

ScholarWorks@GSU

Evaluating Traditional and Newer Indices of Attention Bias Among People with Social Anxiety Disorder

Authors	Molloy, Anthony
Citation	Molloy, Anthony. "Evaluating Traditional and Newer Indices of Attention Bias Among People with Social Anxiety Disorder." 2019. Thesis, Georgia State University. https://doi.org/10.57709/14244720
DOI	https://doi.org/10.57709/14244720
Download date	2026-06-06 20:10:23
Link to Item	https://hdl.handle.net/20.500.14694/13466

EVALUATING TRADITIONAL AND NEWER INDICES OF ATTENTION BIAS AMONG
PEOPLE WITH SOCIAL ANXIETY DISORDER

by

ANTHONY MOLLOY

Under the Direction of Page Anderson, PhD

ABSTRACT

The dot probe task has become the most widely used experimental paradigm for empirically investigating attention biases in anxiety. However, this task's classic reaction time index for calculating mean attention bias toward threat has shown very low reliability. In response to these findings, new indices of attention bias variability and trial-level bias score (TL-BS) strategies have been developed. The purpose of the current study is to evaluate the reliability of all dot probe attention bias indices in the scientific literature. Forty-four community participants with social anxiety disorder completed a facial dot probe task at two sessions one week apart. TL-BS indices of mean attention bias and attention bias variability demonstrated superior reliability to other indices. The findings demonstrate that the TL-BS method is psychometrically superior to other indices for measuring mean attention bias to threat and attention bias variability in people with social anxiety disorder.

INDEX WORDS: Anxiety, Social anxiety disorder, Attention bias, Dot probe, Reliability, Psychometrics

EVALUATING TRADITIONAL AND NEWER INDICES OF ATTENTION BIAS AMONG
PEOPLE WITH SOCIAL ANXIETY DISORDER

by

ANTHONY MOLLOY

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2019

Copyright by
Anthony Molloy
2019

EVALUATING TRADITIONAL AND NEWER INDICES OF ATTENTION BIAS AMONG
PEOPLE WITH SOCIAL ANXIETY DISORDER

by

ANTHONY MOLLOY

Committee Chair: Page Anderson

Committee: Erin Tone

David Washburn

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2019

DEDICATION

To my friends, family, labmates, and student colleagues whose support over the past years made this project possible. Special thanks to my parents and to Kaitlyn.

ACKNOWLEDGEMENTS

I would like to acknowledge my mentor and committee chair Page Anderson for her encouragement, support, and advocacy on my behalf since I came to Georgia State University. I am grateful to have an advisor who is so personally invested in all of her students' success.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS		V
LIST OF TABLES		X
LIST OF FIGURES		XI
1 INTRODUCTION		1
1.1 Calculating Attention Bias Using the Dot Probe Task: Aggregated Mean Indices		2
<i>1.1.1 Original aggregated mean bias index</i>		<i>2</i>
<i>1.1.2 New aggregated mean indices</i>		<i>3</i>
1.2 Findings on Attention Bias among Anxious Populations Using Aggregated Mean Indices		4
<i>1.2.1 Vigilance to threat vs. difficulty disengaging from threat</i>		<i>6</i>
1.3 Reliability of Aggregated Mean Indices		8
1.4 Critique of Aggregated Mean Bias Indices and the Development of Indices for the Dynamics of Attention Bias: Attention Bias Variability and Trial-Level Bias Scores		11
<i>1.4.1 Attention bias variability</i>		<i>11</i>
<i>1.4.2 Trial-level bias scores</i>		<i>12</i>
1.5 Findings on Attention Bias Using Attention Bias Variability and Trial-Level Bias Scores		13
<i>1.5.1 Attention bias variability</i>		<i>13</i>
<i>1.5.2 Peak and mean trial-level bias score indices</i>		<i>15</i>

1.6	Systematically Improving Reliability of the Dot Probe	16
1.6.1	<i>Binning approaches to attention bias variability</i>	17
1.6.2	<i>Trial-level bias score: Peak and mean indices and attention bias variability</i>	19
1.6.3	<i>Outlier handling</i>	21
1.6.4	<i>Other analytic decisions</i>	22
1.6.5	<i>Summary.....</i>	22
1.7	Attentional Biases within Social Anxiety Disorder	23
1.8	Neurocognitive Models of Attention Bias and Attentional Control	24
1.8.1	<i>Attentional control in social anxiety disorder</i>	25
1.8.2	<i>Attentional control and attention bias variability in social anxiety disorder</i>	26
1.9	The Current Study: Reliability of Dot Probe Attention Bias Variability and Trial-Level Bias Score Indices in Social Anxiety Disorder	28
1.9.1	<i>Hypotheses</i>	29
2	METHODS	33
2.1	Participants.....	33
2.2	Measures	34
2.2.1	<i>Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Edition (SCID-I; First, Spitzer, Gibbon, & Williams, 2002)</i>	34
2.2.2	<i>Facial dot probe task.....</i>	34
2.3	Procedure	35

2.3.1	<i>Screening</i>	35
2.3.2	<i>Time 1 assessment</i>	36
2.3.3	<i>Time 2 assessment</i>	36
2.4	Data Analysis	37
2.4.1	<i>Overview</i>	37
2.4.2	<i>MacLeod et al. (1986) aggregated mean bias index</i>	37
2.4.3	<i>Koster et al. (2004) aggregated mean bias indices</i>	37
2.4.4	<i>Zvielli et al. (2015) trial-level bias indices of mean bias, peak bias, and attention bias variability</i>	38
2.4.5	<i>Iacoviello et al. (2014) attention bias variability index</i>	38
2.4.6	<i>Naim et al. (2015) attention bias variability index</i>	39
2.4.7	<i>Arbitrary outlier cutoffs</i>	39
2.4.8	<i>Data-driven cutoffs</i>	40
3	RESULTS	43
3.1	Aggregated Mean Indices (MacLeod et al., 1986; Koster et al., 2004)	43
3.2	Trial-Level Bias Scores (Zvielli et al., 2015): Mean and Peak Attention Bias	44
3.3	Attention Bias Variability Indices (Iacoviello et al., 2014; Naim et al., 2015; Zvelli et al., 2015)	44
3.4	Magnitude of Test-Retest and Split-Half Reliability	45
3.5	Outlier Cutoffs with Highest Reliability Estimates	46

4	DISCUSSION	74
4.1	Study Hypotheses	74
4.2	Consistency of Reliability across Outlier Cutoffs	76
4.3	Implications for Cognitive Models of Social Anxiety Disorder	78
4.4	Limitations and Future Directions	79
4.5	Conclusions and Recommendations	81
	REFERENCES.....	82

LIST OF TABLES

Table 1.1 Split-half reliability of trial-level bias score indices in the scientific literature (Pearson's r).....	31
Table 1.2 Test-retest reliability of trial-level bias score indices in Rodebaugh et al. (2016; Intra- class correlation coefficient).....	32
Table 2.1 Sample demographics.....	41
Table 3.1 MacLeod et al. (1986) index: test-retest and split-half reliability by outlier cutoffs....	48
Table 3.2 Koster et al. (2004) vigilance index: test-retest and split-half reliability by outlier cutoffs	50
Table 3.3 Koster et al. (2004) slow disengagement index: test-retest and split-half reliability by outlier cutoffs	52
Table 3.4 Zvielli et al. (2015) trial-level bias score indices: test-retest reliability by outlier cutoffs	54
Table 3.5 Iacoviello et al. (2014) attention bias variability index: test-retest and split-half reliability by outlier cutoffs	56
Table 3.6 Naim et al. (2015) attention bias variability index: test-retest and split-half reliability by outlier cutoffs	58
Table 3.7 Magnitude of test-retest reliability estimates.....	60
Table 3.8 Magnitude of split-half reliability estimates, time 1.....	61
Table 3.9 Magnitude of split-half reliability estimates, time 2.....	62
Table 3.10 Outlier cutoffs with best test-retest reliability across indices	63

LIST OF FIGURES

Figure 2.1 A diagram of the facial dot-probe task used in the current study.....	42
Figure 3.1 A histogram of all test-retest reliability estimates for the MacLeod et al. (1986) attention bias index	64
Figure 3.2 A histogram of all test-retest reliability estimates for the Koster et al. (2004) attention bias index for vigilance to threat.....	65
Figure 3.3 A histogram of all test-retest reliability estimates for the Koster et al. (2004) attention bias index for slow disengagement from threat	66
Figure 3.4 A histogram of all test-retest reliability estimates for the Iacoviello et al. (2014) index of attention bias variability	67
Figure 3.5 A histogram of all test-retest reliability estimates for the Naim et al. (2015) index of attention bias variability.....	68
Figure 3.6 A histogram of all test-retest reliability estimates for the Zvielli et al. (2015) trial-level bias score index of peak bias toward threat (Peak TL-BS _{POSITIVE}).....	69
Figure 3.7 A histogram of all test-retest reliability estimates for the Zvielli et al. (2015) trial-level bias score index of peak bias away from threat (Peak TL-BS _{NEGATIVE})	70
Figure 3.8 A histogram of all test-retest reliability estimates for the Zvielli et al. (2015) trial-level bias score index of mean bias toward threat (Mean TL-BS _{POSITIVE})	71
Figure 3.9 A histogram of all test-retest reliability estimates for the Zvielli et al. (2015) trial-level bias score index of mean bias away from threat (Mean TL-BS _{NEGATIVE}).....	72
Figure 3.10 A histogram of all test-retest reliability estimates for the Zvielli et al. (2015) trial- level bias score index of attention bias variability (TL-BS _{VARIABILITY}).....	73

1 INTRODUCTION

Decades of research support the idea that biases in cognitive processing play a central role in psychopathology. The cognitive paradigm, first developed in the mid-20th century by theorists like Bower (1981) and Beck, Emery, and Greenberg (1985), explains how biases in information processing contribute to anxiety, depression, and other mental health issues. Cognitive models argue that anxiety is partially caused and maintained by preferential attention to threatening environmental stimuli. Bower (1981), for example, proposed an associative network model in which anxious thoughts and emotions are connected to threat-relevant environmental stimuli via learned associations. The anxious person learns to associate feelings of anxiety with the presence of threat, so when one is present they search for the other. Beck et al. (1985) similarly proposed that states of anxiety create a specific “cognitive set” in an individual, which, when activated, leads the anxious individual to expect to be threatened and seek out stimuli that confirm this expectation, resulting in biased attention. Both models suggest that people with pathological anxiety expect environmental threat when they feel anxious, which influences the ways that anxious people attend to the world around them.

The cognitive model of anxiety inspired a wealth of empirical research. Experimental tasks using a variety of methods have identified attentional biases to threatening stimuli among anxious populations, as theorized by cognitive models. Auditory tasks demonstrated that anxious individuals are able to detect and recall threat-related words more accurately than controls (Burgess, Jones, Robertson, Radcliffe, & Emerson, 1981). Paradigms like the emotional Stroop task demonstrated that, among anxious people, threat-related stimuli result in slower performance on a color-naming task than neutral stimuli, which was understood as an indication that threat-related words demanded more attentional processing resources (Mathews &

MacLeod, 1985; Ray, 1979). However, MacLeod, Mathews, and Tata (1986) argued that anxious peoples' more accurate detection of threatening words on auditory tasks may be due to a mood-dependent bias in participants' guesses for words of which they are unsure. They also criticized interference-based paradigms like the emotional Stroop, arguing that presentation of threatening words could slow performance by increasing negative affect as opposed to demanding more processing resources. To improve on these approaches, MacLeod et al. (1986) developed the visual dot probe task. This task has since become the most widely used cognitive-experimental paradigm for investigating attentional biases among anxious populations.

1.1 Calculating Attention Bias Using the Dot Probe Task: Aggregated Mean Indices

1.1.1 Original aggregated mean bias index

The dot probe task was first developed in 1986 (MacLeod et al., 1986). It was designed as a novel paradigm for assessing visual attentional biases toward threatening stimuli among anxious and depressed samples. During the task, participants are seated in front of a screen and view two stimuli, one near the top and one near the bottom of the screen. These stimuli then quickly disappear, followed by a dot or "probe" that appears at the same location as one of the stimuli. Participants are instructed to press a button to indicate the location of the probe as quickly as possible in each trial. MacLeod et al. (1986) assessed attention bias by comparing reaction times on "threat-congruent" trials, in which the probe replaces a threatening stimulus opposite a neutral stimulus, and "threat-incongruent" trials, in which the probe replaces a neutral stimulus opposite a threatening stimulus. Quicker mean response times for threat-congruent trials than for threat-incongruent trials were conceptualized as evidence of attentional bias toward threatening stimuli because participants whose attention was directed to threatening stimuli would identify probes replacing threatening stimuli (threat-congruent trials) more quickly

than probes replacing neutral stimuli (threat-incongruent trials). Scores calculated using this method are referred to as “aggregated mean” bias indices because they average all scores for each type of trial across the experimental session.

1.1.2 New aggregated mean indices

Koster, Crombez, Verschuere, and De Houwer (2004) argued that MacLeod et al.’s (1986) index was ambiguous because anxious participants could show quicker responses to threat-congruent trials for two different reasons: 1) preferential attention for threat stimuli and/or 2) difficulty disengaging from threat stimuli. From this point of view, faster response times to threat-congruent trials are caused not only by preferential attention to threatening stimuli, but also by difficulty disengaging from threatening stimuli, which would lead to slower responses to probes replacing non-threatening stimuli. Koster et al. (2004) addressed this problem by designing two new aggregated mean bias indices to separately assess vigilance for and difficulty disengaging from threat stimuli. Instead of comparing reaction times for threat-congruent trials to threat-incongruent trials, threat-congruent trials and threat-incongruent trials were each compared to trials consisting of two neutral stimuli, or “neutral-neutral” trials. The investigators interpreted faster responses for threat-congruent vs. neutral-neutral trials as specific evidence of vigilance for threat, whereas slower responses for threat-incongruent vs. neutral-neutral trials were interpreted as specific evidence of difficulty disengaging from threat. These indices are valuable because vigilance for threat and difficulty disengaging from threat may represent distinct patterns of attention bias among anxious populations.

1.2 Findings on Attention Bias among Anxious Populations Using Aggregated Mean Indices

To date, over a dozen meta-analyses and systematic reviews have synthesized empirical research using the dot probe task to investigate attention bias among different clinical populations of interest. Generally, these reviews examine four distinct questions: 1) whether clinical populations show preferential attention to threat, 2) whether clinical populations show stronger attentional preferences to threat as compared to controls, 3) whether specific methodological choices moderate these effects, and 4) whether the specific type of clinical population moderates these effects. A majority of meta-analyses and systematic reviews on attention bias have focused on anxiety, including heterogeneous samples of anxiety disorders (e.g. Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007; Frewen, Dozois, Joanisse, & Neufeld, 2008), as well as specific anxiety disorders, like posttraumatic stress disorder (Hayes, VanElzakker, & Shin, 2012) and social anxiety disorder (Bantin, Stevens, Gerlach, & Hermann, 2016), or anxiety and depression together (Teachman, Joormann, Steinman, & Gotlib, 2012).

Bar-Haim et al. (2007) conducted the most comprehensive meta-analysis of dot probe studies for anxious populations to date and concluded definitively that preferential attention towards threatening stimuli is a robust phenotypic feature of anxiety disorders. They found a significant, moderate within-subjects effect such that anxious participants had quicker response times to threat-congruent trials than to threat-incongruent trials across 35 studies (Cohen's $d = .37$). Using data from 44 studies, they found a significant moderate between-subjects effect, such that anxious participants had quicker relative reaction times to threat-congruent vs. threat-incongruent trials than controls (Cohen's $d = .38$). Duration of stimulus presentation also had a

significant effect; in studies that presented dot probe stimuli for ≥ 1000 milliseconds, there was no significant between-subjects effect on bias. Finally, no significant differences were found between those diagnosed with anxiety disorders and those with high self-reported anxiety. Although the Bar-Haim et al. (2007) meta-analysis included studies investigating a range of anxiety disorders using a variety of methodologies for assessing attention bias (e.g. varying stimulus durations and types of stimuli), their analysis of effect sizes exclusively used the MacLeod et al. (1986) index for attention bias. For studies using clinical populations, they did not compare dot probe results between specific anxiety disorders. Thus, although Bar-Haim et al. (2007) provide compelling evidence of significant preferential attention towards threatening stimuli both within heterogeneous anxious disorder populations and between anxious populations and healthy controls using the MacLeod et al. (1986) index, it is necessary to turn to other reviews for examination of specific disorders and newer indices.

Bantin, Stevens, Gerlach, and Hermann (2016) recently conducted a systematic review of dot probe studies of attention bias among people with social anxiety disorder and high self-reported social anxiety. The authors exclusively reviewed dot probe studies that used facial stimuli instead of words. This form of the dot probe was first developed by Bradley et al. (1997), and is commonly used in social anxiety studies because facial stimuli are potent signals of the core feature of the disorder, negative evaluation from others (Mogg & Bradley, 1998; Rapee & Heimberg, 1997). Bantin et al. (2016) found a significant small within-subjects effect on the MacLeod et al. (1986) index across 10 studies, such that socially anxious participants responded more quickly to threat-congruent than to threat-incongruent trials (Hedge's $g = .21$). Using the same 10 studies, they also found a significant moderate between-subjects effect, such that socially anxious participants had quicker relative reaction times to threat-congruent vs.

threat-incongruent trials than controls (Hedge's $g = .53$). Like in Bar-Haim et al. (2007), findings differed across studies with different stimulus durations. In a single study using a 17 millisecond subliminal duration and in studies using ≥ 1000 milliseconds, there were no significant within-subjects effects. However, these durations still demonstrated significant between-subjects effects for socially anxious participants vs. controls. Within-subjects and between-subjects effects were found for both people diagnosed with social anxiety disorder and for analogue socially anxious samples.

1.2.1 Vigilance to threat vs. difficulty disengaging from threat

Recall that Koster et al. (2004) criticized the MacLeod et al. (1986) index because it did not distinguish between vigilance for threat and difficulty disengaging from threat. They used alternative indices that compared threat-congruent trials and threat-incongruent trials to neutral-neutral trials (rather than to each other) to better distinguish between vigilance for threat and difficulty disengaging from threat, respectively. It should therefore be noted that the meta-analyses conducted by Bar-Haim et al. (2007) and Bantini et al. (2016) were exclusively based on the MacLeod et al. (1986) index. Bar-Haim et al. (2007) did not make the distinction between vigilance for threat and difficulty disengaging from threat, nor discuss ways in which reaction time data from the dot probe task could be analyzed to try to differentiate between the two. Bantini et al. (2016) did make this distinction. Although their quantitative analysis exclusively used the MacLeod et al. (1986) index, their review discussed the results of two studies that used the Koster et al. (2004) indices. Both of these studies found that socially anxious samples scored higher than controls on Koster et al.'s (2004) index of vigilance to threat, but found no within-subjects effects on difficulty disengaging from threat (Klumpp & Amir, 2009; Stevens, Rist, & Gerlach, 2009). An earlier systematic review of the dot probe literature conducted by Cisler,

Bacon, and Williams (2009) also distinguished between vigilance for threat and difficulty disengaging from threat. This review included three studies of high trait-anxious individuals that used the Koster et al. (2004) indices (Koster et al., 2004; Koster, Crombez, Verschuere, & De Houwer, 2006; Salemink, van den Hout, & Kindt, 2007). The review concluded that all three studies found evidence for difficulty disengaging from threat, but not for vigilance towards threat. These conclusions directly contradict the ways in which attention bias using the MacLeod et al. (1986) index have commonly been interpreted. However, two studies reviewed by Bantini et al. (2016) found a within-subjects effect on Koster et al.'s (2004) index of vigilance to threat and not on their index of difficulty disengaging from threat in an analogue socially anxious sample (Klumpp & Amir, 2009) and a sample with social anxiety disorder (Stevens et al., 2009). This may indicate a difference in attention bias between people with high levels of nonspecific trait anxiety and high levels of social anxiety.

In summary, aggregated mean attention bias indices are capable of detecting consistent, statistically significant patterns in reaction times on the dot probe task for anxious samples. Specifically, anxious people tend to react more quickly to threat-congruent trials than threat-incongruent trials and this difference is greater in anxious samples than in controls (Bar-Haim et al., 2007). A small number of studies suggest this may be due in part to difficulty disengaging from threat and find that high trait-anxious people have slower reaction times for threat-incongruent trials than neutral-neutral trials (Koster et al., 2004; Koster, Crombez, Verschuere, & De Houwer, 2006; Salemink, van den Hout, & Kindt, 2007). However, this finding was not replicated in two studies using socially anxious samples (Klumpp & Amir, 2009; Stevens, Rist, & Gerlach, 2009). Evidence suggests that the strongest reaction time effects are observed when stimuli are presented for 500 milliseconds and that effects are weaker for subliminal durations

and durations of 1000 milliseconds or longer. Additionally, all available evidence indicates that within-subjects and between-subjects reaction time differences are consistent between clinically anxious samples and samples with high self-reported anxiety.

1.3 Reliability of Aggregated Mean Indices

The body of research described above strongly indicates that aggregated mean indices are sensitive to patterns of preferential attention associated with anxiety. However, these indices' ability to find moderate effects across large groups of participants does not demonstrate that they are precise, i.e. that a high proportion of variance in their outcomes is driven by preferential attention to threat as opposed to measurement error from irrelevant variables. A tool with any degree of sensitivity to a construct of interest can capture group differences given sufficient power, but careful interpretation of aggregated mean indices' results and decisions about appropriate uses for these indices must be informed by an understanding of their precision in measuring preferential attention. It is therefore important to examine these indices' internal consistency and test-retest reliability, because these properties are completely dependent on two factors: the indices' precision in measuring the construct of interest, and the true stability of the construct over time (Thissen, 2000). Therefore, examining the psychometric properties of aggregated mean indices could shed light not only on their precision, but on the stability of true preferential attention in the following ways: 1) a high degree of internal consistency for these indices would suggest that true preferential attention to threat is stable within a session and that aggregated mean indices are capable of precisely measuring it, and 2) a high degree of test-retest reliability would suggest that preferential attention to threat is stable between assessment time points and provide additional evidence of the task's precision. Findings of low internal consistency and test-retest reliability would be ambiguous, however, because they could be

caused by imprecision in the indices' assessments of preferential attention, true variability in preferential attention, or both. Thus, low internal consistency and test-retest reliability would indicate a need for new analytical approaches for dot probe reaction time data in order to disentangle these possibilities.

