The Pennsylvania State University
The Graduate School

IMPLEMENTATION OF THE DIFFUSION MODEL ON DOT-PROBE TASK PERFORMANCE IN CHILDREN WITH BEHAVIORAL INHIBITION

A Thesis in

Psychology

by

Shane Wise

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

December 2019

The thesis of Shane Wise was reviewed and approved* by the following:

Cynthia Huang-Pollock
Professor of Clinical Psychology
Thesis Adviser

Koraly Pérez-Edgar
Professor of Developmental Psychology

Peter Arnett
Professor of Clinical Psychology

Melvin Mark
Department Head, Penn State Department of Psychology

*Signatures are on file in the Graduate School

Abstract

OBJECTIVE: Attentional bias to threat, the process of attending toward potentially threatening environmental stimuli over neutral stimuli, is positively associated with trait anxiety and may underlie the relationship between behavioral inhibition (BI) and anxiety disorders. However, the most commonly used measure of attentional bias to threat, the dot-probe task, has recently been criticized for demonstrating poor reliability. The present study aimed to assess whether the one-week test-retest reliability of the dot-probe task could be improved by utilizing computational parameters of performance. The drift diffusion model is a computational model of perceptual decision making that combines both RT and accuracy into a single set of performance indices. **METHOD:** The drift diffusion model was applied to longitudinal task data of 244 children with and without BI, aged 8-12, to produce three diffusion parameters: drift (*v*), non-decision time (*ter)*, and relative start-point (*z*). Event Related Potentials (ERPs) based on Electroencephalograph (EEG) data were also obtained as indices of direct, time-sensitive neurological markers of attention. **RESULTS:** We observed a non-significant relationship between BI and dot-probe performance as measured by both traditional RT measures and diffusion parameters. It was also found that diffusion model parameters of performance did not significantly improve test-retest reliability of the dot-probe task. Lastly, no significant correlations were found between traditional scoring measures or mean diffusion parameters and ERPs of early attention. **CONCLUSIONS:** This study shows that although diffusion modeling has been utilized to improve reliability and interpretability of findings in other cognitive tasks, it does not improve reliability of the dot-probe task. These results confirm recommendations to move away from using the dot probe task as the sole reliable index of attentional bias, even with indices of performance that relay more accurate summaries of performance. Alternative methods of studying attentional bias to threat are recommended.

Keywords: Threat bias, Behavioral inhibition, diffusion model

TABLE OF CONTENTS

Introduction

Behavioral Inhibition (BI) is an aspect of early temperament in which an individual is wary and avoidant of novel stimuli (Coll, Kagan, & Reznick, 1984; La Greca & Lopez, 1998). It is stable across the lifespan (Clauss & Blackford, 2012) and has been associated with long-term negative outcomes including social rejection (Wichmann, Coplan, & Daniels, 2004), social skills deficits, social anxiety (Biederman et al., 2001; Chronis-Tuscano et al., 2009; Hirshfeld-Becker et al., 2007), and anxiety disorders more generally (Fox & Kalin, 2014; Hirshfeld et al., 1992). In fact, children who are behaviorally inhibited in early childhood are at 7.59 times greater risk of developing an anxiety disorder than children who are not (Clauss & Blackford, 2012).

One possible mechanism that explains BI stability over the lifespan and its association with anxiety disorders is attentional bias to threat. Attentional bias to threat refers to the process whereby a potentially threatening stimulus is preferentially attended to over other, less threatening environmental cues. Threat bias is believed to be an evolutionary adaptation to a world full of stimuli representing varying levels of danger. The more threat a stimulus represents, the more evolutionarily beneficial it is to be aware of, and the more people bias their attention towards it (MacLeod, Mathews, & Tata, 1986).

In the lab, threat bias is often assessed in children via the dot-probe task, a computerized paradigm in which two faces – one angry and one neutral – appear on the screen for 500 ms. After the faces offset, an arrow then appears for 500 ms where one of the two faces had previously been displayed and the child is asked to indicate the direction the arrow is pointing. If a child is selectively attending to threat, then he/she will be consistently faster at responding to arrows that appear behind the angry face (i.e. threat congruent condition) than to arrows behind the neutral face (i.e. threat incongruent condition) because attention will already have oriented to that face. Threat bias is traditionally operationalized by a difference score calculated by subtracting a participant's mean reaction time (RT) on congruent trials from their mean reaction time on incongruent trials. The larger (and more positive) the difference score, the more the participant is biased toward threat.

An impressive body of research has documented that clinically anxious and/or behaviorally inhibited children/adolescents (Pérez-Edgar et al., 2010; Roy et al., 2008) and adults (Bradley, Mogg, White, Groom, & Bono, 1999; Mogg, Millar, & Bradley, 2000) selectively attend to angry vs. neutral faces on a dot-probe task (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & Van Ijzendoorn, 2007; Goodwin, Yiend, & Hirsch, 2017). There is also evidence that threat bias *moderates* the longitudinal relationship between BI and anxiety. Specifically, children who are behaviorally inhibited in toddlerhood are more likely to have anxiety and to be socially withdrawn later in childhood if they also demonstrated threat bias on a dot-probe task (Pérez-Edgar et al., 2011; L. K. White et al., 2017).

In addition, time-sensitive, neurological markers of attention are commonly used to understand the neurologic mechanisms that support attentional orientation, including orientation to threat. In particular, electroencephalography (EEG) is often applied in anxiety research to study attentional monitoring (Dennis & Chen, 2009) and attentional bias to threat (Bar-Haim, Lamy, & Glickman, 2005). There are several event-related potentials (ERPs) of the EEG signal that have been of particular interest within the attentional bias literature. The P1, which primarily reflects activity within the occipital lobe, and N1, which can be seen in both the occipital and parietal lobes, have been shown to be sensitive to early visual and spatial processing (M. J. Taylor, 2002). The amplitudes of both signals are larger following valid as opposed to invalid visual cues (Hillyard, Luck, & Mangun, 1994), and P1 amplitude is larger for visually salient vs.

non-salient stimuli (M. Taylor & Khan, 2000). Both P1 and N1 are also sensitive to face processing, in which amplitudes are larger when viewing happy or angry faces than neutral faces (Eimer & Holmes, 2007). P1 amplitude (Mueller et al., 2009; Rossignol, Campanella, Bissot, & Philippot, 2013) but not N1 amplitude (Eldar, Yankelevitch, Lamy, & Bar-Haim, 2010) has also been found to be higher among anxious adults.

The P2 ERP reflects activity in the frontal and parietal lobes and is central to the processing of emotionally valanced stimuli (Kanske, Plitschka, & Kotz, 2011). Mean P2 amplitude is higher in response to emotion words versus neutral words on a lexical decision task (Kanske & Kotz, 2007), in response to negative versus positive pictures (Carretié, Mercado, Tapia, & Hinojosa, 2001), as well as to angry faces in anxious versus control individuals in a dot-probe task (Eldar et al., 2010) and an attention-shifting paradigm (Bar-Haim et al., 2005). Last of the early appearing waveforms, N2, reflects activity in the frontal and parietal lobes. It has been observed following errors (Van Veen & Carter, 2002), and the amplitude of the N2 is associated with higher anxiety in adults (Lamm et al., 2014) and children (Henderson, 2010).

Despite a strong history, recent studies have been unable to replicate the link between BI and dot-probe performance (Morales, Taber‑Thomas, & Pérez‑Edgar, 2017; Pérez-Edgar et al., 2011; Thai, Taber-Thomas, & Pérez-Edgar, 2016; L. K. White et al., 2017) and there is accruing evidence that the reaction-time dependent performance of the dot-probe task of threat bias has poor reliability (Rodebaugh et al., 2016; Waechter & Stolz, 2015). Split-half reliability and test-retest reliability with an inter-test interval of one week were tested on the threat bias index of two dot-probe tasks (one using words, the other images) and showed that for all reliability indices, $r < .20$ (Schmukle, 2005). This finding of unacceptable reliability has since been replicated using several standard dot-probe tasks of happy and angry faces in healthy (all $r < .30$, $p > .05$;

Staugaard, 2009) and socially anxious individuals (all $r < .15$, $p > .05$; Waechter & Stolz, 2015).

