Psychological Research (2022) 86:831–843
https://doi.org/10.1007/s00426-021-01532-3

ORIGINAL ARTICLE

Implementation of the diffusion model on dot-probe task performance in children with behavioral inhibition

Shane Wise[1] · **Cynthia Huang-Pollock**[1] · **Koraly Pérez-Edgar**[1]

Received: 7 January 2021 / Accepted: 12 May 2021 / Published online: 28 May 2021

Abstract

Attentional bias to threat, the process of preferentially attending to potentially threatening environmental stimuli over neutral stimuli, is positively associated with behavioral inhibition (BI) and trait anxiety. However, the most used measure of attentional bias to threat, the dot-probe task, has been criticized for demonstrating poor reliability. The present study aimed to assess whether utilizing a sequential sampling model to describe performance could detect adequate test–retest reliability for the dot-probe task, demonstrate stronger cueing effects, and improve the association with neural signals of early attention. One hundred and twenty children aged 9–12 years completed the dot-probe task twice. During the second administration, event-related potentials (ERPs) were obtained as time-sensitive neural markers of attention. BI was not associated with traditional or diffusion model measures of performance. Traditional and diffusion model measures of performance were also not associated with N1, P2, or N2 ERP amplitude. There were main effects of Visit, in which RTs were faster and standard deviation of RT smaller during the second administration due to an increase in drift rate and a decrease in non-decision time. The traditional RT bias score ($r=0.06$) and bias scores formed via diffusion model parameters (all r's <0.40) all demonstrated poor reliability. Results confirm recommendations to move away from using the dot-probe task as the primary or sole index of attentional bias.

Introduction

Behavioral Inhibition (BI) is an early temperamental trait marked by wariness and avoidance of novel stimuli, and has been associated with an increased risk for anxiety disorders in later life (Blackford & Pine, 2012; Coll et al., 1984). BI is commonly measured via parent report of temperament (Chronis-Tuscano et al., 2009; Hudson et al., 2011) or lab-based behavioral measures of approach (e.g. a play-based session with parent and experimenter (Schwartz et al., 1999) which are both generally well correlated (Bishop et al., 2003).

Attentional bias to threat is one possible mechanism that explains BI stability over the lifespan, as well as BI's association with anxiety disorders. Attentional bias to threat refers to a phenomenon in which preferential attention is given to a potentially threatening stimulus over other, less threatening, environmental cues (Morales et al., 2017). It is believed to be an evolutionary adaptation to a world full of stimuli representing varying levels of danger. However, if a bias is applied across stimuli and contexts without regard to reasonable or objective threat levels, it may limit the child's ability to explore the environment, thereby leading to the subjective conclusion that the world is threatening.

In the lab, threat bias is most commonly 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 a brief period (Ehrenreich & Gross, 2002). After the faces offset, an arrow briefly appears 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, they will then be consistently faster to respond 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). Threat bias is traditionally operationalized by a difference score calculated by subtracting a participant's mean reaction

✉ Shane Wise
smw66@psu.edu

Cynthia Huang-Pollock
clh39@psu.edu

Koraly Pérez-Edgar
kxp24@psu.edu

[1] Department of Psychology, The Pennsylvania State University, State College, USA

time (RT) on congruent trials from their mean RT on incongruent trials. The larger (and more positive) the difference score, the more the participant is biased toward the threat. Using the dot-probe task, an impressive body of research has documented that clinically anxious and/or behaviorally inhibited (BI) children, adolescents, and adults selectively attend to angry versus neutral faces on a dot-probe task (Bar-Haim et al., 2007). Children who are behaviorally inhibited as toddlers are more likely to feel anxious and to be socially withdrawn later in childhood if they also demonstrated threat bias on a dot-probe task (Pérez-Edgar et al., 2011).

To better understand the biological correlates of this attentional bias, several event-related potentials (ERPs) of the EEG signal have been of particular interest. The P1 and N1 reflect activity in the occipital and parietal lobes, and are known to be sensitive to early visual spatial as well as face processing (Eimer & Holmes, 2007; Mueller et al., 2009; Rossignol et al., 2013; Taylor, 2002). The P2 reflects activity in the frontal and parietal lobes, is central to the processing of emotionally valanced stimuli, and its amplitude has been found to be higher for angry faces in anxious individuals (Bar-Haim et al., 2005; Carretié et al., 2001; Eldar et al., 2010; Kanske & Kotz, 2007; Kanske et al., 2011). The N2, last of the early appearing waveforms, reflects activity in the frontal and parietal lobes, and its amplitude is positively associated with anxiety in adults and children (Henderson, 2010; Lamm et al., 2014). In children, the N2 moderates the relation between threat bias and BI (Thai et al., 2016).

However, recent studies have had difficulty replicating the BI/dot-probe performance link (Morales et al., 2017; Pérez-Edgar et al., 2011; Thai et al., 2016), and there is accruing evidence that the split-half and test–retest reliability of reaction-time dependent performance of the dot-probe task is quite poor (Kappenman et al., 2014; Molloy & Anderson, 2020; Rodebaugh et al., 2016; Waechter & Stolz, 2015; Waechter et al., 2014). One possible reason for this weakness is that solely relying on mean RT to index performance excludes task accuracy and ignores the shape of the RT distribution. This, therefore, provides an incomplete description of performance that may result in misleading interpretations. RT distributions and accuracy are also together influenced by multiple interactive processes, including how cautious one tends to be in responding, 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 (Ratcliff & Tuerlinckx, 2002). The broader anxiety and BI literature has, in fact, suggested that performance profiles may be explained by cautious response strategies and idiosyncratic response patterns (Bar-Haim et al., 2007). However, the traditional single RT difference score is unable to provide this level of distinction. A measure of task performance that incorporates both RT and accuracy for all trials is necessary to provide a more complete description of task performance and to potentially explain the cognitive processes that produce the phenomenon.

The diffusion model (DM) is a sequential sampling model of perceptual decision making that combines both RT and accuracy into a single set of performance indices. It thereby provides a more comprehensive and nuanced approach to documenting performance on two-alternative forced-choice tasks like the dot-probe (Ratcliff & Tuerlinckx, 2002). The DM assumes that RTs are determined by several interactive factors during a forced-choice decision. First, the time needed to encode a stimulus and for the motor response to be prepared/executed is referred to as non-decision time (and represented in the parameter, Ter). Second, the speed with which information is accumulated towards a decision (e.g., is the arrow pointing right or left?) is represented as drift rate (v). Third, the amount of information a participant requires before coming to an answer (that is, how sure they need to be) is referred to as boundary separation (a). Lastly, response bias, or the predisposition towards a particular response, is represented as z. It is the start point of the decision process and represents the response expectancy bias, or how much evidence needs to be sampled or accumulated for any given decision. Within the diffusion model, bias is commonly identified in one of two ways. In the first, bias is identified if the relative start point (za, in which z is divided by a: Voss et al., 2013) is not equidistant, but is closer to one of the two boundaries. It can be understood as a response expectancy bias (Leite & Ratcliff, 2011; Mulder et al., 2012; White & Poldrack, 2014). A second way that bias is commonly identified is if the drift rate to one decisional boundary is faster than the drift rate towards the other (Leite & Ratcliff, 2011; Mulder et al., 2012; White & Poldrack, 2014). In this latter case, bias occurs because the quality of evidence, or the rate at which evidence accumulates, for one stimulus is faster than for another stimulus.