Schmukle (2005) was the first to examine the dot probe task's psychometric properties. Schmukle (2005) developed three versions of the task, two using verbal stimuli and one using pictures of scenes and other images. Using the original MacLeod et al. (1986) aggregated mean index of attention bias, the study found poor test-retest reliability for all versions of the task. They also found zero to unacceptably low internal consistency, as measured by split-half reliability and Cronbach's alpha. Staugaard (2009) was the first to test the psychometrics of a dot probe task using facial stimuli. Consistent with Schmukle's (2005) results, the facial dot probe showed unacceptably low test-retest reliability and internal consistency using the MacLeod et al. (1986) index. Since these initial investigations into the psychometric properties of the dot probe task, versions of the task with a variety of stimuli have been tested for reliability in samples with nicotine dependence (Spiegelhalder et al., 2011), chronic pain (Dear, Sharpe, Nicholas, & Refshauge, 2011), social anxiety (Waechter, Nelson, Wright, Hyatt, & Oakman, 2014) alcoholism (Van Duijvenbode, Didden, Korzilius, & Engels, 2016), depression (Zvielli, Vrijzen, Koster, & Bernstein, 2016), and others. Rodebaugh et al. (2016), in a review of 13 studies that assess internal consistency and reliability of the MacLeod et al. (1986) index for the dot probe, found that estimates for both properties are almost uniformly low when the index is unmodified. They review two studies that find moderate internal consistency for the MacLeod et al. (1986) index, but even these exceptions to the rule find split-half reliability under $r = .5$.

Waechter et al. (2014) reported the first study to assess the internal consistency of Koster et al.'s (2004) indices of attentional vigilance to threat and difficulty disengaging from threat. They also assessed the original MacLeod et al. (1986) index, representing the first direct investigation into the effects of analytic strategy on the reliability of the dot probe. The investigators administered a dot probe task with facial stimuli to a sample with high self-reported social anxiety and a control group. All data was collected in a single session. The investigators found low, insignificant split-half reliability and Cronbach's alpha for both MacLeod et al.'s (1986) and Koster et al.'s (2004) indices of attention bias. Despite the proposed conceptual advantages of Koster et al.'s (2004) indices for assessing attentional vigilance to threat and difficulty disengaging from threat relative to MacLeod et al.'s (1986) original approach, Waechter et al. (2014) show that Koster et al.'s (2004) indices do not demonstrate better internal consistency.

In summary, aggregated mean indices for the dot probe task are sensitive to group differences in preferential attention. However, aggregated mean indices consistently demonstrate low internal consistency and test-retest reliability and it is important to understand why. One reason may be that these indices lack precision in their ability to measure preferential attention to threat and are therefore highly influenced by irrelevant variables. Another possibility is that studies using aggregated mean indices are based on a flawed assumption—that preferential attention to threat is a static trait rather than a fluctuating state. Exploration of methods that can differentiate between these two possibilities is important. The aggregated mean approach is inadequate to this task, as it is inherently insensitive to fluctuations in preferential attention because it collapses across all trials over the course of a session.

1.4 Critique of Aggregated Mean Bias Indices and the Development of Indices for the Dynamics of Attention Bias: Attention Bias Variability and Trial-Level Bias Scores

1.4.1 Attention bias variability

Iacoviello et al. (2014) reviewed evidence that post-traumatic stress disorder is characterized by both attentional vigilance and attentional avoidance of threatening stimuli. They proposed that a measure of variability in attention bias may be a useful metric for attention-related symptoms of the disorder because it could capture fluctuations between both types of bias. To address this possibility, Iacoviello et al. (2014) developed the first index to test attention bias variability during the dot probe task. This index measured attention bias dynamically, by examining the time course of reaction time differences between threat-congruent trials and nearby threat-incongruent trials. It accomplished this by splitting trials into eight consecutive bins, then calculating the MacLeod et al. (1986) aggregated mean bias index for each bin and the standard deviation of these eight values. The final attention bias variability score was then calculated by dividing this standard deviation by each participant's average reaction time for all trials to control for differences in overall reaction speed. Naim et al. (2015) soon adapted this method into a more sensitive "moving average" approach. Instead of splitting trials into non-overlapping blocks, Naim et al. (2015) compared each consecutive overlapping 10-trial block of threat-congruent trials to each consecutive 10-trial block of threat-incongruent trials (e.g. threat-congruent trials 1-10 vs. threat-incongruent 1-10, threat-congruent trials 2-11 vs. threat-incongruent 2-11, etc.). The MacLeod et al. (1986) aggregated mean bias index was calculated for each pair of blocks, then the standard deviation of these values was divided by each participant's overall mean reaction time for a final estimate of attention bias variability.

The development of attention bias variability indices allowed researchers to measure the stability of attention bias for the first time. This not only provides new insights into a potentially important factor in anxiety psychopathology, but also allows for new insights into the reasons that aggregated mean indices like that of MacLeod et al. (1986) show such poor internal consistency and test-retest reliability. As described above, the low internal consistency and test-retest reliability found for aggregated mean indices of preferential attention to threat may be due to an implicit, incorrect assumption that it is a static trait. Evidence that measures of attention bias variability have good internal consistency and test-retest reliability would support the proposition that aggregated mean indices will always be internally inconsistent and unreliable because they are designed to measure an unstable phenomenon.

1.4.2 Trial-level bias scores

Soon after Iacoviello et al. (2014) and soon before Naim et al. (2015) introduced attention-bias variability measures, Zvielli, Bernstein, and Koster (2015) designed the trial-level bias score, another major contribution to the dot probe literature. Trial-level bias score values are calculated by finding the reaction time difference between each individual threat-congruent trial and its nearest threat-incongruent trial. This allows for a much more sensitive assessment of attention bias than aggregated mean indices, because point estimates of bias can be examined at any given time throughout a session. To assess overall bias toward and away from threat, Zvielli et al. (2015) recommended separately assessing trial pairs that indicate preferential attention toward threat (i.e. faster reaction to the threat-congruent trial relative to the threat-incongruent trial) and trial pairs that indicate preferential attention away from threat (i.e. slower reaction to the threat-congruent trial relative to the threat-incongruent trial). Mean and peak trial-level bias

score values can then be calculated separately for these two types of scores, providing separate indices for biased attention toward and away from threat for each participant.

This method is similar to Koster et al.'s (2004) aggregated mean indices in that it attempts to disentangle different dimensions of preferential attention that are conflated in the MacLeod et al. (1986) index. However, it differentiates between preferential attention toward threat, commonly interpreted as vigilance, and preferential attention away from threat, commonly interpreted as avoidance, as opposed to using neutral-neutral trials to specifically examine the construct of difficulty disengaging from threat. Examining individual pairs of trials represents a major methodological improvement over Koster et al.'s (2004) aggregated mean method because this creates the ability to measure biased attention as a dynamic construct.

The development of trial-level bias scores also represents a direct improvement on Iacoviello et al. (2014) and Naim et al. (2015) because trial-level bias scores can be used to produce a highly sensitive index of attention bias variability. This is done by calculating the standard deviation of all trial-level bias scores in a session. This creates a superior attention bias variability index because it captures variability in bias on a trial to trial level, instead of splitting trials into arbitrary groups and aggregating trials within these groups.

1.5 Findings on Attention Bias Using Attention Bias Variability and Trial-Level Bias Scores

1.5.1 Attention bias variability

Although no reviews or meta-analyses have been conducted, studies using Iacoviello et al.'s (2014) binning method, Naim et al.'s (2015) moving average method, and Zvielli et al. (2015)'s trial-level bias score method of calculating attention bias variability have demonstrated significant group differences in attention bias variability between healthy controls and people

with many types of psychopathology. Iacoviello et al. (2014) found that attention bias variability was significantly higher in participants with posttraumatic stress disorder compared to a group of healthy controls and nonclinical participants who had been exposed to trauma. Naim et al. (2015) similarly found that participants with posttraumatic stress disorder had a higher level of attention bias variability than healthy controls, healthy combat-exposed soldiers, a sample with high trait anxiety, and a sample with social anxiety disorder. None of these four groups differed in attention bias variability. Both Iacoviello et al. (2014) and Naim et al. (2015) additionally found a positive association between posttraumatic stress disorder symptoms and attention bias variability. Using Zvielli et al.'s (2015) trial-level bias score method, Schäfer et al. (2016) investigated the value of attention bias variability scores in predicting posttraumatic stress response to experiences in combat. They found that attention bias variability scores recorded before and after soldiers were deployed significantly interacted with the number of traumatic events experienced during combat, such that higher attention bias variability and higher number of traumatic experiences predicted higher levels of posttraumatic stress. Using Naim et al.'s (2015) index, Schneier et al. (2016) found that a sample with obsessive-compulsive disorder had significantly lower attention bias variability than healthy controls, but did not find this difference for samples with social anxiety disorder and anorexia nervosa. The researchers noted, however, that the sample with social anxiety disorder was smaller and had the same mean and standard deviation for attention bias variability as the obsessive-compulsive disorder sample, suggesting participants with social anxiety disorder might have had significantly lower attention bias variability than controls if there were a larger sample. Using hierarchical logistic regression, Zvielli et al. (2015) found that their trial-level bias score index of attention bias variability explained more variance than the MacLeod et al. (1986) index in discriminating between spider

phobia and healthy controls. Zvielli et al. (2016) subsequently used this index to demonstrate significantly higher attention bias variability in remitted depression as compared to healthy controls. Bardeen, Daniel, Hinnant, and Orcutt (2017) also found a significant positive correlation between the trial-level bias score index of attention bias variability and self-reported emotional dysregulation in a healthy sample.

These findings clearly indicate that attention bias variability plays a significant role in a range of psychopathology. In the brief time since Iacoviello et al.'s (2014) first attention bias variability index, elevated attention bias variability has been robustly found in posttraumatic stress disorder and even found to predict posttraumatic stress response to later trauma in healthy soldiers. Elevated attention bias variability has also been found in spider phobia and remitted depression compared to controls. Additionally, lower attention bias variability has been found in obsessive-compulsive disorder compared to controls. Two studies that examined attention bias variability in social anxiety disorder found no difference between social anxiety disorder and healthy controls and found that attention bias variability was lower in social anxiety disorder than in posttraumatic stress disorder (Naim et al., 2015; Schneier et al., 2016). Neither of these studies, however, used the trial-level bias score method.

1.5.2 Peak and mean trial-level bias score indices

Peak and mean trial-level bias score indices also differentiate between people with and without psychopathology. Zvielli et al. (2015) first developed these indices and found that participants with spider phobia had higher peak and mean trial-level bias scores toward threat as well as lower mean trial-level bias score away from threat as compared to healthy controls. They also demonstrated with hierarchical logistic regression that mean trial-level bias score toward threat explained more variance than the MacLeod et al. (1986) index in discriminating between

spider phobia and healthy controls. Zvielli et al. (2016) subsequently found that peak and mean trial level bias scores both toward and away from threat were higher in remitted depression compared to controls. They also found that all four of these indices were positively associated with number of past depressive episodes. Schäfer et al. (2016) found that mean trial-level bias scores both toward and away from threat recorded before and after soldiers were deployed to combat significantly interacted with the number of traumatic events experienced during combat, such that higher attention bias variability and a higher number of traumatic experiences predicted higher levels of posttraumatic stress. Conversely, aggregated mean indices of attention bias did not predict posttraumatic stress. These findings are important, as the vast majority of research into attention bias using the dot probe has relied on MacLeod et al.'s (1986) aggregated mean bias index. Since Zvielli et al. (2015), there has been compelling evidence that aggregated mean indices have poor predictive validity for psychopathology compared to mean trial-level bias scores toward and away from threat.

1.6 Systematically Improving Reliability of the Dot Probe

In order to improve reliability and validity of the dot probe task and circumvent problems associated with key-pressing, some investigators have assessed reliability of attention bias indices based on eyetracking (Burriss, Barry-Anwar, & Rivera, 2017; Price et al., 2015), EEG (Kappenman, Farrens, Luck, & Proudfit, 2014), and fMRI (White et al., 2016). However, investigators like Waechter et al., (2014), Price et al. (2015), and Zvielli et al. (2015) have also focused on systematically investigating ways in which task design and analytic strategies can be optimized to reliably assess attention bias using reaction time data. This includes using peak and mean trial-level bias scores toward and away from threat, attention bias variability indices

including the trial-level bias score variability index, and manipulation of other analytic decisions like handling of outliers.

1.6.1 Binning approaches to attention bias variability

Price et al. (2015) reported a comprehensive attempt to maximize the dot probe task's reliability by systematically manipulating analytic strategy. They conducted dot probe experiments using three small samples across multiple time points—a group of adults diagnosed with social anxiety disorder assessed at 12 sessions over six weeks, a group of adults diagnosed with social anxiety disorder assessed at eight sessions over four weeks, and a healthy group of children assessed at five sessions over 14 weeks. Stimulus duration and other methodological details varied between these experiments. The investigators then tested the effects of different analytic strategies across the three groups to determine whether any given choice maximized reliability for all groups. They assessed reliability using two types of intra-class correlation coefficients: 1) a single-measure intra-class correlation coefficient, which estimates the test-retest reliability of scores obtained in a single session, and 2) an average-measures intra-class correlation coefficient, which estimates test-retest reliability for the mean score across all sessions. Although a test-retest reliability estimate for the mean score across many sessions lacks clinical utility, Price et al. (2015) present the average-measures intra-class correlation coefficient as a useful metric of internal consistency, since its value is usually nearly identical to Cronbach's alpha.

Price et al. (2015) calculated these reliability metrics for the MacLeod et al. (1986) aggregated mean index, the Koster et al. (2004) aggregated mean index for vigilance to threat, and the Iacoviello et al. (2014) attention bias variability index, representing the first assessment of the Iacoviello et al. (2014) attention bias variability index's psychometric properties. Even

using the most advantageous set of methodological choices in each experiment (e.g. outlier cutoffs, exclusively using dot-top or dot-bottom trials), single measure intra-class correlation coefficients were uniformly low for MacLeod et al. (1986) and Koster et al. (2004) indices (single-measure ICC range: .04-.19), indicating low test-retest reliability. For non-optimal methodological approaches, single measure intra-class correlation coefficients for these indices were mostly insignificant. Single-measure correlation coefficients for Iacoviello et al.'s (2014) attention bias variability index were inconsistent, ranging from low (single-measure ICC = .13) to moderate (single-measure ICC = .65). However, coefficients were significant across all experiments and methodological approaches, indicating generally superior test-retest reliability as compared to aggregated mean indices. Using the most advantageous set of methodological choices for each experiment, average-measures intra-class correlation coefficients ranged from low to moderate for the MacLeod et al. (1986) and Koster et al. (2004) indices (average-measures ICC range: .45-.65). Once again, the non-optimal methodological approaches led mostly to very low or insignificant coefficients, indicating poor internal consistency in most circumstances.

In summary, Price et al.'s (2015) results indicated that the test-retest reliability and internal consistency of Iacoviello et al.'s (2014) attention bias variability index was consistently statistically significant and higher than than MacLeod et al.'s (1986) and Koster et al.'s (2004) aggregated mean indices. Naim et al. (2015) subsequently examined the test-retest reliability of the “moving average” attention bias variability index over two sessions. They found significant but low estimates of test-retest reliability in a healthy sample ($r = .29$) and one diagnosed with posttraumatic stress disorder ($r = .40$), but did not report estimates of internal consistency. This

study provides initial evidence that Iacoviello et al.'s (2014) and Naim et al.'s (2015) "binning" approaches to measuring attention bias variability may have similar psychometric properties.

1.6.2 Trial-level bias score: Peak and mean indices and attention bias variability

Using split-half reliability, five studies to date have assessed the internal consistency of Zvielli et al.'s (2015) trial-level bias score indices of peak bias, mean bias, and attention bias variability (Amir, Zvielli, & Bernstein, 2016; Rodebaugh et al., 2016; Schäfer et al., 2016; Zvielli et al., 2015; Zvielli et al., 2016). Split-half reliability results for these studies are presented in Table 1.1. Although these studies used a range of populations and dot probe stimuli, all split-half reliability results were statistically significant. This is in stark contrast to internal consistency findings for aggregated mean indices, most of which are low or statistically insignificant. Across these studies, the most robust internal consistency has been found for mean trial-level bias score toward threat ($r = .66-.83$). The highest split-half reliability estimates for mean trial-level bias score toward threat were found in Rodebaugh et al.'s (2016) study of heterogeneous anxiety disorders (Time 1; $r = .83$), Amir et al.'s (2016) study of high trait anxiety ($r = .81$), and Zvielli et al.'s (2016) study of remitted depression ($r = .80$). These findings suggest that internal consistency for mean trial-level bias score toward threat may be higher for samples with psychopathology, as opposed to other groups like Zvielli et al.'s (2015) sample of nicotine-deprived smokers ($r = .67$) and Schafer et al.'s (2016) sample of soldiers ($r = .66$). Rodebaugh et al.'s (2016) split-half reliability findings during the second session with their anxious sample, however, is an exception ($r = .68$). Whereas Zvielli et al. (2015) found low split-half reliability for their trial-level bias score variability index in a sample of nicotine-deprived smokers ($r = .46$), all other studies have found higher split-half reliability for this index ($r = .69-.95$). Other indices show less consistent results. Mean trial-level bias score away from

threat and peak trial-level bias scores toward and away from threat have good split-half reliability in some studies, but low split-half reliability in others. However, they are consistently statistically significant and a substantial improvement on the internal consistency of aggregated mean indices.

Rodebaugh et al. (2016) reported the first and only study to date to examine test-retest reliability for trial-level bias score indices of mean and peak bias and attention bias variability. They administered a dot probe task with verbal stimuli to a small sample of adults with heterogeneous anxiety disorders. Participants attended two sessions that were three weeks apart on average and completed four blocks of the task, totaling 1,216 trials. This unusually high number of trials was used so that the researchers could compare reliability of different subsets of trials. Test-retest reliability findings from Rodebaugh et al. (2016) are presented in Table 1.2. The highest estimates of test-retest reliability were found for mean trial-level bias score toward threat, mean trial-level bias score away from threat, and trial-level bias score attention bias variability. Lower estimates were found for peak trial-level bias score indices, with half of reliability estimates for peak bias toward threat and all estimates for peak bias away from threat falling below statistical significance. Results for the entire first block (304 trials) consistently indicated the best reliability, with excellent reliability for attention bias variability ($ICC = .90$), good reliability for mean trial-level bias score toward threat ($ICC = .86$), and moderate reliability for mean trial-level bias score away from threat ($ICC = .63$). These results are very encouraging, as they indicate that a single session of 304 trials would be sufficient to capture highly reliable trial-level bias score estimates of mean bias toward threat, mean bias away from threat, and attention bias variability. However, further studies of test-retest reliability are warranted before

it is possible to generalize these findings to other populations or versions of the task with different stimuli.

1.6.3 Outlier handling

Price et al. (2015) also explored how different strategies for handling outliers affect reliability. Researchers using the dot probe task have typically handled reaction time outliers by setting arbitrary cutoffs and simply excluding all values outside this cutoff range. However, Price et al. (2015) argued that cutoffs could be determined by more data-driven methods, namely using the interquartile ranges of all participants' reaction times in an experiment. As an alternative to discarding values outside of the cutoff range, Price et al. (2015) also tested the effects of "Winsorizing" or changing outliers to the nearest value within the cutoff range. This is recommended as a robust method for handling outliers in psychological research (Erceg-Hurn & Mirosevich, 2008), as it reduces outliers' effects on the data but avoids discarding data points in order to maximize power and accuracy. For aggregated mean indices, test-retest reliability estimates were consistently low or insignificant, with no clear advantage of any outlier-handling strategy. For Iacoviello et al.'s (2014) attention bias variability index, however, the commonly used approach of excluding outliers with discretionary cutoffs produced higher test-retest reliability than the Winsorizing method. Thus, the limited evidence on the effects of outlier-handling strategy on test-retest reliability seems to support using the traditional methods of deleting outliers outside of arbitrary cutoffs. However, manipulation of outlier-handling strategy has yet to be systematically tested for test-retest reliability using trial-level bias score indices of mean bias, peak bias, and Naim et al.'s (2015) and Zvielli et al.'s (2015) approaches to measuring attention bias variability.

1.6.4 Other analytic decisions

For aggregated bias indices, Price et al. (2015) also tested the effects of only including trials in which the dot probe appeared on the top or bottom of the display on test-retest reliability. Exclusively analyzing dot-bottom trials improved reliability of the MacLeod et al. (1986) and Koster et al. (2004) indices to statistical significance, but these reliability estimates remained low. Analyzing all dot-top and dot-bottom trials together reduced test-retest reliability to insignificance for both indices in almost all cases. Lastly, the investigators tested the effect of averaging consecutive pairs of sessions together, in order to see if pairs of sessions administered several days apart are a more reliable metric than individual sessions. This approach produced only marginal improvements in test-retest reliability for aggregated mean indices.