One possible reason for this weakness is that using reaction times to measure performance excludes the role of task accuracy, providing an incomplete description of performance. In addition, RT is influenced by multiple interactive processes. For example, RT is influenced by individual differences in response caution, the time it takes to prepare or execute a motor response, the time it takes to encode the stimulus, as well as whether one is predisposed to respond in a particular manner (i.e. if someone tended to anticipate the arrow would point left). Because of this, reaction time alone remains a nonspecific measure of cognition, making true conditional effects difficult to observe. A measure of task performance which incorporates both reaction time and accuracy of all trials is necessary to more effectively study attentional bias to threat via the dot-probe task.

The *diffusion model* (DM; Ratcliff, 1978), is a computational model of perceptual decision making that combines both RT and accuracy into a single set of performance indices. DM provides a more comprehensive approach to understanding performance on a task such as dot-probe. The DM assumes that during a forced-choice decision, RTs are determined by several factors. First, the amount of time it takes for a stimulus to be encoded, and for a motor response to be prepared/executed is represented by the DM as Non-decision Time (*Ter*). Second, the speed with which information is accumulated and a decision is made is represented as Drift Rate (v). Lastly, response bias, or the predisposition towards picking one particular answer, is represented as Relative Start Point (*za*). Mathematically, it represents a person's start point (how likely they are to make one decision vs another) over their boundary separation (i.e. how much information they require to make each decision, as a representation of response caution).

Diffusion models have been used to study threat bias in anxious adults. In one study, participants were shown a string of letters and asked to respond as quickly as possible whether the string constituted a real word or not. Some of the real words represented threat (e.g., “cancer,” “embarrassment”) and others were neutral (e.g., “planet”, “avocado”). Reaction time and accuracy analyses did not reveal any differences in performance between anxious and non-anxious participants, but drift rate was consistently faster in response to threatening words versus neutral words in the anxious group versus the control group (C. N. White, Ratcliff, Vasey, & McKoon, 2010). In another study, participants were asked to decide if a presented word was threatening or neutral as quickly as possible. Traditional measures of RT indicated that anxious participants responded faster to threatening words than non-anxious participants did. DM parameters broke this result down further and showed that anxious participants had a stronger expectancy bias (*za*) that a word would be threatening, and had faster drift rates in response to threatening compared to non-threatening words (C. N. White, Ratcliff, Vasey, & McKoon, 2010; C. N. White, Skokin, Carlos, & Weaver, 2016). This work indicates promise that diffusion modeling is a more effective method for analyzing attentional bias than is a traditional RT bias score.

Building on this work, the present study aims to determine if diffusion modeling can overcome the known shortcomings of the dot-probe task.

<u>Hypothesis 1</u>. Based on previous findings, we anticipate a main effect of group on the bias score in which behaviorally inhibited children will exhibit a more positive attentional bias score than controls, whose index of attentional bias will not be significantly different from zero. It is expected that this main effect will be stronger when performance is measured by DM parameters (focusing on drift and relative start point) as opposed to standard measures.

Furthermore, if DM parameters are better indices of threat bias than the traditional RT measures, we expect to see a stronger main effect of Cue and a stronger Group x Cue interaction on performance as measured by drift and relative starting point than when performance is measured by RT and accuracy. In the Group x Cue interaction, it is expected that behaviorally inhibited children will demonstrate greater preference for the threat congruent Cue condition than non-behaviorally inhibited children.

Hypothesis2. If the DM parameters are better indices of threat bias, the diffusion parameters will be more strongly correlated with the amplitude of recorded ERPs known to be associated with early attentional processes (N1, P1) and emotional processing (N2, P2) than traditional RT difference scores and accuracy.

Hypothesis3. If DM parameters are more consistent measures of performance on the dot-probe task, they are expected to produce better test-retest reliability than standard indices of performance.

Methods

Participants

A total of 244 children ages 9-12 years old (M = 10.87, SD = .97) were recruited via the FIRSt Families database, a database of families who are interested in participating in Pennsylvania State University research, through community outreach, and through word of mouth in central Pennsylvania as part of a larger study on the development of anxiety in behaviorally inhibited school-aged children (BRAINS, Perez-Edgar).

Potential participants were screened for the study via the administration of the Behavioral Inhibition Questionnaire (BIQ), a parent-report questionnaire designed to identify BI. The BIQ consists of 30 questions about a child’s responses to social or situational novelty on a 1-7 point Likert scale. The sum of all items produces an overall BI score, while the sum of all items within each subscale (social novelty and situational novelty) produces subscale scores. Children scoring ≥ 119 on the BIQ or ≥ 59 on the Social Novelty subscale of the BIQ (Bishop, Spence, & McDonald, 2003) were identified as being behaviorally inhibited (*n* = 83). Children scoring below each of these markers were identified as behaviorally normative (*n* = 161).

As a validation of our definition of BI, we conducted a Maximum Likelihood exploratory factor analysis on the percentage of symptoms that parents endorsed on the Social Anxiety, Generalized Anxiety, and Separation Anxiety disorder modules of computer assisted Diagnostic Interview Schedule for Children version IV (C-DISC 4; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000), which was conducted during participants' first lab visit. Only one factor was extracted with an eigenvalue over 1, and the factor score was output. Total BIQ score was moderately correlated with this Anxiety factor score ($r^2 = .39$, $p < .001$), and children identified as BI scored 0.66 points higher on the factor, $t(199) = 5.04$, $p < .001$, 95% CI: [.39, .91].

Procedures

Children and their parents attended two separate visits spaced approximately one week apart. During the first visit, trained research staff administered the C-DISC 4 to parents. During both visits, children completed a dot-probe task which was written in the E-Prime software package version 2.0 (Psychology Software Tools, Pittsburgh, PA). One of the visits was conducted in the Pennsylvania State University Child Studies Center (CSC), and the other was conducted at the Pennsylvania State University Human Electrophysiology Facility (HEF). During visits conducted at the HEF, children completed the dot-probe task while an electroencephalogram (EEG) was concurrently administered and event-related potentials (ERPs) were recorded. For the majority of children, EEG data was collected during the second visit. Tables 1 and 2 provide a breakdown of participants and the order in which the visits occurred.

Dot-Probe Task. Children saw a fixation cross for 500 ms. This was followed by a pair of faces – one on top and one on bottom – for 500 ms. An arrow that was facing either left or right then appeared in place of one of the faces and remained present for 500 ms (Figure 1). In the Neutral-Neutral (NN) condition, both faces had neutral expressions. In the other two conditions, one face was angry while the other was neutral. In the Neutral-Threat Incongruent (NTi) condition, the target arrow appeared in the neutral face's location. In the Neutral-Threat Congruent (NTc) condition, the target arrow appeared in the angry face's location. Participants were given 2000 ms to indicate the direction of the arrow by clicking either the left or right mouse button, and were asked to do so as quickly as possible. After the 2000 ms response window elapsed, the next trial began,

Equipment

EEG data was collected continuously using a 128-channel geodesic sensor net (Electrical Geodesics Inc., Eugene, Oregon). Vertical eye movements were monitored by electrodes 1cm above and below each eye, while horizontal eye movements were monitored by electrodes 1cm to the outside of each eye. All impedances were kept below 50 kΩ. Electrodes were referenced to Cz during collection and re-referenced to the average of the left and right mastoid during pre-processing. ERPs were recorded at a 1000 Hz sampling rate starting at 100ms before stimulus onset through 500 ms after stimulus onset, in order to allow for a 100 ms baseline correction. Brain Vision Analyzer (Brain Products GmbH, Germany) was used to pre-process and process the data. A high-pass frequency of .1 Hz and a low pass frequency of 40 Hz were used to filter the data. Eye movement artifacts were removed using the Gratton method (Gratton, Coles, & Donchin, 1983). ERPs in response to face prompts were calculated by mean amplitude of either occipital electrodes (65, 66, 69, 70, 71, 74, 76, 82, 83, 84, 89, 90) or fronto-central electrodes (3, 4, 5, 9, 10, 11, 12, 16, 18, 19, 20, 22, 23, 24, 27, 28, 33, 117, 118, 122, 123, 124). Occipital ERPs included the P1 (40-140ms) and N170 (120-220ms). Fronto-central ERPs included the N1 (60-140 ms), P2 (140-240ms), and N2 (260-360ms) (see figure 2).

