Published in final edited form as:

Cognit Ther Res. 2022 February ; 46(1): 146–160. doi:10.1007/s10608-021-10212-w.

The Reliability and Validity of Response-Based Measures of Attention Bias

Emily E. E. Meissel[1], **Huiting Liu**[2], **Elizabeth S. Stevens**[3], **Travis C. Evans**[4], **Jennifer C. Britton**[5], **Allison M. Letkiewicz**[6], **Stewart A. Shankman**[6]

[1]San Diego State University/UC San Diego Joint Doctoral Program in Clinical Psychology, 6363 Alvarado Ct. San Diego, CA 92120

[2]Evidence Based Treatment Centers of Seattle, 1200 5th Avenue, Seattle, WA 98101

[3]Jesse Brown VA Medical Center, Chicago, IL, 820 S. Damen Avenue, Chicago, IL 60612

[4]VA Boston Healthcare System and Massachusetts General Hospital, Boston, MA, 150 S. Huntington Avenue Boston, MA 02130

[5]University of Miami, Coral Gables, FL, P.O. Box 248185 Coral Gables, FL 33124

[6]Northwestern University, Chicago, IL, 446 East Ontario #7-200 Chicago, IL 60611

Abstract

Background: Attention bias to threat is a fundamental transdiagnostic component and potential vulnerability factor for internalizing psychopathologies. However, the measurement of attentional bias, such as traditional scores from the dot-probe paradigm, evidence poor reliability and do not measure intra-individual variation in attentional bias.

Methods: The present study examined, in three independent samples, the psychometric properties of a novel attentional bias (AB) scoring method of the dot-probe task based on responses to individual trials. For six AB scores derived using the response-based approach, we assessed the internal consistency, test-retest reliability, familial associations, and external validity (using Social Anxiety Disorder, a disorder strongly associated with attentional bias to threatening faces).

Results: Compared to traditional AB scores, response-based scores had generally better internal consistency (range of Cronbach's alphas: 0.68–0.92 vs. 0.41–0.71), higher test-retest reliabilities (range of Pearson's correlations: 0.26–0.77 vs. −0.05–0.35), and were more strongly related in family members (range of ICCs: 0.11–0.27 vs. 0–0.05). Furthermore, three response-based scores added incremental validity beyond traditional scores and gender in the external validators of current and lifetime Social Anxiety Disorder.

Conclusions: Findings indicate that response-based AB scores from the dot-probe task have better psychometric properties than traditional scores.

Correspondence concerning this article should be addressed to Stewart A. Shankman, Department of Psychiatry, Northwestern University, Chicago, IL, 60611. stew.shankman@northwestern.edu.

Keywords

Attention Bias; Psychometrics; Social Anxiety Disorder; Masked Faces; Replication

Attentional bias is the preferential allocation of attention to threatening over neutral stimuli in the environment. While attending to potential threats can be protective, undue biases towards threat are a fundamental component of anxiety disorders (for a review, see; Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007). Indeed, attentional bias contributes to the etiology and maintenance of a variety of anxiety disorders, including social anxiety disorder (Amir, Prouvost, & Kuckertz, 2012; Amir, Elias, Klumpp, & Przeworski, 2003; MacLeod & Hagan, 1992; MacLeod, Rutherford, Campbell, Ebsworthy, & Holker, 2002).

The dot-probe task is one of the most widely used paradigms for the behavioral assessment of attentional bias (MacLeod, Mathews, & Tata, 1986), being employed in thousands of studies in over 30 years of research. Research on the dot-probe has also led to several randomized clinical trials to modify attentional biases as potential treatment for internalizing disorders (e.g. Hakamata et al., 2010). There is meta-analytic evidence of robust associations between threat-related attention bias from the dot-probe task and clinical anxiety (d=0.37; see meta-analysis by Bar-Haim et al., 2007). However, the magnitude of the effect of anxiety on attentional biases varies across studies (Shechner, Britton, Pérez-edgar, & Bar-haim, 2012; Bar-Haim et al., 2007) with some dot-probe studies failing to find effects (Kappenman et al., 2014; Kruijt, Parsons, & Fox, 2019). Additionally, traditional measures of attentional bias derived from the dot-probe task have been shown to have poor internal consistency (Evans, Walukevich, Seager, & Britton, 2018; Kappenman, Farrens, Luck, & Proudfit, 2014; Price et al., 2015; Schmukle, 2005; Staugaard, 2009; Waechter, Nelson, Wright, Hyatt, & Oakman, 2014) and weak associations with the N2pc, an event-related potential component that measures the allocation of visual attention to a particular location in the visual field (Kappenman et al., 2014; Kappenman, Macnamara, & Proudfit, 2015). Indeed, the NIH workgroup on the negative valence system of the RDoC Matrix placed the dot-probe task under “Tasks that require further evaluation,” citing these questionable psychometric properties (National Advisory Mental Health Council Workgroup on Tasks and Measures for RDoC, 2016; see a similar conclusion made by Kruijt, Parsons, & Fox, 2019).

Several reasons may explain the dot-probe task’s questionable convergent validity and reliability. First, questionable validity of the dot-probe task (e.g., weak N2pc associations and variable associations between anxiety and attentional biases) may result from the traditional computation of bias scores. In the dot-probe task (MacLeod et al., 1986), pairs of threatening and neutral stimuli (e.g. words or images) are presented, followed by a probe replacing one of the stimuli. Traditionally, attention bias is measured by subtracting an individual’s (a) average reaction time (RT) on trials in which the probe replaces the threatening stimulus (i.e., congruent trials) from the (b) average RT on trials in which the probe replaces the neutral stimulus (i.e., incongruent trials). A positive attention bias (AB) score (congruent < incongruent RT) reflects a bias toward threat and a negative AB score (congruent > incongruent RT) reflects a bias away from threat. Notably, however, two separable attentional bias processes that may have unique relationships with anxiety, orientation toward and disengagement from threat (Cisler & Koster, 2010; Evans, Walukevich, & Britton, 2016), cannot be readily disentangled using this approach (for discussion, see Evans & Britton, 2018). Orientation refers to the relative speed at which attention is drawn toward a threatening stimulus and disengagement refers to the duration that attention is captured by a threatening stimulus, impairing the ability to switch attention toward another stimulus (Cisler & Koster, 2010). For example, individuals who exhibit *both* fast orientation toward and difficulty disengaging from threat cannot be differentiated from individuals who *only* exhibit difficulty disengaging from threat once their attention has been captured, as both individuals will have a positive AB score (see Evans & Britton, 2018 for further discussion). Collapsing across attention components may contribute to the poor convergent and questionable criterion validity of the traditional dot-probe AB computation.

In an effort to better capture and separate orientation and disengagement biases, researchers have included a neutral condition in the dot-probe, in which a pair of two identical neutral stimuli are presented simultaneously and the dot-probe replaces one of the neutral faces. The orientation score is defined as the difference between the average RT on the neutral trials and the average RT on the congruent trials, and disengagement is defined as the difference in the average RT on incongruent and the average RT on neutral trials. Although this approach attempts to better capture distinct attentional bias components, studies using this method also evidences poor psychometric properties (Schmukle, 2005; Staugaard, 2009).

A second potential reason for the questionable psychometric properties of the dot-probe task is that the expression of biases by an individual may vary *during* a task (Evans & Britton, 2018; Iacoviello et al., 2014; Zvielli, Bernstein, & Koster, 2015). While attentional bias processes may be expected to be engaged in a generally consistent (i.e., reliable) manner across a task, attention processes are dynamic and can shift over time (Eysenck, Santos, & Calvo, 2007). Notably, there is evidence that the consistency with which attention-related processes are exhibited across a task differs by process. For example, Evans & Britton (2018) found that some response-based scores were less consistent than others. Because individuals may exhibit automatic orienting toward threat to a different degree than avoidance of threat over time and across repeated tests, combining processes into a single score can yield less reliable results. For example, an individual may be slow to orient toward threat during one trial, but quick to orient to threat on the next trial because of the dynamic nature of attention allocation and control (Eysenck, Santos, & Calvo, 2007). Furthermore, this intra-individual variation may be clinically meaningful. An individual who is fast to orient to threat may exhibit heightened and automatic detection of threats, while an individual with slow orientation to threat may be avoidant of the negative affect elicited by the threat and deviate their attention from the threat (Cisler & Koster, 2010; Evans et al., 2020). Relatedly, individuals with certain anxiety phenotypes exhibit differential patterns of attention-related task engagement to threat, with reduced early attention followed by increased later attention (Spielberg et al., 2013). In short, while traditional bias scores were designed to capture the general tendency to exhibit an attentional bias, averaging RT across trials can obscure attentional bias processes that vary in meaningful ways with anxiety symptoms across time (see Evans & Britton, 2018 for similar arguments).