1.6.5 Summary

In summary, studies using MacLeod et al.'s (1986) and Koster et al.'s (2004) aggregated mean bias indices have overwhelmingly found low or insignificant test-retest reliability and internal consistency. Price et al. (2015) found that exclusively using dot-bottom trials and averaging pairs of experimental sessions together marginally improved reliability for aggregated mean indices, but these reliability estimates were still very low or insignificant. Results from Price et al. (2015) indicate that the traditional approach of excluding outliers using discretionary cutoffs optimizes test-retest reliability for Iacoviello et al.'s (2014) attention bias variability index. The effects of outlier-handling strategy on reliability for Naim et al.'s (2015) index of attention bias variability and Zvielli et al.'s (2015) trial-level bias score indices have yet to be explored. Comparison between Price et al. (2015), Naim et al., (2015), and Rodebaugh et al. (2016) provides initial evidence that the trial-level bias score method of calculating attention bias variability is superior in terms of test-retest reliability. Both Price et al. (2015) and Rodebaugh et

al. (2016) found higher test-retest reliability for attention bias variability in sessions of approximately 300 trials as compared to shorter sessions. Evidence from Rodebaugh et al. (2016) also indicates good to excellent test-retest reliability for mean trial-level bias score toward threat and moderate to good test-retest reliability for mean trial-level bias score away from threat.

1.7 Attentional Biases within Social Anxiety Disorder

In the current study, the author examines indices of attentional biases within individuals diagnosed with social anxiety disorder, one of the most commonly diagnosed anxiety disorders (Kessler, Chiu, Demler, & Walters, 2005). Social anxiety disorder has been associated with a range of negative outcomes including lower educational achievement, unemployment and underemployment, and higher risk for suicide (Nock, Hwang, Sampson, & Kessler, 2010; Stein & Kean, 2000). Since social anxiety disorder first appeared in the DSM-III (American Psychiatric Association, 1980) as social phobia, it has been the focus of a large body of attention bias research.

Clark and Wells (1995) and Rapee and Heimberg (1997) proposed influential models of social anxiety disorder in which strong biases in attention co-occur with somatic, affective, and behavioral symptoms to reinforce and maintain anxiety. Clark and Wells (1995) argue that when an anxious person starts to experience somatic symptoms of anxiety like blushing, sweating, or hyperventilating, they turn their attention inwards towards their physical sensations—“self-focused” attention—and develop expectations that others will see their anxiety symptoms and judge them negatively as a result. Self-focused attention is maladaptive because it precludes socially anxious individuals from attending to cues in the environment, which are likely to be better indicators of performance than their own anxiety symptoms. Rapee and Heimberg’s

(1997) model of social anxiety also includes biased attention towards the self, in the form of a mental picture of how one is viewed by others. They argue that biased attention to external environmental cues of social threat create cognitive demands that reduce available cognitive processing resources for any other tasks (e.g. holding a conversation, giving a presentation). In both models, however, exaggerated attention toward internal and external indicators of social threat play an important role in social anxiety disorder.

These seminal models have an enormous influence on how attention biases have been investigated empirically in social anxiety disorder, particularly in how results from the dot probe have been conceptualized. Specifically, these models argue that there is an overall bias in attention toward threat in people with social anxiety disorder as compared to people without the disorder. These are the exact hypotheses that aggregated mean indices of attention bias are designed to test. Although these indices have predominated in the literature to date, newer measures of attention bias variability may provide an opportunity to re-examine and update our theoretical understanding of attention in social anxiety disorder, particularly given the initial evidence that newer indices have far better psychometric properties.

1.8 Neurocognitive Models of Attention Bias and Attentional Control

Neurocognitive models of anxiety integrate experimental and neuroscience findings into theoretical frameworks. Many of these models focus on dysfunction in attentional control due to weakened connectivity between prefrontal cortical areas and the amygdala. Eysenck, Derakshan, Santos, and Calvo's (2007) attentional control theory provides a comprehensive theoretical model of impaired executive control of attention in anxiety. They propose that there are two competing attentional systems in the brain—a “top-down” executive system that focuses attention in pursuit of conscious goals and a “bottom-up” stimulus-driven system that focuses

attention on salient features of the environment. According to this theory, attentional biases toward threat associated with anxiety are the result of relative weakness in top-down attentional control. This causes problems inhibiting attention to distracting stimuli, focusing on goal-relevant stimuli, and updating information in working memory. Empirical evidence for this phenomenon is summarized by Eysenck et al. (2007), Bishop (2007), and Mujica-Parodi, Cha, and Gao (2017). These reviews provide detailed summaries of specific neurological pathways by which prefrontal brain regions regulate amygdala activity. They also detail fMRI evidence that dysfunction in these pathways is associated with trait anxiety.

1.8.1 Attentional control in social anxiety disorder

Consistent with Eysenck et al.'s (2007) attentional control theory of anxiety, there is a range of empirical evidence that symptoms of social anxiety disorder are associated with impairments in attentional control. This is particularly evident in the presence of threatening social stimuli (i.e. pictures of faces). Multiple studies have found that self-reported social anxiety is negatively correlated with self-reported attentional control in analogue social anxiety samples (Moriya & Tanno, 2008; Morrison & Heimberg, 2013). Experimental studies have found that individuals with social anxiety disorder perform worse than controls on tasks that require inhibiting attention to threatening faces, including on eyetracking tasks (Mohlman & DeVito, 2017; Wieser, Pauli, Weyers, Alpers, & Mühlberger, 2009), although this may be driven by comorbid depression (Morrison et al., 2016). fMRI studies have found that people with social anxiety disorder show patterns of prefrontal brain activity indicating impairment during tasks that require participants to inhibit attention to threatening faces. Specifically, during these tasks, researchers have found reduced activity in the anterior cingulate cortex (Blair et al., 2012; Swartz et al., 2014), an area associated with attentional control. Additionally, higher levels of activity in

the anterior cingulate cortex and the dorsolateral prefrontal cortex during attentional control tasks have positively predicted response to cognitive behavioral therapy in social anxiety disorder (Klumpp et al., 2017; Klumpp, Fitzgerald, Angstadt, Post, & Phan, 2014), indicating that attentional control contributes to treatment response. This is consistent with a study that assessed attentional control over the course of cognitive behavioral therapy for social anxiety disorder and found that it was significantly related to anxiety symptoms (McEvoy & Perini, 2009).

1.8.2 Attentional control and attention bias variability in social anxiety disorder

To date, few studies have directly examined the role of attention bias variability in social anxiety disorder. As mentioned above, only two studies have compared dot probe indices of attention bias variability between people with social anxiety disorder and healthy controls and both found no significant difference (Naim et al., 2015; Schneier et al., 2016). However, both of these studies used the “moving average” method of assessing attention bias variability developed by Naim et al. (2015), which has been demonstrated to have low test-retest reliability. Research using the more reliable trial-level bias score method may yield different results. Additionally, Schneier et al. (2016) argued that attention bias variability may be an important mechanism of anxiety in some people with social anxiety disorder and not in others. This subgroup may benefit from interventions designed to reduce attention bias variability, even if other people diagnosed with social anxiety disorder would not. So far, a single treatment study has assessed attention bias variability over the course of cognitive behavioral therapy for social anxiety disorder and found that attention bias variability decreased over the course of treatment (Davis et al., 2016). Although the authors demonstrate that reductions in attention bias variability did not drive anxiety symptom reduction, these results suggest that attention bias variability and social

anxiety disorder may be related, perhaps due to a connection between attention bias variability and attentional control.

No study to date has directly examined the connection between attention bias variability and attentional control in social anxiety disorder, but two studies have examined this relationship in posttraumatic stress disorder. Bardeen, Tull, Daniel, Evenden, and Stevens (2016) found that performance on an attentional control task moderated the difference in attention bias variability between participants with posttraumatic stress disorder and controls, such that participants with posttraumatic stress disorder had higher attention bias variability than controls when attentional control was low. This task required participants to classify images based on sets of rules that shift quickly between trials, thus controlling and inhibiting their attention to sets of rules that do not apply in a given trial. Consistent with these results, Badura-Brack et al. (2015) demonstrated that attentional control training for individuals with posttraumatic stress disorder reduced attention bias variability and posttraumatic stress symptoms.

These findings cannot be generalized to social anxiety disorder, especially because posttraumatic stress disorder has been robustly associated with increased attention bias variability (Iacoviello et al., 2014; Naim et al. 2015; Schafer et al., 2016) and this association has not been found in social anxiety disorder. However, given the clearly demonstrated deficits in attentional control in social anxiety disorder, the face similarity between the constructs of poor attentional control and high attention bias variability, the unreliable methods that have been used to assess attention bias variability in social anxiety disorder thus far, and the demonstration that poor attentional control and attention bias variability are related in posttraumatic stress disorder, it seems clear that more research into the role of attention bias variability in social anxiety disorder is warranted. This research would be especially valuable because findings of elevated

attention bias variability in social anxiety disorder could drive a theoretical shift in how attention bias is conceptualized in this disorder. Classic cognitive models emphasizing mean differences in bias to socially threatening stimuli between people with social anxiety disorder and others have been repeatedly investigated with aggregated mean indices for the dot probe task, but these indices are completely insensitive to fluctuations in attention bias over time. Evidence of elevated attention bias variability in social anxiety disorder would suggest that, like in posttraumatic stress disorder, people with social anxiety disorder fluctuate more dramatically between preferential attention toward and away from external threat than the general population. This would provide an important foundation for continuing research into the interacting roles of attentional control and attention bias variability in social anxiety disorder in order to form updated neurocognitive models of its etiology, maintenance, and treatment. Although the two studies that have examined attention bias variability in social anxiety disorder to date found null results, the significant methodological limitations of these studies make any conclusions premature. It is therefore necessary to explore the psychometric properties of new attention bias variability indices for the dot probe task, particularly Zvielli et al.'s (2015) trial-level bias score indices, and to compare them to earlier methods to evaluate whether these indices are better suited for future studies investigating attention bias variability in social anxiety disorder.

1.9 The Current Study: Reliability of Dot Probe Attention Bias Variability and Trial-Level Bias Score Indices in Social Anxiety Disorder

For the role of attention bias variability in neurocognitive models of social anxiety disorder to be properly explored, data analytic strategies for assessing this construct should conform to evidence-based standards that maximize reliability. Initial estimates of reliability for Iacoviello et al.'s (2014) original attention bias variability index, Naim et al.'s (2015) "moving

average” method, and Zvielli et al.’s (2015) trial-level bias score method vary considerably. This means that more studies are needed like Price et al. (2015) that compare estimates of reliability across these indices. Whereas Rodebaugh (2016) was the first to examine test-retest reliability for Zvielli et al.’s (2015) trial-level bias score indices, the current study is the first to do so in a sample diagnosed with social anxiety disorder. Given that many studies still exclusively use the MacLeod et al. (1986) index of attention bias, including very recent publications in high-impact journals (e.g. Salum et al., 2017), direct comparisons of reliability between trial-level bias score indices and traditional aggregated mean indices may also provide a valuable contribution to the attention bias literature.

In the current study, the author examined test-retest and split-half reliability for the dot probe task based on data from two different time points in an adult sample with social anxiety disorder. Test-retest and split-half reliability were compared for MacLeod et al.’s (1986) and Koster et al.’s (2004) aggregated mean indices as well as Zvielli et al.’s (2015) trial-level bias score indices of mean and peak bias toward and away from threat. Test-retest and split-half reliability of Iacoviello et al.’s (2014), Naim et al.’s (2015), and Zvielli et al.’s (2015) indices of attention bias variability were also compared. Lastly, reliability of the different outlier handling strategies tested by Price et al. (2015) were re-examined.

1.9.1 Hypotheses

Based on existing research into psychometric properties of the dot probe, the author’s hypotheses were as follows:

- Aggregated mean indices have nonsignificant or very low test-retest and split-half reliability, regardless of outlier-handling strategy.

- Iacoviello et al.'s (2014) and Naim et al.'s (2015) indices of attention bias variability have significant test-retest reliability in the $r = .1$ to $r = .6$ range. They have higher split-half reliability in the $r = .5$ to $r = .9$ range. Split-half and test-retest reliability are stronger with deletion of values outside arbitrary outlier cutoffs, as opposed to the data-driven Winsorizing method.
- Zvielli et al.'s (2015) peak TL-BS_{POSITIVE} and peak TL-BS_{NEGATIVE} have split-half reliability in the $r = .3$ to $r = .8$ range. These indices have nonsignificant or very low test-retest reliability.
- Zvielli et al.'s (2015) mean TL-BS_{NEGATIVE} has split-half reliability in the $r = .3$ to $r = .9$ range. This index has test-retest reliability in the $r = .5$ to $r = .7$ range.
- Zvielli et al.'s (2015) mean TL-BS_{POSITIVE} has split-half and test-retest reliability in the $r = .6$ to $r = .8$ range.
- TL-BS_{VARIABILITY} has the highest test-retest and split-half reliability of all indices, in the $r = .7$ to $r = .9$ range.

Table 1.1 Split-half reliability of trial-level bias score indices in the scientific literature (Pearson's r)

Study	Sample used	Sample size (N)	Dot probe stimuli	Mean bias toward threat	Peak bias toward threat	Mean bias away from threat	Peak bias away from threat	Attention bias variability
Zvielli et al. (2015)	Nicotine-deprived smokers	45	Smoking-related images	.67	.44	.58	.31	.46
Amir et al., (2016)	High trait anxiety	59	Threatening images (non-facial)	.81	.85	.30	.40	.73
Schafer et al. (2016)	Soldiers (pre-deployment)	144	Facial stimuli	.66	.67	.68	.55	.69
Zvielli et al., (2016)	Remitted depression	328	Facial stimuli	.80	.69	.80	.72	.87
	Healthy controls	82	Facial stimuli	.75	.70	.70	.59	.87
Rodebaugh et al. (2016)	Heterogeneous anxiety disorders (Time 1)	24	Threatening words	.83	.31	.87	.56	.95
	Heterogeneous anxiety disorders (Time 2)	24	Threatening words	.68	.59	.92	.65	.88

Note: All findings were statistically significant, $p < .05$.

Table 1.2 Test-retest reliability of trial-level bias score indices in Rodebaugh et al. (2016; Intra-class correlation coefficient)

Blocks/Trials Used	Mean bias toward threat	Peak bias toward threat	Mean bias away from threat	Peak bias away from threat	Attention bias variability
First block (304 trials)	.86	.49	.63	.20 (n.s.)	.90
Fourth block (304 trials)	.73	.47	.74	.15 (n.s.)	.79
First half of first block (152 trials)	.66	.26 (n.s.)	.51	.22 (n.s.)	.83
Second half of fourth block (152 trials)	.64	.37 (n.s.)	.64	.16 (n.s.)	.69

2 METHODS

2.1 Participants

As part of a larger study on cognitive biases in social anxiety disorder, participants were recruited with flyer and radio advertisements in the Atlanta community. In order to reduce financial barriers to participation and make the study accessible to a wider range of people, participants were offered vouchers for free parking adjacent to the University building in which the study was conducted. Additionally, participants were compensated \$50 for their time. For students at the university where the study was conducted, class credit was available as an alternative to cash compensation.

To be included in the study, participants needed to: meet DSM-IV criteria for a primary diagnosis of social anxiety disorder, be 18 years of age or older, be literate in English, and (if applicable) be taking the same dose of psychoactive medication for at least three months. Exclusion criteria for the study were any history of mania or psychosis, active suicidal or homicidal ideation, and current alcohol or substance abuse or dependence.

141 people expressed interest in the study and completed a phone screen. Of these, 115 respondents were judged potentially eligible and completed an in-person assessment of study inclusion and exclusion criteria. Forty-four participants were enrolled in the study and completed both study sessions. Demographic data for participants who completed the study are presented in Table 2.1.

2.2 Measures

2.2.1 Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Edition (SCID-I; First, Spitzer, Gibbon, & Williams, 2002)

The SCID-I is a widely used structured interview for the diagnosis of DSM-IV-TR Axis I disorders and demonstrates reliability and validity for a range of Axis I diagnoses (e.g. Lobbestael, Leurgans, & Arntz, 2010; DeMarce et al., 2013). The SCID-I is often used to assess social anxiety disorder in research settings (Crippa et al., 2008; Fracalanza et al., 2014), including studies specifically examining attention bias in socially anxious samples using visual dot-probe paradigms (Schneier et al., 2016). In the current study, the SCID-I was used during the in-person assessment to determine whether or not participants met criteria for primary social anxiety disorder, as well as to rule out history of mania and current alcohol or substance abuse or dependence. All assessments were administered by doctoral students in clinical psychology. These students were trained using instructional videos and via direct instruction from a licensed clinical psychologist. Assessors were also regularly supervised by a licensed clinical psychologist.

2.2.2 Facial dot probe task

A modified version of the facial dot probe task was developed based on a task and stimuli used by Bradley et al. (1997), the first reported dot probe study to use emotional face stimuli. Stimuli consist of 128 pictures of faces depicting 64 different people with neutral, happy, or threatening expressions. In total, the task uses 32 happy faces, 32 threatening faces, and 64 matching neutral faces.

A diagram of the task is presented in Figure 2.1. Before beginning the task, participants listened to a description of the task, read instructions for the task on a desktop computer, and

completed five practice trials while the experimenter stood by to answer any questions.

Participants then completed the actual task on the computer, which used EPrime software to present the stimulus trials.

During each trial two faces were presented side by side for 500 milliseconds. The faces were then replaced with an asterisk probe on either the left or right side of the screen for 1100 milliseconds. Participants used the keyboard to indicate the location of the probe as quickly as possible, pressing the “1” key if the probe appeared on the left and the “2” key if the probe appeared on the right. Following the probe, participants viewed a blank computer screen for a period of time that varied randomly in duration from 500 to 1250 milliseconds before beginning the next trial.

Sixty-four trials of emotional-neutral pairs (32 happy-neutral and 32 threatening-neutral) were presented twice for a total of 128 emotional-neutral trials. The side of the screen on which the emotional and neutral faces were presented was switched from the first to the second presentation of each face pair (i.e. threat on the left the first time, threat on the right the second time). Thirty-two trials in which participants viewed neutral-neutral pairs were combined with the 128 emotional-neutral trials for a total of 160 trials. The order of the trials was randomly generated for each new participant.

2.3 Procedure

2.3.1 Screening

All procedures for the current study were approved by the Georgia State University Institutional Review Board. The study was advertised via flyer and radio announcements. Interested potential participants were encouraged to contact the lab via phone or email in order to

begin the process of informed consent and a two-part screening process. All screening procedures were conducted by trained doctoral students in clinical psychology.

During the initial stage, potentially eligible participants were screened by phone for basic inclusion and exclusion criteria (e.g., obviously does not meet criteria for social anxiety disorder, medication changes within the last 3 months). Following the phone screening, potentially eligible individuals were invited for an in-person screening. At the beginning of this screening, respondents were consented and informed about their rights as research participants, all study procedures, and all risks and benefits of participation in the study. Those individuals who provided written consent to participate completed the SCID-I to establish primary diagnosis of social anxiety disorder, identify any comorbid disorders, and assess inclusion and exclusion criteria. Once enrolled, each participant was assigned a study ID number.

2.3.2 Time 1 assessment

Immediately following enrollment in the study, participants began the Time 1 assessment. During the Time 1 assessment, participants completed a demographic questionnaire and the facial dot probe task. At the end of this session, participants were compensated \$20. A Time 2 assessment was scheduled for approximately one week later and participants were offered a free parking pass for this session. They were also given a list of referrals for psychotherapy and other mental health resources.

2.3.3 Time 2 assessment

Participants returning for the Time 2 session re-completed the facial dot probe task. They were compensated \$30 for their completion of the study and offered a second copy of the mental health referral sheet from Time 1.

2.4 Data Analysis

2.4.1 Overview

In the current study, the author assessed test-retest reliability for each of the reaction time-based attention bias indices that have been developed to date for the dot probe task. Test-retest reliability was estimated using Pearson's r correlation. Internal consistency was also assessed for each time point using split-half reliability.

For all analyses, trials in which the participant responded with the wrong key press were excluded. This is in keeping with the majority of dot probe studies (e.g. Koster et al., 2004; Iacoviello et al., 2014). Participants for whom more than 20% of trials were invalid due to nonresponse, incorrect response, or reaction time outside of outlier cutoffs were also excluded from each analysis.

2.4.2 MacLeod et al. (1986) aggregated mean bias index

An aggregated mean index of attention bias was calculated from the task's 32 threat-congruent trials and 32 threat-incongruent trials based on the methodology of MacLeod et al. (1986). The author subtracted mean reaction time for threat-congruent trials from mean reaction time for threat-incongruent trials. Shorter mean reaction time for threat-congruent trials was interpreted to indicate attentional bias toward threatening faces.

2.4.3 Koster et al. (2004) aggregated mean bias indices

Two aggregated mean indices of attention bias were calculated from the task's threat-congruent trials, threat-incongruent trials, and 32 neutral-neutral trials based on the methodology of Koster et al. (2004). Mean reaction time for threat-congruent trials was subtracted from mean reaction time for neutral-neutral trials to create an index of vigilance to threat. Mean reaction

time for threat-incongruent trials was subtracted from mean reaction time for neutral-neutral trials to create an index of slowed disengagement from threat.

2.4.4 Zvielli et al. (2015) trial-level bias indices of mean bias, peak bias, and attention bias variability

Zvielli et al.'s (2015) trial-level bias score (TL-BS) method was used to calculate indices of mean bias toward threatening stimuli (Mean TL-BS_{POSITIVE}), peak bias toward threatening stimuli (Peak TL-BS_{POSITIVE}), mean bias away from threatening stimuli (Mean TL-BS_{NEGATIVE}), peak bias away from threatening stimuli (Peak TL-BS_{NEGATIVE}), and attention bias variability (TL-BS_{VARIABILITY}). These indices were calculated from the task's 32 threat-congruent trials and 32 threat-incongruent trials. TL-BS's were calculated for each participant by subtracting reaction time of each threat-congruent trial from its closest temporally contiguous threat-incongruent trial within a maximum distance of five trials. Trials were re-used if they were the closest eligible match to multiple other trials. Trials with no match within five trials were not included in analyses. TL-BS's were classified as "positive" if reaction time was shorter for the threat-congruent trial and "negative" if reaction time was shorter for the threat-incongruent trial. Mean TL-BS_{POSITIVE} and mean TL-BS_{NEGATIVE} were calculated by taking the average value for positive and negative TL-BS's, respectively. Peak TL-BS_{POSITIVE} and peak TL-BS_{NEGATIVE} were the most extreme positive and negative TL-BS's, respectively. To calculate TL-BS_{VARIABILITY} for each participant, the standard deviation was calculated for all of their TL-BS's.