Diffusion models are increasingly being used to study threat bias in anxious adults. In one study, participants were shown a string of letters and asked to indicate whether the string constituted a real word (White et al., 2010). Some of the real words represented threat (e.g., "cancer," "embarrassment") and others were neutral (e.g., "planet," "avocado"). RT and accuracy analyses did not discriminate between anxious and non-anxious participants, but v was consistently faster in response to threatening words versus neutral words among anxious participants. In another study, participants were asked to decide if a presented word was threatening or neutral (White et al., 2016). 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 demonstrated that faster response times were due to both a stronger expectancy bias (za) that a word would be threatening and a faster drift rate to threatening compared to non-threatening words.

In these tasks, the explicit decision (threat or non-threat?) allows bias to be measured directly and facilitates data interpretation, including the interpretations associated with DM parameters. In the dot-probe task, however, bias is measured indirectly: the face serves as the cue, but the actual decision is whether the arrow points left or right. The design of the dot-probe, therefore, places the measurement of potential processing bias at a one-step remove, complicating the interpretation of performance data. However, a very similar and commonly studied cognitive phenomenon in which the diffusion model has been applied, contextual cueing, provides some guidance.

In contextual cueing tasks, participants are asked to indicate whether the target, the letter “T” placed among a set of distractor letter “L”s, is rotated to the left or right. Response time benefits are observed for stimuli in which the position of distractors relative to targets is repeated (Chun & Jiang, 1998, 1999). Three causes of the contextual cueing effect have been identified. In the first, the memory of repeated displays efficiently guides attention during the search process to the target location. Because the onset of the decision (i.e., left or right?) is dependent upon the length of the search time, in the contextual cueing task, the cueing effect of repeated displays is implemented in the diffusion model through the non-decision time parameter Ter (Sewell et al., 2018). In analogy to the dot-probe, if the cue is effective, one way that attentional bias might be reflected is by shorter Ter to angry vs. neutral faces. The argument for Ter to be a measure of attentional bias has also previously been made for a study of anxious adults (Price et al., 2019).

Once attention is focused, two other processes come into play to produce the reduced RT observed in cueing effects. Recognition of the context within which the target is embedded (i.e., the angry face) allows individuals to reduce the amount of information needed to confirm that the target has been identified and attention correctly directed (Kunar & Wolfe, 2011; Kunar et al., 2007; Schankin & Schubo, 2009, 2010; Zhao et al., 2012). In the diffusion model, this process is implemented as decreases in boundary separation (Sewell et al., 2018; Weigard & Huang-Pollock, 2014).[1] Therefore, attentional bias among behaviorally inhibited children would be observed as shorter Ter and smaller boundary separation in a dot-probe task.

DM parameters have also demonstrated satisfactory test–retest reliabilities in lexical decision and recognition memory tasks (Lerche & Voss, 2017). Although test–retest reliability of a difference score of the Ter parameter from a dot-probe task using threat versus non-threat words in a sample of clinically anxious adults produced a low test–retest reliability of ICC = 0.25, it was still better than that for the traditional bias scored based on RTs where ICC = 0.001 (Price et al., 2019). It is not known whether a similar improvement would be found for children. Therefore, building on this newer body of work demonstrating the potential utility of the DM to better understand the cognitive mechanisms supporting attentional bias to threat, and the potential to improve the reliability of performance indices, the present study aims to determine if DM can improve the known shortcomings of the dot-probe task in a sample of behaviorally inhibited (BI) children.

Hypothesis 1 If the diffusion model provides a more accurate index of performance, then a significant main effect of Group (BI vs. non-BI) would be observed in which children who are behaviorally inhibited would have faster Ter and smaller boundary separations in the dot-probe task. They would also be more likely to demonstrate larger attentional bias difference scores when calculated via diffusion model parameters as opposed to traditional RT difference scores.

Hypothesis 2 Similarly, if the diffusion model parameters more accurately describe behavioral performance on the dot-probe task, those 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 performance indexed by RT and Accuracy.

Hypothesis 3 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.

Materials and methods

Participants

A total of 120 children ages 9–12 years old ($M = 10.82$ SD = 1.01) 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. Children were recruited as part of a larger study on the relations between attention and anxiety in behaviorally inhibited school-aged children (Thai et al., 2016) and reflect the ethnic and racial makeup of the region. The ethnicity of the sample was 68.6% White, 3.3% Asian/Pacific Islander,

[1] The information used to judge the orientation of the target is also influenced by learned factors that are independent of the perceptual qualities of the target itself (e.g. including learning to suppress distractors, see Sewell et al. (2018)). This latter process, however, is less likely to be observed in a task like the dot-probe, where explicit distractors are not present.

2.5% Hispanic, 0.8% African American, 2.5% mixed race, and 22.3% unreported.

Potential participants were screened for the study using the Behavioral Inhibition Questionnaire (BIQ), a parent-report 30-item questionnaire of their child's response to social or situational novelty based on a 1–7 point Likert scale (Bishop et al., 2003). The sum of all items produces an overall BI score, in addition to subscale scores for sensitivity to social novelty and situational novelty. Based on published cut-offs, children scoring ≥ 119 on the BIQ total score or ≥ 59 on the Social Novelty subscale were identified as being behaviorally inhibited. This resulted in an $n=43$ who were considered behaviorally inhibited [BI; 53% girls, average age = 10.71 (0.99) years, average IQ = 113.44 (12.53), average BIQ = 126.59 (19.04)]. Children scoring below each of these markers were identified as behaviorally non-inhibited (BN, $n=77$; 53% girls, average age = 10.89 (1.02) years, average IQ = 110.70 (14.28), average BIQ = 73.62 (20.08). There were no group differences in age, IQ, or sex distribution (all p's > 0.30).

A post hoc power analysis was conducted in GPOWER (Faul et al., 2007) assuming $\alpha=0.05$, $N=120$, and $f=0.25$ for main effects and interactions. With these assumptions, power was 0.88, 0.99, and 0.99 to detect the main effect of BI, Visit, and the BI x Visit interactions, respectively. Power was 0.80 to detect an $f=0.22$ for BI, and an $f=0.13$ for the main effect of Visit and the BI × Visit interaction.

Procedures

Once screened into the study, children and their parents attended two separate visits spaced approximately one week apart (average = 9.35 days). During the first visit, trained research staff administered the Computerized Diagnostic Interview Schedule for Children Version IV (C-DISC IV) to parents. During both visits, children completed a dot-probe task presented with the E-Prime software package version 2.0. Electroencephalogram (EEG) recordings were obtained during the second administration only.

