did not respond, the trial was considered incorrect and classified as a response omission and processed separately from incorrect trials due to errors of commission. Participants were given short breaks in between each test block and the entire task took approximately 15 minutes to complete.

During the flanker task, electroencephalogram (EEG) was collected using the International 10-20 System (7). Data from three midline sites (Fz, Cz & Pz) was processed in accordance with previous work demonstrating strong amplitudes of the ERN as well as the Pe across these three sites (see 8-11 for examples). In addition to the EEG, electrooculogram (EOG) was recorded from two mini-electrodes, one placed on the outer canthus and one placed on the supra orbit (above) the right eye, in order to monitor participant eye blinks and artifact score the ERP data. Both EEG and EOG channels were amplified by SA Instrumentation Bioamplifiers by factors of 5000 and 1000 respectively. Filters were set at 0.1 Hz (high pass) and 100 Hz (low pass). Data was digitized on-line with customized acquisition software, sampled at a rate of 512 Hz with an Iotech Daqbook A/D converter, and then re-referenced with the average mastoid configuration.

The EEG was artifact scored with the ERP Analysis System (James Long Company, Caroga Lake, NY). Epochs containing signals +/- 200 μV were removed from analyses while eye movement artifact was regressed. All data channels were baseline corrected using a window from -150 to –50 ms prior to response. Specifically, the baseline window occurred during a blank screen period after the target stimulus has been removed from the screen and just prior to the participant's response. All data channels

were digitally re-filtered with a 15-Hz low-pass filter.

Grand mean as well as the individual waveforms for the response-locked EEG were assessed in order to maximize the selection of the peak of the component of interest across all participants. This approach yielded an ERN window starting at -20 to 120 ms (with the early timing potentially due to the initiation of the ERN prior to the button press registering) and a window of 100-400 ms for the Pe. Data were scored for both peak and mean amplitudes and the results did not differ between these approaches. Participants with less than 10 usable error trials were excluded from the analyses. The number of usable ERP trials ranged from 10 to 176 (M = 47, SD =34.2).

15-year Anxiety Assessment - Adolescents and their parent were separately administered the K-SADS (12), a semi-structured diagnostic interview assessing DSM-IV disorders, including: anxiety disorders (separation anxiety disorder, generalized anxiety disorder, social phobia, specific phobia, obsessive-compulsive disorder), mood disorders (major depression and dysthymia), and disruptive behavior disorders (attention-deficit/hyperactivity disorder, oppositional defiant disorder, conduct disorder). Detailed safety protocols were developed to address reports of abuse and suicidal ideation. Discrepancies between reports on the K-SADS were resolved by bringing the parent and child together to discuss discrepant perspectives, and the interviewer made a final determination based on clear evidence of functional impairment. Kappas were .84 for any disorder, .92 for anxiety disorders, and .90 for disruptive behavior disorders. In the current study, the anxiety measure included generalized anxiety, separation anxiety and social phobia. The rational for this approach reflected the fact that these three anxiety disorders accounted for the majority of anxiety diagnoses in the sample, and prior

research on pediatric anxiety has combined these diagnoses into a single category (13, 14). Presence of a single diagnosis or any combination of these disorders resulted in the classification of clinically significant anxiety.

Data Reduction and Analysis

Creation of the Behavioral Inhibition (BI) Composite – To create the BI composite used for the current study a mean score was calculated based on children's standardized inhibition scores at each of the four assessment time points (14- and 24-months, 4-and 7-years of age). The BI composite of the four time points serves as a more objective assessment of observed BI across childhood among different contexts (i.e. experimenter versus peer interaction) as opposed to relying on a single time-point of assessment. To examine the construct of behavioral inhibition over time there is a need to use age appropriate assessments. BI is assessed in infants and toddlers by measuring reactivity to strangers and novel items. However, for older children who have had more experience with novel items, a more accurate context for assessing inhibition is in play situations with unfamiliar peers. Thus developmental change in stimuli/contexts which more clearly elicit inhibited behavior necessitates the use of composites to assess continuity in the construct of BI over time. This approach has been used frequently in longitudinal research on inhibition (e.g. 4, 15-18) and was used in the current study to assess a more stable measure of behavioral inhibition. The current study also relied on observational, as opposed to maternal ratings of BI, which helped to avoid issues of shared method variance since the psychopathology assessment included maternal report.