A third possible explanation for the questionable validity of the dot-probe task is that studies have used different task parameters that may tap different attentional bias and underlying processes (Shechner et al., 2012). For example, many studies examine bias to emotional stimuli with long presentations (>450 ms; Schmukle, 2005; Staugaard, 2009), while other studies use briefer, potentially subliminal stimulus presentations (e.g. stimuli presented <100 ms that are then masked; Fox, 2002; Mogg & Bradley, 1999). Longer task presentations may confound *automatic* processing of threat, which characterizes certain anxiety phenotypes (Bar-Haim et al., 2007), with more elaborative or later cognitive processes that may play a different role in the experience of anxiety. Supporting this idea, Bar-Haim et al. (2007) found that subliminal exposures were related to a significantly larger effect of anxiety on attentional bias (subliminal, <500 msec: d=0.65 versus supraliminal, >500 msec: d=0.31). Indeed, some attention researchers have argued that in order to assess where attention is automatically drawn, stimuli must be presented rapidly to reduce potential interference by other information (e.g., distractors) that may redirect or alter attentional deployment (Lavie, 1995). Although the use of longer stimulus presentations is not an invalid approach to studying attention processes in anxiety, confounding cognitive processes that emerge during longer durations may introduce greater variation in the degree to which individuals engage different attention-related processes over time (e.g., automatic orientation versus avoidance).

In an effort to enhance the validity and reliability of the dot-probe task measures, Evans and Britton (2018) developed and tested a novel scoring method based on each *individual's specific responses* to particular trials – that is, scores are binned and computed as a function of participants' trial-level responses (see also Evans et al., 2020). The response-based approach requires as many trials as traditional scores, thereby keeping task length and participant burden to a minimum, but provides greater specificity and captures meaningful performance variability better than the traditional approach by decomposing the three traditional attentional bias components (overall attentional bias, orientation, disengagement) into separate measures (though see Kruijt, Field, & Fox, 2016). Rather than collapsing across trials for each condition, as is the case with the traditional approach, the RT for each trial is compared to a reference condition. For example, the RT to each congruent trial can be subtracted from the mean of all of the neutral trials. Congruent trials that are faster the mean of neutral trials (i.e., "fast orientation trials," or Fast $Orient_{RB}$) reflect rapid, automatic detections of threat (Evans et al., 2020). Congruent trials that are slower than the mean of neutral trials (i.e., "slow orientation trials," or Slow $Orient_{RB}$) reflect greater regulation of the negative affect elicited by threat and/or avoidance of threat (Cisler & Koster, 2010; see Evans et al., 2020 and Supplementary Table 1 for interpretations of the different metrics). In a large college sample with elevated levels of social anxiety, all of the response-based scores reached acceptable levels of internal consistency (>.70) and low to moderate test-retest reliability, while traditional AB scores did not. This response-based approach also yielded stronger associations with state anxiety than traditional scoring metrics, independent of RT variability (Evans & Britton, 2018). Additionally, in a recent report, the response-based but not traditional metrics dissociated different neural circuitry (Evans et al., 2020).

Building on previous work, the aim of the present series of studies was to assess the psychometric properties of response-based measures of attentional bias from the *masked* dot-probe, which targets automatic threat processing (Lavie, 1995), in three independent samples, including a large community sample. In Study 1, we examined the internal consistency of the traditional and response-based AB measures. In Study 2, we examined both internal consistency and test-retest reliability of the traditional and response-based AB measures. In Study 3, we examined the internal consistency of the response-based approach and its validity in two ways. Because anxiety disorders (Hettema, Neale, & Kendler, 2001) and mechanisms of anxiety disorders (Gorka et al., 2016) aggregate in families, we investigated the correlation of response-based AB among siblings, which would support the hypothesis that attentional bias is familial, a necessary requisite for determining a vulnerability marker for anxiety (Zubin & Spring, 1977). We also tested the validity of response-based approaches by examining their associations with social anxiety disorder, a condition that is often associated with attentional bias to threatening faces (Bantin, Stevens, Gerlach, & Hermann, 2016; Mansell et al., 1999). Importantly, we also assessed the incremental validity of response-based scores over and above traditional AB measures. Based on preliminary psychometric findings using response-based approaches (Evans & Britton, 2018), we hypothesized that response-based AB measures would demonstrate superior reliability as well as stronger social anxiety-related associations than traditional AB measures.

Method

Study 1

Participants—Study 1 (N = 30) included undergraduate participants who participated in a larger psychophysiological study of emotional and social processing. Exclusion criteria included inability to read/write English, left-handedness, and history of head trauma. Procedures were approved by University of Illinois at Chicago's Institutional Review Board. After providing consent, participants completed the study requirements, including questionnaires, electroencephalogram (EEG) set up and social gambling task, followed by the dot-probe task (the latter of which is included in the present analyses). Table 1 includes demographic information for each of the samples.

Measures

Dot-Probe—Each trial began with a 1-s, centered fixation, followed by two faces of the same person presented simultaneously and briefly (33-ms) to the left and right of the fixation[1]. In the mixed emotion trials, one face was threatening (angry) and one was neutral. Angry faces were used as threatening stimuli given prior attentional bias studies (Mogg, Philippot, & Bradley, 2004). The location of the threatening face was counterbalanced. In neutral trials, both faces were of the same model making a neutral facial expression. The faces then disappeared and were replaced with a mask (100-ms) of two happy faces of the same person as the ones in the subliminally presented photos. Happy face masks were selected, as opposed to neutral masks, because if the latter were used following the presentation of threatening and neutral face pairing, the subject would perceive a change in the threatening side of the screen but not the neutral side (as a neutral face would be replaced

[1]This Stimulus Onset Asynchrony was established from previous research (20–50 ms; Egloff & Hock, 2003; Mathews, Ridgeway, & Williamson, 1996) and piloted before data collection to ensure that the images disappeared before conscious awareness.

with the same neutral face), confounding the experimental conditions. There were an equal number of male and female faces, and faces with open and closed mouths. Faces were from the NimStim databank (see Tottenham et al., 2009). When creating the NimStim databank, participants viewed each stimulus and selected which emotion was being expressed. To ensure emotional valence of stimuli for this version of dot-probe, threatening and neutral faces (and happy masks) were selected if they had greater than 60% agreement within and between raters. After the happy face mask, a dot was immediately presented in the location of where the left or right face were and the reaction time (RT) of participants' detection of the dot's location was recorded. Participants were instructed to press a button corresponding to the side of the screen on which the dot appeared as quickly and accurately as possible. Face pairings established three conditions of RT: congruent (threatening vs. neutral, with dot replacing threatening), incongruent (threatening vs. neutral, with dot replacing neutral), and neutral (neutral vs. neutral, with dot replacing one of the neutral faces). Twenty-four trials of each condition were presented across two blocks, resulting in a total of 72 trials.

Data Analytic Approach

As per standard practice, all incorrect trials (i.e., when the subject pressed the button on the wrong side of where the dot was) were discarded (2.3%). To account for outliers in dot-probe, RT values outside 2.5 standard deviations from each individual's average RT for each trial type 1.5 interquartile ranges were winsorized (see Price et al., (2015) for a similar approach). A Condition (congruent, incongruent, neutral) x Context (safe, threat) repeated measures ANOVA was conducted on RT to investigate the effects of threating context on RT and confirm differences across conditions. [2]

Traditional AB Computation Approach—Average condition scores were calculated for the three conditions. The following traditional AB scores (Cisler & Koster, 2010; MacLeod et al., 1986) were calculated: (1) *Attention Bias (AB)*$_{Trad}$ (average incongruent RT - average congruent RT), reflecting attentional vigilance toward (positive scores) and attentional avoidance away (negative scores) from the emotional face; (2) *Orientation*$_{Trad}$ (average neutral RT - average congruent RT), reflecting orientation to threat; (3) *Disengagement*$_{Trad}$ (average incongruent RT - average neutral RT), reflecting disengagement from threat.

Response-based Approach—Unlike traditional computation methods, a *response-based* computation method (Evans & Britton, 2018) separately compares the response on each trial against a participant's mean reference reaction time (RT). For instance, RT from each congruent trial is *individually* indexed against the participant's *mean* RT of incongruent trials as a reference (i.e., mean Incongruent – [Trial 1 congruent RT…Trial 2 congruent RT…Trial 3 congruent RT]) to obtain a distribution of *Fast AB*$_{RB}$ (i.e., RTs that are faster on congruent trials than incongruent mean, difference in RT > 0ms) and *Slow AB*$_{RB}$ (i.e., RTs that are slower on congruent trials than incongruent mean, difference in RT

[2]We explored whether AB would vary under 'threat' and 'safe' contexts (order of safe and threat were counterbalanced). During the threatening context, participants heard random presentations of a woman screaming or nails scraping on slate (see Neumann, Waters, & Westbury, 2008; who used these sounds as unconditioned stimuli in Pavlovian conditioning). The traditional and response-based measures of AB were comparable under threat vs. safe contexts (all p's > 0.16) as were the reliabilities. Thus, for ease of presentation of results and to increase power, all analyses combined trials from threat and safe contexts.