2.4.5 Iacoviello et al. (2014) attention bias variability index

An index of attention bias variability was calculated based on the methodology of Iacoviello et al. (2014). All valid trials were split into eight sequential bins and the MacLeod et al. (1986) aggregated mean index were calculated separately for each bin. The standard

deviation of these values was calculated, then divided by mean reaction time of all valid threat-neutral trials for each participant. The resulting value was interpreted as an index for within-session attention bias variability.

2.4.6 Naim et al. (2015) attention bias variability index

A “moving average” index of attention bias variability was calculated from the task’s 32 threat-congruent trials and 32 neutral-neutral trials based on the methodology of Naim et al. (2015). The mean reaction time for each consecutive, overlapping set of 10 valid threat-congruent trials was subtracted from mean reaction time for each consecutive, overlapping set of 10 valid neutral-neutral trials (e.g. 1-10, 2-11, etc.). The standard deviation of difference values for all blocks was calculated, then divided by mean reaction time of all valid threat-congruent and neutral-neutral trials for each participant. The resulting value was interpreted as an index for within-session attention bias variability.

2.4.7 Arbitrary outlier cutoffs

Different strategies for handling outliers were compared for test-retest reliability using methodology based on Price et al. (2015) and Rodebaugh et al. (2016). Outliers were excluded below 200 ms and 300 ms and above 1000 and 1100 ms from onset of the probe. These low cutoff points were a replication of Price et al. (2015), the high 1000 ms cutoff was a replication of Rodebaugh et al. (2016), and the high 1100 ms cutoff represented the entire duration that face stimuli were presented in each trial. Low and high cutoffs were also examined at two and three standard deviations from the mean reaction time for each participant. Replicating Price et al. (2015), point cutoffs and standard deviation cutoffs were used individually and together in all possible combinations (e.g. low cutoffs at 200 ms, mean-2SD, 200 ms and mean-2SD, etc.). This resulted in 54 reliability estimates for each index.

2.4.8 Data-driven cutoffs

The author also tested the effect of Winsorizing outliers using Price et al.'s (2015) interquartile range cutoffs. These cutoffs were set at 1.5 interquartile ranges outside the 25th and 75th percentiles of all reaction times across all participants. Reaction time values outside of these cutoffs were changed to the cutoff value instead of being excluded. Reliability for each index described above were calculated using all outlier handling strategies, resulting in 55 total estimates per index.

Table 2.1 Sample demographics

Characteristic	M(SD)	Range
Age	22.61(6.69)	18-60
	<i>N</i>	%
Gender		
Male	13	29.5
Female	30	68.2
Other	1	2.3
Race/Ethnicity		
Black/African-American	24	54.5
White/Caucasian	12	27.3
Hispanic	2	4.5
Asian-American	5	11.4
Asian	1	2.3
Education		
Completed high school	6	13.6
Some college (1-2 years)	14	31.8
Some college (3+ years)	15	34.1
Completed college degree	6	13.6
Some graduate school	2	4.5
Completed graduate degree	1	2.3
Marital Status		
Single	40	90.9
Married	1	2.3
Living with someone	3	6.8
Income Level		
<\$5,000	10	22.7
\$5,000-\$10,000	5	11.4
\$10,000-\$20,000	8	18.2
\$20,000-\$30,000	8	18.2
\$30,000-\$50,000	5	11.4
>\$50,000	8	18.2

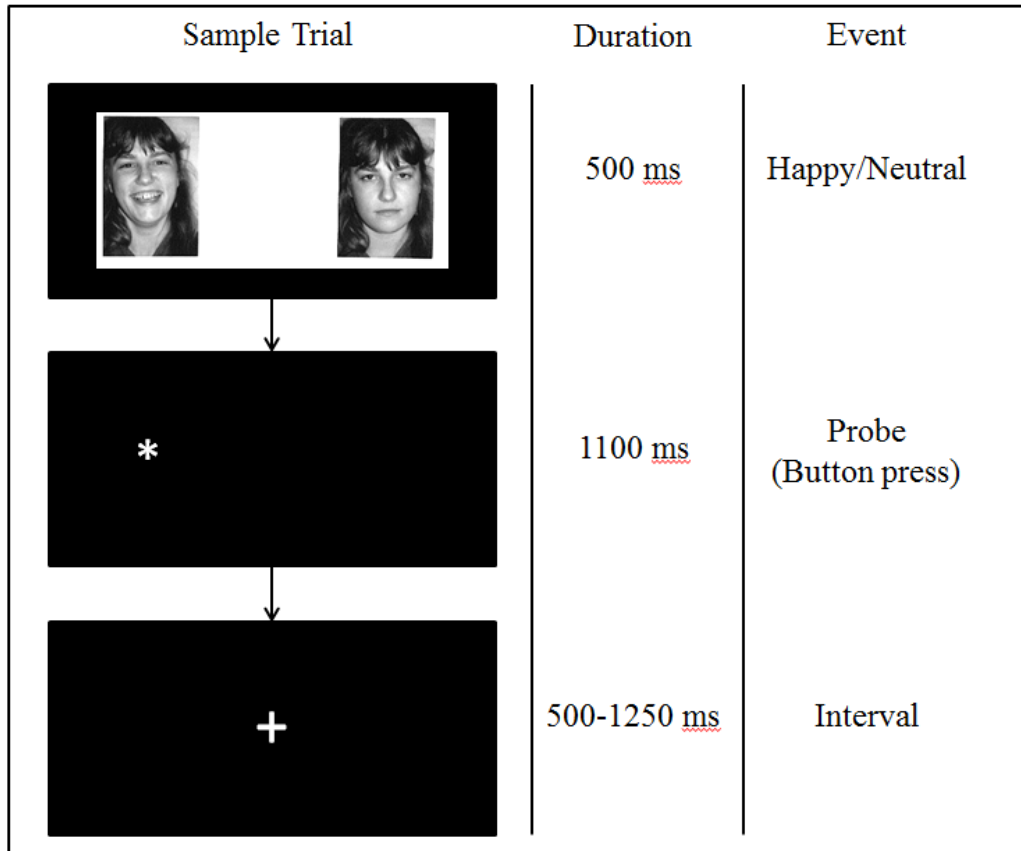


Figure 2.1 A diagram of the facial dot-probe task used in the current study

3 RESULTS

3.1 Aggregated Mean Indices (MacLeod et al., 1986; Koster et al., 2004)

None of the aggregated mean indices of attention bias consistently produced statistically significant estimates of test-retest reliability or split-half reliability across the 55 sets of outlier cutoffs examined. Table 3.1 displays reliability estimates for the MacLeod et al. (1986) index. There were 8 sets of outlier cutoffs that produced statistically significant test-retest reliability, 6 sets of outlier cutoffs that produced statistically significant split-half reliability estimates at Time 1, and none that produced statistically significant split-half reliability at Time 2.

Recall that Koster et al. (2004) designed indices to measure two types of attention bias: attentional vigilance for threat and slow disengagement from threat. Reliability estimates for attentional vigilance are shown in Table 3.2. There were no significant test-retest reliability estimates, 13 sets of outlier cutoff scores that produced statistically significant split-half reliability at Time 1, and 2 sets of outlier cutoff scores that produced statistically significant split-half reliability estimates at Time 2.

Reliability estimates for slow disengagement from threat are shown in Table 3.3. For this index, there was one statistically significant test-retest reliability estimate, 18 sets of outlier cutoff scores that produced statistically significant split-half reliability at Time 1, and no statistically significant split-half reliability estimates at Time 2.

Across the MacLeod et al. (1986) and the Koster et al.'s (2004) indices, there was not a single set of outlier cutoffs that produced significant test-retest reliability or significant Time 2 split-half reliability. Two sets of cutoffs (0 ms, mean-2SD; 1100 ms, mean+3SD and 200 ms, mean-2SD; 1100ms, mean+3SD) produced significant negative split-half reliability at Time 1 across all three indices.

Price et al.'s (2015) data-driven outlier handling method did not produce statistically significant estimates of test-retest or split-half reliability for any of the aggregated mean bias indices.

3.2 Trial-Level Bias Scores (Zvielli et al., 2015): Mean and Peak Attention Bias

Unlike the aggregated mean indices of attention bias, trial-level bias scores produced statistically significant estimates of test-retest reliability at most sets of the 55 outlier cutoffs examined. As shown in Table 3.4, Mean TL-BS_{POSITIVE} and Mean TL-BS_{NEGATIVE} showed statistically significant test-retest reliability across all sets of outlier cutoffs. Peak TL-BS_{POSITIVE} and Peak TL-BS_{NEGATIVE} showed statistically significant test-retest reliability for 35 and 36 of the outlier cutoffs, respectively. Test-retest reliability for mean and peak TL-BS indices are also shown in Figures 3.6-3.9.

Split-half reliability was not calculated for trial-level bias score indices. This is because Zvielli et al. (2015) recommended a minimum of 80 valid trial-level bias scores within a session for valid computation of split-half reliability. The current study's data did not meet this requirement.

Price et al.'s (2015) data-driven cutoff method produced significant test-retest reliability for Mean TL-BS_{POSITIVE}, Mean TL-BS_{NEGATIVE}, and Peak TL-BS_{POSITIVE}, but not Peak TL-BS_{NEGATIVE} (see Table 3.4).

3.3 Attention Bias Variability Indices (Iacoviello et al., 2014; Naim et al., 2015; Zvelli et al., 2015)

The number of statistically significant reliability estimates varied widely across the three indices of attention bias variability (Iacoviello et al., 2014; Naim et al., 2015; and Zvelli et al., 2015). Iacoviello et al.'s (2014) attention bias variability index produced statistically significant

test-retest reliability with 11 of the 55 sets of outlier cutoffs (see Table 3.5). At Time 1, two sets of cutoffs produced statistically significant split-half reliability estimates. At Time 2, 46 sets of outlier cutoffs (including the two sets of cuts-off producing significant split-half reliability at Time 1), produced statistically significant split-half reliability estimates, with the highest estimate at $r=.404, p=.006$. Price et al.'s (2015) data-driven cutoff method produced significant test-retest reliability, but not significant split-half reliability at Time 1 or Time 2.

Naim et al.'s (2015) attention bias variability did not produce statistically significant test-retest reliability nor Time 1 split-half reliability estimates at any set of outlier cutoffs (see Table 3.6). There were 10 statistically significant test-retest reliability estimates at Time 2. Price et al.'s (2015) data-driven cutoff method did not yield significant test-retest reliability, nor significant Time 1 or Time 2 split-half reliability.

Zvelli et al.'s (2015) TL-BS_{VARIABILITY} showed statistically significant test-retest reliability across all sets of outlier cutoffs, including Price et al.'s (2015) data-driven cutoff method (see Table 3.4).

3.4 Magnitude of Test-Retest and Split-Half Reliability

Table 3.7 shows the magnitudes of significant test-retest reliability estimates across all indices. Ranges were calculated for statistically significant reliability estimates and for all reliability estimates regardless of significance. The distributions of test-retest reliability estimates for all indices are presented visually in Figures 3.1-3.10.

TL-BS_{VARIABILITY} demonstrated the highest test-retest reliability, $r = .630, p < .001$, followed closely by Mean TL-BS_{POSITIVE}, $r = .627, p < .001$, and Peak TL-BS_{POSITIVE}, $r = .626, p < .001$. The smallest ranges for all test-retest reliability estimates were found for Mean TL-BS_{NEGATIVE}, Range = .175, TL-BS_{VARIABILITY}, Range = .183, and Mean TL-BS_{POSITIVE}, Range =

.219. All three of these indices had statistically significant test-retest reliability for all sets of outlier cutoffs.

Tables 11-12 show the magnitudes of significant split-half reliability estimates across all indices, for Time 1 and Time 2, respectively. At Time 1, Iacoviello et al. (2014) displayed significant positive split-half reliability for two sets of outlier cutoffs, both $r = .339$, $p = .041$. No other index displayed significant positive split-half reliability at Time 1. However, MacLeod et al. (1986) and both of the Koster et al. (2004) indices displayed significant negative split-half reliability for some sets of outlier cutoffs.

At Time 2, Iacoviello et al. (2014) demonstrated the highest split-half reliability, $r = .491$, $p = .001$. Naim et al.'s (2015) highest significant split-half reliability estimate was $r = .299$, $p = .035$, and Koster et al.'s (2004) vigilance index had its highest estimate at $r = .293$, $p = .039$. No other indices displayed significant split-half reliability at Time 2.

3.5 Outlier Cutoffs with Highest Reliability Estimates

Table 3.10 shows the 10 sets of outlier cutoffs that produced significant test-retest reliability estimates for at least 6 of the indices. Five of these cutoffs produced significant estimates across the MacLeod et al. (1986) aggregated mean index, Iacoviello et al. (2014) index, and all TL-BS indices for a total of seven significant estimates. Five of these cutoffs produced significant estimates for either the MacLeod et al. (1986) aggregated mean index or the Iacoviello et al. (2014) attention bias variability index and were significant for all TL-BS indices for a total of six significant estimates. Of the 10 sets of outlier cutoffs shown in Table 3.10, all used lenient cutoffs at the high end: eight used 1100 ms and mean+3SD as high-end cutoffs, whereas the remaining two used 1000 ms and mean+3SD. Low-end cutoffs were more varied, ranging from no cutoffs at all (i.e. 0 ms) to more restrictive sets of cutoffs like 300 ms and mean-

3SD. Two sets of cutoffs produced the highest test-retest reliability estimates measured in the current study across Iacoviello et al., (2014), Mean TL-BS_{TOWARD}, Peak TL-BS_{TOWARD}, and TL-BS_{VARIABILITY}. These were low-end 300ms, mean-3SD, high-end 1100ms, mean+3SD and low-end cutoffs 300ms, high-end 1100ms, mean+3SD.

Table 3.1 MacLeod et al. (1986) index: test-retest and split-half reliability by outlier cutoffs

Low-End Cutoffs	High-End Cutoffs	Test-Retest Reliability			Split-Half Reliability (Time 1)			Split-Half Reliability (Time 2)		
		N	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>p</i>
0ms	1000ms	44	.189	.22	47	-.256	.083	50	.067	.646
0ms	1000ms, +2SD	44	.244	.111	46	-.25	.094	50	.065	.655
0ms	1000ms, +3SD	44	.329	.029*	47	-.261	.076	50	.023	.876
0ms	1100ms	44	.042	.788	47	-.273	.064	50	-.046	.750
0ms	1100ms, +2SD	44	.244	.111	46	-.25	.094	50	.065	.655
0ms	1100ms, +3SD	44	.386	.010*	47	-.259	.079	50	-.053	.717
0ms, -2SD	1000ms	44	.121	.433	47	-.279	.058	50	-.040	.781
0ms, -2SD	1000ms, +2SD	44	.134	.386	46	-.257	.084	50	-.064	.658
0ms, -2SD	1000ms, +3SD	44	.242	.113	47	-.300	.041*	50	-.086	.553
0ms, -2SD	1100ms	44	-.049	.751	47	-.304	.038*	50	-.144	.320
0ms, -2SD	1100ms, +2SD	44	.134	.386	46	-.257	.084	50	-.064	.658
0ms, -2SD	1100ms, +3SD	44	.301	.047*	47	-.303	.038*	50	-.150	.298
0ms, -3SD	1000ms	44	.178	.247	47	-.256	.083	50	.004	.980
0ms, -3SD	1000ms, +2SD	44	.227	.139	46	-.250	.094	50	.011	.941
0ms, -3SD	1000ms, +3SD	44	.315	.037*	47	-.261	.076	50	-.025	.865
0ms, -3SD	1100ms	44	.027	.860	47	-.273	.064	50	-.098	.498
0ms, -3SD	1100ms, +2SD	44	.227	.139	46	-.250	.094	50	.011	.941
0ms, -3SD	1100ms, +3SD	44	.374	.012*	47	-.261	.076	50	-.094	.514
200ms	1000ms	44	.138	.372	47	-.263	.074	50	-.038	.794
200ms	1000ms, +2SD	44	.196	.203	46	-.237	.112	50	-.046	.751
200ms	1000ms, +3SD	44	.279	.067	47	-.269	.067	50	-.073	.613
200ms	1100ms	44	-.022	.887	47	-.280	.056	50	-.135	.348
200ms	1100ms, +2SD	44	.196	.203	46	-.237	.112	50	-.046	.751
200ms	1100ms, +3SD	44	.343	.023*	47	-.267	.070	50	-.139	.336
200ms, -2SD	1000ms	44	.121	.433	47	-.279	.058	50	-.040	.781
200ms, -2SD	1000ms, +2SD	44	.134	.386	46	-.257	.084	50	-.064	.658

200ms, -2SD	1000ms, +3SD	44	.242	.113	47	-.300	.041*	50	-.086	.553
200ms, -2SD	1100ms	44	-.049	.751	47	-.304	.038*	50	-.144	.320
200ms, -2SD	1100ms, +2SD	44	.134	.386	46	-.257	.084	50	-.064	.658
200ms, -2SD	1100ms, +3SD	44	.301	.047*	47	-.303	.038*	50	-.150	.298
200ms, -3SD	1000ms	44	.138	.372	47	-.263	.074	50	-.038	.794
200ms, -3SD	1000ms, +2SD	44	.196	.203	46	-.237	.112	50	-.046	.751
200ms, -3SD	1000ms, +3SD	44	.279	.067	47	-.269	.067	50	-.073	.613
200ms, -3SD	1100ms	44	-.022	.887	47	-.28	.056	50	-.135	.348
200ms, -3SD	1100ms, +2SD	44	.196	.203	46	-.237	.112	50	-.046	.751
200ms, -3SD	1100ms, +3SD	44	.343	.023*	47	-.267	.070	50	-.139	.336
300ms	1000ms	42	.097	.542	47	-.258	.080	46	-.012	.936
300ms	1000ms, +2SD	39	.099	.547	44	-.131	.397	45	-.004	.979
300ms	1000ms, +3SD	41	.207	.193	47	-.274	.063	45	-.077	.616
300ms	1100ms	42	-.068	.670	47	-.276	.061	46	-.114	.449
300ms	1100ms, +2SD	39	.099	.547	44	-.131	.397	45	-.004	.979
300ms	1100ms, +3SD	41	.288	.068	47	-.27	.066	45	-.156	.306
300ms, -2SD	1000ms	42	.066	.679	47	-.253	.086	46	-.011	.941
300ms, -2SD	1000ms, +2SD	39	.052	.753	44	-.117	.450	45	.001	.993
300ms, -2SD	1000ms, +3SD	41	.164	.305	47	-.276	.060	45	-.074	.631
300ms, -2SD	1100ms	42	-.11	.489	47	-.278	.058	46	-.114	.450
300ms, -2SD	1100ms, +2SD	39	.052	.753	44	-.117	.450	45	.001	.993
300ms, -2SD	1100ms, +3SD	41	.234	.140	47	-.279	.057	45	-.154	.313
300ms, -3SD	1000ms	42	.097	.542	47	-.258	.080	46	-.012	.936
300ms, -3SD	1000ms, +2SD	39	.099	.547	44	-.131	.397	45	-.004	.979
300ms, -3SD	1000ms, +3SD	41	.207	.193	47	-.274	.063	45	-.077	.616
300ms, -3SD	1100ms	42	-.068	.670	47	-.276	.061	46	-.114	.449
300ms, -3SD	1100ms, +2SD	39	.099	.547	44	-.131	.397	45	-.004	.979
300ms, -3SD	1100ms, +3SD	41	.288	.068	47	-.27	.066	45	-.156	.306
Data-Driven	Data-Driven	44	.102	.511	47	-.279	.057	50	-.028	.849

Note. All values of MacLeod et al. (1986) index are included.

* $p < .05$.