Dot-Probe Task During the first visit, children were seated in a comfortable chair facing a computer monitor. When the task started, they 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. A left or right-facing arrow then appeared in place of one of the faces. Participants were given 1500 ms to indicate the direction of the arrow by a finger press.

During the second visit, EEG data were collected during task performance. To allow enough time for ERPs to be recorded, the stimulus presentation time was lengthened to 1000 ms and the response interval was lengthened to 2000 ms. These timing changes may represent an important difference in paradigm construction between the first and second administrations, which we address with respect to reliability estimates in "Discussion".

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 Congruent (NTc) condition, the target appeared in the angry face's location. In the Neutral-Threat Incongruent (NTi) condition, the target appeared in the neutral face's location. After the response window elapsed, the next trial began. Participants were administered 180 trials, split across three blocks of 60 trials each, with trials of each condition split evenly among blocks.

Equipment

EEG data were collected continuously using a 128-channel geodesic sensor net (Electrical Geodesics Inc., Eugene, Oregon). Vertical eye movements were monitored by electrodes 1 cm above and below each eye, while horizontal eye movements were monitored by electrodes 1 cm 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 100 ms before stimulus onset through 500 ms after stimulus onset, 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 0.1 Hz and a low pass frequency of 40 Hz were used to filter the data. Eye movement artefacts were removed using the Gratton method (Gratton et al., 1983). ERPs in response to face prompts were calculated as in Thai et al. (2016) 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). ERPs were calculated as the mean amplitude across the ERP window, which was 50 ms before and after the peak. Specifically, occipital ERPs included the P1 (40–140 ms) and N170 (120–220 ms). Fronto-central ERPs included the N1 (60–140 ms), P2 (140-240 ms), and N2 (260–360 ms). Figure 1 provides Grand Average waveforms.

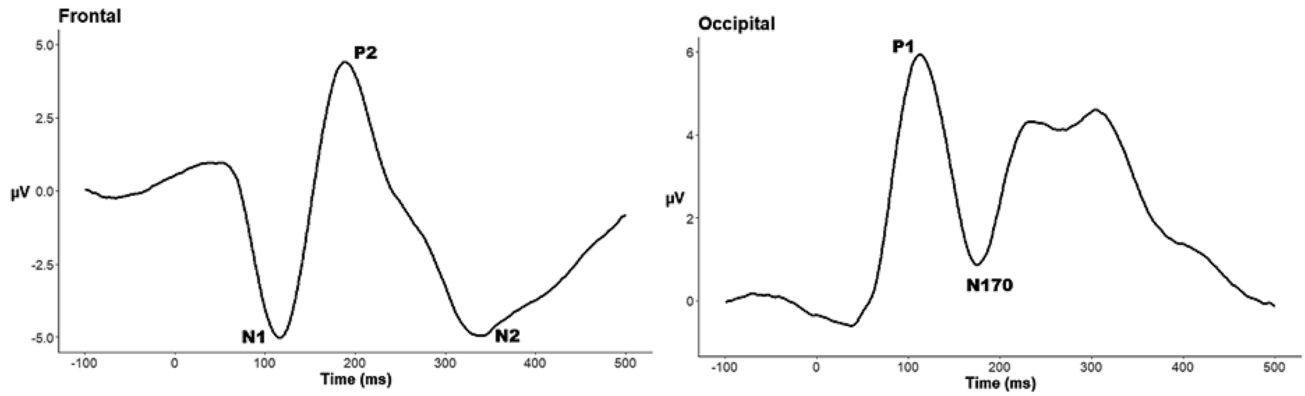


Fig. 1 Grand Average ERPs of interest. The left panel provides grand averages of fronto-central electrodes (3, 4, 5, 9, 10, 11, 12, 16, 18, 19, 20, 22, 23, 24, 27, 28, 33, 117, 118, 122, 123, 124). The right panel provides grand averages of occipital electrodes (65, 66, 69, 70, 71, 74, 76, 82, 83, 84, 89, 90)

Data preparation

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

Diffusion model parameters Anticipatory responses faster than 300 ms were excluded from analysis per convention (Ratcliff & Tuerlinckx, 2002). No other exclusions were set. This resulted in the exclusion of only 4.6% of trials, or an average of 8.33 per participant; the ability to retain more data to provide a more accurate description of the performance is one of the many benefits of the DM. Diffusion model parameters were calculated for each participant using the FastDM software (Voss et al., 2013). Data were computed using the Kolmogorov–Smirnov (ks) optimization criteria, which is the recommended process when the total number of trials ranges from 100 to 400 (Voss et al., 2013). Drift (v), relative starting point (za), non-decision time (Ter) and boundary separation (a) 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. Upper and lower boundaries for the model were coded as correct/incorrect.

Several methods were utilized to assess for model fit. First, KS-tests were conducted on each participant for each task condition. Means and standard deviations are shared in Table 1.

A simulation-recovery study was also conducted using the "rdiffusion()" function from the R package "rtdists". This generates simulated response time data based on the diffusion parameters derived from the empirical response times, which is then fit to the empirical response times. Correlations between the simulated and empirical incongruent and congruent values at both Visit 1 and Visit 2 for v, a, and ter were all above 0.75 with an average of 0.88, indicating the parameters fit well. *Za* was below 0.5 at each data point, indicating poor fit, which we further discuss in "Results".

In addition, cumulative distribution function (CDF) plots were formulated based on the empirical and simulated datasets. Visual inspection of this graphical visualization of the data confirmed adequate fit for each task condition. See Fig. 2 for CDF plots.[3]

For interested readers, Supplemental Table 1 provides a correlation table of traditional RT and accuracy measures with DM parameters.

To evaluate whether results varied by optimization criterion, DM parameters were also computed using Maximum Likelihood (ML) and Chi-Square (CS) methods, but this did not significantly alter the pattern or interpretation

[2] Results did not change when a cutoff of < 300 ms was applied to better approximate cutoffs used for diffusion modeling.

[3] Because the CDF plots suggested the presence of some misfits, 1000 datasets were subsequently simulated. Participants who exhibited a lower model fit (defined as < 10% quantile of the distribution of p values) for any of the four conditions were removed from analysis. This resulted in a reduced $N = 57$ (22 BI, 34 Controls). CDF plots generated from the remaining participants demonstrated improved model fit, but primary results did not change. See Supplementary Table 2 and Supplementary Fig. 1.