Data Analysis Plan

Diffusion model parameters. Trials with responses faster than 300ms were excluded from analysis (Ratcliff & Tuerlinckx, 2002). This resulted in the exclusion of 4.6% of trials, or 8.33 per participant. Diffusion model parameters were calculated for each participant using the FastDM software. Based on the size of the dataset and number of trials per participant, data were computed using the Maximum Likelihood (ML) model (Voss, Nagler, & Lerche, 2013). Drift (*v*), relative starting point (*za*), and non-decision time (*Ter*) were estimated separately for each of the three task conditions (i.e. Neutral, Threat Incongruent, Threat Congruent), and for overall task performance collapsed across cue conditions.

Effects of IV on attentional bias to threat difference scores. To form the attentional bias score for DM parameters, difference scores for each parameter were calculated by subtracting their value on congruent trials from their value on incongruent trials to form the following variables: *v*Diff, *za*Diff, *ter*Diff.

To form the bias score as is standard in the literature, data were processed following Perez-Edgar et al. (2011). Error trials were first removed, as were responses faster than 150 ms and those 2 SDs above or below the participant's mean RT. This resulted in the removal of 14.9% of trials, or 26.98 per participant. Mean reaction time for correct responses on congruent trials were subtracted from mean reaction time to correct responses on incongruent trials to form the following variable: RT Threat bias. Positive values therefore indicate an attentional bias to threat.

A mixed within (Visit, 2 levels: first, second) and between (BI: BI, non) subjects ANOVA was conducted with RT Threat bias, *v*Diff, *za*Diff, and *ter*Diff as DVs.

Repeated Measures ANOVA on performance. Because difference scores may not be the most valid indicator of change when studying group differences (Oakes & Feldman, 2001), an ANOVA with one between subjects factor: BI (2 levels: BI, non-BI) and two within subjects factors: Time (2 levels: cell A vs D, see table 2) and Cue Type (3 levels: Neutral, Threat Incongruent, Threat Congruent) was conducted with RT, SDRT, Accuracy, and the DM parameters as DVs. For these analyses, only RTs < 300ms were excluded to encourage equivalency with the manner in which DM parameters were calculated.

Correlation with electrophysiological indices of attention. Simple regression analyses were conducted to separately assess the association between each DV and N1, P1, N170, N2, and P2 to test whether drift and relative start point are more strongly correlated with neurological markers of attention than standard indices of performance.

Test-retest reliability. To evaluate test-retest reliability, correlations of performance between the first and second task administrations as indexed by DM parameters and by traditional RT parameters were calculated. Test-retest reliability for overall RT, ACC, *v*, *ter*, and *za* was determined, as was the threat bias score, *v*Diff, *ter*Diff, and *za*Diff.

Results

Preliminary analyses. Using all trials from the first testing administration, a mixed within (Cue Type, 3 levels: Neutral; Threat Incongruent; Threat Congruent) and between (Location, 2 levels: cell A vs B, see table 2) subjects ANOVA found a main effect of location on RT (but not accuracy) in which children tested at HEF were faster than those tested at the CSC, $F(1, 239) = 6.01$, $p = .015$, $\eta^2 = .025$. There was no main effect of cue type, or a cue type x location interaction on either RT or accuracy (all $\eta^2 < .13$, all $p > .23$).

Similarly, a GLM comparing Cue Type (3) x Location (2: cells C vs D) also found an effect of location. Children tested at HEF were faster, $F(1, 140) = 5.79$, $p = .017$, $\eta^2 = .04$, and less accurate, $F(1,138) = 7.31$, $p = .008$, $\eta^2 = .05$, than those tested at CSC. Again, there was no main effect of cue type on RT or accuracy, and no Cue x Location interaction (all $\eta^2 < .01$, all $p >$ .31)

Due to these differences in performance, participants were not collapsed across their administration location, and instead participants who completed the task at the CSC first and then at HEF (n = 120) were chosen as the focus for the remainder of the analyses. Subsequent analyses refer to this group.

<u>Attentional bias to threat.</u> Table 3 provides a summary of analyses. A mixed within (Time, 2 levels: first, second) and between (BI: BI, non-BI) subjects ANOVA found no significant main effect of BI or BI x Time interaction on attentional bias to threat, as calculated by the traditional differences score or for any of the three DM parameters of interest.

<u>BI x Cue x Time interactions.</u> Due to concern that difference scores may not be the most valid indicator of change when studying group differences (Oakes & Feldman, 2001; Rodebaugh et al., 2016), a 3-way Cue x Time x BI mixed within and between subjects ANOVA was conducted on RT, SDRT, ACC, *v*, *za*, and *ter*. Results are shown in Table 4.

As before, there was no main effect of BI on any dependent variable (all $p > .21$, all $\eta^2 <$ .007). Similarly, there were no two or three-way interactions of BI with Cue Type or Time on any dependent variable (all $p > .177$, all $\eta^2 < .007$). There was a main effect of time on RT, SDRT, drift rate, and *ter* in which RT, drift rate, and *ter* were all faster and SDRT was smaller on the second vs. first testing visit (all $p < .023$, all $\eta^2 > .021$). There was no main effect of Cue

Type or a Cue Type x Time interaction on any performance variable (all $\eta^2 < .01$, all $p > .10$).

ERP. Table 5 provides a summary of results. RT, RT bias score, and overall diffusion model parameters did not regress significantly onto any ERP. P1 was negatively associated with the *za* difference score and *ter* difference score, and was positively associated with the drift difference score. N1 was positively associated with the *ter* difference score.

Test-retest reliability. Test-retest reliabilities are displayed in Figure 4. Test-retest reliability was higher when examining overall score values as opposed to difference scores. Specifically, the metrics of performance with the highest correlation coefficients were Mean Reaction Time (r = .60, $p < .001$) and Standard Deviation of Reaction Time (r =.58, $p < .001$), followed by Mean overall drift (r = .39, $p < .001$).

Accuracy (r = .275, $p = .001$) and non-decision time (r = .19, $p = .02$) were only weakly correlated over task administrations. The commonly reported threat bias score, (r = -0.08, $p = .37$) and the equivalent drift difference score (r = -0.08, $p = .35$) were not significantly correlated across task administrations. DM parameters were also computed using Kolmogorov-Smirnov (ks), and Chi-Square (cs) methods to assess for reliability of different diffusion calculation methods. The method by which the diffusion model parameters were estimated did not influence this pattern of results (Figure 5).

Test-retest reliability of performance as indexed by accuracy was moderated by age. Across the sample, accuracy increased between the first and second visit, but for each year older a participant was, accuracy increase was 0.12 percent more between visits, t = 2.44, p = .016, 95% CI [.024, .226]. No other moderation effect of age or IQ was found for any variable.

Discussion

Attentional bias to threat, as indexed by performance on the dot-probe task, is commonly used in research to understand the cognitive mechanisms that contribute to behavioral inhibition and anxiety disorders (Bar-Haim et al., 2007; Goodwin et al., 2017; Pérez-Edgar et al., 2010). However, recent research has failed to replicate the link between BI and threat bias (Morales et al., 2017; Pérez-Edgar et al., 2011; Thai et al., 2016; L. K. White et al., 2017), and the dot-probe task has recently drawn criticism for exhibiting poor reliability (Rodebaugh et al., 2016; Waechter & Stolz, 2015). The present study attempted to contribute to the current literature by asking whether the reliability of the dot-probe task, and consistency in findings, could be improved by utilizing computational modeling to break down performance into component parts.