Table 3.2 Koster et al. (2004) vigilance index: test-retest and split-half reliability by outlier cutoffs

Low-End Cutoffs	High-End Cutoffs	Test-Retest Reliability			Split-Half Reliability (Time 1)			Split-Half Reliability (Time 2)		
		N	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>p</i>
0ms	1000ms	44	-.014	.926	47	-.292	.047*	50	.075	.606
0ms	1000ms, +2SD	44	.067	.664	46	-.248	.097	50	.256	.072
0ms	1000ms, +3SD	44	.104	.503	47	-.259	.079	50	.13	.370
0ms	1100ms	44	-.113	.465	47	-.256	.082	50	.088	.543
0ms	1100ms, +2SD	44	.067	.664	46	-.248	.097	50	.256	.072
0ms	1100ms, +3SD	44	.086	.580	47	-.305	.037*	50	.115	.428
0ms, -2SD	1000ms	44	-.047	.761	47	-.289	.048*	50	.076	.602
0ms, -2SD	1000ms, +2SD	44	-.008	.961	46	-.221	.140	50	.252	.078
0ms, -2SD	1000ms, +3SD	44	.073	.636	47	-.252	.087	50	.126	.384
0ms, -2SD	1100ms	44	-.135	.382	47	-.263	.074	50	.109	.450
0ms, -2SD	1100ms, +2SD	44	-.008	.961	46	-.221	.140	50	.252	.078
0ms, -2SD	1100ms, +3SD	44	.056	.719	47	-.314	.031*	50	.139	.337
0ms, -3SD	1000ms	44	-.019	.902	47	-.292	.047*	50	.103	.478
0ms, -3SD	1000ms, +2SD	44	.046	.767	46	-.248	.097	50	.293	.039*
0ms, -3SD	1000ms, +3SD	44	.101	.516	47	-.259	.079	50	.166	.248
0ms, -3SD	1100ms	44	-.118	.447	47	-.256	.082	50	.114	.431
0ms, -3SD	1100ms, +2SD	44	.046	.767	46	-.248	.097	50	.293	.039*
0ms, -3SD	1100ms, +3SD	44	.083	.593	47	-.305	.037*	50	.152	.294
200ms	1000ms	44	-.034	.826	47	-.289	.049*	50	.078	.590
200ms	1000ms, +2SD	44	.023	.883	46	-.222	.139	50	.276	.052
200ms	1000ms, +3SD	44	.084	.588	47	-.248	.093	50	.142	.327
200ms	1100ms	44	-.132	.391	47	-.251	.088	50	.092	.526
200ms	1100ms, +2SD	44	.023	.883	46	-.222	.139	50	.276	.052
200ms	1100ms, +3SD	44	.066	.671	47	-.296	.043*	50	.127	.381
200ms, -2SD	1000ms	44	-.047	.761	47	-.289	.048*	50	.076	.602
200ms, -2SD	1000ms, +2SD	44	-.008	.961	46	-.221	.140	50	.252	.078

200ms, -2SD	1000ms, +3SD	44	.073	.636	47	-.252	.087	50	.126	.384
200ms, -2SD	1100ms	44	-.135	.382	47	-.263	.074	50	.109	.450
200ms, -2SD	1100ms, +2SD	44	-.008	.961	46	-.221	.140	50	.252	.078
200ms, -2SD	1100ms, +3SD	44	.056	.719	47	-.314	.031*	50	.139	.337
200ms, -3SD	1000ms	44	-.034	.826	47	-.289	.049*	50	.078	.590
200ms, -3SD	1000ms, +2SD	44	.023	.883	46	-.222	.139	50	.276	.052
200ms, -3SD	1000ms, +3SD	44	.084	.588	47	-.248	.093	50	.142	.327
200ms, -3SD	1100ms	44	-.132	.391	47	-.251	.088	50	.092	.526
200ms, -3SD	1100ms, +2SD	44	.023	.883	46	-.222	.139	50	.276	.052
200ms, -3SD	1100ms, +3SD	44	.066	.671	47	-.296	.043*	50	.127	.381
300ms	1000ms	42	-.026	.868	47	-.267	.069	46	-.084	.579
300ms	1000ms, +2SD	39	.106	.521	44	-.237	.121	45	.133	.382
300ms	1000ms, +3SD	41	.118	.461	47	-.238	.108	45	-.031	.839
300ms	1100ms	42	-.121	.446	47	-.228	.123	46	-.023	.879
300ms	1100ms, +2SD	39	.106	.521	44	-.237	.121	45	.133	.382
300ms	1100ms, +3SD	41	.096	.549	47	-.285	.052	45	-.017	.909
300ms, -2SD	1000ms	42	-.004	.980	47	-.253	.086	46	-.081	.591
300ms, -2SD	1000ms, +2SD	39	.128	.438	44	-.216	.160	45	.14	.360
300ms, -2SD	1000ms, +3SD	41	.141	.378	47	-.226	.127	45	-.026	.865
300ms, -2SD	1100ms	42	-.103	.515	47	-.226	.127	46	-.02	.893
300ms, -2SD	1100ms, +2SD	39	.128	.438	44	-.216	.160	45	.14	.360
300ms, -2SD	1100ms, +3SD	41	.121	.452	47	-.289	.048*	45	-.013	.935
300ms, -3SD	1000ms	42	-.026	.868	47	-.267	.069	46	-.084	.579
300ms, -3SD	1000ms, +2SD	39	.106	.521	44	-.237	.121	45	.133	.382
300ms, -3SD	1000ms, +3SD	41	.118	.461	47	-.238	.108	45	-.031	.839
300ms, -3SD	1100ms	42	-.121	.446	47	-.228	.123	46	-.023	.879
300ms, -3SD	1100ms, +2SD	39	.106	.521	44	-.237	.121	45	.133	.382
300ms, -3SD	1100ms, +3SD	41	.096	.549	47	-.285	.052	45	-.017	.909
Data-Driven	Data-Driven	44	-.049	.754	47	-.232	.116	50	.104	.472

Note. All values of Koster et al. (2004) vigilance index are included.

* $p < .05$.

Table 3.3 Koster et al. (2004) slow disengagement index: test-retest and split-half reliability by outlier cutoffs

Low-End Cutoffs	High-End Cutoffs	Test-Retest Reliability			Split-Half Reliability (Time 1)			Split-Half Reliability (Time 2)		
		N	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>p</i>
0ms	1000ms	44	.221	.150	50	-.211	.155	50	-.061	.675
0ms	1000ms, +2SD	44	.000	.999	50	-.291	.050	50	.004	.978
0ms	1000ms, +3SD	44	.08	.608	50	-.273	.063	50	-.121	.404
0ms	1100ms	44	.292	.055	50	-.218	.140	50	.015	.916
0ms	1100ms, +2SD	44	.000	.999	50	-.291	.050	50	.004	.978
0ms	1100ms, +3SD	44	.116	.452	50	-.322	.027*	50	-.047	.743
0ms, -2SD	1000ms	44	.195	.204	50	-.205	.168	50	-.05	.733
0ms, -2SD	1000ms, +2SD	44	-.045	.771	50	-.291	.050	50	.033	.822
0ms, -2SD	1000ms, +3SD	44	.069	.658	50	-.273	.064	50	-.120	.406
0ms, -2SD	1100ms	44	.270	.077	50	-.212	.152	50	.032	.826
0ms, -2SD	1100ms, +2SD	44	-.045	.771	50	-.291	.050	50	.033	.822
0ms, -2SD	1100ms, +3SD	44	.107	.491	50	-.321	.028*	50	-.041	.779
0ms, -3SD	1000ms	44	.239	.117	50	-.211	.155	50	-.012	.935
0ms, -3SD	1000ms, +2SD	44	.003	.985	50	-.291	.050	50	.088	.541
0ms, -3SD	1000ms, +3SD	44	.102	.509	50	-.273	.063	50	-.070	.629
0ms, -3SD	1100ms	44	.308	.042*	50	-.218	.140	50	.059	.683
0ms, -3SD	1100ms, +2SD	44	.003	.985	50	-.291	.050	50	.088	.541
0ms, -3SD	1100ms, +3SD	44	.137	.376	50	-.322	.027*	50	.001	.994
200ms	1000ms	44	.206	.179	50	-.208	.160	50	-.018	.902
200ms	1000ms, +2SD	44	-.038	.806	50	-.288	.052	50	.082	.572
200ms	1000ms, +3SD	44	.058	.709	50	-.27	.066	50	-.073	.614
200ms	1100ms	44	.280	.065	50	-.216	.144	50	.053	.713
200ms	1100ms, +2SD	44	-.038	.806	50	-.288	.052	50	.082	.572
200ms	1100ms, +3SD	44	.096	.536	50	-.319	.029*	50	-.003	.985
200ms, -2SD	1000ms	44	.195	.204	50	-.205	.168	50	-.050	.733
200ms, -2SD	1000ms, +2SD	44	-.045	.771	50	-.291	.050	50	.033	.822

200ms, -2SD	1000ms, +3SD	44	.069	.658	50	-.273	.064	50	-.120	.406
200ms, -2SD	1100ms	44	.270	.077	50	-.212	.152	50	.032	.826
200ms, -2SD	1100ms, +2SD	44	-.045	.771	50	-.291	.050	50	.033	.822
200ms, -2SD	1100ms, +3SD	44	.107	.491	50	-.321	.028*	50	-.041	.779
200ms, -3SD	1000ms	44	.206	.179	50	-.208	.160	50	-.018	.902
200ms, -3SD	1000ms, +2SD	44	-.038	.806	50	-.288	.052	50	.082	.572
200ms, -3SD	1000ms, +3SD	44	.058	.709	50	-.27	.066	50	-.073	.614
200ms, -3SD	1100ms	44	.280	.065	50	-.216	.144	50	.053	.713
200ms, -3SD	1100ms, +2SD	44	-.038	.806	50	-.288	.052	50	.082	.572
200ms, -3SD	1100ms, +3SD	44	.096	.536	50	-.319	.029*	50	-.003	.985
300ms	1000ms	42	.187	.235	46	-.204	.170	46	-.164	.277
300ms	1000ms, +2SD	39	-.058	.726	45	-.305	.044*	45	-.009	.956
300ms	1000ms, +3SD	41	.020	.899	45	-.298	.042*	45	-.149	.328
300ms	1100ms	42	.263	.092	46	-.204	.169	46	-.050	.740
300ms	1100ms, +2SD	39	-.058	.726	45	-.305	.044*	45	-.009	.956
300ms	1100ms, +3SD	41	.068	.671	45	-.345	.017*	45	-.044	.773
300ms, -2SD	1000ms	42	.197	.210	46	-.204	.170	46	-.166	.269
300ms, -2SD	1000ms, +2SD	39	-.047	.777	45	-.305	.044*	45	.00	.998
300ms, -2SD	1000ms, +3SD	41	.034	.832	45	-.298	.042*	45	-.147	.336
300ms, -2SD	1100ms	42	.271	.082	46	-.204	.169	46	-.053	.728
300ms, -2SD	1100ms, +2SD	39	-.047	.777	45	-.305	.044*	45	.00	.998
300ms, -2SD	1100ms, +3SD	41	.081	.615	45	-.345	.017*	45	-.042	.784
300ms, -3SD	1000ms	42	.187	.235	46	-.204	.170	46	-.164	.277
300ms, -3SD	1000ms, +2SD	39	-.058	.726	45	-.305	.044*	45	-.009	.956
300ms, -3SD	1000ms, +3SD	41	.020	.899	45	-.298	.042*	45	-.149	.328
300ms, -3SD	1100ms	42	.263	.092	46	-.204	.169	46	-.050	.740
300ms, -3SD	1100ms, +2SD	39	-.058	.726	45	-.305	.044*	45	-.009	.956
300ms, -3SD	1100ms, +3SD	41	.068	.671	45	-.345	.017*	45	-.044	.773
Data-Driven	Data-Driven	44	.248	.105	50	-.142	.340	50	.010	.943

Note. All values of Koster et al. (2004) slow disengagement index are included.

* $p < .05$.

Table 3.4 Zvielli et al. (2015) trial-level bias score indices: test-retest reliability by outlier cutoffs

Low-End Cutoffs	High-End Cutoffs	N	Mean TL-BS Positive		Mean TL-BS Negative		Peak TL-BS Positive		Peak TL-BS Negative		TL-BS Variability	
			<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
0ms	1000ms	44	.458	.002**	.327	.030*	.015	.924	.209	.174	.501	.001**
0ms	1000ms, +2SD	44	.443	.003**	.406	.006**	.243	.111	.471	.001**	.534	<.001**
0ms	1000ms, +3SD	44	.464	.002**	.388	.009**	.425	.004**	.423	.004**	.531	<.001**
0ms	1100ms	44	.543	<.001**	.336	.026*	.141	.362	.263	.085	.534	<.001**
0ms	1100ms, +2SD	44	.443	.003**	.406	.006**	.243	.111	.471	.001**	.534	<.001**
0ms	1100ms, +3SD	44	.549	<.001**	.39	.009**	.532	<.001**	.423	.004**	.580	<.001**
0ms, -2SD	1000ms	44	.412	.005**	.308	.042*	-.018	.907	.180	.242	.461	.002**
0ms, -2SD	1000ms, +2SD	44	.464	.002**	.381	.011*	.332	.028*	.409	.006**	.554	<.001**
0ms, -2SD	1000ms, +3SD	44	.422	.004**	.375	.012*	.392	.009**	.385	.010*	.503	<.001**
0ms, -2SD	1100ms	44	.425	.004**	.320	.034*	.031	.842	.245	.109	.459	.002**
0ms, -2SD	1100ms, +2SD	44	.464	.002**	.381	.011*	.332	.028*	.409	.006**	.554	<.001**
0ms, -2SD	1100ms, +3SD	44	.422	.004**	.376	.012*	.405	.006**	.385	.010*	.506	<.001**
0ms, -3SD	1000ms	44	.470	.001**	.327	.030*	.013	.936	.209	.174	.504	<.001**
0ms, -3SD	1000ms, +2SD	44	.473	.001**	.406	.006**	.299	.049*	.471	.001**	.560	<.001**
0ms, -3SD	1000ms, +3SD	44	.492	.001**	.388	.009**	.448	.002**	.423	.004**	.548	<.001**
0ms, -3SD	1100ms	44	.554	<.001**	.336	.026*	.140	.366	.263	.085	.537	<.001**
0ms, -3SD	1100ms, +2SD	44	.473	.001**	.406	.006**	.299	.049*	.471	.001**	.560	<.001**
0ms, -3SD	1100ms, +3SD	44	.575	<.001**	.390	.009**	.556	<.001**	.423	.004**	.598	<.001**
200ms	1000ms	44	.468	.001**	.301	.047*	.013	.936	.180	.242	.493	.001**
200ms	1000ms, +2SD	44	.473	.001**	.368	.014*	.299	.049*	.409	.006**	.540	<.001**
200ms	1000ms, +3SD	44	.492	.001**	.361	.016*	.448	.002**	.385	.010*	.536	<.001**
200ms	1100ms	44	.552	<.001**	.313	.039*	.140	.366	.245	.109	.528	<.001**
200ms	1100ms, +2SD	44	.473	.001**	.368	.014*	.299	.049*	.409	.006**	.540	<.001**
200ms	1100ms, +3SD	44	.575	<.001**	.362	.016*	.556	<.001**	.385	.010*	.587	<.001**
200ms, -2SD	1000ms	44	.412	.005**	.308	.042*	-.018	.907	.180	.242	.461	.002**
200ms, -2SD	1000ms, +2SD	44	.464	.002**	.381	.011*	.332	.028*	.409	.006**	.554	<.001**

200ms, -2SD	1000ms, +3SD	44	.422	.004**	.375	.012*	.392	.009**	.385	.010*	.503	<.001**
200ms, -2SD	1100ms	44	.425	.004**	.320	.034*	.031	.842	.245	.109	.459	.002**
200ms, -2SD	1100ms, +2SD	44	.464	.002**	.381	.011*	.332	.028*	.409	.006**	.554	<.001**
200ms, -2SD	1100ms, +3SD	44	.422	.004**	.376	.012*	.405	.006**	.385	.010*	.506	<.001**
200ms, -3SD	1000ms	44	.468	.001**	.301	.047*	.013	.936	.180	.242	.493	.001**
200ms, -3SD	1000ms, +2SD	44	.473	.001**	.368	.014*	.299	.049*	.409	.006**	.540	<.001**
200ms, -3SD	1000ms, +3SD	44	.492	.001**	.361	.016*	.448	.002**	.385	.010*	.536	<.001**
200ms, -3SD	1100ms	44	.552	<.001**	.313	.039*	.140	.366	.245	.109	.528	<.001**
200ms, -3SD	1100ms, +2SD	44	.473	.001**	.368	.014*	.299	.049*	.409	.006**	.540	<.001**
200ms, -3SD	1100ms, +3SD	44	.575	<.001**	.362	.016*	.556	<.001**	.385	.010*	.587	<.001**
300ms	1000ms	42	.489	.001**	.366	.017*	.025	.875	.157	.320	.496	.001**
300ms	1000ms, +2SD	39	.490	.002**	.476	.002**	.420	.008**	.444	.005**	.581	<.001**
300ms	1000ms, +3SD	41	.551	<.001**	.412	.007**	.525	<.001**	.376	.015*	.581	<.001**
300ms	1100ms	42	.570	<.001**	.374	.015*	.132	.406	.229	.145	.535	<.001**
300ms	1100ms, +2SD	39	.490	.002**	.476	.002**	.420	.008**	.444	.005**	.581	<.001**
300ms	1100ms, +3SD	41	.627	<.001**	.413	.007**	.626	<.001**	.376	.015*	.630	<.001**
300ms, -2SD	1000ms	42	.408	.007**	.366	.017*	-.035	.825	.157	.320	.447	.003**
300ms, -2SD	1000ms, +2SD	39	.440	.005**	.476	.002**	.404	.011*	.444	.005**	.561	<.001**
300ms, -2SD	1000ms, +3SD	41	.460	.002**	.412	.007**	.473	.002**	.376	.015*	.532	<.001**
300ms, -2SD	1100ms	42	.420	.006**	.374	.015*	.023	.886	.229	.145	.452	.003**
300ms, -2SD	1100ms, +2SD	39	.440	.005**	.476	.002**	.404	.011*	.444	.005**	.561	<.001**
300ms, -2SD	1100ms, +3SD	41	.455	.003**	.413	.007**	.473	.002**	.376	.015*	.532	<.001**
300ms, -3SD	1000ms	42	.489	.001**	.366	.017*	.025	.875	.157	.320	.496	.001**
300ms, -3SD	1000ms, +2SD	39	.490	.002**	.476	.002**	.420	.008**	.444	.005**	.581	<.001**
300ms, -3SD	1000ms, +3SD	41	.551	<.001**	.412	.007**	.525	<.001**	.376	.015*	.581	<.001**
300ms, -3SD	1100ms	42	.570	<.001**	.374	.015*	.132	.406	.229	.145	.535	<.001**
300ms, -3SD	1100ms, +2SD	39	.490	.002**	.476	.002**	.420	.008**	.444	.005**	.581	<.001**
300ms, -3SD	1100ms, +3SD	41	.627	<.001**	.413	.007**	.626	<.001**	.376	.015*	.630	<.001**
Data-Driven	Data-Driven	44	.536	<.001**	.412	.005**	.469	.001**	.150	.332	.506	<.001**

Note. All values for all Zvielli et al. (2015) trial-level bias score indices are included.

* $p < .05$. ** $p < .01$.

Table 3.5 Iacoviello et al. (2014) attention bias variability index: test-retest and split-half reliability by outlier cutoffs

Low-End Cutoffs	High-End Cutoffs	Test-Retest Reliability			Split-Half Reliability (Time 1)			Split-Half Reliability (Time 2)		
		N	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>p</i>
0ms	1000ms	44	.270	.077	47	.052	.728	50	.294	.038*
0ms	1000ms, +2SD	44	.280	.065	46	.266	.074	50	.349	.013*
0ms	1000ms, +3SD	44	.295	.052	47	.147	.324	50	.313	.027*
0ms	1100ms	44	.389	.009**	47	.059	.694	50	.283	.046*
0ms	1100ms, +2SD	44	.280	.065	46	.266	.074	50	.349	.013*
0ms	1100ms, +3SD	44	.441	.003**	47	.125	.403	50	.341	.015*
0ms, -2SD	1000ms	44	.233	.127	47	.043	.774	50	.28	.049*
0ms, -2SD	1000ms, +2SD	44	.222	.148	46	.286	.054	50	.429	.002**
0ms, -2SD	1000ms, +3SD	44	.264	.083	47	.136	.363	50	.341	.015*
0ms, -2SD	1100ms	44	.114	.461	47	.051	.733	50	.251	.079
0ms, -2SD	1100ms, +2SD	44	.222	.148	46	.286	.054	50	.429	.002**
0ms, -2SD	1100ms, +3SD	44	.272	.074	47	.101	.499	50	.369	.008**
0ms, -3SD	1000ms	44	.253	.098	47	.052	.728	50	.263	.065
0ms, -3SD	1000ms, +2SD	44	.276	.070	46	.266	.074	50	.316	.026*
0ms, -3SD	1000ms, +3SD	44	.304	.045*	47	.147	.324	50	.250	.079
0ms, -3SD	1100ms	44	.374	.012*	47	.059	.694	50	.256	.073
0ms, -3SD	1100ms, +2SD	44	.276	.070	46	.266	.074	50	.316	.026*
0ms, -3SD	1100ms, +3SD	44	.451	.002**	47	.125	.403	50	.283	.046*
200ms	1000ms	44	.239	.118	47	.050	.740	50	.333	.018*
200ms	1000ms, +2SD	44	.217	.158	46	.288	.052	50	.415	.003**
200ms	1000ms, +3SD	44	.278	.067	47	.140	.348	50	.337	.017*
200ms	1100ms	44	.362	.016*	47	.057	.704	50	.319	.024*
200ms	1100ms, +2SD	44	.217	.158	46	.288	.052	50	.415	.003**
200ms	1100ms, +3SD	44	.431	.003**	47	.118	.431	50	.360	.010*
200ms, -2SD	1000ms	44	.233	.127	47	.043	.774	50	.280	.049*
200ms, -2SD	1000ms, +2SD	44	.222	.148	46	.286	.054	50	.429	.002**

200ms, -2SD	1000ms, +3SD	44	.264	.083	47	.136	.363	50	.341	.015*
200ms, -2SD	1100ms	44	.114	.461	47	.051	.733	50	.251	.079
200ms, -2SD	1100ms, +2SD	44	.222	.148	46	.286	.054	50	.429	.002**
200ms, -2SD	1100ms, +3SD	44	.272	.074	47	.101	.499	50	.369	.008**
200ms, -3SD	1000ms	44	.239	.118	47	.050	.740	50	.333	.018*
200ms, -3SD	1000ms, +2SD	44	.217	.158	46	.288	.052	50	.415	.003**
200ms, -3SD	1000ms, +3SD	44	.278	.067	47	.140	.348	50	.337	.017*
200ms, -3SD	1100ms	44	.362	.016*	47	.057	.704	50	.319	.024*
200ms, -3SD	1100ms, +2SD	44	.217	.158	46	.288	.052	50	.415	.003**
200ms, -3SD	1100ms, +3SD	44	.431	.003**	47	.118	.431	50	.360	.010*
300ms	1000ms	42	.248	.113	47	.004	.980	46	.274	.065
300ms	1000ms, +2SD	39	.131	.428	44	.296	.051	45	.491	.001**
300ms	1000ms, +3SD	41	.299	.058	47	.095	.526	45	.404	.006**
300ms	1100ms	42	.190	.228	47	.011	.939	46	.328	.026*
300ms	1100ms, +2SD	39	.131	.428	44	.296	.051	45	.491	.001**
300ms	1100ms, +3SD	41	.460	.003**	47	.070	.640	45	.438	.003**
300ms, -2SD	1000ms	42	.206	.191	47	-.010	.949	46	.270	.070
300ms, -2SD	1000ms, +2SD	39	.057	.732	44	.339	.024*	45	.487	.001**
300ms, -2SD	1000ms, +3SD	41	.241	.129	47	.084	.576	45	.401	.006**
300ms, -2SD	1100ms	42	.057	.718	47	-.002	.988	46	.325	.027*
300ms, -2SD	1100ms, +2SD	39	.057	.732	44	.339	.024*	45	.487	.001**
300ms, -2SD	1100ms, +3SD	41	.249	.117	47	.045	.763	45	.435	.003**
300ms, -3SD	1000ms	42	.248	.113	47	.004	.980	46	.274	.065
300ms, -3SD	1000ms, +2SD	39	.131	.428	44	.296	.051	45	.491	.001**
300ms, -3SD	1000ms, +3SD	41	.299	.058	47	.095	.526	45	.404	.006**
300ms, -3SD	1100ms	42	.190	.228	47	.011	.939	46	.328	.026*
300ms, -3SD	1100ms, +2SD	39	.131	.428	44	.296	.051	45	.491	.001**
300ms, -3SD	1100ms, +3SD	41	.460	.003**	47	.070	.640	45	.438	.003**
Data-Driven	Data-Driven	44	.297	.050	47	.128	.391	50	.232	.105

Note. All values of Iacoviello et al. (2014) attention bias variability index are included.