Table 1 Means and SDs of performance, and Pearson's (r) between Congruent (NTc) and Incongruent (NTi) cues

	Visit 1				Visit 2			
	NN	NTc	NTi	r (NTc:Ti)	NN	NTc	NTi	r (NTc:Ti)
MRT								
BI	612.67 (80.79)	611.75 (78.07)	609.35 (73.61)	0.96**	541.18 (81.54)	537.24 (79.93)	539.26 (82.57)	0.94**
BN	632.01 (82.06)	631.81 (80.26)	633.15 (75.80)	0.91**	546.30 (78.39)	543.50 (80.08)	546.69 (85.32)	0.95**
Total	626.03 (81.97)	625.61 (79.95)	625.80 (75.90)	0.92**	544.48 (79.22)	541.28 (79.75)	544.05 (84.08)	0.95**
SDRT								
BI	140.40 (38.41)	141.84 (37.66)	139.20 (29.66)	0.78**	128.27 (47.60)	125.95 (43.04)	132.90 (53.53)	0.64**
BN	144.66 (33.16)	142.28 (34.87)	144.99 (30.85)	0.76**	127.91 (46.93)	128.98 (52.16)	130.44 (53.84)	0.74**
Total	143.34 (34.83)	142.14 (35.68)	143.20 (30.54)	0.77**	128.04 (46.97)	127.91 (48.95)	131.31 (53.52)	0.70**
Acc								
BI	0.91 (0.16)	0.91 (0.15)	0.90 (0.15)	0.98**	0.93 (0.06)	0.92 (0.08)	0.93 (0.06)	0.71**
BN	0.91 (0.11)	0.91 (0.14)	0.91 (0.11)	0.94**	0.92 (0.07)	0.92 (0.07)	0.92 (0.07)	0.73**
Total	0.91 (0.13)	0.91 (0.14)	0.91 (0.13)	0.96**	0.93 (0.07)	0.92 (0.07)	0.93 (0.06)	0.72**
A								
BI	1.25 (0.26)	1.23 (0.23)	1.24 (0.23)	0.67**	1.16 (0.31)	1.13 (0.29)	1.18 (0.31)	0.64**
BN	1.22 (0.18)	1.21 (0.18)	1.23 (0.18)	0.61**	1.09 (0.22)	1.1 (0.24)	1.12 (0.26)	0.64**
Total	1.23 (0.2)	1.22 (0.19)	1.23 (0.2)	0.64**	1.12 (0.26)	1.11 (0.26)	1.14 (0.28)	0.64
V								
BI	2.83 (1.14)	2.86 (1.2)	2.74 (1.10)	0.89**	3.21 (0.73)	3.11 (0.77)	3.25 (0.8)	0.54**
BN	2.82 (0.91)	2.85 (0.93)	2.82 (0.98)	0.78**	3.16 (0.83)	3.18 (0.81)	3.01 (0.81)	0.62**
Total	2.82 (0.99)	2.85 (1.01)	2.8 (1.02)	0.82**	3.18 (0.8)	3.16 (0.79)	3.09 (0.82)	0.58**
Za								
BI	0.40 (0.08)	0.40 (0.08)	0.41 (0.07)	0.43**	0.41 (0.11)	0.42 (0.09)	0.4 (0.1)	0.11
BN	0.38 (0.07)	0.39 (0.08)	0.39 (0.08)	0.10	0.4 (0.08)	0.39 (0.08)	0.43 (0.09)	0.23*
Total	0.39 (0.07)	0.39 (0.08)	0.39 (0.08)	0.21**	0.41 (0.09)	0.4 (0.09)	0.42 (0.1)	0.15
Ter								
BI	0.39 (0.06)	0.39 (0.05)	0.39 (0.06)	0.78*	0.36 (0.04)	0.36 (0.04)	0.35 (0.04)	0.71*
BN	0.4 (0.05)	0.41 (0.05)	0.41 (0.06)	0.83**	0.36 (0.04)	0.36 (0.04)	0.36 (0.04)	0.77**
Total	0.39 (0.08)	0.39 (0.08)	0.4 (0.08)	0.83**	0.37 (0.05)	0.36 (0.05)	0.36 (0.05)	0.75**

MRT mean reaction time, *SDRT* standard deviation of reaction time, *Acc* accuracy, *a* boundary separation, *v* drift rate, *za* start point, *Ter* non-decision time, *BI* behaviorally inhibited, *BN* behaviorally non-inhibited, *NN* neutral–neutral cue, *NTc* neutral-threat congruent cue, *NTi* neutral-threat incongruent cue

*$p < 0.05$

**$p < 0.01$

of results, so results based on KS optimization procedures are reported here.

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 (i.e. NTi–NTc) to form the following variables: *v*Diff, *a*Diff, *za*Diff, *Ter*Diff. Because larger drift rates indicate faster drift, unlike RT difference scores, a negative *v*Diff score would indicate a greater bias to threat. For all other difference scores, a larger positive score indicates bias to threat.

Results

Attentional Bias to Threat See Table 2 for a summary of results. A mixed within (Visit: first, second) and between (BI: BI, non-BI) subjects ANOVA found no significant main effect of BI on attentional bias to threat as calculated by the traditional RT difference score [NTi–NTc, $F(1, 118) = 0.85$, $p = 0.36$, $\eta^2 = 0.01$], or as calculated by any difference scores formed from the DM parameters (*v*Diff, *a*Diff, *za*Diff, and *ter*Diff; all $p > 0.27$, all $\eta^2 < 0.01$). There was also no main effect of Visit on attentional bias

