# Anxiety and attention to threat:
# The psychometric properties of attentional bias scores

by

Stephanie Norma Waechter

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Psychology

Waterloo, Ontario, Canada, 2014

## AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Abstract

Attentional biases to threat are thought to play a central role in the development and maintenance of anxiety disorders. Various measures have been developed to index these biases; unfortunately, the psychometric properties of many common measures of attentional bias have not been thoroughly evaluated, and the reliability of those that have been evaluated is poor.

Three studies assessed the reliability and convergent validity of dot probe and emotional cueing bias scores. Study 1 used a dot probe task and an emotional cueing task that were designed to be as similar as possible to each other (e.g., in terms of display and timing parameters) in an attempt to maximize convergent validity. One hundred fifty-eight participants, selected for high and low levels of trait anxiety, completed the two tasks. The results of Study 1 showed no significant attentional biases to threat, and the psychometric properties of the bias scores were poor.

Study 2 investigated the psychometric properties of measures of early attention to threat using another dot probe task and another emotional cueing task. One hundred twelve participants, again selected for high and low levels of trait anxiety, completed the study. The results of Study 2 again showed no significant attentional biases in the high trait anxiety group, and the psychometric properties of the bias scores were poor.

To increase the likelihood of producing large and potentially reliable attentional biases, Study 3 replicated the methodology from two classic studies in the literature and also included a state anxiety manipulation. One hundred sixty participants completed an anxious or calm mood induction, which consisted of listening to music while thinking of an anxiety-provoking or calming event in their lives, followed by dot probe and emotional cueing tasks. Reliability estimates for bias scores were low, ranging from 0 to .44. The convergent validity estimates were also low. The anxiety induction did not substantially improve reliability or convergent validity of the bias scores.

Overall, the results of these studies improve our understanding of factors that affect the reliability of attentional bias scores. These results also provide some of the first estimates of the reliability and convergent validity of scores from the emotional cueing task. Implications for research and for clinical assessment and intervention are discussed.

# Acknowledgements

First, thank you to all the people over the years who have made the PAS, even with its ugly red carpeting, feel like home. To Andrea, Colleen, Erin, Ester, Grayden, Ivana, Jason, Nate, Paul, Randall, Robyn, Serje, Shannon, Tania, and Tanya—thank you for being such great teachers, co-conspirators, collaborators, office-mates, and friends.

I am grateful for the accomplished and supportive faculty members I have had the opportunity to work with at the University of Waterloo, including David Moscovitch, Derek Besner, Evan Risko, Jonathan Fugelsang, and Jonathan Oakman. I am especially grateful to my supervisor, Jennifer Stolz, for her help, support, and unwavering confidence in me. None of this would be possible without her.

I also want to acknowledge others who have contributed to this research. Thank you to my diligent and talented research assistants for all their help with data collection, including Arief Kartolo, Harjot Dosanjh, Jordan Kemp, Karen Batth, Klint Fung, Marina Ren, Melissa Chartrand, Regina Thurairatnam, and Yuna Lee. Thanks to Mike Lawrence for his assistance with data analysis in R, and to Andrea Nelson for all of our passionate discussions of data and theory. I want to acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research, and the University of Waterloo.

Finally, thank you to my family: Mom and Dad, Laurie, Tony, Dan, Dianne, and Derek. I am so lucky to have each of you in my life.

## Table of Contents

## List of Figures

## List of Tables

# Chapter 1

## General Introduction

Anxiety is thought to be characterized by a cognitive style in which the processing of threatening information is prioritized (e.g., Mathews, 1990; Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van Ijzendoorn, 2007). This attentional bias to threat (ABT) is seen as a causal factor in the development and maintenance of anxiety and anxiety disorders (see Van Bockstaele, Verschuere, Crombez, Tibboel, Koster, & De Houwer, 2014, for a review). For example, Mathews and MacLeod (2002) speculated that for high trait anxious individuals, especially when under stress, attention is captured by even mild threat cues. These individuals frequently attend to possible threats over other types of information, which leads in turn to increased anxiety (Mathews & Macleod, 2002). Other cognitive theories of anxiety disorders also suggest that ABT plays an important role in causing or maintaining anxiety (e.g., Bar-Haim et al., 2007; Beck & Clark, 1997; Mogg & Bradley, 1998; Rapee & Heimberg, 1997).

Multiple theories explaining the phenomenology of ABT in high anxious individuals have been advanced. The two most prominent ABT theories are the vigilance-avoidance hypothesis (e.g., Mogg, Mathews, & Weinman, 1987) and the delayed disengagement hypothesis (e.g., Fox, Russo, Bowles, & Dutton, 2001). The vigilance-avoidance hypothesis posits that individuals initially direct attention to threat stimuli over neutral stimuli; after some time, however, they then *avoid* the threat stimuli. The delayed disengagement hypothesis, in contrast, proposes that attention is not necessarily directed preferentially toward threat initially. Once threat is attended, however, anxious individuals have difficulty diverting their attention away from it. Many studies have focused on understanding the phenomenology of ABT in different populations and under different conditions (e.g., timescale, stimulus conditions, etc.), and support for both of these hypotheses has been found. Thus, ABT can present as facilitated attention, delayed disengagement, or avoidance, depending on the timescale and stimulus conditions (see Weierich, Treat & Hollingworth, 2008, for a review).

### 1.1.1 Measuring ABT

A variety of methodological paradigms have been used to examine anxiety-related ABT. Some of the most common of these are the emotional cueing paradigm, the Stroop paradigm, the dot probe paradigm, and eye-tracking paradigms. A large amount of research has been conducted using these paradigms. For example, over 100 studies on ABT were included in a 2007 review paper (Bar-Haim et al., 2007). Despite their frequent use, however, researchers have only recently begun to evaluate the psychometric properties of ABT measures resulting from these tasks. This dissertation examines the reliability and convergent validity of bias scores from dot probe and emotional cueing tasks in participants with low and high trait anxiety. First, these paradigms are described in some detail.

### 1.1.2 Dot Probe Paradigm

In the dot probe paradigm, participants are presented with a display that includes a threat cue (e.g., a threat word or an angry face) and a neutral cue. After a short interval, the cues disappear and a target appears in the location previously occupied by one of the cues. (See Figure 1 for an illustration). In a congruent trial, the target appears in the location of the threat cue. In an incongruent trial, the target appears in the location of the neutral cue. Typically, half of the trials are congruent and half are incongruent. Participants are asked to respond as quickly and accurately as possible to the target on some dimension (e.g., location or identity). Responses are assumed to be faster when the target appears in an attended location than in an unattended location; thus, faster responses on congruent trials than on incongruent trials are interpreted as an attentional “bias” toward threat.

The overall bias score (mean RT on incongruent trials – mean RT on congruent trials) is the most common method of computing bias in the dot probe task, but it is not the only method. Koster, Crombez, Verschuere, and De Houwer (2004) argued that the bias score described above does not differentiate between facilitated attention towards threat and difficulty disengaging from threat. Facilitated attention refers to accelerated attention toward the threat information over neutral information. Difficulty disengaging refers to difficulty moving attention away from threat information compared with neutral information, once it has been attended. Koster and colleagues (2004) proposed an alternate method of calculating attentional bias scores that requires a neutral condition (two neutral cues; no threat cues) as a baseline. Using this method, faster RTs on congruent trials in comparison with neutral trials indicate facilitated attention towards threat stimuli. In contrast, slower RTs on incongruent trials as compared to neutral trials indicate a difficulty in disengaging from threat stimuli. Cisler, Baker, and Williams (2009) recommend this method of calculating scores in addition to the traditional method, and some recent studies examining ABT in clinical populations have begun to use this method of calculation (e.g., Cooper & Langton, 2006; Stevens, Rist, & Gerlach, 2009; Waechter, Nelson, Wright, Hyatt, & Oakman, 2014) along with the traditional bias score.

Dozens of published studies have found evidence for ABT using the dot probe task (e.g., see Bar-Haim, 2007, for a meta-analysis and review), but a few published studies have found *no evidence* for such biases (e.g., Mogg, Bradley, Dixon, Fisher, Twelftree & McWilliams, 2000 in high trait anxiety; Bradley et al., 1997 in high social anxiety). These contradictory results have typically been attributed to differing experimental methodologies (e.g., Bögels & Mansell, 2004). Given that dot probe tasks in the literature vary widely on many factors, including stimuli (faces, words, scenes or objects), task (detection, localization, or identification), participants (clinical or non-clinical samples), timing parameters, display parameters, and number of trials, this assertion by Bögels and Mansell (2004) is certainly possible. In fact, a meta-analysis on attentional bias (Bar-Haim et al. 2007) found evidence for a moderating role of timing such that differences between anxious and control participants were significant for subliminal (e.g., <60 ms and then masked: McNally, Amir, & Lipke, 1996; Mogg, Bradley, Williams, & Mathews, 1993) and 500 ms exposures, but failed to reach significance for longer (e.g., 1,000 ms; Bar-Haim et al., 2007) exposures. Bar-Haim et al. (2007)'s meta-analysis also suggested that the stimuli (e.g., words vs. pictures) may be an important factor in certain contexts. For example, in studies with clinically anxious participants, there was a significant attentional bias with words but not with pictures. In studies using subclinical anxious participants, the bias was significant with both words and pictures (Bar-Haim et al. 2007).

Thus, there is some strong evidence that dot probe effects are moderated by methodological factors. At least one recent study, however, has found inconsistent results in two experiments using nearly identical methodology (Cooper et al., 2011). This suggests that the mixed results in the dot probe literature may also be in part due to other issues, such as poor reliability (as proposed by Cooper et al., 2011).

### 1.1.3 Emotional Cueing Paradigm

In this paradigm, developed by Fox, Russo, Bowles, and Dutton (2001), participants are presented with a display that includes a single cue (e.g., a word, image, or face) to the left or right of fixation. The cue can be threatening (e.g., a threatening word or angry face) or neutral (e.g., a neutral word or calm face). After a short interval, a target appears either to the left or right of fixation. (See Figure 2 for an illustration). In a congruent trial, the target appears in the location previously occupied by the cue. In an incongruent trial, the target appears in the opposite location. Again, participants are asked to respond as quickly and accurately as possible to the target on some dimension (e.g., location or identity). A congruency (cueing) score can be computed for each emotion by subtracting mean RTs on congruent trials from mean RTs on incongruent trials. This is not typically interpreted as an attentional bias score, as it does not reflect changes in attention across emotions. Instead, it simply reflects differences in RTs between incongruent and congruent trials. In the emotional cueing paradigm, an overall bias score is computed using the difference between the congruency effects for threat and neutral trials. In addition, faster responses on congruent threat-cued trials relative to congruent neutral-cued trials are interpreted as a facilitation bias toward threat. Slower responses on incongruent threat-cued trials relative to incongruent neutral-cued trials are interpreted as a disengagement bias (difficulity disengaging) from threat.