< 0ms). Orientation$_{Trad}$ was similarly separated in to *Fast Orient$_{RB}$* and *Slow Orient$_{RB}$* using the mean RT of the neutral trials as a reference (i.e. mean neutral - [Trial 1 congruent RT…Trial 2 congruent RT…Trial 3 congruent RT]). Finally, Disengagement$_{Trad}$ was separated in to *Fast Disengage$_{RB}$* and *Slow Disengage$_{RB}$* (i.e. [Trial 1 incongruent RT…Trial 2 incongruent RT…Trial 3 incongruent RT] - mean neutral). Individual response-based scores were subsequently averaged *within* response-based conditions to create six separate measures (Fast AB$_{RB}$, Slow AB$_{RB}$, Fast Orient$_{RB}$, Fast Orient$_{RB}$, Fast Disengage$_{RB}$, Slow Disengage$_{RB}$). A summary of the score calculations is located in Supplemental Table 1, as well as interpretation of each of the metrics.

As a result of averaging scores based on participant response patterns, each index is based on a unique number of trials. After collapsing across context given the null effects for context (see footnote above), there were the following average number of trials for each of the six response-based metrics: 28.3 (*sd* = 4.8, range = 20–38) trials for Fast AB$_{RB}$, 19.2 (*sd* = 4.6, range = 10–28) for Slow AB$_{RB}$, 29.5 (*sd* = 4.9, range = 20–40) for Fast Orient$_{RB}$, 18.0 (*sd* = 4.7, range = 8–26) for Slow Orient$_{RB}$, 27.9 (*sd* = 4.7, range = 20–37) for Fast Disengage$_{RB}$, and 19.3 (*sd* = 4.5, range = 11–28) for Slow Disengage$_{RB.}$ See Supplemental Table S2A for correlations between traditional and response-based scores.

Reliability Analyses—Internal consistency for both traditional and response-based attentional bias scores was assessed using Cronbach's alpha and split-half reliability with Spearman-Brown correlations. To assess split-half reliability, trials within each condition were separated into odd or even numbered trials; then each half was correlated with the other and corrected for length using the Spearman-Brown formula (see Kappenman et al., 2014; Schmukle, 2005 for similar approaches). Shrout's (1998) conventions were used to describe reliability values - virtually none: (0–0.10), slight: (0.11–0.40), fair: (0.41–0.60), moderate: (0.61–0.80), and substantial: (0.81–1.0).

Results

Task Effect

Table 1 contains means and standard deviations for RTs, traditional, and response-based scores. A Condition (congruent, incongruent, neutral) repeated measures ANOVA indicated a main effect of Condition, $F(2, 28) = 8.65$, $\hat{\eta}^2 = 0.23$, $p < .001$. Follow-up pairwise comparisons indicated that RTs for congruent were significantly faster than incongruent ($F(1, 29) = 3.15$, $p = .01$) and neutral conditions ($F(1, 29) = 3.74$, $p = .002$). The incongruent condition did not significantly differ from neutral ($F(1, 29) = 1.33$, $p = .580$). All response-based scores were greater than zero (p's <.001).

Internal Consistency

Table 2, Study 1 presents the Cronbach's alpha and split half reliability of traditional and response-based analyses. While the traditional measures yielded moderate to fair internal consistency, the response-based measures yielded moderate to substantial internal consistency. However, none of the traditional scores reached levels of internal consistency necessary for clinical use (Cronbach's alpha >.9; Nunnally & Bernstein, 1994). In comparison, three of the response-based scores reached that cut off.

Discussion

We investigated the internal consistency of traditional and response-based AB scores computed from the dot-probe task in which masked emotional faces were presented in an undergraduate sample. Using two rigorous measures of internal consistency, we found that traditional AB scores did not reach an acceptable level of reliability for clinical tools, whereas three of the response-based scores did. The weakest score of the response-based approach was the Slow Orient_{RB} score (although its Cronbach's alpha and split-half reliability was still higher than the AB_{Trad} and Disengage_{Trad} score). A Slow Orient_{RB} response indicates that the response to a congruent trial was slower than the neutral mean. The reliability for the Slow Orient_{RB} index might have had the lowest reliability because it had fewer trials (M=18, sd=4.7) than any of the other scores. Fast AB_{RB} and Fast Orient_{RB} had the strongest internal consistency, perhaps because they had the highest number of trials included (i.e. on average 28.3 and 29.5, respectively).

Given that reliability must be established before a measure can be used clinically (see Rodebaugh et al., 2016), these results suggest that the reliability of some of the response-based scores fall in the range of clinical utility while the traditional scores do not. Once reliability within a measure is established, retest reliability of the measure over time must be assessed. In Study 2, we therefore investigated whether response-based scores demonstrated stability over traditional AB scores and sought to replicate the internal consistency effects from Study 1.

Study 2

Participants—Study 2 (N = 44) was a sample of undergraduates independent of the one in Study 1 who completed the dot-probe task two times, an average of 7.43 (*sd* = 3.36, range = 5–21) days apart. Six participants were excluded (two each for equipment error, missing questionnaires and only having one assessment) yielding a sample of 38. Procedures were approved by University of Illinois at Chicago's Institutional Review Board. After signing consent, participants completed the study requirements, including questionnaires, another laboratory task, and the dot-probe task. Demographics are in Table 1.

Measures

Dot-Probe—The dot-probe task in Study 2 was identical to the 'safe' context in Study 1, with the exception that there were 40 trials per condition (neutral, congruent, and incongruent).

Data Analytic Approach

The computations of the traditional and response-based AB scores, and analyses of internal consistency (Cronbach's alpha and Spearman-Brown) were similar to Study 1's procedures. To assess test-retest reliability, we used Pearson's correlations scores from Time 1 and Time 2 and used Cohen's (1988) conventions for size of effect (small = .1, medium = .3, large =

.5). At Time 1, on average 24.1 (*sd* = 3.9 range = 17–34) trials were included in the Fast AB_{RB}, 15.8 (*sd* = 4.0, range = 6–23) in Slow AB_{RB}, 24.8 (*sd* = 3.7, range = 17–31) in Fast $Orient_{RB}$, 15.0 (*sd* = 3.9, range = 8–23) in Slow $Orient_{RB}$, 16.5 (*sd* = 4.3, range = 6–27) in Fast $Disengage_{RB}$, and 22.8 (*sd* = 4.1, range = 13–32) in Slow $Disengage_{RB.}$ At Time 2, on average 23.0 (*sd* = 5.7, range = 13–34) trials were included in the Fast AB_{RB}, 17.1 (*sd* = 5.7, range = 6–27) in Slow AB_{RB}, 23.3 (*sd* = 5.7, range = 13–33) in Fast $Orient_{RB}$, 16.7 (*sd* = 5.7, range = 7–27) in Slow $Orient_{RB}$, 16.6 (*sd* = 4.7, range = 7–26) in Fast $Disengag_{RB}$, and 22.4 (*sd* = 5.2, range = 9–31) in Slow $Disengage_{RB}$. See Supplemental Table S2B and C for correlations between traditional and response-based scores.

Results

Task Effect

Table 1 contains means and standard deviations for RTs, traditional, and response-based scores. A Condition (congruent, incongruent, neutral) x Time (Time 1, Time 2) repeated measures ANOVA indicated a main effect of Condition ($F(2, 36) = 4.37$, $\hat{\eta}^2 = 0.10$, $p = .02$) and Time ($F(1, 37) = 19.00$, $\hat{\eta}^2 = 0.34$, $p < .001$) but no Condition by Time interaction ($F(2, 36) = 0.72$, $\hat{\eta}^2 = 0.01$, $p = .48$). For the main effect of Condition, pairwise comparisons indicated that RTs for congruent were significantly faster than neutral ($F(1, 37) = 2.67$, $p = 0.03$), but not incongruent conditions ($p > .05$). The incongruent and neutral condition did not significantly differ ($p > .05$). For the main effect of Time, pairwise comparisons indicated that RTs for Time 2 were significantly faster than for Time 1 ($F(1, 37) = 4.36$, $p < .01$).

For both the time points, one-sample t-tests indicated that all response-based scores were greater than zero (*p*'s <.001). Paired sample t-tests indicated that the response-based scores did not change between Time 1 and Time 2 (*p*'s > .05).