We found no main effect of Behavioral Inhibition or a Time x BI interaction on either the traditional bias difference score or diffusion parameters. Due to concern that difference scores may not be the most valid indicator of change when evaluating group differences, a follow-up analysis was conducted in which Cue Type was entered as a second within subjects factor and the dependent variable was RT, as opposed to a difference score. Results did not change. Traditional measures of task performance and the overall mean values of diffusion parameters did not regress significantly onto the amplitudes of concurrently recorded ERPs (P1, N1, N170, P2, N2) that are known to be associated with early attentional processes. However, *v*Diff was positively associated with P1 and N170, *ter*Diff was positively associated with N1 but negatively associated with P1, and *za*Diff was negatively associated with P1 and P2.

We used a well-validated and reliable measure of behavioral inhibition (Broeren & Muris, 2010; Thai et al., 2016), and children identified as behaviorally inhibited were also more anxious, so our lack of findings were unlikely to be due to sample identification problems. Similarly, to ensure valid ERP data was collected, EEG recording was filtered at below .1Hz and above 40 Hz and a well-established method of eye-blink correction was used (Gratton et al., 1983). Only participants with at least 10 acceptable trails per condition were analyzed.

One possible explanation for the observed lack of association between performance and BI or ERPs is poor task reliability. In the current study, we found evidence of practice effects in which RTs were faster and less variable in the second administration of the task. Diffusion modeling revealed that this change reflected a faster drift rate and non-decision time. Thus, during the second visit, participants came to decisions more quickly once they oriented to the stimulus, which likely reflects increased task familiarity. Faster *Ter* score similarly indicated a decreased need for motor preparation time, which suggests that participants were more accustomed to the pacing of the task and had greater comfort with the procedures.

Practice effects do not necessarily result in low test-retest reliability, however. Reliability, as measured by Pearson correlations, provides information on the strength of the linear association between participants' performance at time 1 and time 2. In the current study, test-retest reliability as indexed by a Pearson's r was unacceptably low, and ranged from -.079-.604. Consistent with previous literature (Waechter & Stoltz, 2015; Enoch et al., 2014), the highest reliability was observed when all trials were averaged together (e.g. Mean RT r = .604). When computed as a difference score between congruent and incongruent trials, however, reliability was quite low (ranging from a low of -.079 for *vDiff* to a high of .040 for *zaDiff*), indicating a nonlinear relationship between performance at time 1 and time 2.

The observed practice effects and low reliability across all metrics suggests that children did tend to improve their performance between the first and second administration, but that the strength of this increase in performance was not consistent across participants.

Because diffusion modeling provides a more accurate and nuanced description of performance, it was initially hypothesized that these parameters might be both more reliable and possibly sensitive to differences in temperament. Indeed, these techniques have shown utility in breaking down components of RT in letter identification and threat detection tasks (Ratcliff & Rouder, 2000; C. N. White et al., 2010; C. N. White et al., 2016) and DM parameters in lexical decision and recognition memory tasks have shown decent test-retest reliabilities (Lerche & Voss, 2017) despite that test-retest reliability has been hypothesized to be susceptible to low reliability due to the measurement error inherent to model fitting (Lerche & Voss, 2017). Here, although reliability of drift rate was higher than that of the standard attentional bias difference score, test-retest reliability of DM parameters was still modest.

Not only do these findings have implications on the interpretability of the relationship between behavioral inhibition and attentional bias, but also for studies interested in studying trait change of behavioral inhibition. For example, recent studies have examined the influence of Cognitive Behavioral Therapy (CBT) on threat bias as indexed by the RT difference score (Calamaras, Tone, & Anderson, 2012), but low reliability suggests that results of the study cannot be attributed to the CBT treatment effects alone, as dot-probe performance is prone to regression to the mean and random error (Rodebaugh et al., 2016).

That being said, the context and environment in which data are collected appear to play important roles in behavioral performance. Children tested at the HEF were faster than those tested in the standard lab setting (e.g. comparison of cells B vs A and D vs C, see Table 2). Due to this finding, data were not collapsed across settings. It is possible that the low reliability observed here is due exclusively to changes in context. It is possible that reliability would have been higher if testing had been administered in the same setting each time.

To address the limitations of the dot-probe task, it may be appropriate to use other paradigms of attentional bias as alternatives to the dot-probe. For example, emotionally valanced versions of the Posner attentional cuing task (Posner, 1980) have been successfully implemented to study attentional bias in various anxious samples (Cisler & Olatunji, 2010; Derryberry & Reed, 2002; Pérez-Edgar & Fox, 2005). Other studies have indicated that selective attention tasks using emotionally valanced target words may be useful for classifying attentional bias to threat (Dodd, Vogt, Turkileri, & Notebaert, 2017; Rinck, Becker, Kellermann, & Roth, 2003). Notably, several groups have begun utilizing eye-tracking technology to directly measure children's pupil dilation and gaze time toward threatening versus neutral stimuli during computerized tasks (In-Albon, Kossowsky, & Schneider, 2010; Price et al., 2016; Seefeldt, Krämer, Tuschen-Caffier, & Heinrichs, 2014). There is also evidence to suggest that naturalistic, mobile, eye-tracking may provide a useful, novel, approach to understanding attentional bias to threat in anxious or behaviorally inhibited children (Fu & Pérez-Edgar, 2019).

Conclusion

Attentional bias to threat is a risk factor for the development of a future anxiety disorders, and over the last 20 years, the dot-probe task has served as the gold standard of such a bias. Recent research has called that body of work into question due to demonstrated poor reliability. The present study aimed to determine if the diffusion model could be used improve reliability of measurement of attentional bias, and relationship with BI as well as electrophysiological indicators of performance. This was not the case. Overall, these results confirm recommendations to move away from using the dot-probe task as the sole reliable index of attentional bias, even with indices of performance that relay more accurate summaries of performance.

Appendix
Figures and Tables

Figure 1. The dot-probe paradigm (Thai et al., 2016)