The emotional cueing paradigm has also produced evidence for ABT in anxious participants (e.g., Yiend & Mathews, 2001; Mogg, Holmes, Garner, & Bradley, 2008). Specifically, a number of studies have demonstrated difficulty disengaging from threat in high trait anxious individuals (Fox, Russo, & Dutton, 2002, Experiment 1; Yiend & Mathews, 2001), whereas other studies with high trait anxious individuals have found both facilitated attention and delayed disengagement (e.g., Van Damme, Crombez, Hermans, Koster, & Eccleston, 2006; Koster, Crombez, Verschuere, Van Damme, and Wiersema (2006) with highly threatening cues. In their meta-analysis, Bar-Haim et al. (2007) found significant threat-related biases for anxious participants using a within-participant analysis in the emotional cueing paradigm. The between-groups effect (comparing high anxious and low anxious participants) did not reach significance, but this analysis was based on only four studies.

Less is known about moderating factors in the emotional cueing paradigm compared with the dot probe paradigm, chiefly because there are fewer studies using emotional cueing (Bar-Haim et al., 2007).

As with the dot probe paradigm, a few studies using the emotional cueing paradigm have failed to find significant anxiety-related ABT (e.g., Broomfield & Turpin, 2005; Koster, Leyman, De Raedt, & Crombez, 2006). As discussed by Cooper, Rowe, Penton-Voak, and Ludwig (2009), poor reliability may also contribute to the mixed and null results in the emotional cueing literature.

### 1.1.4 Psychometric Properties of ABT Measures

#### 1.1.4.1 Reliability

Reliability refers to the proportion of a measure's variance that reflects true score variance, rather than measurement error. A measure's reliability can be estimated in a number of ways, including split-half reliability and internal consistency reliability (e.g., Sattler, 2008). Split-half reliability reflects the extent to which a participant's score on one half of a measure predicts his or her score on the second half of the measure. A related but somewhat more stable measure, internal consistency reliability, reflects the extent to which different items on the same test correlate with each other. Internal consistency reliability is often measured using Cronbach's alpha, which is mathematically equivalent to the average of all possible split halves.

Evaluating reliability is a crucial step in the development of measures in psychology. Low reliability can decrease statistical power to detect correlations between measures and differences between groups (see Kopriva & Shaw, 1991). If measures of ABT have low reliability, this may in part explain some of the null and contradictory findings in the ABT literature. Although there is no strict rule for what level of reliability is acceptable, reliability values below .60 are typically considered very low or unacceptable (Murphy & Davidshofer, 2005; Sattler, 2008). Values above .70 are considered moderate, and values above .90 are considered high or excellent (Murphy & Davidshofer, 2005; Sattler, 2008).

#### 1.1.4.2 Reliability of response time measures

Traditional response time (RT) research examines the effect of experimental manipulations by comparing RTs across two or more conditions. Researchers typically report the mean RT by condition as the main dependent variable, although *difference scores* in RTs between conditions are also sometimes used. Unfortunately, the reliabilities of RT measures often remain unassessed and unreported in traditional RT research (see Stolz, Besner, & Carr, 2005 for a discussion of this issue).

Over the past two decades, however, a small subset of RT researchers have taken an interest in the reliability of RT measures, and a handful of studies on the reliability of RT effects have been published in the cognitive psychology and clinical psychology literatures. These studies generally show acceptable to excellent reliabilities for raw RTs (e.g., all estimates above .8 in

Waechter et al., 2014; all estimates above .7 in Eide, Kemp, Silberstein, Nathan, & Stough, 2002). For RT difference scores, however, studies show both low reliabilities (near 0 for emotional Stroop scores in Eide et al., 2002; below .60 for Alerting and Orienting indices from the ANT: MacLeod et al., 2010; below .60 for two versions of the Numerical Distance effect: Maloney, Risko, Preston, Ansari, & Fugelsang, 2010) and acceptable reliabilities (Spearman-Brown corrected reliabilities of .70 for semantic priming effects in Waechter et al., 2014; corrected reliability of .81 for the Executive Control index from the ANT: MacLeod et al., 2010). Thus, while raw RTs are typically very reliable, the reliability estimates for RT difference scores are more variable and can be unacceptably low.

A recent study by Miller and Ulrich (2013) uses a model based on classical test theory to examine the factors that affect the reliability of RTs and RT difference scores. Miller and Ulrich (2013) demonstrate that the reliability of raw RTs depends strongly on the number of trials per subject and condition; however, the minimum number of trials for acceptable reliability is not high. In fact, the reliabilities of RT measures are generally higher than .85 with even 10–20 trials per participant. RT variability also has an effect on the reliability of RTs; similarly, though, the impact of within-condition RT variability on reliability becomes less strong as the number of trials increases, such that RT variability generally only has a substantial impact with fewer than 20-30 trials per participant and condition. The authors also examined the impact of participants' general processing time (G), their processing time for the particular task at hand (Δ), and their residual sensory-motor processing times (R, which reflects factors such as light transduction processes in the retina and the movement of muscle fibers in the hand for a key press) on RT reliability. They found that more variability across participants on G, Δ, and R results in higher reliability, which is consistent with the Classical Test Theory principle that reliability increases with the amount of true score variance. That said, as long as there is sufficient variation on at least one of G, Δ, and R, raw RTs tend to be reliable (Miller & Ulrich, 2013). Finally, Miller and Ulrich (2013) demonstrate that the reliability of raw RTs is relatively independent of task difficulty. These conclusions are consistent with the pattern observed above from the empirical literature: Except in unusual circumstances (e.g., with very few trials or highly limited variability in participant processing times), raw RTs are quite reliable.

Miller and Ulrich (2013) also examine factors that affect the reliability of RT *difference scores*. Again, the number of trials has a strong effect. Unlike with the raw RTs, 10 trials are typically not sufficient to attain high reliability with difference scores. In fact, in some cases (depending on the factors discussed next), hundreds of trials per participant and condition are needed to exceed a reliability of .80. Miller and Ulrich (2013) also demonstrate that reliability increases with larger effects. Increases in the time taken for both perceptual input and motor output actually *decrease* reliability, largely by increasing trial-to-trial error variance. Similarly, within-condition RT variability also decreases reliability. Increases in G and Δ typically have small positive effects on reliability.

Finally, Miller and Ulrich (2013) note that reliabilities are higher when the difference scores are taken from *opposing task processes* rather than from *common task processes.* For example, traditional (non-emotional) Stroop and cueing effects, as well as the flanker effect, result from opposing task processes. With opposing task processes, task-specific processing times for the two conditions that make up the difference score are strongly negatively correlated. This occurs when the hypothesized process underlying the difference score has opposing effects on the two task-specific processing times. For example, with the Stroop effect, the difference in RTs between incongruent and congruent trials is thought to represent the effect of automatic word reading. When word reading is more automatic, task-specific processing times for congruent trials decrease and task-specific processing times for incongruent trials increase (Miller & Ulrich, 2013); thus, these are opposing task processes. Conversely, reliabilities are lower when the difference scores are taken from *common task processes.* With common task processes, task-specific processing times for the two conditions are strongly positively correlated. Miller and Ulrich (2013) use the example of the difference between mental rotation scores when rotating objects 180 degrees versus 90 degrees. In this case, the difference between the two conditions simply reflects efficiency at mental rotation, and the task-specific processing times are positively correlated. In the case of *unrelated task processes*, in which the task-specific processing times by condition are not theoretically correlated either positively or negatively with each other, reliabilities tend to be intermediate.

Miller and Ulrich's model (2013) clearly demonstrates that the reliabilities of RT difference scores are affected by various factors and can vary widely. The factors that impact the reliability of difference scores include the number of trials, the size of effects, the within-condition RT variability, the variability in G and Δ, and the type of task process (opposing or common). Miller and Ulrich's (2013) findings help to clarify and explain the empirical literature discussed above, which shows that some RT difference scores have acceptable reliability and that some have very low reliability. Next, the literature specifically examining the reliability of bias scores from dot probe tasks and emotional cueing tasks is reviewed.

1.1.4.3 Reliability of dot probe bias scores

Over the past decade, a handful of studies on the reliability of dot probe bias scores have been published. Schmukle (2005) and Staugaard (2009) evaluated the reliability of dot probe bias scores using words and face images, respectively, with university student volunteers. Unlike in most ABT studies, these students were not pre-selected for particular levels of trait anxiety. Both Schmukle (2005) and Staugaard (2009) found that the reliabilities of the bias scores were extremely low. In fact, a number of the estimates were negative and the highest reliability estimate from either study was .32 (Schmukle, 2005). Van Bockstaele and colleagues (2011), Cooper and colleagues (2011), and Waechter et al. (2014) found similarly low reliabilities for dot probe bias scores. In a recent study with children as participants, Brown et al. (2014) also found unacceptably low reliabilities for these bias scores ($r$s < .29).

Two studies, however, have found somewhat higher reliabilities for bias scores using dot probe tasks. Bar-Haim et al. (2010) assessed the reliability of bias scores with participants living under war-related threat near the Gaza Strip in early 2009. The ABT analysis revealed a bias *away* from threat in the high-risk groups. The authors reported a split-half reliability of $r$ = .45, which translates to an acceptable Spearman-Brown corrected reliability of .62. In another study, Enock, Hofmann, and McNally (2014) asked participants to complete three sessions per day of Attentional Bias Modification (ABM) training for 28 days. The active ABM training was designed to teach participants to attend to neutral rather than threat faces. The training paradigm was similar to the dot probe task, except that in the active training condition the target always appeared in the location of the neutral face (i.e., all incongruent trials). As a result, the best strategy for fast responding was to attend to the neutral face on each trial. The researchers also included a control training condition, in which participants completed a traditional dot probe task (half congruent and half incongruent trials) three times a day. ABT was assessed with a traditional dot probe task once a week, and the reliability of these bias scores was examined. Enock and colleagues (2014) found that both groups had very unreliable (near 0) dot probe bias scores initially and in the first week of ABM. However, the reliability of the bias scores increased for both groups as the training continued. At week 4 for the active ABM group, the Spearman-Brown corrected reliability of the bias score was .53.