Internal Consistency—Table 2 presents the Cronbach's alpha and split half reliability of traditional and response-based analyses for Study 2. Similar to Study 1, while the traditional measures yielded slight to moderate internal consistency, the response-based measures yielded moderate to substantial internal consistencies for both time points. None of the traditional measures reached a level of internal consistency acceptable for clinical use, while four of the response-based scores did at Time 1, and three at Time 2.

Test-Retest Reliability—Table 3 presents the test-retest reliability of traditional and response-based analyses after averaging the winsorized scores. While a significant but medium correlation for the traditional calculation for orientation ($Orientation_{Trad}$) was observed ($r = 0.35$, $p < .05$), the other two traditional measures did not demonstrate significant test-rest reliability ($p > .17$). Overall, response-based scores demonstrated superior test-retest reliability to traditional AB scores ($r > 0.44$, $p < .01$), with the exception of the Fast $Disengage_{RB}$ score ($p > .12$).

Discussion

Study 2 investigated the internal consistency and test-retest reliability of traditional and response-based AB scores computed from the dot-probe task in which masked emotional faces were presented in an undergraduate sample. At the two time points, we replicated our internal consistency results from Study 1 such that traditional AB scores did not reach an acceptable level of internal consistency, while the majority of response-based scores did. Again, the weakest internal consistency of the scores of the response-based approach were the Slow AB_{RB} and Slow $Orient_{RB}$ scores and the strongest of the response-based approach were the Fast AB_{RB} and Fast $Orient_{RB}$ scores. This may be because there were fewer participants in this sample who had a large number of trials in which they avoided the threatening stimulus. Studies examining individuals with attentional avoidance patterns may aid to elucidate whether the worse reliability for Slow AB_{RB} and Slow $Orient_{RB}$ was due to the number of trials.

The size of the effect of test-retest reliability for three traditional AB measures - AB_{Trad}, $Orientation_{Trad}$, and $Disengagement_{Trad}$ - were small-to-medium, medium, and very small, respectively. By contrast, 5 of the six response based measures had higher correlations than the highest traditional measure and were of either large effect (Fast AB_{RB}, Slow $AB_{RB,}$ Fast $Orient_{RB,}$ and Slow $Orienti_{B}$) or medium (Slow $Disengage_{RB}$).

Study 1 and Study 2 support the hypothesis that splitting the traditional metrics based on how the individual’s responses more reliably captures variability in attentional bias. Reliability is necessary but not sufficient for a measurement scale to be validated. To test the validity of the response-based scores, in Study 3 we investigated in a community sample (a) the associations between response-based scores and Social Anxiety Disorder diagnoses and (b) associations between scores within siblings.

Study 3

Participants—Study 3 was a family study on transdiagnostic factors of internalizing psychopathology (for additional details see Gorka et al., 2016; Shankman, Lerner, Funkhouser, Klein, & Davila, 2018; Weinberg, Liu, Hajcak, & Shankman, 2016). Study 3 consists of a sample of sibling pairs (total N = 481), ages 18–30, which allowed the comparison of bias and symptomology between siblings (see Table 1 for demographic information). Participants were recruited through advertisement from the community and mental health clinics. A Research Domain Criteria (RDoC) approach was taken to participant recruitment such that recruitment screening was agnostic to Diagnostic and Statistical Manual of Mental Disorders (DSM) diagnostic categories (beyond the exclusion criteria listed below). However, participants with severe internalizing psychopathology were oversampled to ensure that the sample was clinically relevant. Specifically, the Depression, Anxiety, and Stress Scale (DASS; Lovibond & Lovibond, 1995) was administered during the initial phone screen to ensure that the severity of internalizing symptomology within the sample was normally distributed, but also had a higher average symptoms (M = 23.6, SD = 20.45) than the general population (M = 8.3, sd = 9.83; Crawford, Cayley, Lovibond, Wilson, & Hartley, 2011). To be included in the study, participants had to have one full-biological sibling that was also willing to participate. Study 3 also excluded individuals with personal or family histories of psychosis or mania, inabilities to read or write English, histories of serious head trauma, and left-handedness. Procedures were approved by University of Illinois at Chicago's Institutional Review Board. After signing consent, participants completed the study requirements, including questionnaires and clinical interviews, then the dot-probe task.

Measures

Dot-Probe—In Study 3, the dot-probe task was identical to Study 1 (i.e., three conditions [neutral, congruent, incongruent], but the number of trials per condition were unequal due to a computer programming error (Congruent = 48 trials, Incongruent = 58 trials, Neutral = 38 trials). To ensure an equal number of trials included per conditions, 38 trials of each condition were randomly selected for inclusion in the following analyses, resulting in 114 total trials (79% of trials maintained).

Structured Clinical Interview for DSM 5 (SCID-5; First, Williams, Karg, & Spitzer, 2015).

As part of the larger study, participants in Study 3 were assessed for Lifetime and Current Social Anxiety Disorder with the Social Anxiety module of the SCID-5. Although the present study focuses on Social Anxiety Disorder, for descriptive purposes, Table 1 presents the rates of other psychopathologies in the sample, as assessed by the SCID-5.

Data Analytic Approach—The computations of the traditional and response-based scores, and analyses of internal consistency (Cronbach's alpha and Spearman-Brown) were identical to Study 1. However, the sample was randomly split in half to avoid confounding sibling effects and to allow for a further replication of internal consistency effects in each of the sibling groups. Similar to Studies 1 and 2, there were on average 22.0 (*sd* = 4.0, range = 10–32) trials included in the Fast AB_{RB}, 15.7 (*sd* = 4.0, range = 6–28) in Slow AB_{RB}, 22.3 (*sd* = 4.0, range = 11–34) in Fast $Orient_{RB}$, 15.3 (*sd* = 4.0, range = 4–27) in Slow $Orient_{RB}$, 21.0 (*sd* = 4.3, range = 9–32) in Fast $Disengage_{RB}$, and 16.6 (*sd* = 4.2, range = 2–29) in Slow $Disengage_{RB}$. See Supplemental Table S2 for correlations between traditional and response-based scores.

Validity analyses included family concordance and prediction of Social Anxiety Disorder diagnosis. Family concordance of the traditional and response-based approaches were assessed using one-way random, single measure intraclass correlation coefficients (ICCs). For the response-based scores, we sought to examine whether relationships were evident beyond the contribution of traditional scores by identifying whether there was overlap between the confidence intervals of each score (McGraw & Wong, 1996). The association between social anxiety diagnosis and the traditional and response-based scores were examined using mixed logistic regression models. Mixed effects logistic regression models can be used to model binary outcome variables (i.e. Social Anxiety Diagnosis) as a linear combination of continuous predictor variables (i.e. attentional bias scores) when there is shared variance between siblings from the same family. Siblings were nested within a higher order family factor. First, each of the three traditional AB scores and six response-based scores were entered into separate models predicting Lifetime and Current Social Anxiety (present or absent) as assessed with the SCID-5. Second, to assess for the incremental validity of the response-based scores, the response-based models were re-run controlling for the effects of traditional scores (AB_{Trad} for Fast and Slow AB_{RB}, $Orient_{Trad}$ for Fast and Slow $Orient_{RB}$, and $Disengage_{Trad}$ for Fast and Slow $Disengage_{RB}$) and gender.

Results

Task Effect

Replicating the results from Study 1, a Condition (congruent, incongruent, neutral) x Context (threat, safe) intra-individual mixed effects model indicated a main effect of Condition ($\beta = 2.73$, SE = 1.38, $p = 0.05$).

Internal Consistency

Table 2, Study 3 presents the Cronbach's alpha and split half reliability of traditional and response-based analyses in each sibling. Similar to Study 1 and 2, while the traditional measures yielded moderate internal consistency, the response-based measures yielded predominantly substantial internal consistency in both siblings (almost double that of the traditional measures) and reached acceptable levels for clinical use for half of the scores.

Anxiety-related Associations

Ninety-six people (19.96%) met DSM-5 criteria for Lifetime Social Anxiety and fifty-one people (10.60%) met DSM-5 criteria for Current Social Anxiety. Mixed model logistic regressions with each type of bias score predicting anxiety membership revealed that AB_{Trad} significantly predicted likelihood of having a Current Social Anxiety Diagnosis (*OR* = 2.04 [95% CI: 1.13, 3.67]), and Lifetime Social Anxiety Diagnosis (*OR* = 1.37 [95% CI: 1.03, 1.82]). $Orientation_{Trad}$ and $Disengagement_{Trad}$ scores did not predict Current nor Lifetime Social Anxiety Diagnoses (*p*'s > 0.23). Regarding the univariate models for the response-based AB scores, Fast AB_{RB} and Slow $Disengage_{RB}$ significantly predicted Current Social Anxiety Diagnosis (*OR* = 1.49 [95% CI: 1.06, 2.10] and *OR* = 1.61 [95% CI: 1.14, 2.26], respectively). However, only Slow $Disengage_{RB}$ significantly predicted Lifetime Social Anxiety Diagnosis (*OR* = 1.42 [95% CI: 1.08, 1.87]). See left-half of Table 4.