* $p < .05$. ** $p < .01$.

Table 3.6 Naim et al. (2015) attention bias variability index: test-retest and split-half reliability by outlier cutoffs

Low-End Cutoffs	High-End Cutoffs	Test-Retest Reliability			Split-Half Reliability (Time 1)			Split-Half Reliability (Time 2)		
		N	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>p</i>
0ms	1000ms	44	-.112	.471	46	.036	.814	50	.148	.305
0ms	1000ms, +2SD	44	-.042	.788	44	-.076	.624	49	.095	.516
0ms	1000ms, +3SD	44	-.075	.630	46	-.022	.887	50	.224	.118
0ms	1100ms	44	-.102	.511	46	.167	.269	50	.201	.161
0ms	1100ms, +2SD	44	-.042	.788	44	-.076	.624	49	.095	.516
0ms	1100ms, +3SD	44	-.081	.603	46	-.004	.978	50	.230	.108
0ms, -2SD	1000ms	44	-.089	.566	46	.073	.629	50	.141	.329
0ms, -2SD	1000ms, +2SD	44	-.044	.778	44	-.068	.660	49	.098	.504
0ms, -2SD	1000ms, +3SD	44	-.065	.677	46	.013	.933	50	.283	.047*
0ms, -2SD	1100ms	44	-.079	.612	46	.209	.163	50	.174	.227
0ms, -2SD	1100ms, +2SD	44	-.044	.778	44	-.068	.660	49	.098	.504
0ms, -2SD	1100ms, +3SD	44	-.068	.663	46	.027	.858	50	.289	.042*
0ms, -3SD	1000ms	44	-.085	.583	46	.036	.814	50	.141	.328
0ms, -3SD	1000ms, +2SD	44	.004	.977	44	-.076	.624	49	.097	.507
0ms, -3SD	1000ms, +3SD	44	-.039	.803	46	-.022	.887	50	.288	.042*
0ms, -3SD	1100ms	44	-.080	.604	46	.167	.269	50	.190	.187
0ms, -3SD	1100ms, +2SD	44	.004	.977	44	-.076	.624	49	.097	.507
0ms, -3SD	1100ms, +3SD	44	-.044	.777	46	-.004	.978	50	.294	.039*
200ms	1000ms	44	-.080	.607	46	.036	.814	50	.142	.327
200ms	1000ms, +2SD	44	-.029	.851	44	-.076	.624	49	.066	.651
200ms	1000ms, +3SD	44	-.039	.802	46	-.022	.887	50	.294	.038*
200ms	1100ms	44	-.074	.634	46	.167	.269	50	.188	.191
200ms	1100ms, +2SD	44	-.029	.851	44	-.076	.624	49	.066	.651
200ms	1100ms, +3SD	44	-.044	.777	46	-.004	.978	50	.299	.035*
200ms, -2SD	1000ms	44	-.089	.566	46	.073	.629	50	.141	.329
200ms, -2SD	1000ms, +2SD	44	-.044	.778	44	-.068	.660	49	.098	.504

200ms, -2SD	1000ms, +3SD	44	-.065	.677	46	.013	.933	50	.283	.047*
200ms, -2SD	1100ms	44	-.079	.612	46	.209	.163	50	.174	.227
200ms, -2SD	1100ms, +2SD	44	-.044	.778	44	-.068	.660	49	.098	.504
200ms, -2SD	1100ms, +3SD	44	-.068	.663	46	.027	.858	50	.289	.042*
200ms, -3SD	1000ms	44	-.080	.607	46	.036	.814	50	.142	.327
200ms, -3SD	1000ms, +2SD	44	-.029	.851	44	-.076	.624	49	.066	.651
200ms, -3SD	1000ms, +3SD	44	-.039	.802	46	-.022	.887	50	.294	.038*
200ms, -3SD	1100ms	44	-.074	.634	46	.167	.269	50	.188	.191
200ms, -3SD	1100ms, +2SD	44	-.029	.851	44	-.076	.624	49	.066	.651
200ms, -3SD	1100ms, +3SD	44	-.044	.777	46	-.004	.978	50	.299	.035*
300ms	1000ms	42	-.013	.933	46	.119	.432	46	.079	.602
300ms	1000ms, +2SD	39	.119	.472	42	-.064	.687	44	.076	.623
300ms	1000ms, +3SD	41	-.005	.973	46	.048	.753	45	.236	.118
300ms	1100ms	42	-.026	.869	46	.169	.260	46	.177	.240
300ms	1100ms, +2SD	39	.119	.472	42	-.064	.687	44	.076	.623
300ms	1100ms, +3SD	41	-.011	.943	46	.067	.658	45	.243	.108
300ms, -2SD	1000ms	42	-.002	.989	46	.146	.334	46	.092	.544
300ms, -2SD	1000ms, +2SD	39	.12	.467	42	-.058	.713	44	.080	.606
300ms, -2SD	1000ms, +3SD	41	.00	.999	46	.070	.646	45	.248	.101
300ms, -2SD	1100ms	42	-.015	.927	46	.197	.188	46	.190	.207
300ms, -2SD	1100ms, +2SD	39	.120	.467	42	-.058	.713	44	.080	.606
300ms, -2SD	1100ms, +3SD	41	-.005	.973	46	.085	.573	45	.254	.092
300ms, -3SD	1000ms	42	-.013	.933	46	.119	.432	46	.079	.602
300ms, -3SD	1000ms, +2SD	39	.119	.472	42	-.064	.687	44	.076	.623
300ms, -3SD	1000ms, +3SD	41	-.005	.973	46	.048	.753	45	.236	.118
300ms, -3SD	1100ms	42	-.026	.869	46	.169	.260	46	.177	.240
300ms, -3SD	1100ms, +2SD	39	.119	.472	42	-.064	.687	44	.076	.623
300ms, -3SD	1100ms, +3SD	41	-.011	.943	46	.067	.658	45	.243	.108
Data-Driven	Data-Driven	44	-.100	.517	46	.164	.276	50	.225	.116

Note. All values of Naim et al. (2015) attention bias variability index are included.

* $p < .05$.

Table 3.7 Magnitude of test-retest reliability estimates

Index	Sets of Outlier Cutoffs with Significant Reliability	Highest Significant Estimate		Lowest Significant Estimate		Range of Significant Reliability Estimates	Range of All Reliability Estimates
		<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>		
MacLeod et al. (1986)	8	.386	.010	.301	.047	.085	.496
Koster et al. (2004) Vigilance	0	-	-	-	-	-	.276
Koster et al. (2004) Slow Disengagement	1	.308	.042	.308	.042	-	.366
Iacoviello et al. (2014)	11	.460	.003	.304	.045	.156	.403
Naim et al. (2015)	0	-	-	-	-	-	.232
Mean TL-BS _{POSITIVE}	55	.627	<.001	.408	.007	.219	.219
Mean TL-BS _{NEGATIVE}	55	.476	.002	.301	.047	.175	.175
Peak TL-BS _{POSITIVE}	35	.626	<.001	.299	.049	.327	.661
Peak TL-BS _{NEGATIVE}	36	.471	.001	.376	.015	.095	.321
TL-BS _{VARIABILITY}	55	.630	<.001	.447	.003	.183	.183

Table 3.8 Magnitude of split-half reliability estimates, time 1

Index	Sets of Outlier Cutoffs with Significant Reliability	Highest Significant Estimate		Lowest Significant Estimate		Range of Significant Reliability Estimates	Range of All Reliability Estimates
		<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>		
MacLeod et al. (1986)	6	-.300	.041	-.304	.038	.004	.187
Koster et al. (2004) Vigilance	13	-.289	.049	-.314	.031	.025	.098
Koster et al. (2004) Slow Disengagement	18	-.298	.042	-.345	.017	.047	.203
Iacoviello et al. (2014)	2	.339	.024	.339	.024	-	.340
Naim et al. (2015)	0	-	-	-	-	-	.285

Table 3.9 Magnitude of split-half reliability estimates, time 2

Index	Sets of Outlier Cutoffs with Significant Reliability	Highest Significant Estimate		Lowest Significant Estimate		Range of Significant Reliability Estimates	Range of All Reliability Estimates
		<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>		
MacLeod et al. (1986)	0	-	-	-	-	-	.223
Koster et al. (2004) Vigilance	2	.293	.039	.293	.039	-	.377
Koster et al. (2004) Slow Disengagement	0	-	-	-	-	-	.254
Iacoviello et al. (2014)	46	.491	.001	.280	.049	.211	.259
Naim et al. (2015)	10	.299	.035	.283	.047	.016	.233

Table 3.10 Outlier cutoffs with best test-retest reliability across indices

Low-End Cutoffs	High-End Cutoffs	Number of Significant Indices	MacLeod et al. (1986)	Iacoviello et al. (2014)	Mean TL-BS Positive	Mean TL-BS Negative	Peak TL-BS Positive	Peak TL-BS Negative	TL-BS Variability
0 ms	1000 ms, + 3SD	6	.329	-	.464	.388	.425	.423	.531
0 ms	1100 ms, + 3SD	7	.386	.441	.549	.390	.532	.423	.580
0 ms, - 2SD	1100 ms, + 3SD	6	.301	-	.422	.376	.405	.385	.506
0 ms, - 3SD	1000 ms, + 3SD	7	.315	.304	.492	.388	.448	.423	.548
0 ms, - 3SD	1100 ms, + 3SD	7	.374	.451	.575	.390	.556	.423	.598
200 ms	1100 ms, + 3SD	7	.343	.431	.575	.362	.556	.385	.587
200 ms, - 2SD	1100 ms, + 3SD	6	.301	-	.422	.376	.405	.385	.506
200 ms, - 3SD	1100 ms, + 3SD	7	.343	.431	.575	.362	.556	.385	.587
300 ms	1100 ms, + 3SD	6	-	.460	.627	.413	.626	.376	.630
300 ms, - 3SD	1100 ms, + 3SD	6	-	.460	.627	.413	.626	.376	.630

Note. Reliability estimates are Pearson's r values, all $p < .05$.

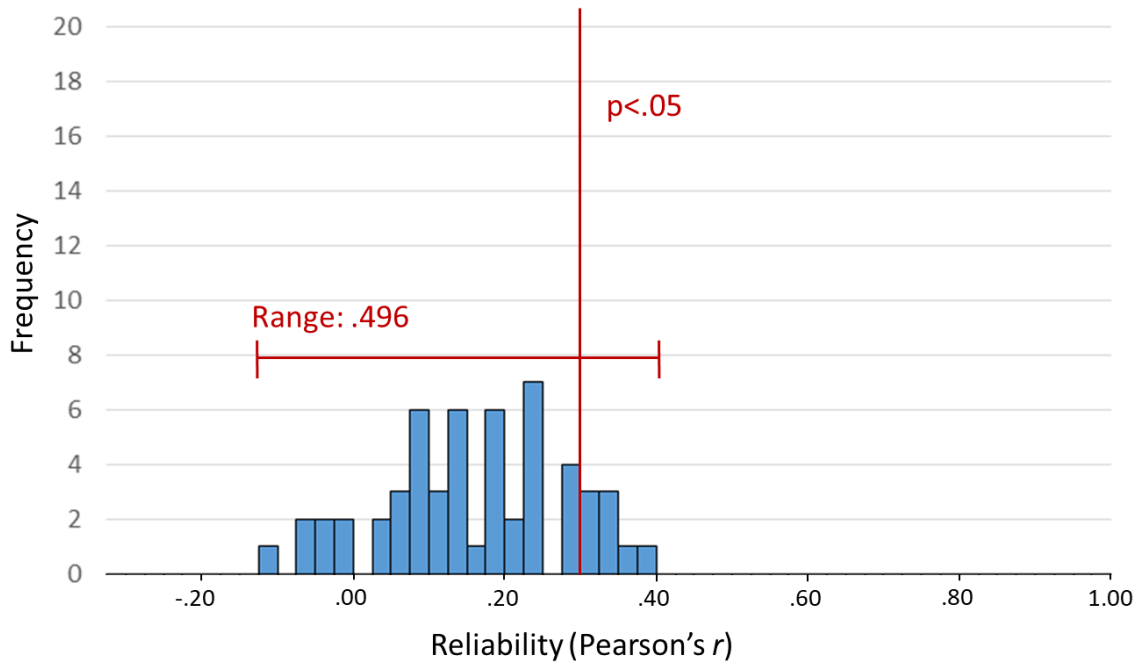


Figure 3.1 A histogram of all test-retest reliability estimates for the MacLeod et al. (1986) attention bias index

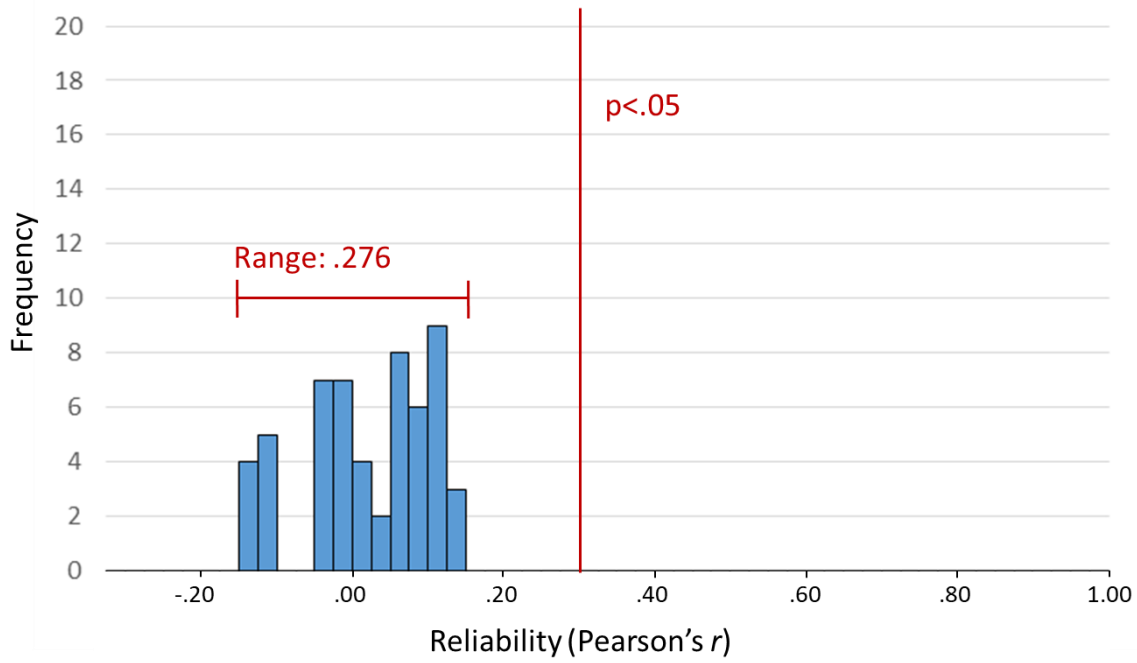


Figure 3.2 A histogram of all test-retest reliability estimates for the Koster et al. (2004) attention bias index for vigilance to threat

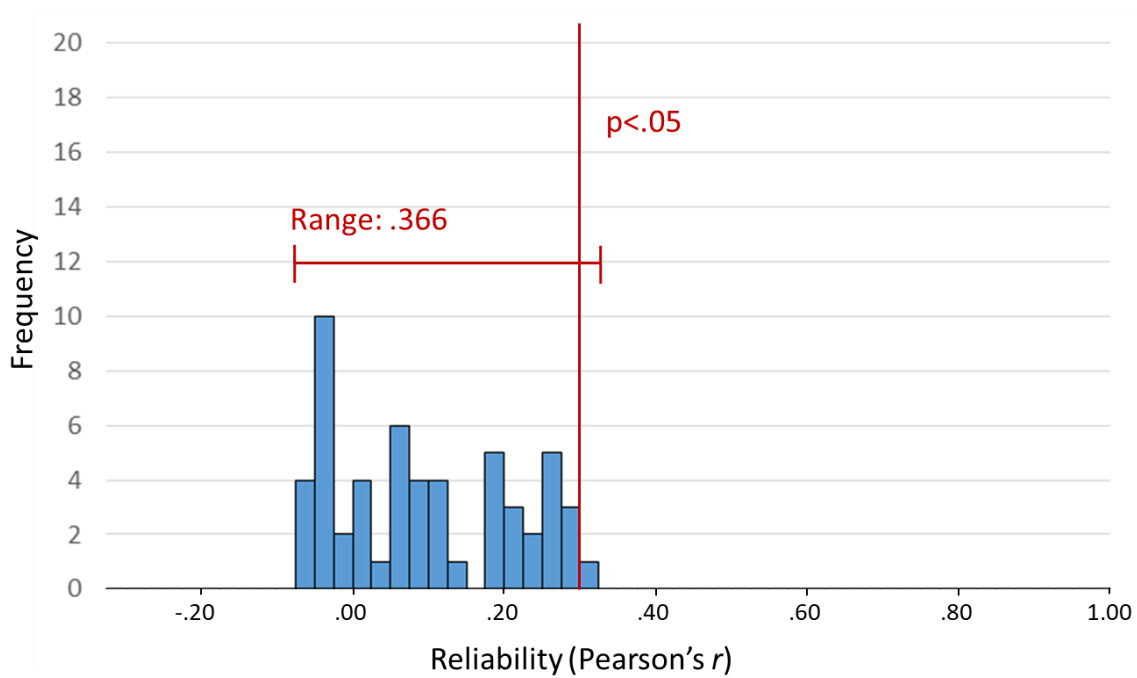


Figure 3.3 A histogram of all test-retest reliability estimates for the Koster et al. (2004) attention bias index for slow disengagement from threat

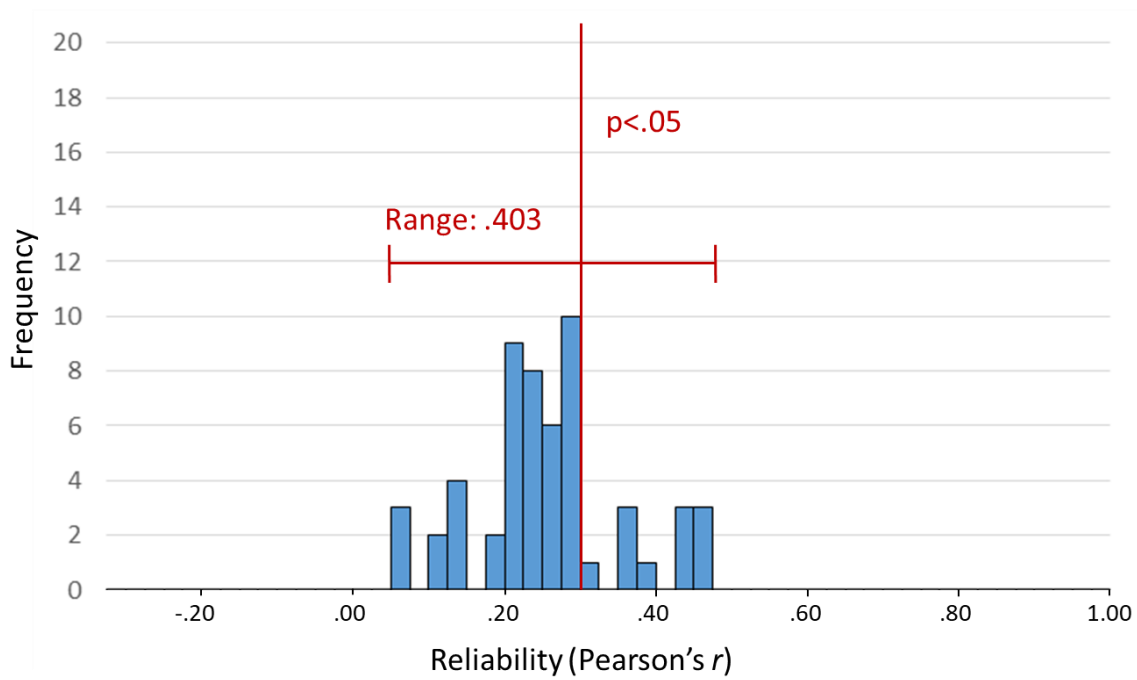


Figure 3.4 A histogram of all test-retest reliability estimates for the Iacoviello et al. (2014) index of attention bias variability