Although the majority of studies have shown that dot probe bias scores have very low reliability, the results of Bar-Haim et al. (2010) and Enock et al. (2014) are promising in that they provide empirical evidence that dot probe scores can be at least somewhat more reliable in certain contexts. Unfortunately, it is not clear from the current literature precisely what those contexts are. Scores may become more reliable after daily repetition of the dot probe task for several weeks (Enock et al., 2014); this repetition is of course not feasible for most ABT studies, which typically take place in a single session.

#### 1.1.4.4 Reliability of emotional cueing bias scores

Very little is known about the reliability of emotional cueing bias scores. Only one published paper (Enock et al., 2014) has examined this question, and that was in the context of an ABM training study. Enock et al. (2014) found Spearman-Brown corrected reliabilities near 0 (-.22 and -.04) for emotional cueing overall bias scores, suggesting that the reliability of these scores is unacceptably low. No study has yet examined the reliability of facilitation and disengagement bias scores in the emotional cueing task.

#### 1.1.4.5 Summary of reliability data

In sum, the reliability of dot probe bias scores is generally very poor, and the reliability of emotional cueing bias scores is largely unknown. Other RT measures of ABT do not fare much better—the reliabilities of emotional Stroop bias scores (e.g., Eide et al., 2002; Brown et al., 2014) and visual search bias scores (Brown et al., 2014) are also very low, and the reliability of ABT measures from the attentional blink task is unknown. Using measures with low reliability in ABT research is problematic in that it (a) reduces power to detect correlations and group differences (e.g., Kopriva & Shaw, 1991); (b) contributes to inconsistencies in the literature (e.g., Cooper et al., 2011); and (c) impedes progress in our understanding of ABT and their association with anxiety disorders (see Van Bockstaele et al., 2014).

#### 1.1.4.6 Validity

Validity refers to the extent to which a test measures the construct that it is intended to measure. That is, to what extent do biases from the dot probe and emotional cueing paradigm actually index ABT? The current study examines evidence for the *convergent validity* of ABT scores from these two tasks. If dot probe and emotional cueing scores index the same construct (ABT), they should in theory be strongly correlated.

To date, there is limited evidence for convergent validity between different measures of ABT in the literature. A few studies have investigated correlations between individuals' dot probe and emotional Stroop bias scores. These studies have revealed either no correlation (Dalgleish et al., 2003; Mogg et al., 2000; Gotlib et al., 2004) or modest correlations (Brosschot, de Ruiter, & Kindt, 1999; Egloff & Hock, 2003). Because of the lack of strong correlations between measures, these results are typically interpreted as evidence that the dot probe and emotional Stroop tasks index separate processes. No published study has yet examined the correlation between dot probe and emotional cueing scores, despite their similar use in indexing biases in spatial attention.

Notably, Miller and Ulrich (2013) urge caution in the interpretation of correlations involving RTs and difference scores. First, correlations between difference scores are limited by the reliability of the difference scores; two measures will not correlate if they are unreliable. Second, Miller and Ulrich's (2013) model shows that correlations involving RT difference scores are influenced by numerous factors and that the *observable* correlation between two RT difference scores is often a poor reflection of their true coherence. That is, the observed correlation between two ABT measures does not necessarily reflect the extent to which they are both measuring the same process (e.g., ABT). Although the present work uses correlations to examine the convergent validity of dot probe and emotional cueing tasks, the results of these analyses should be interpreted with caution in light of Miller and Ulrich's findings.

# Chapter 2
# Psychometric Properties of a Dot Probe and an Emotional Cueing Task

## 2.1 Study 1 Introduction

A review of the literature showed that the reliabilities of RT measures of ABT are either largely unknown (e.g., emotional cueing bias scores) or unacceptably low (e.g., dot probe bias scores). If ABT indices with low reliability continue to be used in research, then progress in understanding the role of ABT in anxiety, anxiety disorders, and anxiety interventions will be necessarily limited. The purpose of Study 1 is to examine the psychometric properties of two of the most common measures of spatial ABT. In doing so, some limitations of the existing literature will be addressed.

First, when the dot probe and emotional cueing measures are used in the anxiety literature, participants are typically recruited based on high and low trait anxiety scores (e.g., Mogg & Bradley, 1999; Fox et al., 2001). No study has yet examined the reliability of dot probe or emotional cueing measures with participants recruited for high and low trait anxiety. Recruiting participants with a wide range of anxiety levels (and therefore presumably a wide range of attentional bias scores) may be especially important because a wider range of measured true scores typically results in higher measured reliability (e.g., Horwitz & Horwitz, 2012).

A second limitation of the existing literature is how little is known about the reliability of emotional cueing scores. Only one recently published study has examined this question (Enock et al., 2014), and this study looked only at the reliability of the overall emotional cueing scores. Thus, the current study is the first to examine the reliability of facilitation and disengagement bias scores from the emotional cueing task.

Third, no other study has computed the convergent validity of dot probe and emotional cueing bias scores.

Study 1 examines the following: (1) the reliability of dot probe and emotional cueing bias scores; and (2) whether individuals' ABT scores are consistent across dot probe and emotional cueing tasks. In Study 1, the dot probe and emotional cueing task procedures were modelled after procedures that have produced significant ABT in previous research (Bradley, Mogg, Falla, & Hamilton, 1998 for the dot probe task; Mogg, Holmes, Garner, & Bradley, 2008 for the emotional cueing task). Because one of the questions of interest in this study was the extent to which the two paradigms measure the same construct, some modifications were implemented to make the procedures for the two tasks as similar as possible. Specifically, the timing parameters, display parameters, response options, and number of trials were set to be identical for the two tasks.

Finally, although Bradley et al. (1998) and Mogg et al. (2008) used angry faces as their threat stimuli, fear faces were also added to Study 1. With this addition, the reliability of biases toward fear faces (e.g., Fox, 2002) can also be examined.

## 2.2 Method

### 2.2.1 Participants

A large sample of undergraduate students from the University of Waterloo completed the State-Trait Inventory of Cognitive and Somatic Anxiety (STICSA; Ree, French, MacLeod, & Locke, 2008; trait version only) as part of a Mass Testing session between 2 and 8 weeks before the Study session. Participants scoring in the bottom 20% (STICSA score below 28) and top 20% (STICSA score above 44) were eligible to participate in the current study.

One hundred fifty-eight students (114 female; 44 male) participated in the study in exchange for course credit. The mean age of participants was 19.85 years. Participants identified primarily as Asian (36.7%) and White (34.8%). Other groups represented in the sample included East Indian (7.0%), Biracial (5.7%), West Indian (4.4%), Middle Eastern (3.8%), Black (1.9%), First Nations (0.6%) and Other (3.2%). Three participants (1.9%) chose not to disclose their race.

### 2.2.2 Measures and Materials

#### 2.2.2.1 Computer Task Materials

Images of 32 actors from the NimStim Face Stimulus Set (Tottenham et al., 2009) were used. Each actor portrayed the following closed-mouth expressions: Anger, Fear, Happiness, Calm, and Neutral. Figure 3 shows sample images for each emotion.

#### 2.2.2.2 State-Trait Inventory of Cognitive and Somatic Anxiety

The STICSA consists of a state version and a trait version. Each version is a 21-item measure of cognitive and somatic symptoms of anxiety with items rated on a 4-point Likert scale. Cronbach’s alphas in a validation study were .91 for the trait scale and .92 for the state scale (Grös, Antony, Simms, & McCabe, 2007).

#### 2.2.2.3 Depression Anxiety and Stress Scale (DASS-21; Lovibond & Lovibond, 1995)

The DASS-21 is a 21-item measure with three 7-item subscales: Depression, Anxiety, and Stress. Items are rated on a 4-point Likert scale. Cronbach’s alphas in a validation study were .94 for the Depression Subscale, .87 for the Anxiety Subscale, and .91 for the Stress subscale (Antony, Bieling, Cox, Enns, & Swinson, 1998). The DASS was included to differentiate any trait anxiety-related biases from biases associated with depression symptoms.

#### 2.2.2.4 Social Desirability Scale Version XX (SDS; Strahan & Gerbasi, 1972)

The SDS is a 20-item measure of the tendency to give socially desirable responses. Cronbach's alpha in a validation study was .87 (Fischer & Fick, 1993). The SDS was included because high social desirability can confound measures of anxiety and attentional bias (e.g., Fox, 1993).

### 2.2.3 Procedure

These procedures were reviewed and approved by the University of Waterloo's Office of Research Ethics. After arriving at the lab and providing informed consent, participants completed the dot probe task and the emotional cueing task in counterbalanced order. They then completed the battery of questionnaires described above.

#### 2.2.3.1 Dot probe task

For the dot probe task, each of the emotion images (one Anger, one Fear, one Happy, and one Calm or Neutral) for each of the 32 actors was paired with a Calm or Neutral image, resulting in 128 pairs. The location of the emotional face and the location of the target were counterbalanced across participants. Each participant saw each pair of faces once in Block 1 and once in Block 2, for a total of 256 trials. Participants completed Block 1, took a self-paced break, and then completed Block 2. Trials were presented in randomized order within blocks. Twelve practice trials (3 trials for each emotion) were created using additional images.

On each trial, participants saw a white fixation cross in the center of a black screen (see Figure 1). The fixation cross remained on the screen throughout the trial. After 500 ms, two faces appeared (one left of fixation, one right of fixation). The face images measured 7 cm wide by 9.5 cm tall. The closest or inner edge of each image appeared 1.0 degrees of visual angle from the center of the screen. The faces remained on the screen for 500 ms, followed by a 50 ms display with only the fixation cross on the screen. Then a white target symbol (: or .. in size 16 Calibri font) appeared 5.3 degrees of visual angle to the left or right of the centre of the screen. Participants were instructed to identify the target by pressing the “m” key (labelled with a .. sticker) or the “k” key (labelled with a : sticker) as quickly and accurately as possible on a standard keyboard. The target remained on the screen until a response was made.