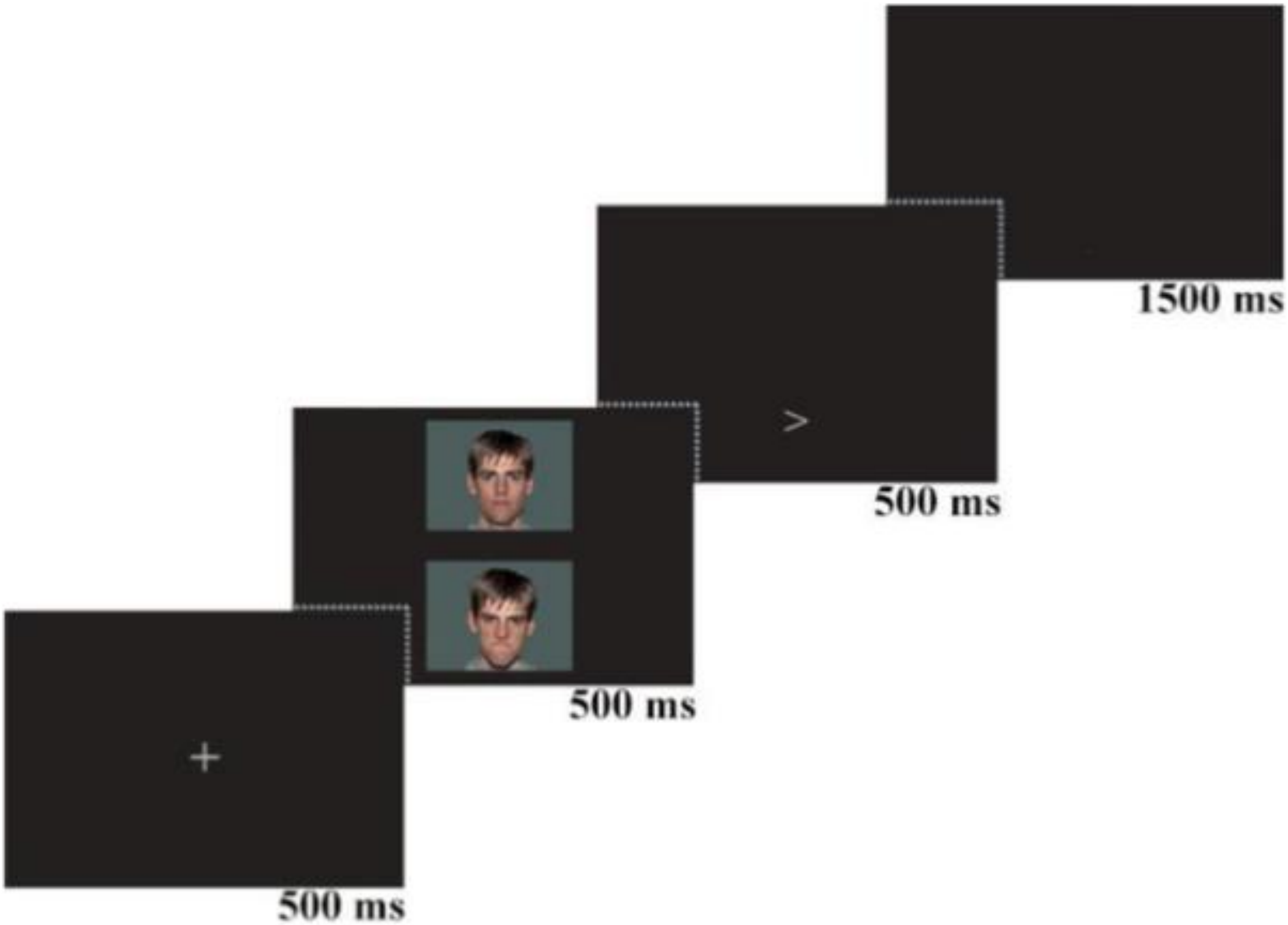


Figure 2. ERPs of interest (Thai et al., 2016)