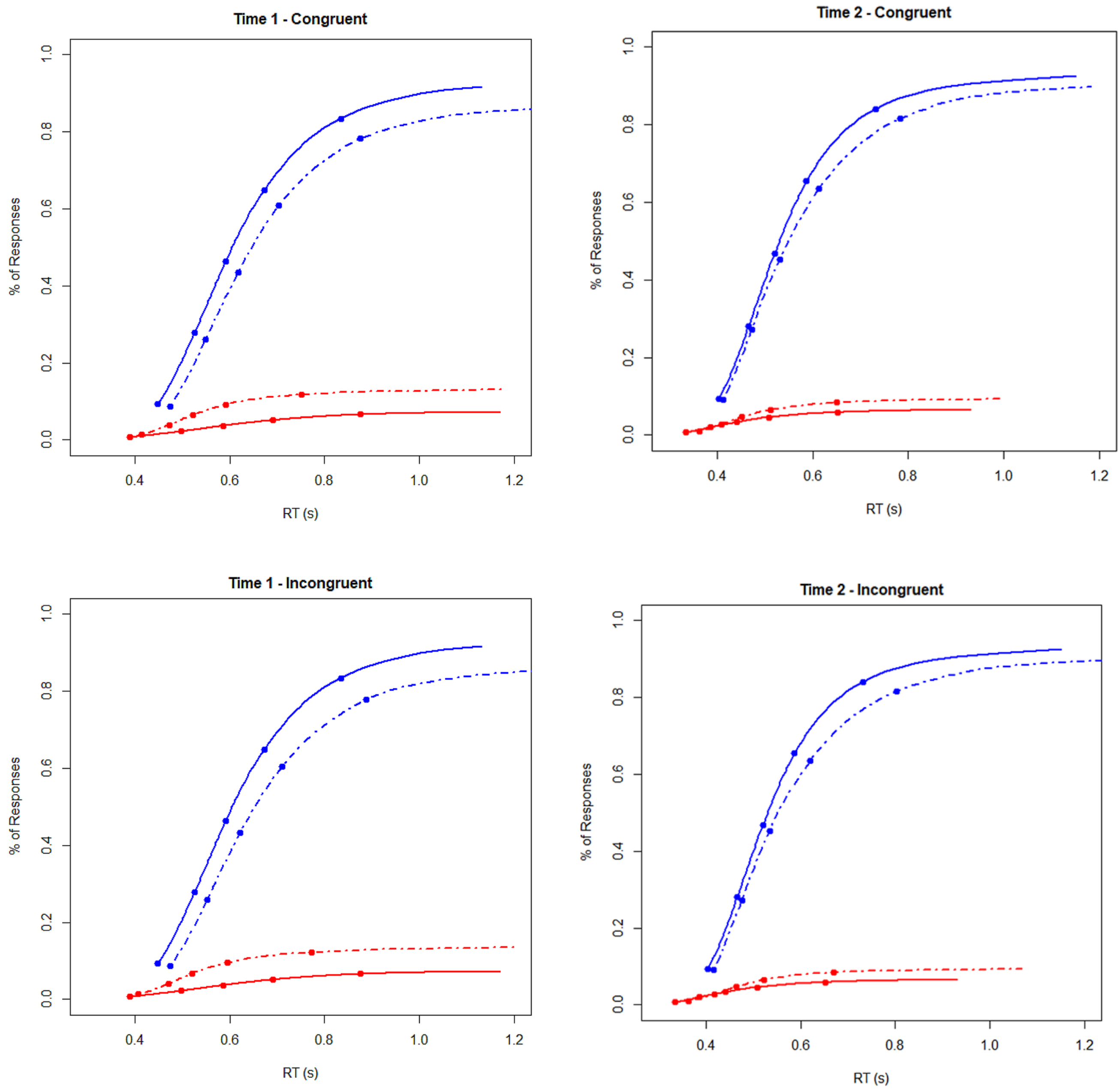


Fig. 2 Cumulative Distrubution Function (CDF) plots of congruent (left panels) and Incongruent (right panels) data at both Visits. Blue lines represent the distribution of accurate responses, while red lines indicate the distribution of inaccurate responses. Continuous lines represent observed data, while dashed lines represent simulated data

to threat regardless of how it was calculated (all $p > 0.11$, all $\eta^2 < 0.02$).

There was, however, a significant BI × Visit interaction on vDiff ($p = 0.01$, $\eta^2 = 0.06$) in which the drift difference score was larger and positive at the second timepoint among those identified as BI ($p = 0.04$, $\eta^2 = 0.10$), but there was no effect of Visit on controls ($p = 0.13$, $\eta^2 = 0.03$). There was also BI × Visit interaction on *za*Diff ($p < 0.01$, $\eta^2 = 0.08$) in which control participants showed a significant increase of *za*Diff in their second visit ($p < 0.01$, $\eta^2 = 0.11$), but BI participants did not ($p = 0.11$, $\eta^2 = 0.06$). We return to these different indices of bias in "Discussion". See Table 2 for a list of model results.[4]

[4] A Cue (3: Neutral, Threat Congruent, Threat Incongruent) Visit (2) BI (2) GLM replicated these effects. However, this GLM identified an additional Visit (2) Cue (3) interaction on Ter ($F(2, 236) = 5.17$, $p = 0.006$, $\eta^2 = 0.042$) in which the Neutral cue trials did not differ between visit 1 and 2 as much as the task condition cue trials did.

Table 2 Visit × BI General Linear Model of difference score values

	Parameter	F (1, 118)	P	η^2
BI	RT threat bias	0.85	0.36	0.01
	SDRT Diff	3.50	0.06	0.03
	ACC Diff	0.09	0.76	0.00
	*a*Diff	0.24	0.63	0.00
	*v*Diff	0.91	0.34	0.01
	*za*Diff	1.21	0.27	0.01
	TerDiff	1.01	0.32	0.01
Visit	RT threat bias	0.02	0.89	0.00
	SDRT Diff	0.59	0.44	0.01
	ACC Diff	2.67	0.11	0.02
	aDiff	1.13	0.29	0.01
	vDiff	0.72	0.40	0.01
	zaDiff	0.62	0.43	0.01
	TerDiff	0.03	0.87	0.00
Visit × BI	RT threat bias	0.47	0.50	0.00
	SDRT Diff	0.25	0.62	0.00
	ACC Diff	0.21	0.65	0.00
	aDiff	0.62	0.43	0.01
	vDiff	7.01	0.009*	0.06
	zaDiff	9.89	0.002*	0.08
	TerDiff	0.97	0.33	0.01

BI behaviorally inhibited, *RT* reaction time, *SDRT* standard deviation of reaction time, *Acc* accuracy, *a* boundary separation, *v* drift rate, *za* start point, *Ter* non-decision time, *Diff* difference score

*$p < 0.05$

**$p < 0.01$

ERP analyses Simple regression analyses were conducted to separately assess the association between attentional bias indices and N1, P1, N170, N2, and P2 (see Table 3). RT difference score, *a*Diff, *v*Diff, and *ter*Diff did not regress significantly onto any ERP (all r's < 0.154, all p's > 0.056). P1 and N170 were negatively associated with the *za*Diff (P1: $r = -0.22$, $p = 0.01$; N170: $r = -0.21$, $p = 0.01$). In a decision task where the judgement is explicit (i.e. is the face a threat or non-threat) such a correlation might suggest stronger expectancy biases are associated with lower state arousal or reduced selective attention (Luck et al., 2000). However, because the attentional bias in the dot-probe task is measured indirectly, and because our simulation recovery study of the *za* parameter indicated poor fit, the interpretability or relevance of this particular association is not clear.

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 re-test reliability of difference scores are maximized when scores to be subtracted from one another are only weakly positively correlated, and if the individual scores are reliable in and of themselves (Rodebaugh et al., 2016). The correlation between performance to incongruent and congruent cues at both time points was moderate to high for most variables (r's 0.61–0.97) except *za*, which was small to moderate (r's 0.10–0.43; see Table 1).