2.2.3.2 Emotional cueing task

The cue faces consisted of one emotion image for each of the 32 actors (one Anger, one Fear, one Happy, and one Calm or one Neutral). The location of the emotional face (left or right) and the location of the target (left or right) were again counterbalanced across participants. Each participant saw each face once in Block 1 and once in Block 2, for a total of 256 trials. Again, trials were presented in randomized order within blocks. Twelve practice trials (3 trials for each emotion) were created using additional NimStim images.

Participants first saw a white fixation cross in the center of a black screen, flanked by two white rectangular boxes measuring 7.5 cm wide by 10 cm tall (see Figure 2). The fixation cross and the boxes remained on the screen throughout the trial. The centre of the images appeared 5.3 degrees of visual angle from the center of the screen. After 500 ms, a single face (7 cm wide by 9.5 cm tall) appeared in one of the boxes. The face remained on the screen for 500 ms, followed by a 50 ms display with only the fixation cross and boxes on the screen. Then a white target symbol (: or .. in size 16 Calibri font) appeared in the centre of one of the boxes, 5.3 degrees of visual angle from the centre of the screen. Participants identified the target using a key press as in the dot probe task. The target remained on the screen until a response was made.

## 2.3 Results

### 2.3.1 Anxiety Groups

Participants were divided into groups based on their trait anxiety (STICSA) scores at the testing session. Participants whose STICSA Trait scores varied between the testing session and the in-lab session to the extent that they switched groups (participants with scores below 27 on the screening session and above 45 on the in-lab session, or vice-versa) were identified (n = 2). To ensure stable and accurate measurement of trait anxiety, these participants were excluded from analysis. Two groups were created based on the median in-lab STICSA Trait score (38). The groups consisted of 79 participants (low trait anxiety; LTA) and 77 participants (high trait anxiety; HTA), respectively.

A small amount of questionnaire data (<1.0%) was missing. These missing data were replaced using Expectation-Maximization imputation (Dempster, Laird, & Rubin, 1977). Demographics and questionnaire scores for the two groups are presented in Table 1. The HTA group was significantly older than the LTA group (see Table 1), but the groups did not differ on sex. The groups differed as expected on trait and state anxiety (STICSA), depression, anxiety, and stress symptoms (DASS), and social desirability scores (SDS).

As in Mogg and Bradley (1999), Social Desirability Scale (SDS) scores were examined for unusually large values. The box-and-whisker plot did not identify any outliers, and when SDS scores were transformed to z-scores no value was above 2.5. Thus, no participants were excluded for scoring unusually high on social desirability.

### 2.3.2 Dot Probe Results

Response time (RT) analyses were conducted for experimental trials in which an accurate response was given. Two participants (one from each group) had fewer than 75% accurate responses ($M$ = .51 and $M$ = .73) and were excluded from the dot probe RT analyses. The other participants' responses were very accurate ($M$ = .946). As is typical in the ABT response time literature, error rates and trials with errors were not analyzed further.

To identify outliers, a non-recursive method was used that adjusted the cut-off criterion depending on the sample size for each participant and condition (Van Selst & Jolicoeur, 1994). This method was chosen over the box and whisker procedure for the RTs (e.g., Mogg & Bradley, 1999), in which all RTs more extreme than a particular value are identified (regardless of participant and condition) for two reasons: (1) outlier procedures with hard cut-offs typically identify a disproportionate amount of outliers in the fastest and slowest conditions, which may artificially attenuate real effects in RT (see Van Selst & Jolicoeur, 1994); and (2) with hard cut-off procedures, the sample size highly influences what the mean value will be after outliers are removed (Miller, 1991). The Van Selst and Jolicoeur (1994) procedure was developed to address these concerns. For the dot probe task, outliers made up 2.88% of correct RT data and were excluded from analysis[1].

A 2 (Anxiety Group: HTA vs. LTA) X 3 (Emotion: Angry vs. Fear vs. Happy) X 2 (Congruency: Incongruent vs. Congruent) mixed ANOVA was conducted on the RT data. In a congruent trial, the target followed the faces in the location of the emotional face; in an incongruent trial, the target followed the faces in the location of the neutral face. Table 2 presents

[1] When these data were analysed using an alternate hard cut-off outlier procedure, the ABT results were similar and the reliability estimates were similar or lower.

the RT data. For this and all subsequent ANOVAs, Greenhouse-Geisser corrected estimates were used when necessary (i.e., when assumptions of sphericity were violated). There were no significant effects of Anxiety Group: $F(1, 152) = 1.11$, $MSE = 55908.28$, $p = .294$, Emotion: $F(1.81, 275.7) = .077$, $MSE = 991.02$, $p = .910$, or Congruency: $F(1, 152) = 1.18$, $MSE = 550.93$, $p = .280$. There were no interactions, *F*s < 1.6, *p*s > .20. Thus, the ANOVA revealed no evidence of anxiety-related ABTs.

Next, bias scores were computed. First, traditional dot probe bias scores were computed for each emotion by subtracting the mean RT on congruent trials from the mean RT on incongruent trials. Then, a facilitated bias score was calculated by subtracting the mean RT from congruent trials from the mean RT on neutral-neutral trials (Koster et al., 2004). A disengagement bias score was computed by subtracting the mean RT from neutral-neutral trials from the mean RT on incongruent trials (Koster et al., 2004). T-tests were computed comparing bias scores for each group and emotion to 0 (see Table 3). There were no significant biases.

Correlations between the bias scores and the questionnaire measures were also computed. The results are presented in Table 4. There were no significant correlations with STICSA scores. There were small but significant correlations between SDS scores and both overall ($r(154) = .22$, $p < .01$) and disengage ($r(154) = .17$, $p < .05$) bias scores for angry faces.

#### 2.3.2.1 Dot probe reliability

To estimate the reliability of the bias scores, a permutation approach was used (e.g., MacLeod et al., 2010; Enock et al., 2014). The data were split into two random halves 1000 times using R (R Development Core Team, 2009). For each of these 1000 permutations, the dot probe bias score was calculated for each participant for Half 1 and Half 2, and these scores were correlated. The reliability estimate was obtained by averaging the 1000 correlations, and the estimate was subsequently corrected for test length using the Spearman-Brown prophecy formula. This approach was used because single split-half reliabilities tend to produce unstable estimates (e.g., Enock et al., 2014), and because traditional reliability estimates like Cronbach's alpha cannot be applied to difference scores in a straightforward way.

Table 5 presents the mean reliability estimates based on the 1000 permutations, 95% confidence intervals based on Fischer's r-to-z transformation, and Spearman-Brown corrected reliability estimates for each bias score. Note that negative reliability estimates are theoretically impossible and are considered misestimates of 0. The corrected reliability estimates ranged from 0 to .50, all of which were below acceptable levels of reliability.

### 2.3.3 Emotional Cueing Results

Data from five participants were excluded from the emotional cueing analyses for low accuracy (< 75% correct trials). The remaining participants were generally accurate ($M$ = .937). Inaccurate trials were excluded from the RT analyses. The same non-recursive method adjusting the cut-off criterion by sample size was used for identifying outliers (Van Selst & Jolicoeur, 1994). Outliers made up 2.79% of correct RT data and were excluded from analysis.

A 2 (Anxiety Group: HTA vs. LTA) X 4 (Emotion: Angry vs. Fear vs. Happy vs. Neutral) X 2 (Congruency: Incongruent vs. Congruent) mixed ANOVA was conducted on the RT data. In a congruent trial, the target appeared in the location previously occupied by the face cue; in an incongruent trial, the target appeared in the opposite location. Table 2 presents RTs by condition. There were no significant effects of Anxiety Group: $F$ (1, 149) = .831, $MSE$ = 40374.71, $p$ = .363, Emotion: $F$ (2.85, 424.21) = .419, $MSE$ = 991.02, $p$ = .729, or Congruency:

*F* (1, 149) = .689, *MSE* = 924.63, *p* = .408. There were no interactions, *F*s < 1, *p*s > .40. Thus, the ANOVA revealed no evidence of anxiety-related ABTs.

Next, three bias indices were calculated for the emotional cueing task (e.g., Koster et al., 2006a): an overall bias, computed by subtracting the congruency effect for neutral trials from the congruency effect for emotional trials; a facilitation bias, computed by subtracting mean RTs for congruent emotional trials from mean RTs for congruent neutral trials; and a disengagement bias, computed by subtracting mean RTs for incongruent neutral trials from mean RTs for incongruent emotional trials. Planned t-tests were conducted comparing these bias scores to 0 for each group and emotion. There were no significant biases (see Table 3).

Correlations between the bias indices and the questionnaire data are presented in Table 5. The angry bias was *negatively* correlated with state anxiety scores (*r* (151) = -.17, *p* < .05) only.

2.3.3.1 Emotional cueing reliability

Table 5 presents the reliabilities for the difference scores (congruency effects, overall biases, facilitation biases, and disengage biases) from the emotional cueing paradigm. All corrected reliability estimates for the bias scores were low, ranging from 0 to .43. Even the congruency scores (which are not a measure of bias but rather measure differences in RTs between incongruent and congruent cues) showed low reliability, ranging from .17 to .44.

### 2.3.4 Between Task Correlations

To determine whether bias scores on the two tasks were related, Pearson’s correlations were computed between bias scores on the dot probe and emotional cueing tasks for each emotion. There were no significant correlations, *p*s > .05.

## 2.4 Discussion

### 2.4.1 Reliability

In Study 1, the reliability estimates for dot probe bias scores ranged from 0 to .43, and the reliability estimates for emotional cueing bias scores ranged from 0 to .50. When only the estimates for angry and fear trials (as opposed to happy trials, which do not measure biases to *threat*) are examined, these estimates ranged from .11 to .50 in the dot probe task and from 0 to .43 in the emotional cueing task. In addition, many of the 95% confidence intervals around the reliability estimates overlapped with 0. Thus, the results of this study are consistent with previous findings that dot probe bias scores have unacceptably low reliability (e.g., Schmukle, 2005; Staugaard, 2009; Waechter et al., 2014). This study is also among the first to examine the reliability of emotional cueing bias scores. Although one published study (Enock et al., 2014) included a reliability analysis for overall bias scores for the emotional cueing task, this is the first study to examine the reliability of the facilitation and disengagement bias scores from this task. The current results suggest that the facilitation and disengagement bias scores from the emotional cueing task also have unacceptably low reliability.