To test the incremental validity of the response-based scores over traditional AB score, the right half of Table 4 contains the results of the mixed model logistic regressions controlling for gender and AB_{Trad}. Interestingly, for these incremental models, Fast AB_{RB}, Slow $Disengage_{RB}$ and Fast $Disengage_{RB}$ predicted Current and/or Lifetime Social Anxiety Diagnosis after adjusting for gender and traditional scores (AB_{Trad} for Fast and Slow AB_{RB}, $Orientation_{Trad}$ for Fast and Slow $Orientat_{RB}$, and $Disengagement_{Trad}$ for Fast and Slow $Disengagement_{RB;}$ see Table 4 for OR statistics). Using McFadden's pseudo R^2 (Mittlböck & Schemper, 1996) to examine the incremental validity of Slow $AB_{RB,}$ Fast $Disengage_{RB,}$ and Slow $Disengage_{RB}$, Slow AB_{RB} and Fast $Disengage_{RB}$ explained 1% and Slow $Disengage_{RB}$ explained 2% more of the variance in the current Social Anxiety model compared to the model with just the traditional scores and gender included. For lifetime Social Anxiety Diagnosis, Slow AB_{RB} and Slow $Disengage_{RB}$ each explained 1% more of the variance in the model compared to the model with the traditional scores and gender included. In comparison, the AB_{Trad} and gender model explained 1% more of the variance compared to an intercept only model for Current Social Anxiety Diagnosis, and 2% more for Lifetime. Thus, in both cases, the response-based scores nearly doubled the predictive value of the model, albeit modestly.

Family Concordance

One-way single measure ICC's between sibling 1 and sibling 2's bias scores indicated no relationship for the traditional measures. As shown in Table 5, confidence intervals (CIs) were computed for each of the traditional and response-based scores and were compared to identify whether there was overlap between the intervals (McGraw & Wong, 1996). There was no overlap between the AB_{Trad} and Fast AB_{RB}, indicating that Fast AB_{RB} exhibits greater familiality than the traditional AB score. Additionally, there was no overlap between the $Disengagement_{Trad}$ and either Fast $Disengage_{RB}$ or Slow $Disengage_{RB}$, indicating that both disengagement-based RB scores exhibit greater familiality than the traditional disengagement score. Using an alternative approach to compare the ICCs, we used 1,000 bootstrap replications to construct 95% CIs around ICC difference scores (traditional minus response-based). This approach yielded comparable results, whereby the upper boundary of the CI of the difference between the traditional and response-based ICC was less than zero for AB_{Trad} minus Fast AB_{RB} [−.41, −.08], $Disengagement_{Trad}$ minus Fast $Disengage_{RB}$ [−.48, −.06], and $Disengagement_{Trad}$ minus Slow $Disengage_{RB}$ [−.44, −.14].

Discussion

In Study 3, we replicated the results of Study 1 and 2 that response-based scores demonstrated better internal consistency than traditional bias scores gleaned from the dot-probe task in two samples of siblings. In both siblings, we found that the levels of internal consistency of traditional AB scores were not acceptable, while the majority of response-based scores were in the acceptable range or higher for internal consistency. Unlike Study 1 and Time 2 of Study 2, we did not find that the Cronbach's alphas of Slow AB_{RB} and Slow $Orient_{RB}$ scores were less reliable than the other response-based scores. This may be attributed to the fact that there were major differences in this sample, namely it was nearly four times the size of Study 1 and Study 2. Further, Study 3 was not an unselected undergraduate sample and was recruited from the community to contain elevated levels of psychopathology. As such, we anticipated higher variability in the patterns of responding to the dot-probe (e.g., there were on average a higher number of Slow $Orient_{RB}$ responses, improving the reliability of this score).

Importantly, the Slow AB_{RB}, Slow $Disengage_{RB}$ and Fast $Disengage_{RB}$ score had incremental validity for predicting Current and/or Lifetime Social Anxiety Diagnoses, independent of the effects of gender and traditional measure of AB (AB_{Trad}). Finally, Study 3 also investigated the concordance of bias scores within sibling pairs. Interestingly, we found no familial associations for the three traditional AB scores, but all six RB scores demonstrated significant associations. Furthermore, the Fast AB_{RB}, Fast $Disenge_{RB}$e, and Slow $Disengage_{RB}$ scores demonstrated greater familiality than the traditional scores.

General Discussion

Overall, in three separate studies, the present report demonstrated that a response-based approach exhibited stronger psychometric properties than traditional attentional-bias calculations. Specifically, compared to traditional bias scores, several of the response-based scores demonstrated (a) higher internal consistency in five sets of analyses – Study 1, Study 2 (at two time points) and Study 3 (for each sibling separately); (b) higher test-retest reliability; (c) significantly stronger relationships with Social Anxiety Disorder (although this was only evidenced for the Fast AB_{RB}, Slow $Disengage_{RB,}$ and Fast $Disengage_{RB}$ scores); and (d) significant (albeit small) familial associations. These results replicate and extend a prior study that found superior psychometric properties for a response-based approach in capturing attentional-bias to *overtly* presented threats (Evans & Britton, 2018), to attentional bias to brief, *masked* presentations of threats.

There are several reasons that a response-based approach to quantifying attentional bias may have exhibited better reliability than traditional approaches. Response-based methods assess the multifaceted nature of attentional bias as they compare the pattern of responding for each trial to the individuals' own average responding. Further, they capture the dynamic nature of threat-related attention, even while maintaining a low participant burden with the relatively few number of trials included. The low levels of internal consistency of traditional measures raises concerns about their clinical utility and the improved internal consistency and retest reliability may be a step towards a better prediction tool.

In all three of the present studies, we used brief (33 ms), masked presentations of the stimuli to target early attentional biases, which are thought to relate more to anxiety disorders than depression (MacLeod et al., 1986). All three samples demonstrated a similar pattern of results despite having slightly different trial lengths, contexts, and stimulus sizes. Moreover, our results are consistent with previous work using overt faces (Evans & Britton, 2018). Taken together, the convergence of results across studies with varying parameters and methods highlights the robustness of this computational approach to measuring attentional bias. Future research should replicate these results in various popular versions of the task (e.g., with fearful or disgusted facial expressions as threats, words as stimuli, or various stimulus presentation duration), and using other measures of attentional bias, such as eye-tracking and event-related potentials.

These inter-individual results support several theoretical models of social anxiety (e.g., Clark & Wells, 1995), which posit that social anxiety is maintained in part by attention being biased towards threatening information in the environment. Although the AB_{Trad} score had the strongest univariate relationship with social anxiety in our sample, it also had substantially lower internal consistency and test retest reliability than response-based scores across three studies. Low internal consistency and test-retest reliability do not rule out the possibility that there is a true effect of AB_{Trad} on social anxiety, but instead mean that some studies will find an effect and some studies will not as has been the case with the dot-probe task thus far. If measures of attentional bias are to be reliably incorporated into the assessment and treatment of patients, it is necessary to ensure that scores are being measured reliably.

Only certain response-based scores were related to social anxiety, which indicates that attentional-biases are dynamic and can tap multiple components of attention that might be masked by traditional scoring. For example, the widely used AB_{Trad} cannot separate initial orientation to threat and disengagement from threat (although it should be noted that it was not intended to do so); but parsing these processes within individuals is important as they may reflect different psychological and neural processes and be differentially associated with psychopathologies (Evans et al., 2020). In the case of social anxiety, the present data suggest that the tendency to have a slower RT to incongruent trials relative to both an individual’s congruent (i.e., Slow AB_{RB}) and neutral (i.e., Slow $Disengage_{RB}$) averages may be an important predictor of current social anxiety disorder. Slow $Disengage_{RB}$ likely reflects elaborative processing of threats that may stem from poor regulatory attentional processes, while Slow AB_{RB} likely reflects an avoidance of threat bias (Cisler & Koster, 2010; Evans et al., 2020). In comparison, Evans and Britton (2018) did not find any anxiety associations using the Leibowitz Social Anxiety Scale and the State-trait Anxiety Inventory. The difference in paradigm or sample characteristics are both viable explanations for these results. Specifically, it is possible that the masked presentation of stimuli (compared to their overt version) increased the internal consistency and test-retest reliability in our sample to detect the relationship. Additionally, the differences in results could be due to our larger sample size or higher level of overall psychopathology in our sample.