These strong positive correlations between cue conditions not surprisingly lead to uniformly low test–retest reliabilities of all difference scores (see Fig. 3). Neither the commonly reported threat bias score, ($r = 0.06$, $p = 0.52$), nor any of the

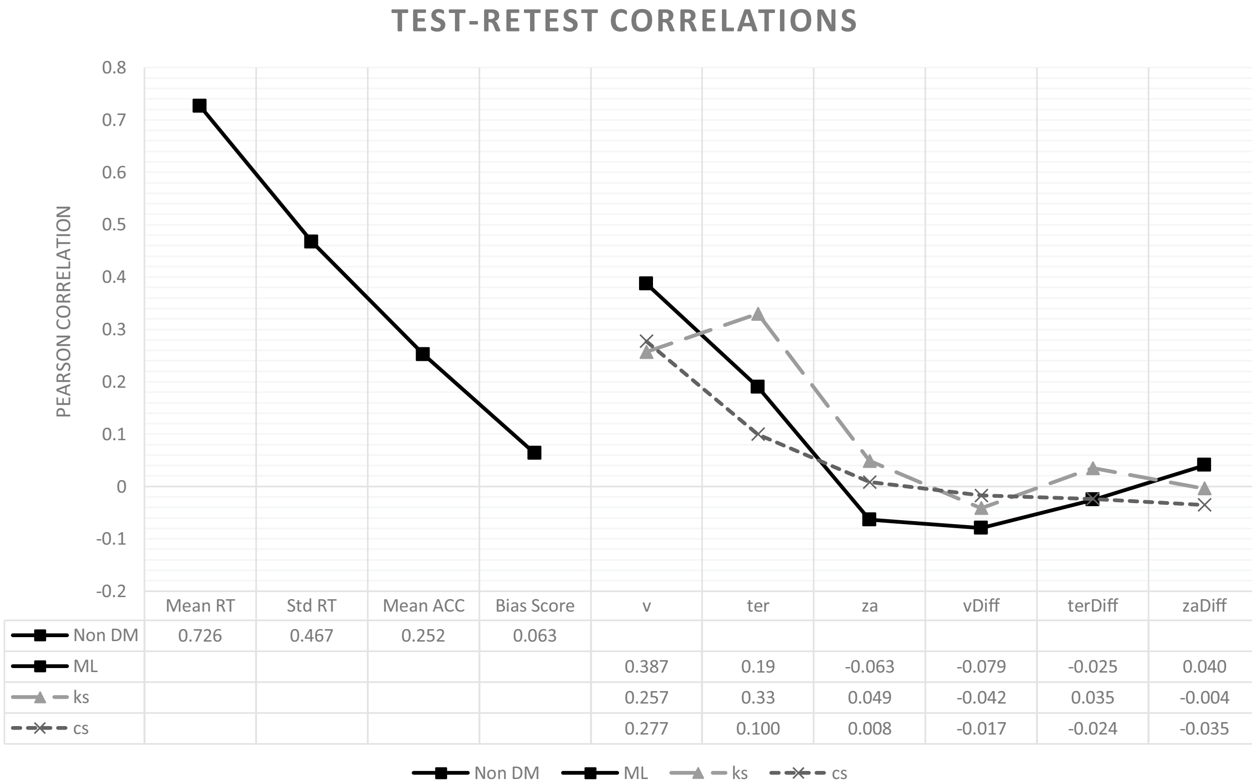
Table 3 Pearson correlation of ERP values with traditional and diffusion model indices of performance

Performance parameter	N1	P1	N170	P2	N2
RT	0.135	−0.046	0.03	0.054	0.042
ACC	−0.045	0.023	0.037	−0.023	−0.104
ACC difference score	−0.133	0.092	0.052	−0.238	−0.110
RT difference score (bias)	0.064	0.069	0.122	0.098	0.133
A	−0.015	0.180*	−0.014	0.075	−0.082
*a*Diff	0.097	−0.020	−0.090	−0.004	−0.072
V	−0.044	−0.032	0.097	−0.083	0.097
*v*Diff	< 0.001	0.154	0.148	−0.043	−0.068
Za	−0.017	0.142	−0.007	0.110	−0.108
*za*Diff	0.107	−0.216*	−0.211*	−0.018	−0.032
Ter	0.013	0.037	−0.043	−0.026	−0.071
terDiff	−0.021	−0.102	0.095	0.021	0.039

RT reaction time, *a* boundary separation, *v* drift rate, *za* start point, *Ter* non-decision time, *Diff* difference score

*$p < 0.05$

**$p < 0.001$



Note: KS = Kolmogorov-Smirnov optimization criteria, ML = Maximum Likelihood optimization criteria, CS = chi-square optimization criteria.

Fig. 3 Test–retest reliability of various performance measures

equivalent difference scores formed via the DM parameters (r's -0.08 to 0.04, all p's > 0.35) were significantly correlated across task administrations. DM parameters were also computed using Maximum Likelihood), and Chi-Square (cs) methods (Voss et al., 2013), but this did not alter the pattern of results.

Because of the strong positive correlation in performance between cue conditions, test–retest reliability was higher when examining overall score values collapsed across cue type. Metrics of performance with the highest correlation coefficients were mean RT ($r=0.73$, $p<0.001$) and standard deviation of RT ($r=0.47$, $p<0.001$), followed by mean overall v ($r=0.39$, $p<0.001$). Accuracy across administrations was generally high (~92%); both accuracy and Ter were only weakly correlated over task administrations (Accuracy $r=0.25$, $p=0.01$; Ter $r=0.33$, $p=0.02$).

Test–retest reliability of performance as indexed by mean 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% more between visits, $t=2.44$, $p=0.016$, 95% CI [0.024, 0.226]. No other moderating effect of age, IQ, or severity of BI was found for any variable.

Discussion

Attentional bias to threat is a core cognitive construct that is believed to contribute to the development and maintenance of behavioral inhibition and anxiety disorders. However, recent research has been unable to replicate the link between behavioral inhibition and threat bias as indexed by the RT difference score on the dot-probe task. Difference scores typically have low test–retest reliability (Peter et al., 1993), leading some to question their ability to be used as a marker of individual differences in cognition (Enkavi et al., 2019; Rodebaugh et al., 2016).

Consistent with those concerns, in a large and well-characterized school-aged sample of children, we found no main effects of BI on the difference score for any index of performance, using either traditional indices or any diffusion model parameter. Angry faces did not differentially cue attention to a location in space for either group of children. There was, however, a significant BI × Visit interaction for *v*Diff; bias was somewhat smaller among children identified as BI in the second vs. first visit. There was also a BI × Visit interaction for *za*Diff in which zaDiff increased for controls at Visit 2. Recall that bias may occur either from an increase in information accumulation (i.e. faster drift rate) or through changes in response expectancy (i.e. changes in relative start point). Taken at face value, these results would suggest somewhat different processes occurring between visits 1 and 2 depending on BI status. In visit 2, information accumulation for angry faces normalizes for children with BI (*v*Diff), but controls have a slightly stronger response expectancy bias (*za*Diff) for angry faces in the second visit.