HTA and LTA participants were selected for this study based on the hypothesis that more variation in trait anxiety might lead in turn to more variation in ABT scores, which should improve reliability (e.g., Horwitz & Horwitz, 2012). As discussed above, however, the observed reliability estimates from the current study were in the unacceptable range and not substantially higher than other estimates in the literature (e.g., Staugaard, 2009). Therefore, the selection of HTA and LTA participants likely had no effect on the reliability of ABT scores.

### 2.4.2 ABT and Response Times

The current study failed to find significant attentional biases. The HTA participants did not show ABT in either the dot probe or the emotional cueing task. There was a significant correlation between state anxiety scores and the emotional cueing bias for angry faces, but this correlation was not in the expected direction and thus does not reflect anxiety-related ABT.

It is unclear why the current study failed to replicate the biases found in many dot probe and emotional cueing studies (e.g., as reviewed by Bar-Haim et al., 2007); however, some other studies have also failed to find significant ABT in anxious populations (e.g., Bradley et al., 1997; Mogg et al., 2000). In the literature, these null and mixed results are typically attributed to methodological differences between studies. It is possible that the methodological changes made to two tasks (e.g., the addition of fear faces; addition of more trials; changes to the timing parameters and the display sizes in order to make the two tasks as similar as possible) may have resulted in the observed null findings. Alternately, the low reliability of the measures may contribute to these null findings, as reliability affects power to detect differences between groups (Kopriva & Shaw, 1991).

Notably, Study 1 failed to find an overall congruency (cueing) effect in the emotional cueing task. Typically in emotional cueing studies, participants respond more quickly to congruent trials compared to incongruent trials (but see Broomfield & Turpin, 2005, for the opposite pattern). In the case of Study 1, the lack of a congruency effect was likely due to the relatively long stimulus onset asynchrony (SOA; 550 ms) between the cue and the target. Cueing effects in response to peripheral cues (such as those in Study 1) are most robust at relatively short SOAs. For example, Muller and Rabbitt (1989) examined participants' accuracy on valid trials with both peripheral and central cues in a standard Posner cueing paradigm. They found that responses to peripheral cues showed a fast rise in accuracy that peaked at the 175 ms SOA—much shorter than the 550 ms SOA used in Study 1. In the existing literature, the emotional cueing paradigm typically involves SOAs of 150-300 ms (Fox et al., 2001; Mogg et al., 2001) although SOAs of up to 500 ms have also been used (Koster, Crombez, Verschuere, Van Damme, & Wiersema, 2006). Like the congruency effect, ABT in the emotional cueing task may also be more readily found with a shorter SOA after the onset of the emotional face.

### 2.4.3 Convergent Validity

Despite the clear methodological similarities between the two tasks, the bias scores resulting from the dot probe and emotional cueing tasks were unrelated. It is likely that the correlations between tasks were limited (at least to some extent) by the poor reliability of the bias scores or by other factors that affect these correlations (see Miller & Ulrich, 2013). Alternatively, despite the fact that the two tasks in Study 1 used very similar displays, identical timing parameters, *and* that both the dot probe and emotional cueing tasks putatively measure ABT, the two bias scores may not actually be indexing the same processes.

# Chapter 3

# Psychometric Properties of Measures of Early ABT

## 3.1 Study 2 Introduction

Study 1 examined the reliability and convergent validity of dot probe and emotional cueing bias scores. Consistent with previous research, the results showed unacceptably low reliability for dot probe bias scores. This study was also among the first to examine the reliabilities of emotional cueing scores, which were also unacceptably low in Study 1.

Importantly, Study 1 failed to replicate the typical congruency (cueing) effect in the emotional cueing task. The lack of congruency effect was likely due to the atypically long SOA between the onset of the face cues and the targets in Study 1. In an attempt to obtain the congruency effect seen in previous research, Study 2 is identical to Study 1 but uses an SOA of 150 ms.

The shorter SOA in Study 2 also allows for the investigation of the psychometric properties of *early* attention to threat. In typical dot probe and emotional cueing tasks, information about attention is captured only at a single time point. With the 500 ms SOA, for example, bias scores are intended to capture attention to threatening information 500 ms after the threat onset. To date, a number of studies have used multiple SOAs to better understand the time course of attention to threat in both unselected and anxious participants (e.g., Mogg, Bradley, De Bono, & Painter, 1997; Cooper & Langton, 2006; Fox, Russo, & Dutton, 2002; Koster, Verschuere, Crombez, & Van Damme, 2005), with some interesting results. For example, using a dot probe task, Cooper and Langton (2006) found an attentional bias to angry faces with a 100 ms SOA. With a 500 ms SOA, they found a negative attentional bias, which can be interpreted as the avoidance of angry faces (Cooper & Langton, 2006).

Study 2 will allow for better understanding of the psychometric properties of measures of early attention (i.e., 150 ms after threat onset) to threat.

## 3.2 Method

### 3.2.1 Participants

The recruitment procedures were identical to Study 1. No individuals who had completed Study 1 were permitted to participate in Study 2.

One hundred twelve undergraduate students (86 female; 26 male) participated in exchange for credit in a psychology course. The mean age of participants was 20.40 years. Participants identified primarily as Asian (48.2%) and White (30.4%). Other groups represented in the sample included East Indian (7.1%), Biracial (7.1%), West Indian (2.7%), Middle Eastern (2.7%), and Other (0.9%). One participant (0.9%) chose not to disclose her race.

### 3.2.2 Measures and Materials

The measures and materials used were identical to those used in Study 1.

### 3.2.3 Procedure

The procedure was identical to Study 1 with one exception. In the computer tasks, the cue stimuli (two faces for the dot probe task; one face for the emotional cueing task) appeared for 100 ms rather than 500 ms.

### 3.3 Results

### 3.3.1 Anxiety Groups

Participants were again divided into groups based on their trait anxiety (STICSA) scores at the testing session. Participants whose STICSA Trait scores varied between the testing session and the in-lab session to the extent that they switched groups (participants with scores below 27 on the screening session and above 45 on the in-lab session, or vice-versa) were identified (n = 3). To ensure stable and accurate measurement of trait anxiety, these participants were excluded from analysis. Two groups were created based on the median in-lab STICSA Trait score (37). The groups consisted of 54 participants (LTA) and 55 participants (HTA), respectively.

A small amount of questionnaire data (<1.0%) was missing; these missing data were replaced using Expectation-Maximization imputation. Demographics and questionnaire scores for the two groups are presented in Table 6. The groups did not differ on age or sex, but differed as expected on trait and state anxiety (STICSA), depression, anxiety, and stress symptoms (DASS), and social desirability scores (SDS).

As in Mogg and Bradley (1999), Social Desirability Scale (SDS) scores were examined for unusually large values. The box-and-whisker plot identified one outlier with a score of 20. When SDS scores were transformed to z-scores, this participant's z-score was greater than 2.5 ($z$ = 2.67). The outlier was excluded from the remaining analyses.

### 3.3.2 Dot Probe Results

RT analyses were conducted for experimental trials in which an accurate response was given. Participants' responses were very accurate ($M$ = .946), and no participant had less than 75% correct responses. The same non-recursive method was used to identify outliers (Van Selst

& Jolicoeur, 1994) as in Study 1[2]. Outliers made up 2.81% of correct RT data and were excluded from analysis.

A 2 (Anxiety Group: HTA vs. LTA) X 3 (Emotion: Angry vs. Fear vs. Happy) X 2 (Congruency: Incongruent vs. Congruent) mixed ANOVA was conducted on the RT data (see Table 7). There were no significant effects of Anxiety Group: $F$ (1, 106) = 0.20, $MSE$ = 42368.22, $p$ = .656, or Congruency: $F$ (1, 106) = .491, $MSE$ = 626.12, $p$ = .485. There was a significant effect of Emotion, $F$ (2, 212) = 3.52, $MSE$ = 608.24, $p$ = .031, $\eta_p^2$ = .032. Follow-up t-tests revealed that participants responded more quickly to happy trials ($M$ = 516) than angry trials ($M$ = 522), $t$ (107) = 2.18, $p$ = .032, $d$ = .209, or fear trials (($M$ = 521), $t$ (107) = 2.19, $p$ = .031, $d$ = .210. Finally, there was a significant Anxiety Group X Congruency interaction, $F$ (1, 106) = 6.61, $MSE$ =626.12, $p$ = .012, $\eta_p^2$ = .059. The LTA group had larger Congruency effects ($M$ = 6.5 ms) than the HTA group ($M$ = -3.7 ms), $t$ (106) = 2.59, $p$ = .011, $d$ = .249. There were no other interactions, $F$s < 1, $p$s > .40; thus, the Congruency effect did not vary by emotion.

As in Study 1, overall, facilitation, and disengagement bias scores were computed. T-tests were computed comparing bias scores for each group and emotion to 0 (see Table 8). Contrary to expectations, the *LTA group* showed difficulty disengaging from fear faces. There were no other significant biases.

Correlations between the bias scores and the questionnaire measures were also computed. The results are presented in Table 9. There were no significant correlations.

[2] When these data were analysed using an alternate hard cut-off outlier procedure, the ABT results were similar and the reliability estimates were similar or lower.

3.3.2.1 Dot Probe Reliability

The permutation approach used in Study 1 was also used here. Table 10 presents the mean reliability based on the 1000 permutations, 95% confidence intervals based on Fischer's r-to-z transformation, and Spearman-Brown corrected reliability estimates. Corrected reliabilities ranged from 0 to .53, which are once again below acceptable levels.

### 3.3.3 Emotional Cueing Results

Data from two participants were excluded from the emotional cueing analyses for low accuracy (< 75% correct trials). The remaining participants were generally accurate (*M* = .947). Inaccurate trials were excluded from the RT analyses. The same method was used for identifying outliers as in Study 1 (Van Selst & Jolicoeur, 1994). Outliers made up 2.69% of correct RT data and were excluded from analysis.