Demonstrating that a disorder, or mechanism of a disorder, runs in families has long been viewed as an important indicator of its validity (Kendler, 2006; Klein, Shankman, Lewinsohn, Rohde, & Seeley, 2004; Robins & Guze, 1970). The traditional scores did not exhibit significant relationships within sibling pairs, but the response-based scores did (and three response-based scores exhibited greater familiality than the traditional score), albeit with small effects. It is not surprising that the effects for familial concordance were not as large as prior twin studies (which unlike family studies, can disambiguate the effects of genetics and environment) that have shown that anxiety is due to a large percentage of unique environmental variance (Hettema et al., 2001). Thus, as the adult siblings in this study likely experienced different environmental events, the effects for familial concordance were likely attenuated. However, the increased psychometric properties of the response-based score relative to traditional scoring will likely help to elucidate the familial (and ultimately genetic) relationship between attentional biases and anxiety.

Our study had several notable limitations. First, we used a community sample for Study 3 to assess the validity of response-based scores to predict social anxiety. Further, the individuals with social anxiety in Study 3 had several other DSM diagnoses and comorbidities (~45% of the sample had 2 or more diagnoses in their life, and ~13% of the sample had 2 or more diagnoses at the time of the interview). Future studies should replicate the validity findings in a more homogeneous sample of clinically socially anxious adults as well as in other populations. However, given the high comorbidity rates between social anxiety and other psychopathologies in the general population, examining samples that allow comorbidities is more ecologically valid (Shankman & Klein, 2003). Second, we did not find an effect for threatening context (shock vs. safe – see footnote 2). Although other studies have used aversive noises to manipulate defensive responding (Neumann & Waters, 2006), one possibility is that the sounds that were used were not aversive enough to create an enhanced state of anxiety. We did not measure state anxiety during each of the blocks and thus cannot confirm whether the manipulation was effective at affecting subjective anxiety. Finally, given that the dot-probe task is commonly used to assess attentional biases in anxiety (Bar-Haim et al., 2007), we selected this task to identify whether an alternative scoring approach improves reliability and validity. However, the dot-probe is limited in its ability to capture specific attention-related processes in a relatively "pure" manner. Indeed, compared to measures derived from other paradigms, it does not adequately separate engagement from disengagement processes and does not meet the criteria proposed by Clarke at al. (2013) as a valid measure of biased attentional engagement and disengagement.

A potential limitation related to computing the response-based scores is fact that we had unequal number of trial types in Study 3 resulting in data loss (i.e. only 38 trials per condition). However, the response based-scores still reached high levels of internal consistency, test-retest reliability, and validity. These results suggest that with more reliable calculation methods, fewer trials may be necessary to detect an effect. Relatedly, differences in the number of trials per response-based AB score across individuals may be considered a potential limitation of response-based approach; however, given that Evans and Britton (2018) found that most response-based AB scores exhibited acceptable reliability with very few trials (~6–8), there was more than adequate number of trials to calculate the response-based metrics across all studies given. More specifically, across all 3 of our studies, nearly all participants had a sufficient number of trials (i.e., enough trials to reach an acceptable level of reliability) for the computation of each response-based score, and the majority of participants had enough trials for good to excellent reliability for each response-based score. Finally, prior work raises some concerns about a trial-level, nearest-neighbor approach to AB calculation (Kruijt et al., 2016), and further work should be done to confirm that related concerns do not extend to the response-based scoring approach.

In summary, in three studies, the present study demonstrated improved reliability and validity of response-based scores over and above traditional bias scores, increasing the utility of the dot-probe task. These results suggest that the response-based scores offer a potential solution to the questionable psychometrics of the traditional dot-probe scoring metrics (Kappenman et al., 2014; Schmukle, 2005; Staugaard, 2009) and provide an important tool for researchers seeking to test cognitive models of attentional bias.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institute of Mental Health grant R01 MH098093 awarded to Dr. Shankman. Declarations of interest: None.

References

Amir N, Elias J, Klumpp H, & Przeworski A. (2003). Attentional bias to threat in social phobia: facilitated processing of threat or difficulty disengaging attention from threat? Behaviour Research and Therapy, 41, 1325–1335. 10.1016/S0005-7967(03)00039-1 [PubMed: 14527531]

Amir N, Prouvost C, & Kuckertz JM (2012). Lack of a benign interpretation bias in social anxiety disorder. Cognitive Behaviour Therapy, 41(2), 119–129. 10.1080/16506073.2012.662655 [PubMed: 22545788]

Bantin T, Stevens S, Gerlach AL, & Hermann C. (2016). What does the facial dot-probe task tell us about attentional processes in social anxiety? A systematic review. Journal of Behavior Therapy and Experimental Psychiatry, 50, 40–51. 10.1016/j.jbtep.2015.04.009 [PubMed: 26042381]

Bar-Haim Y, Lamy D, Pergamin L, Bakermans-Kranenburg MJ, & van IJzendoorn MH (2007). Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study. Psychological Bulletin, 133(1), 1–24. 10.1037/0033-2909.133.1.1 [PubMed: 17201568]

Cisler JM, & Koster EHW (2010). Mechanisms of attentional biases towards threat in anxiety disorder: An integrative review. Clinical Psychology Review, 30(2), 1–29. 10.1016/j.cpr.2009.11.003.Mechanisms [PubMed: 19751958]

Clark D, & Wells A. (1995). A Cognitive Model of Social Phonia. In Social Phobia (p. 69-). New York.

Clarke PJF, MacLeod C, & Guastella AJ (2013). Assessing the role of spatial engagement and disengagement of attention in anxiety-linked attentional bias: A critique of current paradigms and suggestions for future research directions. Anxiety, Stress & Coping, 26(1), 1–19.

Crawford J, Cayley C, Lovibond PF, Wilson PH, & Hartley C. (2011). Percentile Norms and Accompanying Interval Estimates from an Australian General Adult Population Sample for Self-Report Mood Scales (BAI, BDI, CRSD, CES-D, DASS, DASS-21, STAI-X, STAI-Y, SRDS, and SRAS). Australian Psychologist, 46(1), 3–14. 10.1111/j.1742-9544.2010.00003.x

Egloff B, & Hock M. (2003). Assessing attention allocation toward threat-related stimuli: a comparison of the emotional Stroop task and the attentional probe task. Personality and Individual Differences, 35, 475–483. Retrieved from http://www.sciencedirect.com.ezproxy.wesleyan.edu:7790/science/article/pii/S0191886902002659

Evans TC, & Britton JC (2018). Improving the psychometric properties of dot-probe attention measures using response-based computation. Journal of Behavior Therapy and Experimental Psychiatry, (March 2017), 0–1. 10.1016/j.jbtep.2018.01.009

Evans TC, Walukevich KA, & Britton JC (2016). Vigilance-avoidance and disengagement are differentially associated with fear and avoidant behaviors in social anxiety. Journal of Affective Disorders, 199, 124–131. 10.1016/j.jad.2016.04.003 [PubMed: 27131063]

Eysenck MW, Santos R, & Calvo MG (2007). Anxiety and Cognitive Performance : Attentional Control Theory, 7(2), 336–353. 10.1037/1528-3542.7.2.336

First MB, Williams JBW, Karg RS, & Spitzer RL (2015). Structured clinical interview for DSM-5: Research version (SCID-5 for DSM-5, research version; SCID-5-RV). American Psychiatric Association.

Fox E. (2002). Processing emotional facial expressions: the role of anxiety and awareness. Cognitive, Affective & Behavioral Neuroscience, 2(1), 52–63. 10.1080/02699930143000527

Gorka SM, Hee D, Lieberman L, Mittal VA, Phan KL, & Shankman SA (2016). Reactivity to uncertain threat as a familial vulnerability factor for alcohol use disorder. Psychological Medicine, 46(16), 3349–3358. 10.1017/S0033291716002415 [PubMed: 27624067]

Gotlib IH, Mclachlan AL, & Katz AN (1988). Biases in Visual Attention in Depressed and Nondepressed Individuals. Cognition and Emotion, 2(3), 185–200. Retrieved from https://www.researchgate.net/profile/Ian_Gotlib/publication/247496656_Biases_in_Visual_Attention_in_Depressed_and_Nondepressed_Individuals/links/00b495223b3dca46b2000000.pdf

Hakamata Y, Lissek S, Bar-Haim Y, Britton JC, Fox N, Leibenluft E, … Pine DS (2010). Attention Bias Modification Treatment: A meta-analysis towards the establishment of novel treatment for anxiety. Biol Psychiatry, 76(October 2009), 211–220. 10.1007/s11103-011-9767-z.Plastid

Hettema JM, Neale MC, & Kendler KS (2001). A Review and Meta-Analysis of the Genetic Epidemiology of Anxiety Disorders. American Journal of Psychiatry, 158(10), 1568–1578. 10.1176/appi.ajp.158.10.1568

Iacoviello BM, Wu G, Abend R, Murrough JW, Feder A, Fruchter E, … Charney DS (2014). Attention bias variability and symptoms of posttraumatic stress disorder. Journal of Traumatic Stress, 27, 232–239. 10.1002/jts. [PubMed: 24604631]

Kappenman ES, Farrens JL, Luck SJ, & Proudfit GH (2014). Behavioral and ERP measures of attentional bias to threat in the dot-probe task : poor reliability and lack of correlation with anxiety, 5(December), 1–9. 10.3389/fpsyg.2014.01368

Kappenman ES, Macnamara A, & Proudfit GH (2015). Electrocortical evidence for rapid allocation of attention to threat in the dot-probe task. Social Cognitive and Affective Neuroscience, 10, 577–583. 10.1093/scan/nsu098 [PubMed: 25062842]

Kendler KS (2006). Reflections on the Relationship Between Psychiatric Genetics and Psychiatric Nosology. American Journal of Psychiatry, (July), 1138–1146.