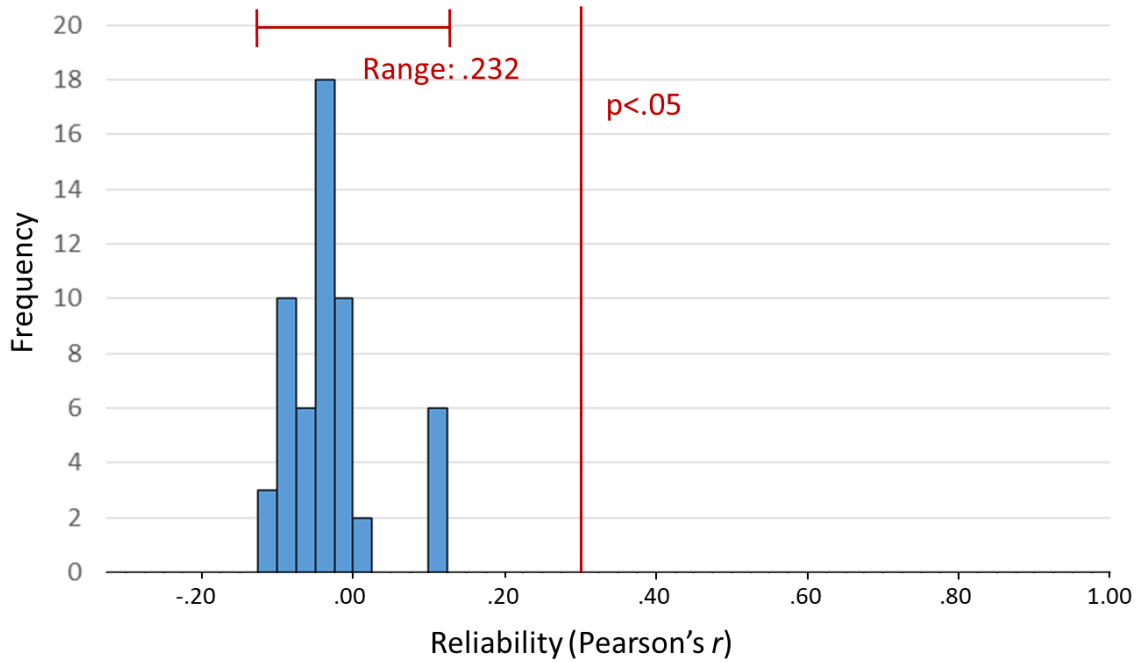


Figure 3.5 A histogram of all test-retest reliability estimates for the Naim et al. (2015) index of attention bias variability

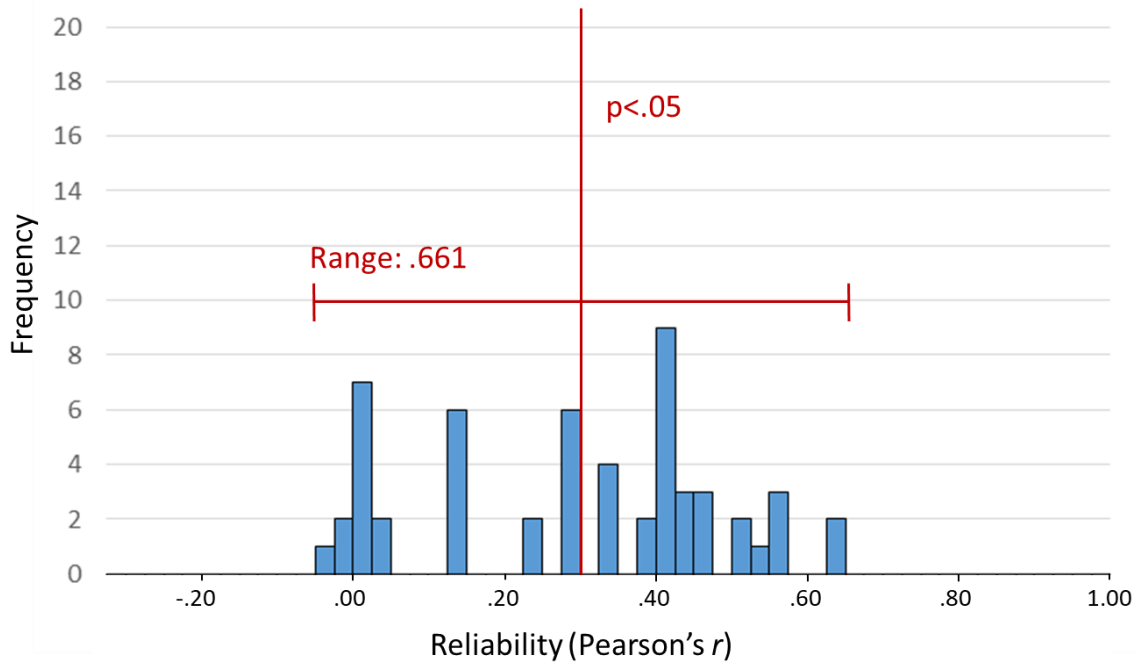


Figure 3.6 A histogram of all test-retest reliability estimates for the Zvielli et al. (2015) trial-level bias score index of peak bias toward threat ($Peak TL-BS_{POSITIVE}$)

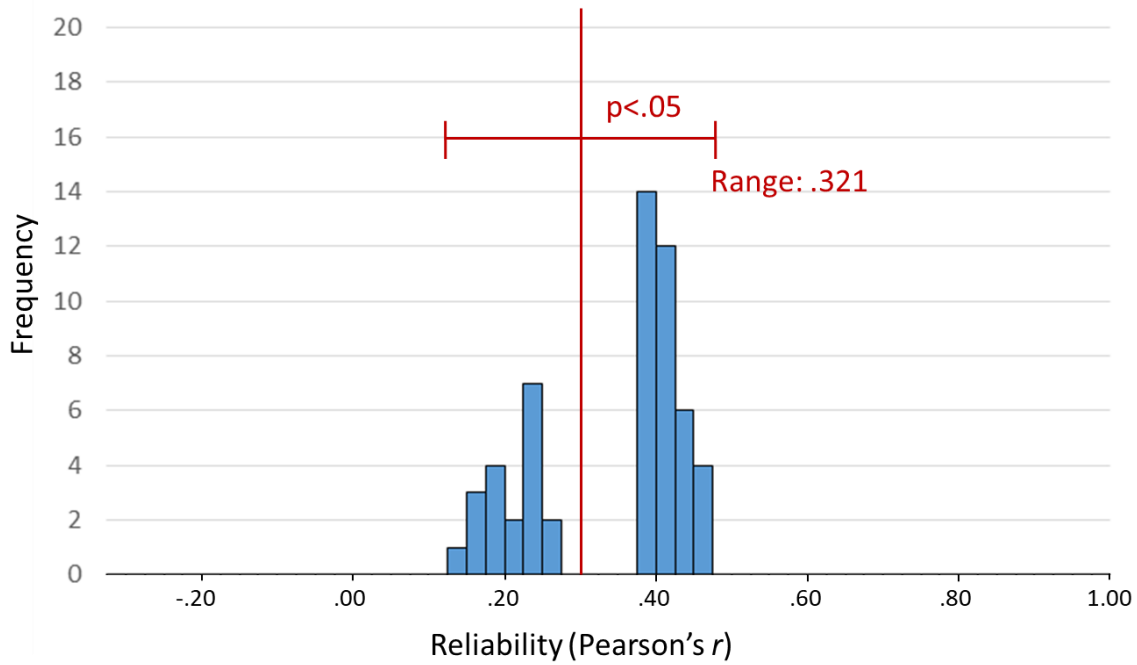


Figure 3.7 A histogram of all test-retest reliability estimates for the Zvielli et al. (2015) trial-level bias score index of peak bias away from threat (Peak TL-BS_{NEGATIVE})

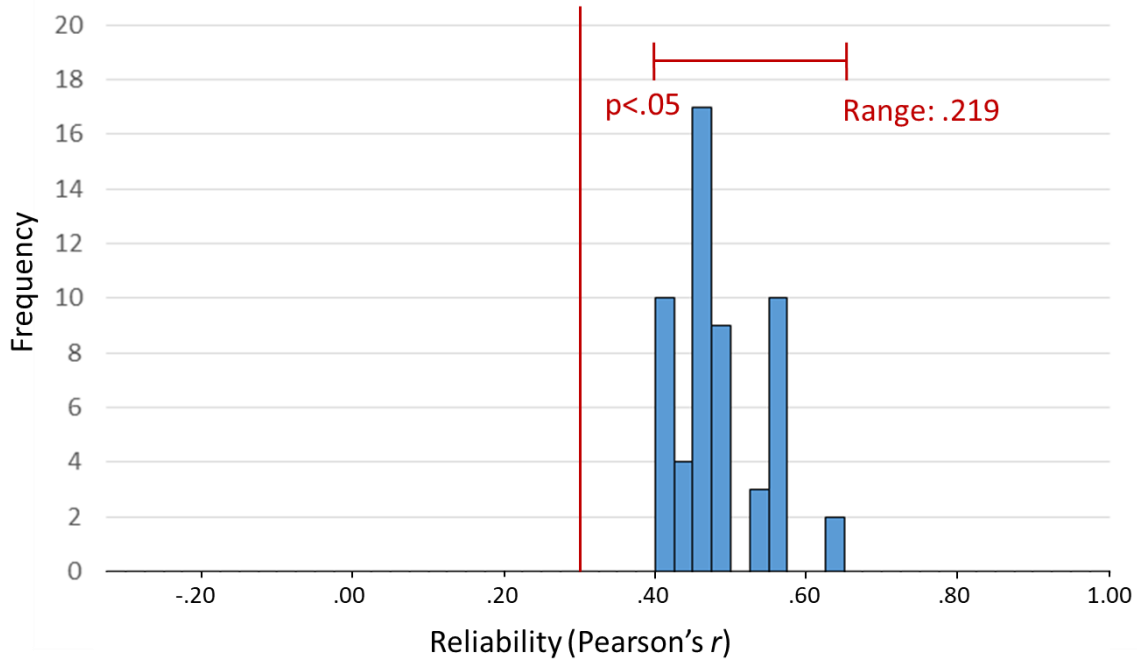


Figure 3.8 A histogram of all test-retest reliability estimates for the Zvielli et al. (2015) trial-level bias score index of mean bias toward threat (Mean $TL-BS_{POSITIVE}$)

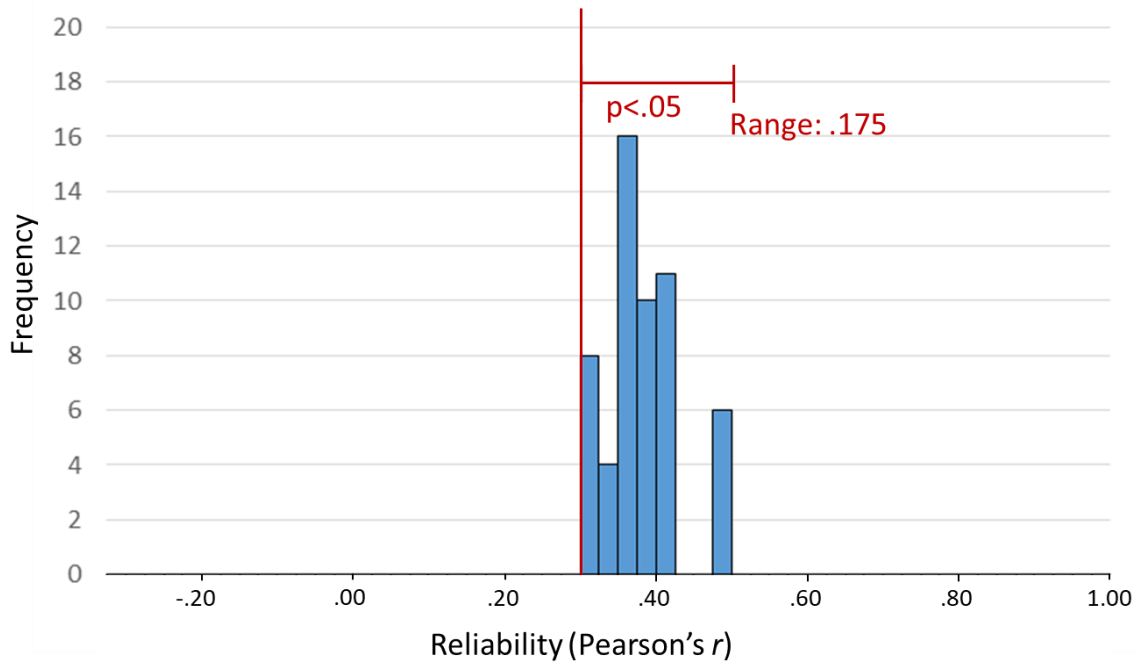


Figure 3.9 A histogram of all test-retest reliability estimates for the Zvielli et al. (2015) trial-level bias score index of mean bias away from threat ($Mean TL-BS_{NEGATIVE}$)

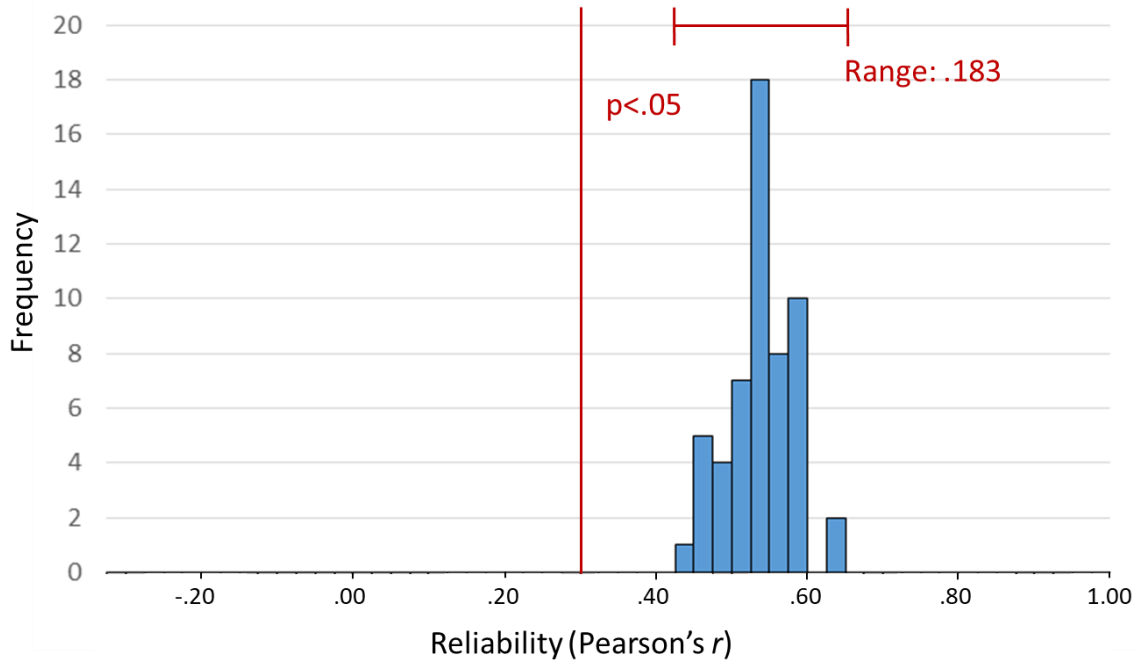


Figure 3.10 A histogram of all test-retest reliability estimates for the Zvielli et al. (2015) trial-level bias score index of attention bias variability (TL-BS_{VARIABILITY})

4 DISCUSSION

In the current study, the author compared reliability of all analytic strategies for the dot probe task that have been published in the scientific literature. By examining reliability, the author also hoped to gain insight into the nature of attention bias to threat in social anxiety disorder. Specifically, indices that assume attention bias is a stable phenomenon were compared to indices that measure fluctuations over the course of the task in order to determine the stability of attention bias within and between dot probe sessions. Findings indicate that whereas attention bias does fluctuate over the course of a dot probe session, both mean bias to threat and attention bias variability can be measured with statistically significant reliability using trial-level bias scores. However, aggregated mean indices, which have been used in the vast majority of the dot probe literature, showed mostly insignificant reliability.

4.1 Study Hypotheses

TL-BS_{VARIABILITY}, Mean TL-BS_{POSITIVE}, and Mean TL-BS_{NEGATIVE} demonstrated generally superior test-retest reliability to all other indices of attention bias, though they underperformed study hypotheses. Test-retest reliability estimates for all non-TL-BS indices were mostly or completely insignificant. Similarly, split-half reliability across all non-TL-BS indices were either negative, insignificant, or lower than hypothesized. Individual study hypotheses and outcomes are listed below:

- The hypothesis that aggregated mean indices would have insignificant or very low test-retest and split-half reliability was confirmed for the MacLeod et al. (1986) index and both Koster et al. (2004) indices.
- For the Iacoviello et al. (2014) attention bias variability index, 11 sets of outlier cutoffs produced significant reliability estimates within the hypothesized range of $r = .1$ to $r = .6$,

whereas reliability for 44 other sets of cutoffs were not significant. Split-half reliability for this index at both study sessions fell below hypothesized range of $r = .5$ to $r = .9$.

- The Naim et al. (2015) attention bias variability index had no significant test-retest reliability and no significant split-half reliability for the first study session. It demonstrated significant estimates of split-half reliability for 10 sets of outlier cutoffs at the second study session, but these were below the hypothesized range of $r = .5$ to $r = .9$.
- Winsorizing outliers outside of data-driven cutoffs, as opposed to deletion outside of arbitrary cutoffs, did not produce any statistically significant estimates of test-retest or split-half reliability for the Iacoviello et al. (2014) or Naim et al. (2015) indices. As many other sets of arbitrary cutoffs also produced insignificant reliability, the comparison between data-driven and arbitrary outlier-handling methods is inconclusive.
- Peak TL-BS_{POSITIVE} outperformed study hypotheses of low or insignificant test-retest reliability, with 35 sets of outlier cutoffs producing significant reliability. Peak TL-BS_{NEGATIVE} similarly outperformed study hypotheses of low or insignificant test-retest reliability, with 36 sets of outlier cutoffs producing significant reliability. Split-half reliability could not be tested for these indices due to insufficient trials.
- Although all test-retest reliability estimates for mean TL-BS_{NEGATIVE} were statistically significant, they fell below the hypothesized range of $r = .5$ to $r = .7$. Split-half reliability could not be tested for this index due to insufficient trials.
- Although all test-retest reliability estimates for mean TL-BS_{POSITIVE} were statistically significant, only two estimates out of 55 reached the hypothesized range of $r = .6$ to $r = .8$. Split-half reliability could not be tested for this index due to insufficient trials.

- Although all test-retest reliability estimates for TL-BS_{VARIABILITY} were statistically significant, they fell below the hypothesized range of $r = .7$ to $r = .9$. Split-half reliability could not be tested for this index due to insufficient trials.

It is important to note that the current study was not sufficiently powered to test some of the hypotheses listed above, as it was unable to detect statistically significant reliability below $r = .3$. More significant test-retest reliability estimates could have been found in the lower end of the hypothesized range for the Iacoviello et al. (2014) and Naim et al. (2015) attention bias variability indices with increased power.

4.2 Consistency of Reliability across Outlier Cutoffs

Ten sets of outlier cutoffs produced significant test-retest reliability for either six or seven dot probe indices in the current sample (see Table 3.10). All other sets of cutoffs produced significant test-retest reliability for fewer than six indices. There were some observable patterns in these cutoffs; for example, eight used the same high-end cutoff of 1100 ms, mean+3SD. Low-end cutoffs were inconsistent, which seems to reflect that all sets with the best high-end cutoffs fared similarly. Consistent with Price et al. (2015), the strategy of Winsorizing outliers beyond cutoffs based on interquartile ranges did not produce higher reliability than deletion outside arbitrary cutoffs.

The most important finding in the current study with regard to outlier cutoffs is not that specific cutoffs fared better than others, rather that outlier cutoff selection as a whole had a much greater effect on reliability for some indices than others. The narrowest ranges of test-retest reliability estimates were found for indices with the highest reliability: TL-BS_{VARIABILITY}, Mean TL-BS_{POSITIVE}, and Mean TL-BS_{NEGATIVE}. These were the only three indices that displayed significant test-retest reliability across all 55 sets of outlier cutoffs used. In contrast, less reliable

indices such as the MacLeod et al. (1986) index, the Iacoviello et al. (2014) index, and Peak TL-BS_{POSITIVE} had much wider ranges of reliability estimates (see Figures 3.1-3.10).

This variation in reliability estimates due to changes in outlier cutoffs is an important, understudied characteristic of the dot probe. It demonstrates that reliability estimates from the vast majority of dot probe studies, which use aggregated mean indices, are likely to be highly influenced by arbitrary cutoff selection. For example, Schmukle et al. (2005), the first study to assess test-retest reliability for the dot probe task, used two sets of arbitrary outlier cutoffs that produced insignificant test-retest reliability. In the current study, the two sets of cutoffs closest to those for the MacLeod et al. (1986) index also produced insignificant results, but eight different sets of cutoffs produced significant results with a peak reliability of $r=.386, p=.010$. Given the high range of reliability estimates reported here for the MacLeod et al. (1986) index, it seems likely that Schmukle et al.'s (2005) study and many like it did not account for substantial effects of outlier cutoffs on their results.

Whereas a lenient high-end cutoff improved reliability in our sample, work with the dot probe (Price et al. 2015) and other reaction time paradigms (Ratcliff, 1993) indicate that the optimal cutoffs in any single reaction time study are not likely to generalize to others. Although searching for broadly optimal cutoffs is unlikely to yield a consistent result, however, use of attention bias indices with relatively consistent reliability across differing sets of cutoffs like TL-BS_{VARIABILITY}, Mean TL-BS_{POSITIVE}, and Mean TL-BS_{NEGATIVE} makes finding universally optimal cutoffs unnecessary. Instead, the author recommends using methodology drawn from the literature on reaction time distributions. Ratcliff (2012) argued that to reduce low-end outliers, experimenters should examine the accuracy of reaction time data at different intervals, e.g. 1-100 ms, 101-200 ms. When the proportion of accurate vs. inaccurate responses is roughly

1:1, an experimenter can infer that reaction times during this interval reflect measurement error and not the cognitive processes under scrutiny. When the proportion of accurate responses begins to rise, the experimenter can infer that some responses in a given interval are valid. High-end cutoffs can be selected arbitrarily, so long as it is clear that small changes in cutoffs do not have a dramatic effect on reliability.