A 2 (Anxiety Group: HTA vs. LTA) X 4 (Emotion: Angry vs. Fear vs. Happy vs. Neutral) X 2 (Congruency: Incongruent vs. Congruent) mixed ANOVA was conducted on the RT data. Table 7 presents RTs by condition. There were no significant effects of Anxiety Group: $F$ (1, 104) = .748, *MSE* = 45935.84, $p$ = .389 or Emotion: $F$ (3, 312) = 1.82, *MSE* = 576.27, $p$ = .143. As expected with the shorter SOA, there was a significant effect of Congruency, $F$ (1, 104) = 37.04, *MSE* = 1554.06, $p < .001$, $\eta_p^2 = .263$. Participants responded more quickly to congruent trials (*M* = 519 ms) compared to incongruent trials (*M* = 503 ms).

There was a marginally significant Anxiety Group X Emotion interaction, $F$ (3, 312) = 2.51, *MSE* = 576.27, $p$ = .059; however, paired samples t-tests revealed no differences between the LTA and HTA groups for angry, fear, happy, or neutral RTs, *t*s < 1.5, *p*s > .20. There were no other interactions, *F*s < 1.5, *p*s > .15. Thus, the ANOVA revealed no evidence of anxiety-related ABTs.

As in Study 1, the three bias indices were calculated for the emotional cueing task (e.g., Koster et al., 2006a). Planned t-tests were conducted comparing these bias scores to 0 for each group and emotion. The LTA group showed significant *negative* facilitation biases for angry and happy faces and a significant disengage bias for fear faces. The HTA group showed no significant biases.

Correlations between the bias indices and the questionnaire data are presented in Table 10. There were no significant correlations.

#### 3.3.3.1 Emotional Cueing Reliability

Table 9 presents the reliabilities for the difference scores from the emotional cueing paradigm. The overall, facilitation, and disengage bias scores had low but somewhat variable reliability, with corrected estimates between .03 and .58. In general, higher reliability estimates were observed for the happy trials compared with the angry and fear trials. The congruency scores were somewhat more consistent, with corrected reliabilities ranging from .39 to .47.

### 3.3.4 Between Task Correlations

To determine whether bias scores on the two tasks were related, Pearson's correlations were computed between bias scores on the dot probe and emotional cueing tasks for each emotion. There were no significant correlations for angry trials, *p*s > .05. For fear trials, the following positive correlations were observed: between the two overall fear biases, $r$ (106) = .34, $p < .01$; between the emotional cueing overall bias and the dot probe facilitation bias, $r$ (106) = .25, $p < .01$; between the emotional cueing facilitation bias and the dot probe overall bias, $r$ (106)

= .22, $p < .05$; between the emotional cueing disengage bias and the dot probe overall bias, $r$ (106) = .29, $p < .01$; and between the emotional cueing disengage bias and the dot probe facilitation bias, $r$ (106) = .21, $p < .05$. For happy trials, the following negative correlations were observed: between the two overall happy biases, $r$ (106) = -.20, $p < .05$; between the two disengage biases, $r$ (106) = -.29, $p < .01$; and between the emotional cueing overall bias and the dot probe disengage bias, $r$ (106) = -.25, $p = .01$.

## 3.4 Discussion

### 3.4.1 Response Times and Reliability

As predicted, the shorter SOA in Study 2 resulted in a significant overall congruency (cueing) effect in the emotional cueing task; however, this cueing effect did not vary as a function of anxiety group or emotion. Overall, the results of Study 2 are consistent with Study 1. Specifically, there was no evidence of ABT for high anxious participants with either the dot probe or emotional cueing task, and the reliability of the bias scores was again unacceptably low (ranging from .01 to .53 for the dot probe task, and ranging from .03 to .58 for the emotional cueing task). Thus, measures of early attention to threat also appear to have poor reliability.

Contrary to expectations, there were significant attentional biases in the *LTA group* in Study 2 (see Table 8). Specifically, the LTA group showed difficulty disengaging from fear faces in both tasks, and showed negative facilitation biases to both angry and happy faces in the emotional cueing task. It is difficult to directly compare the current dot probe results to the rest of the literature, as relatively few studies have used fear faces and even fewer studies have used Koster et al. (2004)'s delayed disengagement indices. Generally, though, when biases for fear faces (and other threat stimuli) are found in dot probe tasks, these biases tend to occur in the

HTA rather than the LTA group (Fox, 2002), which is inconsistent with the current findings. Similarly, most biases in the emotional cueing literature occur in the HTA group and not the LTA group (Fox et al., 2002: Experiment 1; Fox et al., 2001: Experiments 2, 3, and 4). Mogg et al. (2008) did find negative 12-ms overall bias to angry faces in their LTA group in an emotional cueing task, but they did not test this effect for significance. Though it is unclear how the current results fit in the existing literature, Study 2 provides some evidence that LTA participants are faster to initially move their attention to neutral faces compared with emotional faces (in the emotional cueing task) and that they may have some difficulty disengaging their attention from fear faces (in both tasks).

### 3.4.2 Convergent Validity

Unlike in Study 1, in Study 2 there were small but significant correlations between dot probe and emotional cueing bias scores for the fear faces. There were also significant correlations between dot probe and emotional cueing scores for happy faces, but these correlations were not in the expected direction and are therefore difficult to interpret.

On the whole, these results remain consistent with previous findings on the convergent validity of RT measures of ABT, in that other studies have shown no correlations (Dalgleish et al., 2003; Mogg et al., 2000; Gotlib et al., 2004) or small correlations (Brosschot, de Ruiter, & Kindt, 1999; Egloff & Hock, 2003).

### 3.4.3 Attentional Bias

Why did Study 2 fail to find anxiety-related ABT in the HTA group? One possibility is that the low reliability of the measures reduced the power to detect differences (Kopriva & Shaw, 1991) and resulted in the lack of ABT findings. Another possibility is that the differences in methodology between previously published tasks and the tasks used in Study 1 and Study 2 resulted in the lack of ABT in the current tasks. Although these changes in methodology were made purposefully to improve the tasks (e.g., adding fear faces) and to make dot probe and emotional cueing tasks as similar as possible (e.g., the changes in timing and display parameters), the literature clearly demonstrates that methodological factors can moderate ABT effects, at least in the dot probe paradigm. This limitation will be addressed in Study 3.

Finally, a third possibility is that the failure to observe ABT with HTA participants was due to a lack of a state anxiety induction. Several prominent theoretical models of anxiety-related ABT specify a role for state anxiety. For example, Williams, Watts, MacLeod, & Mathews (1988, 1997) propose that whereas trait anxiety determines whether attention is directed *toward* threat (with HTA) or *away* from threat (LTA), state anxiety affects the extent to which stimuli are perceived as threatening. In a later model by Mathews and Macintosh (1998), high state anxiety again lowers the threshold at which stimuli are perceived as threatening and thus affects the threshold at which attentional capture occurs. High trait anxiety is proposed to heighten this effect. With relatively weak threat cues, attentional biases in non-clinically anxious populations may therefore only be seen when both trait anxiety and state anxiety are high (e.g., Mathews & Macintosh, 1998).

A number of empirical studies also suggest that ABT might be more likely to occur when participants are in an anxious state. Quigley, Nelson, Carriere, Smilek, and Purdon (2012) sought to examine the relative contributions of state and trait anxiety to attentional biases using an eye-tracking paradigm. They recruited HTA and LTA participants and asked them to view threat and neutral photos while their eye movements were monitored. The eye tracking task occurred both before and after state anxiety induction. The authors found no effect of trait anxiety, but a significant effect of *state anxiety* such that participants in both groups spent more time viewing the threat images after the state anxiety induction. These results were recently replicated by Nelson, Purdon, Quigley, Carriere, and Smilek (2014).

In a much earlier study, MacLeod and Mathews (1988) examined ABT under high and low state anxiety. HTA and LTA participants were recruited under low state anxiety conditions (12 weeks before a major exam) or high state anxiety conditions (1 week before the exam). In the low state anxiety testing session, there was no significant ABT for either group. In the high state anxiety testing session, however, they found that HTA individuals preferentially attended to exam-related threat words while LTA individuals avoided such words.

In sum, theoretical accounts and empirical evidence suggest that ABT may be more likely to occur in high state anxiety contexts. Interestingly, the empirical evidence also suggests that the *reliability* of ABT scores may also be higher under conditions of high state anxiety. Bar-Haim et al. (2010) is one of the only published studies to both observe ABT and to assess the reliability of ABT scores. The Spearman-Brown corrected reliability of dot probe bias scores in that study was higher than most other estimates (.62, in the acceptable range). The Bar-Haim et al. (2010) study was unique in that participants were living in a region with frequent rocket attacks. Many participants lived so close to the origin of the attacks that they had only 15 seconds to seek safety after siren alerts sounded. The high levels of state anxiety experienced by these participants may have contributed to the higher observed reliability of their dot probe scores, perhaps by increasing ABT effect sizes (e.g., Miller & Ulrich, 2013). Thus, the absence of ABT for HTA participants *and* the low reliability of ABT scores in Study 2 may both be due, at least in part, to the absence of a state anxiety manipulation. Study 3 addresses this limitation.

# Chapter 4
# State Anxiety, Trait Anxiety, and the Psychometric Properties of ABT

## 4.1 Study 3 Introduction

The purpose of Study 3 is to examine the reliability of dot probe and emotional cueing scores under conditions that are likely to produce large and potentially reliable ABT effects. To this end, Study 3 uses "classic" procedures that have produced significant biases with an undergraduate student population. Participants also undergo a state anxiety manipulation before completing the ABT tasks.

### 4.1.1 Choosing ABT Task Procedures

To choose the particular ABT tasks, all published studies examining trait anxiety-related ABT in adults using emotional faces as stimuli were reviewed. Because minor changes in methodology between studies are very common, very few exact replications exist in the literature. In an attempt to choose classic or common procedures, the dot probe task methodology from Mogg and Bradley (1999) was selected. Notably, these procedures were successfully replicated by Bradley, Mogg, and Millar (2000), although the latter study found only state-anxiety related ABT rather than trait anxiety-related ABT.

For the emotional cueing task, the procedures from Mogg, Holmes, Garner, & Bradley (2008) were chosen. The authors found a significant disengagement bias for the state anxious group. In a very similar study using schematic faces, Fox, Russo, and Dutton (2002) found the same pattern of results in a high trait anxious group.