Klein DN, Shankman SA, Lewinsohn PM, Rohde P, & Seeley JR (2004). Family Study of Chronic Depression in a Community Sample of Young Adults. American Journal of Psychiatry, 161(4), 646–653. 10.1176/appi.ajp.161.4.646

Kruijt AW, Field AP, & Fox E. (2016). Capturing dynamics of biased attention: Are new attention variability measures the way forward?. PloS one, 11(11), e0166600.

Kruijt AW, Parsons S, & Fox E. (2019). A meta-analysis of bias at baseline in RCTs of attention bias modification: No evidence for dot-probe bias towards threat in clinical anxiety and PTSD. Journal of Abnormal Psychology, 128(6), 563–573. 10.1037/abn0000406 [PubMed: 31368735]

Lavie N. (1995). Perceptual Load as a Necessary Condition for Selective Attention. Journal of Experimental Psychology: Human Perception and Performance, 21(3), 451–468. 10.1037//0096-1523.21.3.451 [PubMed: 7790827]

Lovibond S, & Lovibond P. (1995). Manual for the depression anxiety stress scales. Sydney, Australia: The Psychology Foundation of Australia, Inc.

MacLeod C, & Hagan R. (1992). Individual differences in the selective processing of threatening information, and emotional responses to a stressful life event. Behaviour Research and Therapy, 30(2), 151–161. 10.1016/0005-7967(92)90138-7 [PubMed: 1567344]

MacLeod C, Mathews A, & Tata P. (1986). Attentional bias in emotional disorders. Journal of Abnormal Psychology, 95(1), 15–20. 10.1037/0021-843X.95.1.15 [PubMed: 3700842]

MacLeod C, Rutherford E, Campbell L, Ebsworthy G, & Holker L. (2002). Selective attention and emotional vulnerability: Assessing the causal basis of their association through the experimental manipulation of attentional bias. Journal of Abnormal Psychology, 111(1), 107–123. 10.1037/0021-843X.111.1.107 [PubMed: 11866165]

Mansell W, Clark DM, Ehlers A, Chen Y, Mansell W, Clark DM, … Social YC (1999). Social Anxiety and Attention away from Emotional Faces. Cognition and Emotion, 13(6), 673–690. 10.1080/026999399379032

Mathews A, Ridgeway V, & Williamson DA (1996). Evidence for attention to threatening stimuli in depression. Behaviour Research and Therapy, 34(9), 695–705. 10.1016/0005-7967(96)00046-0 [PubMed: 8936752]

McGraw KO, & Wong SP (1996). Forming inferences about some intraclass correlation coefficients. Psychological Methods, 1(1), 30–46.

Mittlböck M, & Schemper M. (1996). Explained variation for logistic regression. Statistics in Medicine, 15(19), 1987–1997. 10.1002/(SICI)1097-0258(19961015)15:19<1987::AID-SIM318>3.0.CO;2-9 [PubMed: 8896134]

Mogg K, & Bradley BP. (1999). Orienting of Attention to Threatening Facial Expressions Presented under Conditions of Restricted Awareness. Cognition and Emotion, 13(6), 713–740. 10.1080/026999399379050

Mogg K, Philippot P, & Bradley BP (2004). Selective attention to angry faces in clinical social phobia. Journal of Abnormal Psychology, 113(1), 160. [PubMed: 14992669]

National Advisory Mental Health Council Workgroup on Tasks and Measures for Research Domain Criteria. (2016). Behavioral assessment methods for RDoC constructs. Retrieved from https://www.nimh.nih.gov/about/advisory-boards-and-groups/namhc/reports/behavioral-assessment-methods-for-rdoc-constructs.shtml

Neumann DL, & Waters AM (2006). The use of an unpleasant sound as an unconditional stimulus in a human aversive Pavlovian conditioning procedure. Biological Psychology, 73, 175–185. 10.1016/j.biopsycho.2006.03.004 [PubMed: 16698165]

Price RB, Kuckertz JM, Siegle GJ, Ladouceur CD, Silk JS, Ryan ND, … Amir N. (2015). Empirical recommendations for improving the stability of the Dot-Probe Task in clinical research. Psychological Assessment, 27(2), 365–376. 10.1037/pas0000036 [PubMed: 25419646]

Robins E, & Guze SB (1970). Establishment of diagnostic validity in psychiatric illness: Its application to schizophrenia. American Journal of Psychiatry, 126(7).

Rodebaugh TL, Scullin RB, Langer JK, Dixon DJ, Huppert JD, Bernstein A, … Lenze EJ (2016). Unreliability as a Threat to Understanding Psychopathology: The Cautionary Tale of Attentional Bias. Journal of Abnormal Psychology, 125(6), 840–851. 10.1037/abn0000184 [PubMed: 27322741]

Schmukle SC (2005). Unreliability of the dot probe task. European Journal of Personality, 19(7), 595–605. 10.1002/per.554

Shankman SA, & Klein DN (2003). The relation between depression and anxiety: an evaluation of the tripartite, approach-withdrawal and valence-arousal models. Clinical Psychology Review, 23, 605–637. 10.1016/S0272-7358(03)00038-2 [PubMed: 12788112]

Shankman SA, Lerner D, Funkhouser CJ, Klein DN, & Davila J. (2018). Reliability and validity of severity dimensions of psychopathology assessed using the Structured Clinical Interview for DSM ‐ 5 (SCID). International Journal of Methods in Psychiatry Research, 1–12. 10.1002/mpr.1590

Shechner T, Britton JC, Pérez-edgar K, & Bar-haim Y. (2012). Attention biases, anxiety, and development: Toward or away from threats or rewards? Depress Anxiety, 29(4), 282–294. 10.1002/da.20914.Attention [PubMed: 22170764]

Spielberg JM, De Leon AA, Bredemeier K, Heller W, Engels AS, Warren SL, … Miller GA (2013). Anxiety type modulates immediate versus delayed engagement of attention‐related brain regions. Brain and Behavior, 3(5), 532–551. 10.1002/brb3.157 [PubMed: 24392275]

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

Tottenham N, Tanaka J, Leon A, McCarry T, Nurse M, Hare T, … Nelson C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. Psychiatry Research, 168(3), 242–249. 10.1016/j.psychres.2008.05.006.The [PubMed: 19564050]

Waechter S, Nelson AL, 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. 10.1007/s10608-013-9588-2

Weinberg A, Liu H, Hajcak G, & Shankman SA (2016). Blunted Neural Response to Rewards as a Vulnerability Factor for Depression: Results from a Family Study. Journal of Abnormal Psychology, 124(4), 878–889. 10.1037/abn0000081.Blunted

Zubin J, & Spring B. (1977). Vulnerability: A New View of Schizophrenia. Journal of Abnormal Psychology, 86(2), 103–126. 10.1037//0021-843X.86.2.103 [PubMed: 858828]

Zvielli A, Bernstein A, & Koster EHW (2015). Temporal dynamics of attentional bias. Clinical Psychological Science, 3(5), 772–788. 10.1177/2167702614551572

Table 1.