Reliability is not task-specific but is a function of the sample and measurement (Ross et al., 2015). Because diffusion modelling provides a more accurate and nuanced description of the performance, it was hoped that these parameters, including those like Ter and boundary separation that are most theoretically related to the cueing effect, might prove to be more reliable and possibly more sensitive to differences in temperament. DM parameters in lexical decision and recognition memory tasks have, in fact, demonstrated satisfactory test–retest reliabilities (Lerche & Voss, 2017). We did not find this to be the case for the current dot-probe data. Consistent with existing literature, reliability for the standard RT difference score between congruent and incongruent trials was $r = 0.03$. It did not improve with difference scores created from DM parameters, ranging from a low of − 0.079 for *v*Diff to a high of 0.040 for *za*Diff. Thus, despite the somewhat intriguing interpretation of the Visit × BI interaction for *v*Diff and *za*Diff, such interpretation should be qualified given the low reliability of those indices. Stronger reliability coefficients were observed when we averaged across cue conditions.

The presence of strong positive correlations in performance between conditions inherently reduces the reliability of a difference score. To our knowledge, Price et al. (2019) provide the only other analysis of test–retest reliability of a dot-probe task using DM parameters, albeit in a young adult sample. Although they reported a low test–retest reliability (ICC = 0.25) of their Ter difference score across three-time points, they reported a moderate Pearson's correlation ($r = 0.63$) across the first two. The correlation coefficients of Ter and RT for each condition at each of these first two time points was not reported, which might have provided some explanation for the difference in findings.

Use of diffusion model parameters also did not improve the association between performance and temporally sensitive indices of attentional functioning. Neither traditional measures of task performance nor the overall mean values of the diffusion model parameters regressed significantly onto the amplitudes of concurrently recorded ERPs (P1, N1, N170, P2, N2), with the exception of *za*Diff. The P1 is believed to be modulated by arousal or selective attention (Luck et al., 2000), so in an explicit bias task, such a correlation might be interpreted as indicating strong expectancy biases are associated with reduced arousal or selective attention. However, because the dot-probe task is an indirect measure of attentional bias, the interpretational relevance of this particular association is not clear and may not be meaningful.

It bears mentioning that EEG data collection only occurred during the second but not the first administration. In addition, due to the requirements of ERP data collection, slight differences in the duration of stimulus (500 vs 1000 ms) and the amount of time participants were given to make a response (1500 vs. 2000 ms) varied between visits. The non-random variance in these administration procedures may have introduced systematic error to our estimates and together contributed to the poor reliability coefficients reported here.

However, even if it were the case, it is unlikely that such modifications could have completely accounted for such low test–retest reliability coefficients. We are reassured in the overall interpretive accuracy of our results because the reliability of performance collapsed across cue conditions was much larger (in the small to moderate range) compared to the difference scores (the foremost index of attentional bias to threat) which had negligible reliability. If these administrative differences had influenced our reliability coefficients, it would have done so across the board.

A better explanation for low reliability in this and other studies is likely due to low cue validity. The dot-probe task was modelled off cueing tasks that were originally developed within the cognitive literature to study visuospatial attentional control (Huang-Pollock & Nigg, 2003). Cuing is not an all-or-nothing process. The ability of a cue to orient attention is partially dependent upon how relevant it is to accomplish the goal of a task (Victeur et al., 2020), as well as the cue's validity (Vossel et al., 2014). Cues that are probabilistic and predict target location are more effective in orienting attention than cues that are not (Vossel et al., 2014). However, in most dot-probe tasks, the facial cues are non-probabilistic and appear behind the angry face 50% of the time. It is also not uncommon for cues in dot-probe tasks to be negatively probabilistic—in these designs, neutral–neutral face pairings, or catch trials in which no target appears at all, are added alongside threat-neutral face pairings (in which the target appears behind the angry face 50% of the time). In these designs, the angry face predicts target location on only 25% of trials. These design flaws are therefore likely causing inconsistent cueing effects that contribute to the low reliability of the difference score. Some have argued that attentional bias to threat might best be conceptualized as a dynamic process in which attention may first be directed towards and then away from an angry face during any given trial (Rodebaugh et al., 2016). However, even if it were the case that attention shifts first to and then away from the angry face, whether attention is initially oriented to the angry face would still be heavily dependent upon the probability of the cue predicting the location of the target.

Given the fragility of the cueing response, ongoing exploration of other paradigms of attentional bias, rather than reliance on the dot-probe, are increasingly appropriate. First, as with work by White et al. (2016), we recommend adopting tasks in which bias can be directly measured during the decision process (e.g., is this a threat or not-threat?) as opposed to indirectly measured via cueing methods. Several groups have used eye-tracking technology to directly measure children's pupil dilation and gaze time toward threatening versus neutral stimuli during computerized tasks (Lisk et al., 2019; Price et al., 2016). However, the low between subjects variability that is often the goal when developing and designing experimental measures of cognition may also make them unsuitable for individual differences research (Enkavi et al., 2019; Hedge et al., 2018). Thus, mobile eye-tracking measures in more naturalistic contexts that are not dependent upon robust experimental manipulations may ultimately provide a useful approach to understanding attentional bias to threat in anxious or behaviorally inhibited children (Fu & Pérez-Edgar, 2019).

Conclusion

Over the last 20 years, the dot-probe task has served as the gold standard laboratory task for capturing attentional bias to threat. However, 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 to improve the reliability of the measurement of attentional bias, as well as its relation with BI and 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 typically provide more accurate summaries of performance.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00426-021-01532-3.

Acknowledgements The authors would like to thank the Pennsylvania State University Social, Life, and Engineering Sciences Imaging Center (SLEIC) Human Electrophysiology Facility for supporting data collection, the TAU/NIMH ABMT Initiative for providing the task toolkit, and the many individuals who contributed to data collection and data processing. We would especially like to thank the parents of the children who participated and continue to participate in our studies.

Funding This work was supported by the National Institutes of Health under Grant [R01MH094633] to Koraly Pérez-Edgar.

Availability of data and material Raw study data used in this publication is available upon request.

Declarations

Conflict of interest There are no known conflicts of interest for any author.

Ethics approval Study procedures were approved by the Institutional Review Board (IRB number PRAMS00038109) of The Pennsylvania State University. Families were compensated for their participation.

Consent to participate Participants and their parents provided written assent and consent for data collection.

Consent for publication Participants and their parents provided written assent and consent for potential data sharing.

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.

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. https://doi.org/10.1046/j.1467-8624.2003.00645.x

Blackford, J. U., & Pine, D. S. (2012). Neural substrates of childhood anxiety disorders: a review of neuroimaging findings. *Child and Adolescent Psychiatric Clinics, 21*(3), 501–525.

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., et al. (2009). Stable early maternal report of behavioral inhibition predicts lifetime social anxiety disorder in adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry, 48*(9), 928–935. https://doi.org/10.1097/CHI.0b013e3181ae09df

Chun, M. M., & Jiang, Y. H. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*(1), 28–71.

Chun, M. M., & Jiang, Y. H. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science, 10*(4), 360–365.

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

Ehrenreich, J. T., & Gross, A. M. (2002). Biased attentional behavior in childhood anxiety: A review of theory and current empirical investigation. *Clinical Psychology Review, 22*(7), 991–1008.

Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia, 45*(1), 15–31. https://doi.org/10.1016/j.neuropsychologia.2006.04.022

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.

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences, 116*(12), 5472–5477.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146

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

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.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166–1186.

Henderson, H. A. (2010). Electrophysiological correlates of cognitive control and the regulation of shyness in children. *Developmental Neuropsychology, 35*(2), 177–193. https://doi.org/10.1080/87565640903526538

Huang-Pollock, C., & Nigg, J. T. (2003). Searching for the attention deficit in attention deficit hyperactivity disorder: the case of visuospatial orienting. *Clinical Psychology Review, 23*(6), 801–830. https://doi.org/10.1016/s0272-7358(03)00073-4

Hudson, J. L., Dodd, H. F., Lyneham, H. J., & Bovopoulous, N. (2011). Temperament and family environment in the development of anxiety disorder: Two-year follow-up. *Journal of the American Academy of Child and Adolescent Psychiatry, 50*(12), 1255–1264. https://doi.org/10.1016/j.jaac.2011.09.009

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.

Kappenman, E. S., Farrens, J. L., Luck, S. J., & Proudfit, G. H. (2014). Behavioral and ERP measures of attentional bias to threat in the dot-probe task: Poor reliability and lack of correlation with anxiety. *Frontiers in Psychology, 5*, 1368.

Kunar, M. A., & Wolfe, J. M. (2011). Target absent trials in configural contextual cuing. *Attention Perception and Psychophysics, 73*(7), 2077–2091. https://doi.org/10.3758/s13414-011-0164-0

Kunar, M. A., Flusberg, S., Horowitz, T. S., & Wolfe, J. M. (2007). Does contextual cuing guide the deployment of attention? *Journal of Experimental Psychology-Human Perception and Performance, 33*(4), 816–828. https://doi.org/10.1037/0096-1523.33.4.816

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.

Leite, F. P., & Ratcliff, R. (2011). What cognitive processes drive response biases? A diffusion model analysis. *Judgment and Decision Making, 6*(7), 651–687.

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

Lisk, S., Vaswani, A., Linetzky, M., Bar-Haim, Y., & Lau, J. Y. F. (2019). Systematic review and meta-analysis: Eye-tracking of attention to threat in child and adolescent anxiety. *Journal of the American Academy of Child and Adolescent Psychiatry*. https://doi.org/10.1016/j.jaac.2019.06.006

Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences, 4*(11), 432–440. https://doi.org/10.1016/s1364-6613(00)01545-x

Molloy, A., & Anderson, P. L. (2020). Evaluating the reliability of attention bias and attention bias variability measures in the dot-probe task among people with social anxiety disorder. *Psychological Assessment, 32*(9), 883.

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. M., Hofmann, S. G., Santesso, D. L., Meuret, A. E., Bitran, S., & Pizzagalli, D. A. (2009). Electrophysiological evidence of attentional biases in social anxiety disorder. *Psychological Medicine, 39*(7), 1141–1152. https://doi.org/10.1017/s0033291708004820

Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the brain: A diffusion model analysis of prior probability and potential payoff. *Journal of Neuroscience, 32*(7), 2335–2343.

Pérez-Edgar, K., Reeb-Sutherland, B. C., McDermott, J. M., White, L. K., Henderson, H. A., Degnan, K. A., et al. (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.

Peter, J. P., Churchill, G. A., & Brown, T. J. (1993). Caution in the use of difference scores in consumer research. *Journal of Consumer Research, 19*(4), 655–662. https://doi.org/10.1086/209329

Price, R. B., Rosen, D., Siegle, G. J., Ladouceur, C. D., Tang, K., Allen, K. B., et al. (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.

Price, R. B., Brown, V., & Siegle, G. J. (2019). Computational modeling applied to the dot-probe task yields improved reliability and mechanistic insights. *Biological Psychiatry, 85*(7), 606–612.

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

Rodebaugh, T. L., Scullin, R. B., Langer, J. K., Dixon, D. J., Huppert, J. D., Bernstein, A., et al. (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. https://doi.org/10.1016/j.bandc.2013.05.008

Ross, D. A., Richler, J. J., & Gauthier, I. (2015). Reliability of composite-task measurements of holistic face processing. *Behavior Research Methods, 47*(3), 736–743.

Schankin, A., & Schubo, A. (2009). Cognitive processes facilitated by contextual cueing: Evidence from event-related brain potentials. *Psychophysiology, 46*(3), 668–679. https://doi.org/10.1111/j.1469-8986.2009.00807.x

Schankin, A., & Schubo, A. (2010). Contextual cueing effects despite spatially cued target locations. *Psychophysiology, 47*(4), 717–727. https://doi.org/10.1111/j.1469-8986.2010.00979.x

Schwartz, C. E., Snidman, N., & Kagan, J. (1999). Adolescent social anxiety as an outcome of inhibited temperament in childhood. *Journal of the American Academy of Child and Adolescent Psychiatry, 38*(8), 1008–1015. https://doi.org/10.1097/00004583-199908000-00017

Sewell, D. K., Colagiuri, B., & Livesey, E. J. (2018). Response time modeling reveals multiple contextual cuing mechanisms. *Psychonomic Bulletin and Review, 25*(5), 1644–1665. https://doi.org/10.3758/s13423-017-1364-y

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.

Victeur, Q., Huguet, P., & Silvert, L. (2020). Attentional allocation to task-irrelevant fearful faces is not automatic: Experimental evidence for the conditional hypothesis of emotional selection. *Cognition and Emotion, 34*(2), 288–301. https://doi.org/10.1080/02699931.2019.1622512

Vossel, S., Mathys, C., Daunizeau, J., Bauer, M., Driver, J., Friston, K. J., & Stephan, K. E. (2014). Spatial attention, precision, and bayesian inference: A study of saccadic response speed. *Cerebral Cortex, 24*(6), 1436–1450. https://doi.org/10.1093/cercor/bhs418

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology, 60*(6), 385.

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. https://doi.org/10.1007/s10608-015-9670-z

Waechter, S., Nelson, A. L., Wright, C., Hyatt, A., & Oakman, J. (2014). Measuring attentional bias to threat: Reliability of dot probe and eye movement indices. *Cognitive Therapy and Research, 38*(3), 313–333.

Weigard, A., & Huang-Pollock, C. L. (2014). A diffusion modeling approach to understanding contextual cueing effects in children with ADHD. *Journal of Child Psychology and Psychiatry, 55*(12), 1336–1344. https://doi.org/10.1111/jcpp.12250

White, C. N., & Poldrack, R. A. (2014). Decomposing bias in different types of simple decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(2), 385.

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.

Zhao, G., Liu, Q., Jiao, J., Zhou, P. L., Li, H., & Sun, H. J. (2012). Dual-state modulation of the contextual cueing effect: Evidence from eye movement recordings. *Journal of Vision, 12*(6), 1–13. https://doi.org/10.1167/12.6.11