### 4.1.2 State Anxiety Manipulation

A number of different procedures have been developed to conduct research with participants experiencing specific mood states. Some studies have used naturally-occurring mood states by having participants complete the task before an important exam, for example (e.g., MacLeod & Mathews, 1988; Mogg, Bradley, & Hallowell, 1994). Unfortunately, naturally-occurring mood states typically do not allow for random assignment to condition and thus can make it more difficult to separate the influence of state anxiety and trait anxiety. These procedures are also logistically difficult to implement, as they often require that many participants complete the study in a very short amount of time (e.g., in the week before an exam).

To avoid these problems, mood induction procedures have also been developed for use in the lab. In the anxiety literature in particular, common procedures include the following: asking participants to perform a video-taped speech (Phillips & Giancola, 2008; Sayette, Martin, Perrott, Wertz, & Hufford, 2001); having participants read Velten emotional statements such as 'I am so worried that I can't concentrate on anything' (York, Borkovec, Vasey, & Stern, 1987); and asking participants to focus on an anxiety-provoking event in their past while listening to music (Jefferies, Smilek, Eich, & Enns, 2008). The last procedure has been particularly well-validated by Eich and colleagues at the University of British Columbia (see Eich, Ng, Macaulay, Percy, & Grebneva, 2007, for a review). They label this method the *MCI technique,* as it uses Music (M) and Contemplation (C) in an Idiographic (I) context. Although there is a paucity of research comparing different mood induction techniques, especially with regard to anxiety, there is some evidence that MCI-like procedures may be more effective than other procedures in inducing sad mood (Van der Does, 2002). MCI has high success in inducing mood states that seem authentic to participants and the same mood can be induced repeatedly (Eich et al., 2007; Hernandez, Vander Wal, & Spring, 2011).

The MCI procedure was chosen for Study 3 because it is well-validated, can be implemented consistently in a lab setting, and has been successfully used in previous studies of ABT (Quigley et al., 2012; Nelson et al., 2014). The effects of the state anxiety manipulation on ABT and on the reliability of the biases will be examined.

## 4.2 Method

### 4.2.1 Participants

The recruitment procedures were identical to the previous two studies. No individuals who had completed Study 1 or 2 were permitted to participate in Study 3.

A power analysis (Faul, Erdfelder, Lang, & Buchner, 2007) was conducted using the average within-group effect size for ABT from the Bar-Haim et al. (2010) meta-analysis ($d$ = .45). Assuming a two-tailed alpha of .05, 40 participants per group resulted in power of approximately 80% (79.3%). Because participants would be randomly assigned to one of two state anxiety manipulation conditions, 80 HTA and 80 LTA participants were recruited.

Overall, one hundred sixty students (122 female; 36 male; 2 chose not to disclose their sex) participated in the study in exchange for course credit. The mean age of participants was 20.1 years. Participants identified primarily as White (38%), East Asian (30%) and South Asian (14%). Other groups represented in the sample included Southeast Asian (6%), Black (3%), Middle Eastern (2%), Biracial (2%), Caribbean (1%), and Hispanic (1%). Three participants (2%) chose not to disclose their race.

### 4.2.2 Measures and Materials

The questionnaire materials were identical to those in the previous two studies, with the addition of two measures to determine the effectiveness of the anxiety manipulation: an Affect Grid and a Genuineness Rating.

4.2.2.1 Affect Grid (Russell, Weiss, & Mendelsohn, 1989)

This measure is a 9 X 9 grid with “Affect” on the X axis (Extremely Unpleasant to Extremely Pleasant) and “Arousal” on the Y axis (Extremely Low Energy to Extremely High Energy). Increases in anxiety are indexed by decreases in pleasantness and increases in arousal.

4.2.2.2 Genuineness Rating (Jefferies, Smilek, Eich, & Enns, 2008)

As in previous mood induction research, participants indicated how genuinely they rated their mood in the study on a 9-point scale.

4.2.2.3 Dot probe task

For the dot probe task, each of the two emotion images (one Anger, one Happy) for each of 32 actors was paired with a Neutral image, resulting in 64 pairs. The location of the emotional face and the location of the target were counterbalanced across participants. Each participant saw each pair of faces once in Block 1 and once in Block 2, for a total of 128 trials. Notably, Mogg and Bradley’s (1999) study comprised only 64 trials. To ensure sufficient power to detect reliability, the current task was double its length. Blocks were separated by a self-paced rest break and trials were presented in randomized order within blocks. Sixteen practice trials were created using additional NimStim images.

On each trial, participants saw a fixation cross in the centre of the screen for 500 ms, followed by a face pair for 500 ms. The relative dimensions and the visual angles were retained from Mogg and Bradley (1999), but the actual sizes were modified to fit with the current lab set-up. Participants were seated approximately 70 cm from a 25-inch flat screen computer monitor. The face images subtended 2.6 degrees of visual angle (3.7 cm) wide by 4.0 degrees of visual angle (5 cm) tall and appeared 2.0 degrees of visual angle (2.5 cm) from fixation. Emotional pictures and probes appeared on the left and right with equal frequency. After the offset of the face pair, a single dot probe was presented in the location of one of the faces. Participants were instructed to press “a” or “l” as quickly and as accurately as possible to indicate the location of the probe (left or right). On congruent trials, the probe appeared in the location previously occupied by the emotional face. On incongruent trials, the probe appeared in the location previously occupied by the neutral face. The inter-trial interval varied between 500 and 1250 ms.

#### 4.2.2.4 Emotional cueing task

The cue faces consisted of one emotion image for each of 16 NimStim actors (one Angry, one Happy, and one Neutral). There were 192 experimental trials with a self-paced rest break halfway through the task. Consistent with Mogg et al. (2008), 75% of trials were valid and 25% were invalid. There were 64 trials for each emotion: Anger (48 valid, 16 invalid), Happiness (48 valid, 16 invalid), and Neutral (48 valid, 16 invalid). Each face image was presented four times in the experiment. Trials were presented in randomized order. Twelve practice trials were created using additional NimStim images.

The size of the face images and their distance from fixation were computed using visual angle calculations from Mogg et al. (2008). Each face was presented within one of two gray boxes measuring 7.6 cm tall by 5.7 cm wide. The boxes appeared to the left and right of fixation, with 2.95 cm separating the fixation cross from the inner edge of each box. On each trial, a fixation cross and the boxes were presented for 1000 ms. Then, a photograph of a face appeared in one of the boxes for 200 ms. Fifty milliseconds later, a target arrow measuring 0.6 cm tall appeared to the left or right of fixation. Participants were asked to press “k” or “m” as quickly and accurately as possible to indicate the direction the arrow pointed (up or down). On congruent trials, the target appeared in the box previously occupied by the face; on incongruent trials, the target appeared in the opposite box. The target remained on the screen for 6000 ms or until response. The inter-trial interval varied between 500 and 1050 ms.

**4.2.3 Procedure**

Participants were randomly assigned to the State Anxiety or Calm group. Both groups completed a package of questionnaires: the STICSA Trait, STICSA State, DASS-21, SDS, and Affect Grid. Participants then completed the mood induction. They were instructed to think about an anxious or calm event in their lives while listening to music that promoted an anxious or calm mood (see Jefferies et al., 2008; Quigley et al., 2012). The complete script and the music selections can be found in Appendix A. After five minutes, they again completed the Affect Grid and STICSA State.

Next, based on their counterbalance condition, participants completed the first of the two computer tasks (emotional cueing or dot probe). As in previous research (e.g., Jefferies et al., 2008), the music continued to play during the computer tasks.

Participants completed the Affect Grid and STICSA State again. They then repeated the same mood induction procedure for five minutes to ensure the induced mood did not decay before the second computer task. They completed the Affect Grid, STICSA State, and the second computer task. Finally, they completed the Affect Grid, the STICSA State, and the Genuineness Rating. Participants in the Anxious induction condition watched an amusing video to restore a positive mood before leaving the lab.

## 4.3 Results

### 4.3.1 Anxiety Groups

Participants were divided into groups based on their trait anxiety (STICSA) scores at the testing session. Participants whose STICSA Trait scores varied between the testing session and the in-lab session to the extent that they switched groups (participants with scores below 27 on the screening session and above 45 on the in-lab session, or vice-versa) were identified ($n = 2$). To ensure stable and accurate measurement of trait anxiety, these participants were excluded from analysis. Two groups were created based on the median in-lab STICSA Trait score (34). The groups consisted of 76 participants (LTA) and 82 participants (HTA), respectively.

Demographics and questionnaire scores for the two groups are presented in Table 11. A small amount of questionnaire data (<1.5%) was missing; these missing data were replaced using Expectation-Maximization imputation. The two groups did not differ on age or sex (see Table 11), but differed as expected on trait and state anxiety (STICSA), depression, anxiety, and stress symptoms (DASS), and social desirability scores (SDS).

As in Mogg and Bradley (1999), Social Desirability Scale (SDS) scores were examined for unusually large values. The box-and-whisker plot did not identify any outliers, and when SDS scores were transformed to z-scores no value was above 2.5. Therefore, no participants were excluded for scoring unusually high on social desirability.

### 4.3.2 State Anxiety Manipulation Check

Participants were randomly assigned to the Anxious or Calm mood conditions. In the Calm condition, there were 35 LTA participants and 41 HTA participants. In the Anxious condition, there were 43 LTA participants and 39 HTA participants. All participants completed the STICSA State and Affect Grid five times during the study to assess the effectiveness of the state anxiety manipulation. The first observation (T1) occurred at the beginning of the study; the second (T2) occurred immediately after the first mood induction; the third (T3) occurred after the first computer task; the fourth (T4) occurred after the second mood induction; and the fifth (T5) observation occurred after the second computer task. Mood ratings by Time, Anxiety Group and Condition are presented in Figure 3. Increases in anxiety are represented by increases on the STICSA State, *decreases* in Pleasantness ratings on the Affect Grid, and increases in Arousal ratings on the Affect Grid.