Characterization of Samples

	Study 1	Study 2*	Study 3
N	30	38	481
Mean Age (SD)	19.5 (1.72)	19.03 (1.24)	22.37 (3.19)
Gender	18 female	23 female	307 female
Race (%)			
Black/African American	2 (6.67)	6 (15.79)	75 (15.59)
Asian/Pacific Islander	8 (26.67)	17 (44.74)	58 (12.05)
Caucasian/White	9 (30.00)	5 (13.16)	202 (42.00)
Hispanic/Latino	10 (33.33)	10 (26.32)	102 (21.21)
Other	1 (3.33)	1 (2.63)	2 (0.42)
Middle Eastern	na	na	15 (3.12)
Mixed Race	na	na	27 (5.61)
Mean Congruent RT (SD)	396.78 (54.40)	397.56 (61.74)	392.85 (57.35)
Mean Incongruent RT (SD)	406.19 (59.08)	403.74 (62.21)	396.99 (58.02)
Mean Neutral RT (SD)	411.29 (66.76)	406.83 (63.08)	397.53 (56.82)
Mean AB_{Trad} (SD)	9.55 (16.32)	6.17 (17.19)	4.15 (15.97)
Mean $Orientation_{Trad}$ (SD)	14.58 (20.41)	9.26 (19.10)	4.67 (15.67)
Mean $Disengagement_{Trad}$ (SD)	−5.03 (21.04)	−3.06 (18.49)	−.52 (16.32)
Mean Fast AB_{RB} (SD)	57.59 (17.41)	51.28 (22.40)	45.22 (17.48)
Mean $SlowAB_{RB}$ (SD)	−66.83 (30.91)	−68.17 (43.49)	−55.87 (35.09)
Mean Fast $Orient_{RB}$ (SD)	59.71 (22.60)	53.13 (23.32)	45.24 (17.09)
Mean Slow $Orient_{RB}$ (SD)	−63.81 (26.32)	−66.91 (41.93)	−57.32 (35.64)
Mean Fast $Disengage_{RB}$ (SD)	−54.91 (19.26)	−68.60 (37.86)	−43.94 (17.27)
Mean Slow $Disengage_{RB}$ (SD)	69.28 (33.10)	53.50 (23.50)	56.83 (31.54)
Total LT Social Anxiety Dx (%)	na	na	96 (19.96)
Total Current Social Anxiety Dx (%)	na	na	51 (10.60)
Total LT MDD (%)	na	na	166 (34.5)
Total Current MDD (%)	na	na	22 (4.6)

	Study 1	Study 2*	Study 3
Total LT AUD (%)	na	na	137 (28.5)
Total Current AUD (%)	na	na	26 (5.4)
Total LT SUD (%)	na	na	98 (20.4)
Total Current SUD (%)	na	na	26 (5.4)
Total LT PTSD (%)	na	na	36 (7.5)
Total Current PTSD (%)	na	na	6 (1.2)
Total LT Panic (%)	na	na	36 (7.5)
Total Current Panic (%)	na	na	10 (2.1)
Total LT Specific Phobia (%)	na	na	98 (20.4)
Total Current Specific (%)	na	na	71 (14.8)
Total LT OCD (%)	na	na	29 (6)
Total Current OCD (%)	na	na	17 (3.5)
Total LT GAD (%)	na	na	46 (9.6)
Total Current GAD (%)	na	na	18 (3.7)
Total LT More than One Dx (%)	na	na	216 (44.9)
Total Current More than One Dx (%)	na	na	60 (12.5)

Note. Trad = Traditional, RB=Response-based computation

*Study 2 means are from the first time point.

Table 2.

Internal Consistency

	Study 1		Study 2				Study 3			
			Time 1		Time 2		Sibling 1		Sibling 2	
	Cronbach's alpha	Split-half	Cronbach's alpha	Split-half	Cronbach's alpha	Split-half	Cronbach's alpha	Split-half	Cronbach's alpha	Split-half
AB_{Trad}	0.54	0.76	0.41	0.05	0.69	0.51	0.53	0.49	0.48	0.50
$Orientation_{Trad}$	0.71	0.85	0.52	0.24	0.69	0.51	0.49	0.46	0.47	0.49
$Disengage_{Trad}$	0.67	0.63	0.50	0.47	0.67	0.75	0.46	0.51	0.55	0.58
Fast AB_{RB}	0.88	0.97	0.92	0.97	0.92	0.96	0.87	0.94	0.88	0.95
Slow AB_{RB}	0.74	0.85	0.86	0.92	0.63	0.49	0.74	0.82	0.76	0.73
Fast $Orient_{RB}$	0.90	0.96	0.92	0.97	0.92	0.94	0.84	0.94	0.87	0.95
Slow $Orient_{RB}$	0.68	0.83	0.85	0.80	0.69	0.77	0.77	0.82	0.77	0.74
Fast $Disengage_{RB}$	0.83	0.83	0.77	0.85	0.74	0.87	0.77	0.84	0.81	0.87
Slow $Disengage_{RB}$	0.80	0.93	0.90	0.92	0.88	0.94	0.77	0.85	0.81	0.90

Note. Trad = Traditional, RB=Response-based computation, Orient=Orientation, Disengage=Disengagement

Table 3.

Test-Retest Reliability for Study 2

	Pearson's Correlation
AB_{Trad}	0.23
$Orientation_{Trad}$	0.35^{*}
$Disengagement_{Trad}$	−0.05
Fast AB_{RB}	0.58^{**}
Slow AB_{RB}	0.70^{**}
Fast $Orient_{RB}$	0.53^{**}
Slow $Orient_{RB}$	0.77^{**}
Fast $Disengage_{RB}$	0.26
Slow $Disengage_{RB}$	0.44^{**}

Note.

* $p < .05$

** $p < .01$, Trad = Traditional, RB=Response-based computation

Table 4.

Study 3 Incremental validity of response-based scores predicting SAD

	Logistic regression models: Single attention-bias score as predictor					Logistic regression models: Traditional and response-based scores as simultaneous predictors (controlling for gender)			
	Current SAD		Lifetime SAD			Current SAD		Lifetime SAD	
	OR	***95% CI***	***OR***	***95% CI***	**IVs**	***OR***	***95% CI***	***OR***	***95% CI***
AB_{Trad}	2.04 *	1.13 3.67	1.37 *	1.03 1.82		--	--	--	--
$Orientation_{Trad}$	1.83 *	1.02 3.29	1.19	0.90 1.58		--	--	--	--
$Disengagement_{Trad}$	1.23	0.87 1.73	1.16	0.88 1.52		--	--	--	--
Fast AB_{RB}	1.49 *	1.06 2.10	1.27 ^{t}	0.97 1.65	AB_{Trad}	1.22	0.83 1.79	1.27	0.93 1.72
					Fast AB_{RB}	1.35	0.95 1.91	1.13	0.84 1.15
Slow AB_{RB}	-- ^{i}	--	0.82	0.64 1.06	AB_{Trad}	1.48 *	1.04 2.10	1.41 *	1.07 1.87
					Slow AB_{RB}	0.75 *	0.56 1.00	0.76 *	0.59 0.99
Fast $Orient_{RB}$	1.36 ^{t}	0.97 1.9	1.25	0.96 1.63	$Orientation_{Trad}$	1.02	0.71 1.47	1.08	0.80 1.46
					Fast $Orient_{RB}$	1.33	0.94 1.88	1.20	0.89 1.60
Slow $Orient_{RB}$	-- ^{i}	--	0.91	0.7 1.17	$Orientation_{Trad}$	1.21	0.85 1.72	1.21	0.91 1.60
					Slow $Orient_{RB}$	0.81	0.60 1.09	0.87	0.67 1.13
Fast $Disengage_{RB}$	0.75 ^{t}	0.55 1.04	0.87	0.67 1.14	$Disengagement_{Trad}$	1.31	0.95 1.81	1.20	0.91 1.57
					Fast $Disengage_{RB}$	0.72 *	0.53 0.98	0.85	0.65 1.11
Slow $Disengage_{RB}$	1.61 *	1.14 2.26	1.42 *	1.08 1.87	$Disengagement_{Trad}$	1.05	0.75 1.47	1.04	0.79 1.37
					Slow $Disengage_{RB}$	1.56 *	1.13 2.15	1.39 *	1.06 1.84

Note.

* $p<.05$

$^{t}p<.01$

^{i}model did not converge, Trad = Traditional, RB=Response-based computation. Odds-Ratios presented in the right side of the table are controlling for gender and include respective traditional and response-based score in same model.

Table 5.

Study 3 Familial Concordance

		95% CI	
	ICC	Lower bound	Upper bound
AB_{Trad}	0.01	−0.10	0.12
$Orientation_{Trad}$	0.05	−0.06	0.16
$Disengagement_{Trad}$	−0.05	−0.18	0.08
Fast AB_{RB}	0.27**	0.16	0.37
Slow AB_{RB}	0.11*	0.00	0.22
Fast $Orient_{RB}$	0.14*	0.03	0.25
Slow $Orient_{RB}$	0.16**	0.05	0.26
Fast $Disengage_{RB}$	0.22**	0.11	0.32
Slow $Disengage_{RB}$	0.24**	0.13	0.34

Note.

** p <.01

* *p*<.05

t *p*<.1, Trad = Traditional, RB = Response-based computation.