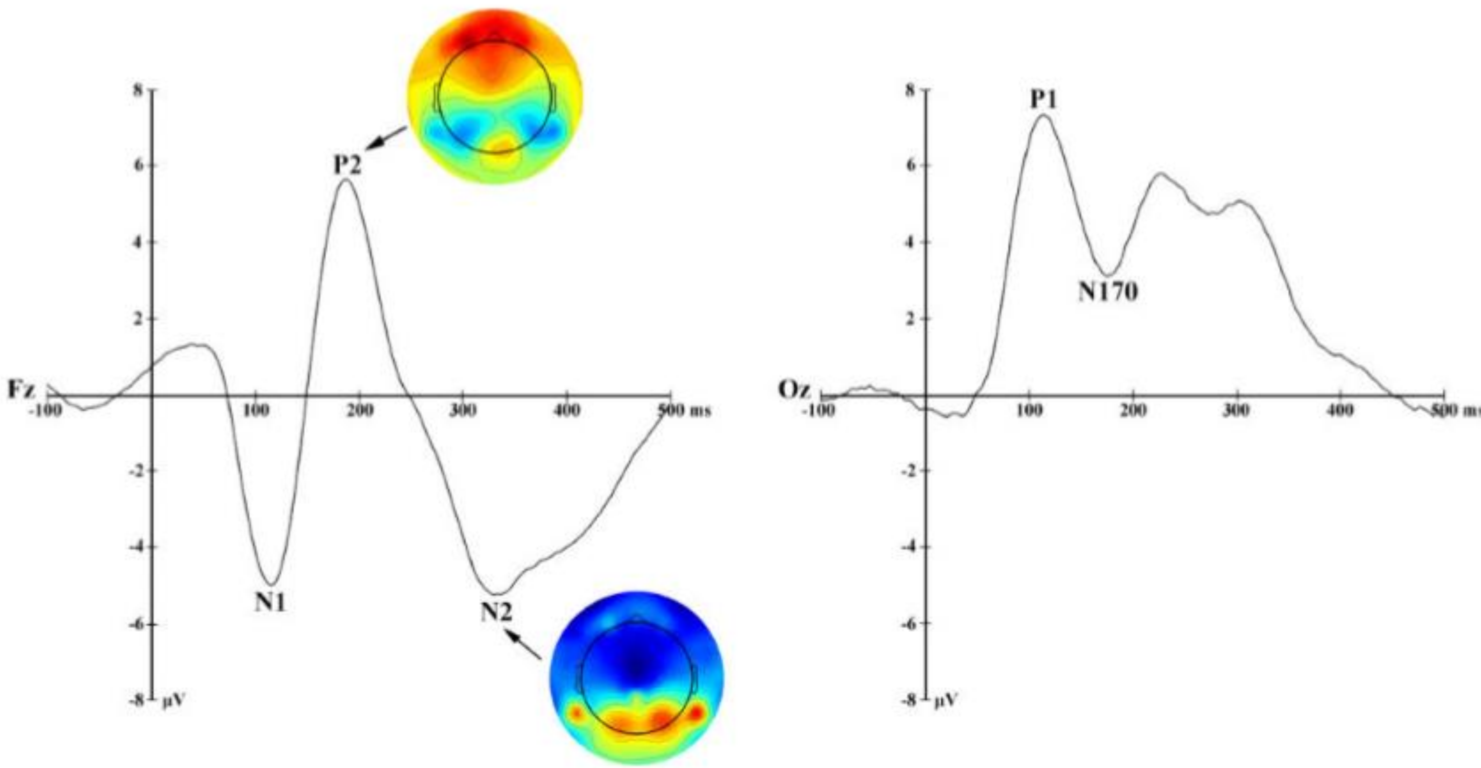


Figure 3. The diffusion model

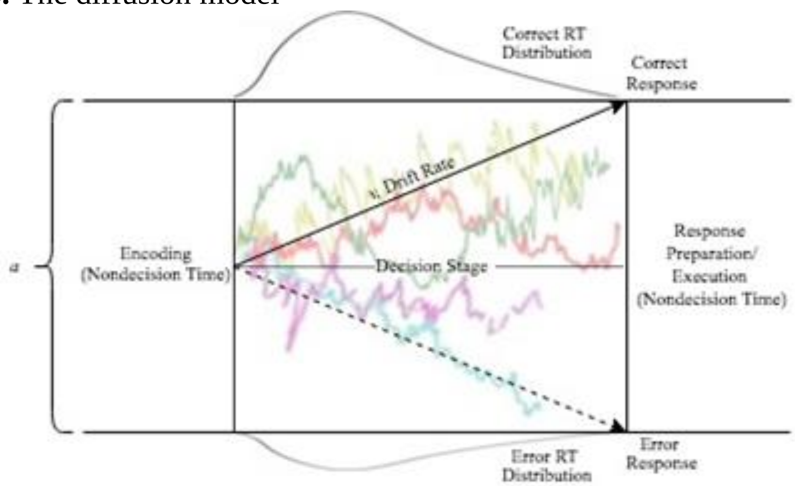
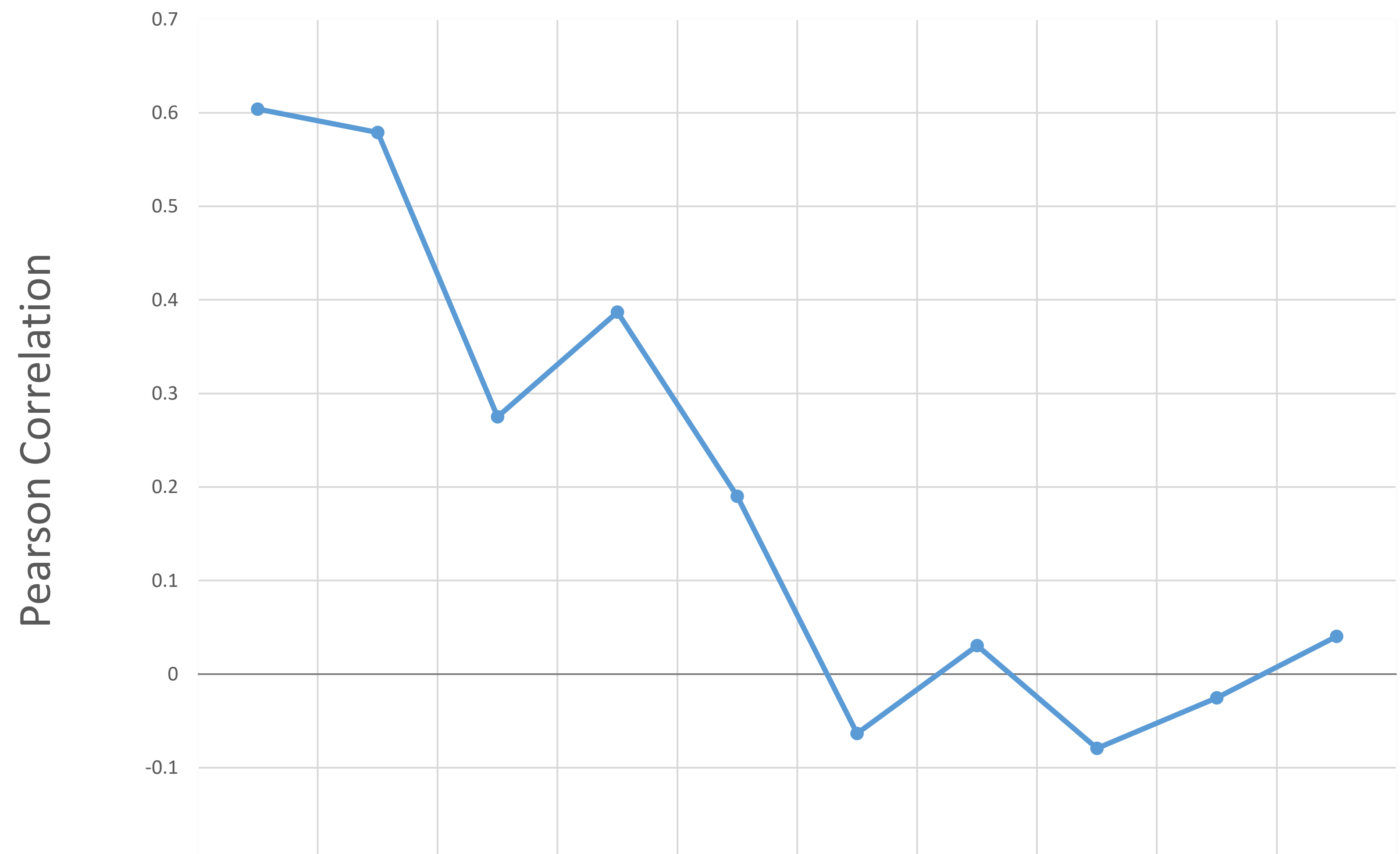
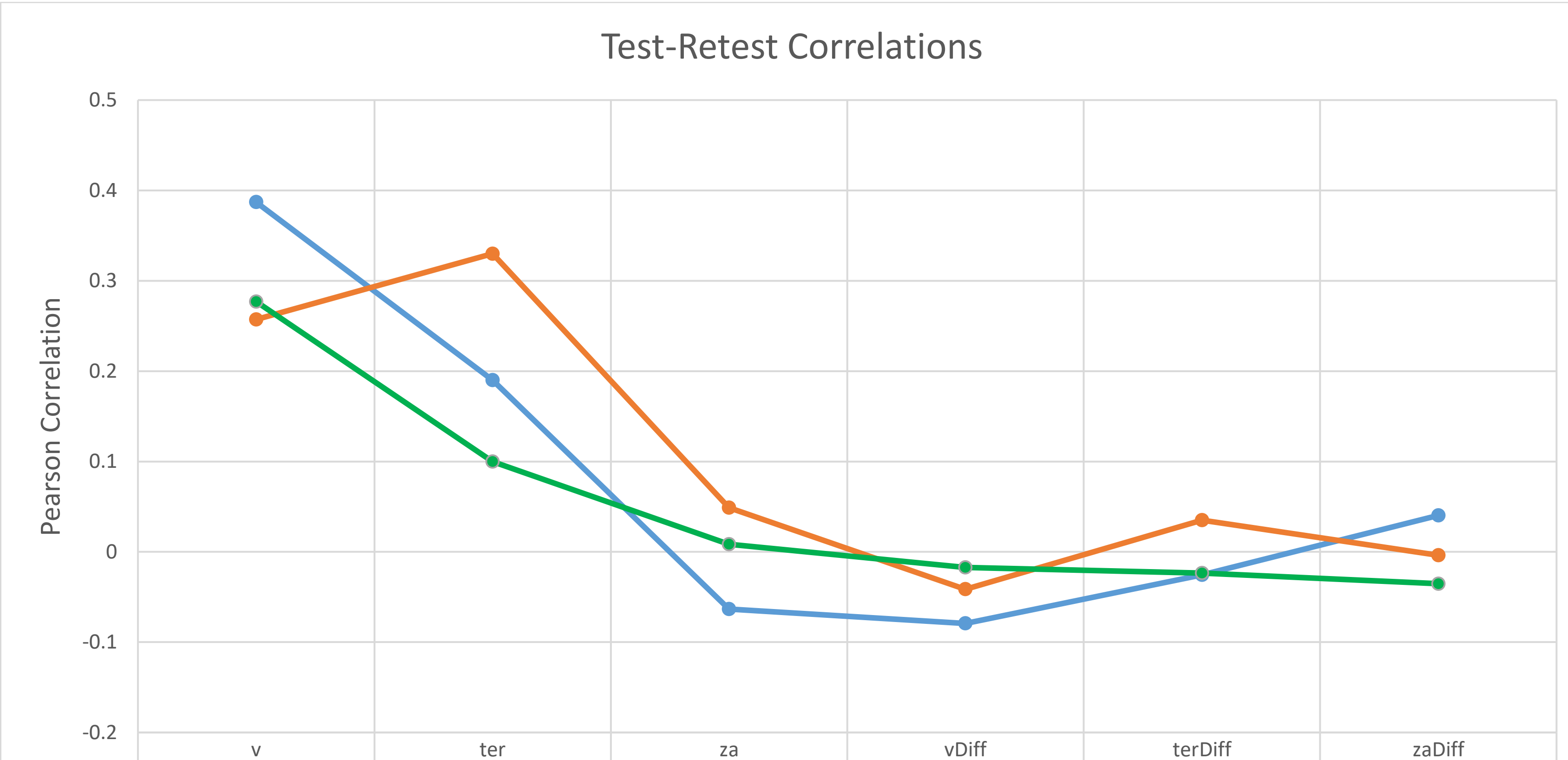


Figure 4. Test-retest reliability of various performance measures.



	Mean RT	Std RT	Mean ACC	v	ter	za	Bias Score	vDiff	terDiff	zaDiff
Pearson Correlation	0.604	0.579	0.275	0.387	0.19	-0.063	0.030	-0.079	-0.025	0.040
Sig. (2-tailed)	0.000	0.000	0.001	0.000	0.024	0.455	0.719	0.350	0.765	0.635

Figure 5. Test-retest reliability of diffusion parameters as calculated by the Kolmogorov-Smirnov (ks), and Chi-Square (cs) methods in addition to the already reported Maximum Likelihood (ML) method.



	v	ter	za	vDiff	terDiff	zaDiff
ML	0.387	0.19	-0.063	-0.079	-0.025	0.040
ks	0.257	0.33	0.049	-0.042	0.035	-0.004
cs	0.277	0.100	0.008	-0.017	-0.024	-0.035

Table 1. Participant Demographics

	Non-BI	BI	Total
N	152 (69.1%)	68 (30.9%)	220
Age (SD)	10.91 (.96)	10.80 (.99)	10.87 (.97)
IQ (SD)	109.87 (13.47)	111.89 (12.80)	110.66 (13.21)
BIQ raw score (SD)	74.59 (21.24)	128.28 (18.47)	92.86 (32.58)
Anxiety Factor Score	-0.26 (.77)	0.40 (1.08)	0.00 (.97)
Female (% of total)	83 (54.6%)	37 (54.4%)	120 (54.5%)
Asian/Pacific Islander	1 (.6%)	3 (3.6%)	4 (1.6%)
Hispanic	0	3 (3.6%)	3 (1.2%)
White	123 (76.4%)	63 (75.9%)	186 (76.2%)
African American	0	4 (4.8%)	4 (1.6%)
Native American	1 (0.7%)	0	1 (0.5%)
Mixed Race	3 (1.9%)	4 (4.8%)	7 (2.9%)
Unknown Race	33 (20.5%)	6 (7.2%)	39 (16.0%)

Table 2. Visit order. Two hundred twenty participants completed their first visit at the CSC. One hundred twenty-one of those returned for a second visit at the HEF. Twenty-four participants completed their first visit at the HEF. Twenty-one of those returned for a second visit at the CSC.