Assessment scores were positively correlated across all age points (See Supplemental Table 2). Specifically, the two early assessments of inhibition were significantly correlated ($r = .29$, $p < .05$) as were the social reticence assessments ($r = .46$, $p < .01$). A composite of the inhibition paradigm at 14- and 24-months was also significantly correlated with the composite of the social reticence paradigm at 4- and 7-years of age ($r = .43$, $p < .01$). These patterns are in line with a recent review (19) which highlights moderate stability across temperament paradigms and stronger associations within paradigms.

The overall standardized BI composite was used to median-split participants into two BI groups: high or low. Mean BI values for these groups are presented in Supplemental Table 3. Follow-up analyses indicate that inhibition scores for the high and low BI groups differed significantly across all four assessments (t's $\geq$ -2.69, p's $\leq$.01). BI scores for the infant and childhood assessments are presented in Supplemental Figures 1 and 2, respectively.

Supplemental Table 1. This table depicts the study design including the ages of assessment and the specific measures used at each assessment.

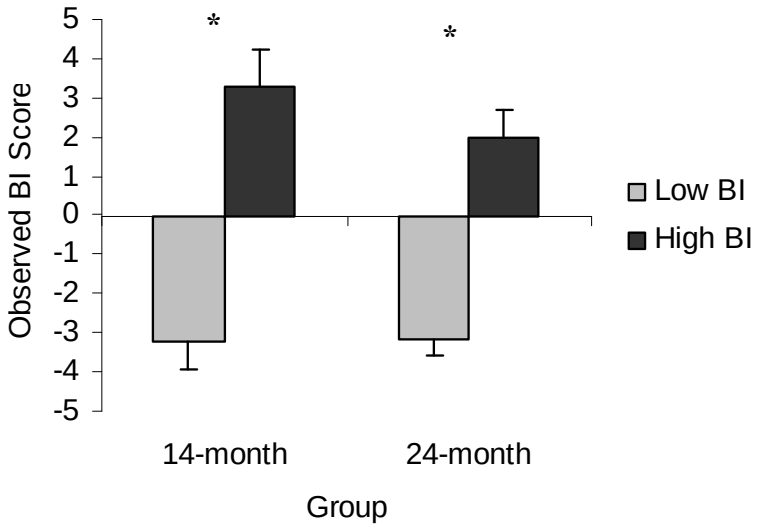
		MEASURE			
		Behavioral Inhibition (BI)	Social Reticence	Response Monitoring	Clinical Anxiety
AGE	14-months	●			
	24-months	●			
	4-years		●		
	7-years		●		
	15-years			●	●

Supplemental Table 2. This table displays the correlations among the inhibition assessments. At 14- and 24-months inhibition was assessed with a laboratory based observational paradigm. For 4- and 7-years of age inhibition was assessed by observing children's play with an unfamiliar peer in the laboratory. Note: $^{*}p < .05$, $^{**}p < .01$.

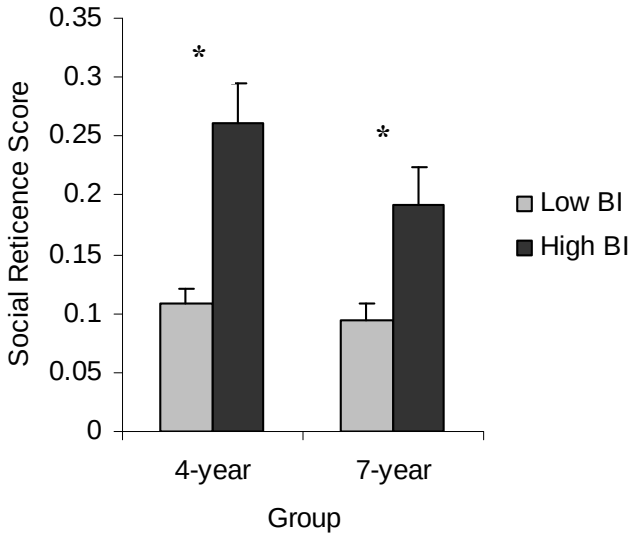
	Laboratory BI		**Social Reticence**	
	14-months	24-months	4-years	7-years
14-months	-	.29*	.19	.28*
24-months	.29*	-	.38**	.34**
4-years	.19	.38**	-	.46**
7-years	.28*	.34**	.46**	-

Supplemental Table 3. This table presents descriptive information for each of the BI composite groups.