4.3 Implications for Cognitive Models of Social Anxiety Disorder

Classic cognitive models of social anxiety disorder posit that consistent attention bias towards socially threatening stimuli is a maintaining mechanism of social anxiety symptoms. However, the original MacLeod et al. (1986) index of attention bias derived from the dot probe task is unreliable and studies using TL-BS have demonstrated that attention bias fluctuates over the course of the dot probe session, suggesting that mean bias to threat may not be a reliable construct in this population. Consistent with previous studies, the author found mostly insignificant reliability for the MacLeod et al. (1986) index. However, Mean TL-BS_{POSITIVE}, an index of mean bias toward threat, had completely significant and roughly equal reliability to TL-BS_{VARIABILITY}. This indicates that mean bias to threat and attention bias variability are both important dimensions of dot probe performance that can be reliably measured. This finding supports classic cognitive models of social anxiety disorder, in that a reliable mean bias to threat can be measured in our sample. However, the reliability of TL-BS_{VARIABILITY} also indicates that these models may need to be updated to include attention bias variability.

To determine whether attention bias variability is a clinical feature of social anxiety disorder, the two studies to date that compare attention bias variability between people with social anxiety disorder and controls using Naim et al.'s (2015) index should be replicated using the TL-BS method. It is possible their null results were due to poor psychometrics of their

attention bias variability index. If a difference in attention bias variability does exist in people with social anxiety disorder, the TL-BS method may be more likely to detect it. These studies should also examine attentional control, as deficits in attentional control have been well demonstrated in social anxiety disorder (e.g. Blair et al., 2012; Swartz et al., 2014) and could be related to attention bias variability. Updated neurocognitive models of social anxiety disorder will need to account for mean attention bias, attention bias variability, and attentional control as potential mechanisms of social anxiety symptoms. Interactions between these constructs, such as the possibility that impaired top-down attentional control increases variability in attention bias to threat, should also be examined.

4.4 Limitations and Future Directions

Although TL-BS methods improve the reliability of the dot probe task, there are several limitations that remain for the dot probe paradigm as a whole. One is related to inherent problems with using indices that have been derived from difference scores. Whereas there is a substantial literature on reliability of difference scores which cannot be fully described here (see Miller & Kane, 2001 for a review), low reliability is generally likely to result when a score is derived from subtracting two distributions (e.g. threat-incongruent and threat-congruent trials) that have similar variance and are highly correlated with each other. Both aggregated mean indices and TL-BS parameters rely on difference scores, albeit in different ways, and thus may be subject to this limitation on reliability. Additionally, using button-press reaction times as an indicator of visual attention introduces sources of measurement error that might be circumvented using other paradigms.

Magnitude of reliability across many of the attention bias indices in this study was lower than hypothesized. The Iacoviello et al. (2014) and Naim et al. (2015) attention bias variability

indices had particularly poor test-retest reliability as compared to previous findings. The dot probe task used for the current study included 64 threat-neutral trials, 64 happy-neutral trials, and 32 neutral-neutral trials, which may explain why reliability for many of the study's indices fell short of study hypotheses, particularly the attention bias variability indices. Because happy-neutral trials were not included in the current analyses, each individual index was calculated using a total 64 trials, or two-fifths of total data collected during a session. Given that attention bias variability indices are designed to be sensitive to fluctuations in bias over the course of a dot probe session, these indices' precision in measuring attention bias variability may have been compromised because many trials were excluded from analyses. Happy-neutral trials also reduced the number of relevant trials for calculating TL-BS indices, which made it impossible to examine split-half reliability for these indices and may have limited their precision in sensitively capturing fluctuations in bias. Zvielli et al. (2015) write that a "bare minimum" of 40 valid trial-level bias scores per session in order to calculate these indices. Our participants barely met this criterion and a higher number of scores may have improved reliability.

Future studies examining attention bias to threat in social anxiety disorder should solely use threat-neutral trials in order to more sensitively measure the time course of attention bias to threat. TL-BS analyses should be used over aggregated mean or other attention bias variability indices. Studies examining samples with social anxiety disorder with healthy controls will be useful in understanding the role of attention in social anxiety disorder. Additionally, the author agrees with Rodebaugh et al.'s (2016) assessment that accurately evaluating the role of attention bias to threat in anxiety disorders will involve using the dot probe alongside other approaches such as eye tracking and neuroimaging in order to provide convergent evidence.

4.5 Conclusions and Recommendations

In summary, the current study demonstrates that it is no longer appropriate to use aggregated mean indices in dot probe research, as they have unacceptably low reliability which seems to result from a high degree of measurement error. It found that Mean TL-BS_{POSITIVE}, Mean TL-BS_{NEGATIVE}, and TL-BS_{VARIABILITY} have superior test-retest reliability to all other indices reported in the literature. The study also found that test-retest reliability estimates for these three indices varied much less than other indices in response to small changes in outlier cutoffs. This demonstrates that both mean attention bias toward threat and attention bias variability can be more precisely and reliably measured in people with social anxiety disorder using TL-BS than with other methods. In future studies, the TL-BS approach will hopefully expand the dot probe's utility in exploring attention biases across a wide range of psychopathology.

REFERENCES

- Amir, I., Zvielli, A., & Bernstein, A. (2016). (De)coupling of our eyes and our mind's eye: A dynamic process perspective on attentional bias. *Emotion, 16*(7), 978–986.
<https://doi.org/10.1037/emo0000172>
- Association, A. P. (1980). *Diagnostic and statistical manual, 3rd edn (DSM-III)*. Washington: American Psychiatric Association.
- Badura-Brack, A. S., Naim, R., Ryan, T. J., Levy, O., Abend, R., Khanna, M. M., ... Bar-Haim, Y. (2015). Effect of attention training on attention bias variability and PTSD symptoms: Randomized controlled trials in Israeli and U.S. Combat Veterans. *American Journal of Psychiatry, 172*(12), 1233–1241. <https://doi.org/10.1176/appi.ajp.2015.14121578>
- Bantin, T., Stevens, S., Gerlach, A. L., & 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.
<https://doi.org/10.1016/j.jbtep.2015.04.009>
- 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), 1–24. <https://doi.org/10.1037/0033-2909.133.1.1>
- Bardeen, J. R., Daniel, T. A., Hinnant, J. B., & Orcutt, H. K. (2017). Emotion dysregulation and threat-related attention bias variability. *Motivation and Emotion, 41*(3), 402–409.
<https://doi.org/10.1007/s11031-017-9604-z>
- Bardeen, J. R., Tull, M. T., Daniel, T. A., Evenden, J., & Stevens, E. N. (2016). A Preliminary Investigation of the Time Course of Attention Bias Variability in Posttraumatic Stress

- Disorder: The Moderating Role of Attentional Control. *Behaviour Change*, 33(02), 94–111.
<https://doi.org/10.1017/bec.2016.5>
- Beck, A., Emery, G., & Greenberg, R. (1985). *Anxiety disorders and phobias: A cognitive approach*. New York: Basic.
- Bishop, S. J. (2007). Neurocognitive mechanisms of anxiety: an integrative account. *Trends in Cognitive Sciences*, 11(7), 307–316. <https://doi.org/10.1016/j.tics.2007.05.008>
- Blair, K. S., Geraci, M., Smith, B. W., Hollon, N., Devido, J., Otero, M., ... Pine, D. S. (2012). Reduced dorsal anterior cingulate cortical activity during emotional regulation and top-down attentional control in generalized social phobia, generalized anxiety disorder, and comorbid generalized social phobia/generalized anxiety disorder. *Biological Psychiatry*, 72(6), 476–482. <https://doi.org/10.1016/j.biopsych.2012.04.013>
- Bower, G. H. (1981). Mood and memory. *Behaviour Research and Therapy*, 25(6), 443–455.
[https://doi.org/10.1016/0005-7967\(87\)90052-0](https://doi.org/10.1016/0005-7967(87)90052-0)
- Bradley, B. P., Mogg, K., Millar, N., Bonham-Carter, C., Fergusson, E., Jenkins, J., & Parr, M. (1997). Attentional biases for emotional faces. *Cognition and Emotion*, 11(1), 25–42.
<https://doi.org/10.1080/026999397380014>
- Burgess, I. S., Jones, L. M., Robertson, S. A., Radcliffe, W. N., & Emerson, E. (1981). The degree of control exerted by phobic and non-phobic verbal stimuli over the recognition behaviour of phobic and non-phobic subjects. *Behaviour Research and Therapy*, 19(3), 233–243. [https://doi.org/10.1016/0005-7967\(81\)90007-3](https://doi.org/10.1016/0005-7967(81)90007-3)
- Burris, J. L., Barry-Anwar, R. A., & Rivera, S. M. (2017). An eye tracking investigation of attentional biases towards affect in young children. *Developmental Psychology*, 53(8), 1418.
- Cisler, J. M., Bacon, A. K., & Williams, N. L. (2009). Phenomenological characteristics of

- attentional biases towards threat: A critical review. *Cognitive Therapy and Research*, 33(2), 221–234. <https://doi.org/10.1007/s10608-007-9161-y>
- Clark, D. M., & Wells, A. (1995). A cognitive model of social phobia. In R. G. Heimberg (Ed.), *Social phobia: Diagnosis, assessment, and treatment*. Guilford Press.
- Davis, M. L., Rosenfield, D., Bernstein, A., Zvielli, A., Reinecke, A., Beevers, G., ... Reinecke, A. (2016). Attention Bias Dynamics and Symptom Severity During and Following CBT for Social Anxiety Disorder Attention Bias Dynamics and Symptom Severity During and Following CBT for Social Anxiety Disorder. *Journal of Consulting and Clinical Psychology*, 84(9), 795–802. <https://doi.org/10.1037/ccp0000125>
- Dear, B. F., Sharpe, L., Nicholas, M. K., & Refshauge, K. (2011). The psychometric properties of the dot-probe paradigm when used in pain-related attentional bias research. *Journal of Pain*, 12(12), 1247–1254. <https://doi.org/10.1016/j.jpain.2011.07.003>
- Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: an easy way to maximize the accuracy and power of your research. *American Psychologist*, 63(7), 591.
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion*, 7(2), 336.
- Frewen, P. A., Dozois, D. J. A., Joanisse, M. F., & Neufeld, R. W. J. (2008). Selective attention to threat versus reward: meta-analysis and neural-network modeling of the dot-probe task. *Clinical Psychology Review*, 28(2), 307–337. <https://doi.org/10.1016/j.cpr.2007.05.006>
- Hayes, J. P., VanElzakker, M. B., & Shin, L. M. (2012). Emotion and cognition interactions in PTSD: a review of neurocognitive and neuroimaging studies. *Frontiers in Integrative Neuroscience*, 6(October), 1–14. <https://doi.org/10.3389/fnint.2012.00089>
- Iacoviello, B. M., Wu, G., Abend, R., Murrough, J., Feder, A., Fruchter, E., ... Charney, D. S.

- (2014). Attention Bias Variability and Symptoms of Posttraumatic Stress Disorder. *Journal of Traumatic Stress*, 27(2), 232–239. <https://doi.org/10.1002/jts>.
- 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(DEC), 1–9. <https://doi.org/10.3389/fpsyg.2014.01368>
- Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, Severity, and Comorbidity of. *Arch Gen Psychiatry*, 62(6), 617–627. <https://doi.org/10.1001/archpsyc.62.6.617>
- Klumpp, H., & Amir, N. (2009). Examination of vigilance and disengagement of threat in social anxiety with a probe detection task. *Anxiety, Stress and Coping*, 22(3), 283–296. <https://doi.org/10.1080/10615800802449602>
- Klumpp, H., Fitzgerald, D. A., Angstadt, M., Post, D., & Phan, K. L. (2014). Neural response during attentional control and emotion processing predicts improvement after cognitive behavioral therapy in generalized social anxiety disorder. *Psychological Medicine*, 44(14), 3109–3121.
- Klumpp, H., Fitzgerald, J. M., Kinney, K. L., Kennedy, A. E., Shankman, S. A., Langenecker, S. A., & Phan, K. L. (2017). Predicting cognitive behavioral therapy response in social anxiety disorder with anterior cingulate cortex and amygdala during emotion regulation. *Neuroimage. Clinical*, 1525–1534.
- Koster, E. H. W., Crombez, G., Verschuere, B., & De Houwer, J. (2004). Selective attention to threat in the dot probe paradigm: Differentiating vigilance and difficulty to disengage. *Behaviour Research and Therapy*, 42(10), 1183–1192.

<https://doi.org/10.1016/j.brat.2003.08.001>

- Koster, E. H. W., Crombez, G., Verschuere, B., & De Houwer, J. (2006). Attention to threat in anxiety-prone individuals: Mechanisms underlying attentional bias. *Cognitive Therapy and Research*, 30(5), 635–643. <https://doi.org/10.1007/s10608-006-9042-9>
- Macleod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology*, 95(1), 15. <https://doi.org/10.1037/0021-843X.95.1.15>
- Mathews, A., & Macleod, C. (1985). Selective Processing of Threat Cues in Anxiety States. *Behaviour Research and Therapy*, 23(5), 563–569. [https://doi.org/10.1016/0005-7967\(90\)90160-K](https://doi.org/10.1016/0005-7967(90)90160-K)
- McEvoy, P. M., & Perini, S. J. (2009). Cognitive behavioral group therapy for social phobia with or without attention training: A controlled trial. , 23(4), 519-528. *Journal Of Anxiety Disorders*, 23(4), 519–528.
- Miller, T. B., & Kane, M. (2001). The precision of change scores under absolute and relative interpretations. *Applied Measurement in Education*, 14(4), 307–327. https://doi.org/10.1207/S15324818AME1404_1
- Mogg, K., & Bradley, B. P. (1998). A cognitive-motivational analysis of anxiety. *Behaviour Research and Therapy*, 36(9), 809–884.
- Mohlman, J., & DeVito, A. (2017). The impact of social threat cues on a card sorting task with attentional-shifting demands. *Journal Of Behavior Therapy And Experimental Psychiatry*, 57, 45–52.
- Moriya, J., & Tanno, Y. (2008). Relationships between negative emotionality and attentional control in effortful control. *Personality And Individual Differences*, 44(6), 1348–1355.
- Morrison, A. S., Brozovich, F. A., Lakhan-Pal, S., Jazaieri, H., Goldin, P. R., Heimberg, R. G.,

- & Gross, J. J. (2016). Attentional blink impairment in social anxiety disorder: Depression comorbidity matters. *Journal Of Behavior Therapy And Experimental Psychiatry*, *50*, 209–214.
- Morrison, A. S., & Heimberg, R. G. (2013). Social Anxiety and Social Anxiety Disorder. *Annual Review of Clinical Psychology*, *9*(1), 249–274. <https://doi.org/10.1146/annurev-clinpsy-050212-185631>
- Mujica-Parodi, L. R., Cha, J., & Gao, J. (2017). From Anxious to Reckless: A Control Systems Approach Unifies Prefrontal-Limbic Regulation Across the Spectrum of Threat Detection. *Frontiers in Systems Neuroscience*, *11*(April), 1–14. <https://doi.org/10.3389/fnsys.2017.00018>
- Naim, R., Abend, R., Wald, I., Eldar, S., Levi, O., Fruchter, E., ... Bar-Haim, Y. (2015). Threat-related attention bias variability and posttraumatic stress. *American Journal of Psychiatry*, *172*(12), 1242–1250. <https://doi.org/10.1176/appi.ajp.2015.14121579>
- Nock, M. K., Hwang, I., Sampson, N. A., & Kessler, R. C. (2010). Mental disorders, comorbidity and suicidal behavior: results from the National Comorbidity Survey Replication. *Molecular Psychiatry*, *15*(8), 868–876.
- Price, R. B., Kuckertz, J. M., Siegle, G. J., Ladouceur, C. D., Silk, J. S., Ryan, N. D., ... Amir, N. (2015). Empirical recommendations for improving the stability of the dot-probe task in clinical research. *Psychological Assessment*, *27*(2), 365–376. <https://doi.org/10.1037/pas0000036>
- Rapee, R. M., & Heimberg, R. G. (1997). A cognitive-behavioral model of anxiety in social phobia. *Behaviour Research and Therapy*, *35*(8), 741–756. [https://doi.org/10.1016/S0005-7967\(97\)00022-3](https://doi.org/10.1016/S0005-7967(97)00022-3)

- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532.
- Ratcliff, R. (2012). Response time distributions. In H. Cooper, P. M. Camici, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 1: Foundations, planning, measures, and psychometrics* (pp. 429–443). Washington, D.C.: American Psychological Association.
- Ray, C. (1979). Examination stress and performance on a color-word interference test. *Perceptual and Motor Skills*, *(5)*, 400–402.
- Rodebaugh, T. L., Scullin, R. B., Langer, J. K., Dixon, D. J., Huppert, J. D., Bernstein, A., ... Lenze, A. (2016). Unreliability as a Threat to Understanding Psychopathology: The Cautionary Tale of Attentional Bias. *Journal of Abnormal Psychology*, *125*(6), 840–851. <https://doi.org/10.1037/abn0000184>
- Salemink, E., van den Hout, M. A., & Kindt, M. (2007). Selective attention and threat: Quick orienting versus slow disengagement and two versions of the dot probe task. *Behaviour Research and Therapy*, *45*(3), 607–615. <https://doi.org/10.1016/j.brat.2006.04.004>
- Schäfer, J., Bernstein, A., Zvielli, A., Höfler, M., Wittchen, H. U., & Schönfeld, S. (2016). Attentional Bias Temporal Dynamics Predict Posttraumatic Stress Symptoms: a Prospective–Longitudinal Study Among Soldiers. *Depression and Anxiety*, *33*(7), 630–639. <https://doi.org/10.1002/da.22526>
- Schmukle, S. C. (2005). Unreliability of the dot probe task. *European Journal of Personality*, *19*(7), 595–605. <https://doi.org/10.1002/per.554>
- Schneier, F. R., Kimeldorf, M. B., Choo, T. H., Steinglass, J. E., Wall, M. M., Fyer, A. J., & Simpson, H. B. (2016). Attention bias in adults with anorexia nervosa, obsessive-

- compulsive disorder, and social anxiety disorder. *Journal of Psychiatric Research*, 79, 61–69. <https://doi.org/10.1016/j.jpsychires.2016.04.009>
- Spiegelhalder, K., Jähne, A., Kyle, S. D., Beil, M., Doll, C., Feige, B., & Riemann, D. (2011). Is smoking-related attentional bias a useful marker for treatment effects? *Behavioral Medicine*, 37(1), 26–34. <https://doi.org/10.1080/08964289.2010.543195>
- Staugaard, S. R. (2009). Reliability of two versions of the dot-probe task using photographic faces. *Psychology Science Quarterly*, 51(3), 339–350.
- Stein, M. B., & Kean, Y. B. (2000). Disability and quality of life in social phobia: epidemiologic findings. , 157(10), 1606-1613. *American Journal of Psychiatry*, 157(10), 1606–1613.
- Stevens, S., Rist, F., & Gerlach, A. L. (2009). Influence of alcohol on the processing of emotional facial expressions in individuals with social phobia. *British Journal of Clinical Psychology*, 48(2), 125–140. <https://doi.org/10.1348/014466508X368856>
- Swartz, J., Phan, L., Angstadt, M., Klumpp, H., Fitzgerald, K., & Monk, C. S. (2014). Altered activation of the rostral anterior cingulate cortex in the context of emotional face distractors in children and adolescents with anxiety disorders., 31(10), 870–879.
- Teachman, B. A., Joormann, J., Steinman, S., & Gotlib, I. H. (2012). Automaticity in Anxiety Disorders and Major Depressive Disorder. *Clinical Psychology Review*, 32(6), 575–603. <https://doi.org/10.1016/j.cpr.2012.06.004>.Automaticity
- Thissen, D. (2000). Reliability and measurement precision. In H. Wainer (Ed.), *Computerized adaptive testing: A primer* (pp. 159–184).
- Van Duijvenbode, N., Didden, R., Korzilius, H. P. L. M., & Engels, R. C. M. E. (2016). The addicted brain: Cognitive biases in problematic drinkers with mild to borderline intellectual disability. *Journal of Intellectual Disability Research*, 60(3), 242–253.

<https://doi.org/10.1111/jir.12244>

- 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. <https://doi.org/10.1007/s10608-013-9588-2>
- White, L. K., Britton, J. C., Sequeira, S., Ronkin, E. G., Chen, G., Bar-Haim, Y., ... Pine, D. S. (2016). Behavioral and neural stability of attention bias to threat in healthy adolescents. *NeuroImage*, 136, 84–93. <https://doi.org/10.1016/j.neuroimage.2016.04.058>
- Wieser, M. J., Pauli, P., Weyers, P., Alpers, G. W., & Mühlberger, A. (2009). Fear of negative evaluation and the hypervigilance-avoidance hypothesis: An eye-tracking study. *Journal of Neural Transmission*, 116(6), 717–723. <https://doi.org/10.1007/s00702-008-0101-0>
- Zvielli, A., Bernstein, A., & Koster, E. H. W. (2015). Temporal Dynamics of Attentional Bias. *Clinical Psychological Science*, 3(5), 772–788. <https://doi.org/10.1177/2167702614551572>
- Zvielli, A., Vrijssen, J. N., Koster, E. H. W., & Bernstein, A. (2016). Attentional Bias Temporal Dynamics in Remitted Depression. *Journal of Abnormal Psychology*, 125(July), 768–776. <https://doi.org/10.1037/abn0000190>