N=244	CSC	HEF
First Administration	A. 220	B. 24
Second Administration	C. 21	D. 121

Table 3. Effects of BI and Time on threat bias

	Between and Mixed Subjects Effects			
	Parameter	F	p	η2
BI	RT Threat Bias	0.848	0.359	0.007
	vDiff	0.019	0.89	< .001
	zaDiff	0.095	0.759	0.001
	terDiff	2.247	0.137	0.019
Time x BI	RT Threat Bias	0.468	0.495	0.004
	vDiff	0.19	0.663	0.002
	zaDiff	0.155	0.695	0.001
	terDiff	0.153	0.696	0.001

	Within Subjects Effects			
	Parameter	F	p	η2
Time	RT Threat Bias	0.02	0.888	< .001
	vDiff	0.735	0.393	0.006
	zaDiff	2.658	0.106	0.022
	terDiff	0.108	0.743	0.001

Table 4. 3 (CueType) x 2 (Time) x 2 (BI) ANOVA

	Within Subjects Effects			
	Parameter	F	*p*	η2
Cue	RT	1.232	0.293	0.005
	stdRT	0.589	0.555	0.002
	ACC	0.624	0.536	0.003
	ter	0.89	0.411	0.004
	v	1.1	0.331	0.005
	za	0.483	0.617	0.002
Time	RT	62.106	< .001 **	0.208
	stdRT	10.122	0.002 **	0.041
	ACC	2.352	0.126	0.01
	ter	5.256	0.023 *	0.022
	v	10.741	0.001 **	0.043
	za	3.249	0.073	0.014
Time x Cue	RT	0.147	0.863	0.001
	stdRT	1.241	0.29	0.005
	ACC	0.797	0.451	0.003
	ter	0.659	0.518	0.003
	v	0.458	0.633	0.002
	za	1.927	0.147	0.008

	Between Subjects and Mixed Effects			
	Parameter	F	*p*	η2
BI	RT	1.604	0.207	0.007
	stdRT	0.507	0.477	0.002
	ACC	0.002	0.963	< .001
	ter	0.104	0.747	< .001
	v	0.011	0.918	< .001
	za	0.447	0.504	0.002
Tme x BI	RT	0.466	0.496	0.002
	stdRT	0.488	0.485	0.002
	ACC	0.183	0.669	0.001
	ter	0.006	0.937	< .001
	v	0.003	0.954	< .001
	za	0.237	0.627	0.001
Cue x BI	RT	0.644	0.526	0.003
	stdRT	0.1	0.904	< .001
	ACC	0.437	0.646	0.002
	ter	1.736	0.177	0.007
	v	0.069	0.933	< .001
	za	0.253	0.776	0.001
Time x Cue x BI	RT	0.117	0.89	< .001
	stdRT	0.342	0.71	0.001
	ACC	0.102	0.9	< .001
	ter	0.217	0.81	0.001
	v	0.413	0.66	0.002
	za	0.107	0.898	< .001

Table 5. Pearson Correlation of ERP values with traditional and computational measures of performance

Performance Parameter	N1	P1	N170	P2	N2
RT	0.032	0.113	-0.037	0.041	-0.114
Accuracy	-0.051	0.043	0.034	-0.027	-0.116
RT Difference score (bias)	0.124	-0.053	-0.069	0.022	0.01
v	0.054	-0.077	-0.002	-0.043	0.127
*v*Diff	-0.17	.228*	.189*	-0.072	-0.031
za	-0.074	0.038	-0.06	0.117	0.132
*za*Diff	0.032	-.238*	-0.164	-0.188*	-0.132
ter	-0.034	0.121	0.133	-0.009	-0.144
*ter*Diff	.192*	-.228*	-0.144	0.091	0.04

Note: * $p < .05$, ** $p < .001$

References

Bar-Haim, Y., Lamy, D., & Glickman, S. (2005). Attentional bias in anxiety: A behavioral and ERP study. *Brain and cognition, 59*(1), 11-22.

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & Van Ijzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study. *Psychological bulletin, 133*(1), 1.

Biederman, J., Hirshfeld-Becker, D. R., Rosenbaum, J. F., Hérot, C., Friedman, D., Snidman, N., . . . Faraone, S. V. (2001). Further evidence of association between behavioral inhibition and social anxiety in children. *American journal of Psychiatry, 158*(10), 1673-1679.

Bishop, G., Spence, S. H., & McDonald, C. (2003). Can parents and teachers provide a reliable and valid report of behavioral inhibition? *Child Development, 74*(6), 1899-1917.

Bradley, B. P., Mogg, K., White, J., Groom, C., & Bono, J. (1999). Attentional bias for emotional faces in generalized anxiety disorder. *British Journal of Clinical Psychology, 38*(3), 267-278.

Broeren, S., & Muris, P. (2010). A psychometric evaluation of the behavioral inhibition questionnaire in a non-clinical sample of Dutch children and adolescents. *Child Psychiatry & Human Development, 41*(2), 214-229.

Calamaras, M. R., Tone, E. B., & Anderson, P. L. (2012). A pilot study of attention bias subtypes: examining their relation to cognitive bias and their change following cognitive behavioral therapy. *Journal of Clinical Psychology, 68*(7), 745-754.

Carretié, L., Mercado, F., Tapia, M., & Hinojosa, J. A. (2001). Emotion, attention, and the 'negativity bias', studied through event-related potentials. *International journal of psychophysiology, 41*(1), 75-85.

Chronis-Tuscano, A., Degnan, K. A., Pine, D. S., Perez-Edgar, K., Henderson, H. A., Diaz, Y., . . . Fox, N. A. (2009). Stable early maternal report of behavioral inhibition predicts lifetime social anxiety disorder in adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry, 48*(9), 928-935.

Cisler, J. M., & Olatunji, B. O. (2010). Components of attentional biases in contamination fear: Evidence for difficulty in disengagement. *Behaviour research and therapy, 48*(1), 74-78.

Clauss, J. A., & Blackford, J. U. (2012). Behavioral inhibition and risk for developing social anxiety disorder: a meta-analytic study. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(10), 1066-1075. e1061.

Coll, C. G., Kagan, J., & Reznick, J. S. (1984). Behavioral inhibition in young children. *Child Development*, 1005-1019.

Dennis, T. A., & Chen, C. C. (2009). Trait anxiety and conflict monitoring following threat: an ERP study. *Psychophysiology, 46*(1), 122-131.

Derryberry, D., & Reed, M. A. (2002). Anxiety-related attentional biases and their regulation by attentional control. *Journal of abnormal psychology, 111*(2), 225.

Dodd, H. F., Vogt, J., Turkileri, N., & Notebaert, L. (2017). Task relevance of emotional information affects anxiety-linked attention bias in visual search. *Biological psychology, 122*, 13-20.

Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia, 45*(1), 15-31.

Eldar, S., Yankelevitch, R., Lamy, D., & Bar-Haim, Y. (2010). Enhanced neural reactivity and selective attention to threat in anxiety. *Biological psychology, 85*(2), 252-257.

Fox, A. S., & Kalin, N. H. (2014). A translational neuroscience approach to understanding the development of social anxiety disorder and its pathophysiology. *American journal of Psychiatry, 171*(11), 1162-1173.