	N	Age (SD)	Mean BI Composite Score (SD)
Low BI	41	14.93 (.97)	-.52 (.25)
High BI	41	15.18 (.97)	.44 (.68)



Supplemental Figure 1. This figure illustrates the mean inhibition scores by BI group for the two infant assessments of inhibition at 14- and 24-months. Note: * $p < .01$.



Supplemental Figure 2. This figure illustrates inhibition scores by BI group for the two childhood assessments of inhibition (i.e. Social Reticence) at 4- and 7-years of age. Note: * $p \leq .01$.

References

1. Calkins, S.D., Fox, N.A., & Marshall, T.R. (1996). Behavioral and physiological antecedents of inhibition in infancy. *Child Development, 67*, 523-540.
2. Fox, N.A., Henderson, H.A., Rubin, K.H., Calkins, S.D., & Schmidt, L.A. (2001). Continuity and discontinuity of behavioral inhibition and exuberance: psychophysiological and behavioral influences across the first four years of life. *Child Development, 72*, 1-21.
3. Kagan, J., Reznick, J.S., Clarke, C., Snidman, N. & Garcia-Coll, C. (1984). The physiology and psychology of behavioral inhibition in children. *Child Development, 58*, 1459-1473.
4. Reznick, J.S., Gibbons, J.L., Johnson, M.O., & McDonough, P.M. (1989). Perspectives on behavioral inhibition. In J.S. Reznick (Ed.) *Behavioral inhibition in normative sample* (pp. 25-49). Chicago: University of Chicago Press.
5. Rubin, K.H. (1989). The Play Observation Scale (POS). Ontario, Canada: University of Waterloo.
6. Eriksen, B.A. & Eriksen, C.W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics, 16*, 143-149.
7. Jasper, H.A. (1958). The ten-twenty system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10,* 371-375.

8. Hajcak, G., Holroyd, C.B., Moser, J.S. & Simons, R.F. (2005a). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology, 42,* 161-170.
9. Hajcak, G., McDonald, N., & Simons, R.F. (2003a). Anxiety and error-related brain activity. *Biological Psychology, 64*, 77-90.
10. Hajcak, G., McDonald, N. & Simons, R.F. (2003b). To err is autonomic: Error-related brain potentials, ANS activity, and post-error compensatory behavior. *Psychophysiology, 40*, 895-903.
11. Hajcak, G., Moser, J.S., Yeung, N. & Simons, R.F. (2005a). On the ERN and the significance of errors. *Psychophysiology, 42*, 151-160.
12. Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P. *et al.*, (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS_PL): Initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*, 980-988.
13. RUPP, The Research Unit on Pediatric Psychopharmacology Anxiety Study Group. (2000). Fluvoxamine for the treatment of anxiety disorders in children and adolescents. *New England Journal of Medicine, 344,* 1279-1285.
14. Roy, A.K., Vasa, R.A., Bruck, M., Mogg, K., Bradley, B., Sweeney, M. *et al.* (2008). Attention bias toward threat in pediatric anxiety disorders. *Journal of American Academy of Child and Adolescent Psychiatry, 47*, 1189-1196.
15. Guyer, A., Nelson, E.E., Perez-Edgar, K., Hardin, M.G., Roberson-Nay, R., Monk, C.S. *et al.* (2006). Striatal functional alteration in adolescents characterized by early childhood behavioral inhibition. *The Journal of Neuroscience, 26*, 6399-6405.
16. Mullen, M., Snidman, N. & Kagan, J. (1993). Free-play behavior in inhibited and uninhibited children. *Infant Behavior and Development, 16*, 383-389.
17. Pérez-Edgar, K., Schmidt, L. A., Henderson, H. A., Schulkin, J., & Fox, N. A. (2008). Salivary cortisol levels and infant temperament shape developmental trajectories in boys at risk for behavioral maladjustment. *Psychoneuroendocrinology, 33,* 916-925.
18. Pérez-Edgar, K., Roberson-Nay, R., Hardin, M. G., Poeth, K., Guyer, A. E., Nelson, E. E., *et al.* (2007). Attention alters neural responses to evocative faces in behaviorally inhibited adolescents. *NeuroImage, 35,* 1538-1546.
19. Henderson, H.A. & Wachs, T.D. (2007). Temperament theory and the study of cognition-emotion interactions across development. *Developmental Review, 27*, 396-427.