Fu, X., & Pérez-Edgar, K. (2019). Threat-related attention bias in socioemotional development: A critical review and methodological considerations. *Developmental Review, 51*, 31-57.

Goodwin, H., Yiend, J., & Hirsch, C. R. (2017). Generalized Anxiety Disorder, worry and attention to threat: A systematic review. *Clinical psychology review, 54*, 107-122.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and clinical neurophysiology, 55*(4), 468-484.

Henderson, H. A. (2010). Electrophysiological correlates of cognitive control and the regulation of shyness in children. *Developmental Neuropsychology, 35*(2), 177-193.

Hillyard, S. A., Luck, S. J., & Mangun, G. R. (1994). The cuing of attention to visual field locations: Analysis with ERP recordings. In *Cognitive electrophysiology* (pp. 1-25): Springer.

Hirshfeld-Becker, D. R., Biederman, J., Henin, A., Faraone, S. V., Davis, S., Harrington, K., & Rosenbaum, J. F. (2007). Behavioral inhibition in preschool children at risk is a specific predictor of middle childhood social anxiety: a five-year follow-up. *Journal of Developmental & Behavioral Pediatrics, 28*(3), 225-233.

Hirshfeld, D. R., Rosenbaum, J. F., Biederman, J., Bolduc, E. A., Faraone, S. V., Snidman, N., . . . Kagan, J. (1992). Stable behavioral inhibition and its association with anxiety disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 31*(1), 103-111.

In-Albon, T., Kossowsky, J., & Schneider, S. (2010). Vigilance and avoidance of threat in the eye movements of children with separation anxiety disorder. *Journal of abnormal child psychology, 38*(2), 225-235.

Kanske, P., & Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain research, 1148,* 138-148.

Kanske, P., Plitschka, J., & Kotz, S. A. (2011). Attentional orienting towards emotion: P2 and N400 ERP effects. *Neuropsychologia, 49*(11), 3121-3129.

La Greca, A. M., & Lopez, N. (1998). Social anxiety among adolescents: Linkages with peer relations and friendships. *Journal of abnormal child psychology, 26*(2), 83-94.

Lamm, C., Walker, O. L., Degnan, K. A., Henderson, H. A., Pine, D. S., McDermott, J. M., & Fox, N. A. (2014). Cognitive control moderates early childhood temperament in predicting social behavior in 7‐year‐old children: an ERP study. *Developmental science*, *17*(5), 667-681.

Lerche, V., & Voss, A. (2017). Retest reliability of the parameters of the Ratcliff diffusion model. *Psychological research*, *81*(3), 629-652.

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of abnormal psychology*, *95*(1), 15.

Mogg, K., Millar, N., & Bradley, B. P. (2000). Biases in eye movements to threatening facial expressions in generalized anxiety disorder and depressive disorder. *Journal of abnormal psychology, 109*(4), 695.

Morales, S., Taber‐Thomas, B. C., & Pérez‐Edgar, K. E. (2017). Patterns of attention to threat across tasks in behaviorally inhibited children at risk for anxiety. *Developmental science*, *20*(2), e12391.

Mueller, E., Hofmann, S., Santesso, D., Meuret, A., Bitran, S., & Pizzagalli, D. A. (2009). Electrophysiological evidence of attentional biases in social anxiety disorder. *Psychological medicine*, *39*(7), 1141-1152.

Oakes, J. M., & Feldman, H. A. (2001). Statistical power for nonequivalent pretest-posttest designs: The impact of change-score versus ANCOVA models. *Evaluation Review, 25*(1), 3-28.

Pérez-Edgar, K., Bar-Haim, Y., McDermott, J. M., Chronis-Tuscano, A., Pine, D. S., & Fox, N. A. (2010). Attention biases to threat and behavioral inhibition in early childhood shape adolescent social withdrawal. *Emotion, 10*(3), 349.

Pérez-Edgar, K., & Fox, N. A. (2005). A behavioral and electrophysiological study of children's selective attention under neutral and affective conditions. *Journal of Cognition and Development, 6*(1), 89-118.

Pérez-Edgar, K., Reeb-Sutherland, B. C., McDermott, J. M., White, L. K., Henderson, H. A., Degnan, K. A., . . . Fox, N. A. (2011). Attention biases to threat link behavioral inhibition to social withdrawal over time in very young children. *Journal of abnormal child psychology, 39*(6), 885-895.

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology, 32*(1), 3-25.

Price, R. B., Rosen, D., Siegle, G. J., Ladouceur, C. D., Tang, K., Allen, K. B., . . . Silk, J. S. (2016). From anxious youth to depressed adolescents: Prospective prediction of 2-year depression symptoms via attentional bias measures. *Journal of abnormal psychology, 125*(2), 267.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review, 85*(2), 59.

Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human perception and performance, 26*(1), 127.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic bulletin & review, 9*(3), 438-481.

Rinck, M., Becker, E. S., Kellermann, J., & Roth, W. T. (2003). Selective attention in anxiety: Distraction and enhancement in visual search. *Depression and anxiety, 18*(1), 18-28.

Rodebaugh, T. L., Scullin, R. B., Langer, J. K., Dixon, D. J., Huppert, J. D., Bernstein, A., . . . Lenze, E. J. (2016). Unreliability as a threat to understanding psychopathology: The cautionary tale of attentional bias. *Journal of abnormal psychology, 125*(6), 840.

Rossignol, M., Campanella, S., Bissot, C., & Philippot, P. (2013). Fear of negative evaluation and attentional bias for facial expressions: an event-related study. *Brain and cognition, 82*(3), 344-352.

Roy, A. K., Vasa, R. A., Bruck, M., Mogg, K., Bradley, B. P., Sweeney, M., . . . Pine, D. S. (2008). Attention bias toward threat in pediatric anxiety disorders. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*(10), 1189-1196.

Schmukle, S. C. (2005). Unreliability of the dot probe task. *European Journal of Personality, 19*(7), 595-605. doi:10.1002/per.554

Seefeldt, W. L., Krämer, M., Tuschen-Caffier, B., & Heinrichs, N. (2014). Hypervigilance and avoidance in visual attention in children with social phobia. *Journal of behavior therapy and experimental psychiatry, 45*(1), 105-112.

Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(1), 28-38.

Staugaard, S. R. (2009). Reliability of two versions of the dot-probe task using photographic faces. *Psychology Science Quarterly, 51*(3), 339-350.

Taylor, M., & Khan, S. (2000). Top-down modulation of early selective attention processes in children. *International journal of psychophysiology, 37*(2), 135-147.

Taylor, M. J. (2002). Non-spatial attentional effects on P1. *Clinical neurophysiology, 113*(12), 1903-1908.

Thai, N., Taber-Thomas, B. C., & Pérez-Edgar, K. E. (2016). Neural correlates of attention biases, behavioral inhibition, and social anxiety in children: An ERP study. *Developmental cognitive neuroscience, 19*, 200-210.

Van Veen, V., & Carter, C. S. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiology & behavior, 77*(4-5), 477-482.

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology. *Experimental psychology*.

Waechter, S., & Stolz, J. A. (2015). Trait Anxiety, State Anxiety, and Attentional Bias to Threat: Assessing the Psychometric Properties of Response Time Measures. *Cognitive Therapy and Research, 39*(4), 441-458. doi:10.1007/s10608-015-9670-z

White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010). Anxiety enhances threat processing without competition among multiple inputs: A diffusion model analysis. *Emotion, 10*(5), 662.

White, C. N., Skokin, K., Carlos, B., & Weaver, A. (2016). Using decision models to decompose anxiety-related bias in threat classification. *Emotion, 16*(2), 196.

White, L. K., Degnan, K. A., Henderson, H. A., Pérez‐Edgar, K., Walker, O. L., Shechner, T., . . . Fox, N. A. (2017). Developmental relations among behavioral inhibition, anxiety, and attention biases to threat and positive information. *Child Development*, *88*(1), 141-155.

Wichmann, C., Coplan, R. J., & Daniels, T. (2004). The social cognitions of socially withdrawn children. *Social Development, 13*(3), 377-392.