To assess the effectiveness of the state anxiety manipulation for the two groups, a 2 (Anxiety Group: LTA vs. HTA) X 2 (Condition: Anxious vs. Calm Induction) X 5 (Time: T1 vs. T2 vs. T3 vs. T4 vs. T4) mixed ANOVA was conducted on the three measures of mood. For the STICSA State, there were main effects of Anxiety Group ($F$ (1, 154) = 64.17, $MSE$ = 242.80, $p$ < .001, $\eta_p^2$ = .294), Condition ($F$ (1, 154) = 81.40, $MSE$ = 242.80, $p$ < .001, $\eta_p^2$ = .346), and Time ($F$ (2.41, 371.23) = 17.84, $MSE$ = 27.30, $p$ < .001, $\eta_p^2$ = .104). Critically, there was a Condition X Time interaction, $F$ (2.41, 371.23) = 73.91, $MSE$ = 2017.57, $p$ < .001, $\eta_p^2$ = .324. Follow-up tests revealed no effect of Condition at T1, $t$ (156) = .23, $p$ = .815, but significant effects of Condition at each time point after the mood induction: T2 ($t$ (156) = 9.23, $p$ < .001, $d$ = 1.47 ), T3 ($t$ (156) =

6.41, $p < .001$, $d = 1.02$), T4 ($t$ (156) = 9.15, $p < .001$, $d = 1.46$), and T5 ($t$ (156) = 7.27, $p < .001$, $d = 1.16$). No other interactions reached significance, $p$s > .05.

For the Pleasantness ratings, there were main effects of Anxiety Group ($F$ (1, 154) = 10.20, $MSE = 6.01$, $p < .01$, $\eta_p^2 = .062$), Condition ($F$ (1, 154) = 172.42, $MSE = 6.01$, $p < .001$, $\eta_p^2 = .528$), and Time ($F$ (3.18, 489.38) = 27.54, $MSE = 1.46$, $p < .001$, $\eta_p^2 = .152$). There was again a Condition X Time interaction, $F$ (3.18, 489.38) = 52.42, $MSE = 1.46$, $p < .001$, $\eta_p^2 = .254$. Follow-up tests revealed no effect of Condition at T1, $t$ (156) = -1.55, $p = .122$ and significant effects of Condition after the mood induction at T2 ($t$ (156) = -16.78, $p < .001$, $d = -2.67$), T3 ($t$ (156) = -8.24, $p < .001$, $d = -1.31$), T4 ($t$ (156) = -14.34, $p < .001$, $d = -2.29$), and T5 ($t$ (156) = -8.84, $p < .001$, $d = -1.41$). No other interactions reached significance, $p$s >. 05.

Lastly, the Arousal ratings revealed only a main effect of Condition ($F$ (1, 154) = 5.96, $MSE = 10.39$, $p < .05$, $\eta_p^2 = .037$) and a Condition X Time interaction, $F$ (3.39, 521.5) = 6.30, $MSE = 1.85$, $p < .001$, $\eta_p^2 = .039$. Follow-up tests revealed no effect of Condition at T1, $t$ (156) = 0.68, $p = .498$, but significant effects of Condition after the mood induction at T2 ($t$ (156) = 2.34, $p < .05$, $d = .37$), T3 ($t$ (156) = 2.27, $p < .05$, $d = .36$), T4 ($t$ (156) = 4.04, $p < .001$, $d = .64$), and T5 ($t$ (156) = 2.35, $p < .05$, $d = .37$). There were no other main effects or interactions, $p$s > .05.

Thus, the state anxiety manipulation had its intended effect. After the manipulation, participants in the Anxious condition reported higher anxious affect compared to participants in the Calm condition, and this difference persisted through the study. Participants also reported their affect ratings were genuine throughout the study ($M = 7.78$ on a 9-point scale, $SD = 1.01$).

### 4.3.3 Dot Probe Results

RT analyses were conducted for experimental trials in which an accurate response was given. One participant from the LTA Group had fewer than 75% accurate responses (*M* = .562) and was excluded from the dot probe RT analyses. The other participants' responses were very accurate (*M* = .989). To identify outliers, the same non-recursive method was used (Van Selst & Jolicoeur, 1994).[3] For the dot probe task, outliers made up 2.96% of correct RT data and were excluded from analysis.

A 2 (Anxiety Group: HTA vs. LTA) X 2 (Condition: Anxious vs. Calm) X 2 (Emotion: Angry vs. Happy) X 2 (Congruency: Incongruent vs. Congruent) mixed ANOVA was conducted on the RT data. Table 12 presents the RT data. There was an effect of Congruency, $F$ (1, 153) = 4.75, *MSE* = 251.46, $p$ = .031, $\eta_p^2$ = .030, such that participants responded more quickly to *incongruent* (*M* = 407 ms) compared to congruent (*M* = 410 ms) trials. There was also a marginally significant Anxiety Group X Emotion X Congruency interaction, $F$ (1, 153) = 2.81, *MSE* = 269.95, $p$ = .096, $\eta_p^2$ = .018. To better understand this interaction, traditional dot probe bias scores were computed by subtracting the mean RT on congruent trials from the mean RT on incongruent trials for both angry and happy face trials. Independent samples t-tests showed no difference in bias scores for angry faces between the LTA (-2.1 ms) and HTA (-4.6 ms) groups, $t$ (155) = .641, $p$ = .523, $d$ = .103, unlike Mogg and Bradley (1999). For happy faces, in contrast, there was a marginally significant difference in bias scores between the LTA (-5.3 ms) and HTA (0.8 ms) groups, $t$ (155) = -1.83, $p$ = .070, $d$ = -.293. There were no other main effects or interactions.

[3] When these data were analysed using a box and whisker plot outlier procedure (Mogg & Bradley, 1999), the ABT results were similar and the reliability estimates were similar or lower.

Next, planned t-tests were conducted comparing bias scores for each group and condition to 0, and the results and effect sizes are presented in Table 13. There was a negative bias (a bias *away*) from angry faces in the HTA group in the Anxious condition. No other effects reached significance.

Correlations between the bias scores and the questionnaire measures were also computed. Neither the angry bias score ($r$ = -.067, $p$ = .408) nor the happy bias score ($r$ = .125, $p$ = .120) was correlated with trait anxiety (STICSA), nor were they correlated with DASS-Depression or SDS scores, *p*s > .05. When the sample was split by Condition (Anxious vs. Calm) and this analysis was repeated, there were again no significant correlations.

Because the original Mogg and Bradley (1999) task was only half as long as the current dot probe task, the analyses above were repeated using only the first 64 trials. The results were the same with the following exceptions: (1) the group difference in bias scores to happy faces was no longer marginally significant, $t$ (155) = -1.30, $p$ = .197; (2) when the biases were compared to 0, there was a significant bias away from happy faces in the LTA Calm condition, $t$ (34) = -3.03, $p$ = .005, $d$ = -.478; and (3) in the Calm condition only, STICSA Trait scores correlated negatively with the angry bias ($r$ = -.235, $p$ < .05) and positively with the happy bias ($r$ = .265, $p$ < .05).

4.3.3.1 Dot Probe Reliability

Table 14 presents the mean reliability based on the 1000 permutations, 95% confidence intervals based on Fischer's r-to-z transformation, and Spearman-Brown corrected reliability estimates by group. The corrected reliability estimates for the overall sample were .44 (angry trials) and .17 (happy trials). The reliability estimates for angry trials were somewhat higher in the Calm condition than the Anxious condition (corrected estimates .59 in the Calm condition and .15 in the Anxious condition).

### 4.3.4 Emotional Cueing Results

Data from five participants were excluded from the emotional cueing analyses for low accuracy (< 75% correct trials). The remaining participants were very accurate (*M* = .952). Inaccurate trials were excluded from the RT analyses. The same non-recursive method adjusting the cut-off criterion by sample size was used for identifying outliers (Van Selst & Jolicoeur, 1994). Outliers made up 2.23% of correct RT data and were excluded from analysis.

A 2 (Anxiety Group: HTA vs. LTA) X 2 (Condition: Anxious vs. Calm Induction) X 2 (Emotion: Angry vs. Happy) X 2 (Congruency: Incongruent vs. Congruent) mixed ANOVA was conducted on the RT data. In a congruent trial, the target appeared in the location previously occupied by the face cue; in an incongruent trial, the target appeared in the opposite location. Table 12 presents RTs by condition.

The ANOVA revealed a significant effect of Congruency, $F$ (1, 149) = 538.52, $MSE$ = 2364.43, $p < .001$, $\eta_p^2 = .783$, such that participants responded more quickly to congruent (*M* = 468 ms) than incongruent (*M* = 543 ms) trials. There was a main effect of Emotion, $F$ (2, 298) = 4.92, $MSE$ = 535.11, $p = .008$, $\eta_p^2 = .032$, such that participants responded more quickly to neutral trials (*M* = 484 ms) than happy trials (*M* = 488 ms), $t$ (152) = 2.64, $p = .009$, $d = .214$. There was also an Emotion X Congruency interaction, $F$ (2, 298) = 10.27, $MSE$ = 620.45, $p < .001$, $\eta_p^2 = .065$. Follow-up tests showed a larger overall Congruency effect for Angry trials (*M* = 85 ms) than Happy trials (*M* = 67 ms), $t$ (152) = 4.23, $p < .001$, $d = .342$, or Neutral trials (*M* = 72 ms) trials, $t$ (152) = 3.16, $p = .002$, $d = .255$. Finally, there was a marginally significant

Anxiety Group X Condition X Emotion X Congruency interaction, *F* (2, 298) = 2.87, *MSE* = 620.45, *p* = .058, $\eta_p^2$ = .019.

To better understand this interaction, follow-up ANOVAs were conducted for each Condition (Anxious vs. Calm) separately. In the Calm condition, the Anxiety Group X Emotion X Congruency interaction was not significant, *F* (2, 146) = 0.58, *MSE* = 614.10, *p* = .943. In the Anxious condition, this interaction was significant, *F* (2, 152) = 4.70, *MSE* = 626.54, *p* = .01, $\eta_p^2$ = .118.

The same bias scores from Study 1 and Study 2 (overall, facilitation, and disengagement) were computed for both angry and happy trials to assist in interpreting this interaction. Notably, the facilitation and disengagement biases simply represent the effects of emotion relative to neutral for congruent (facilitation) and incongruent (disengagement) trials.

Four 2 X 2 ANOVAs were conducted on the Anxious condition data to examine the effect of emotional cues relative to neutral cues separately in the congruent and incongruent cue conditions.

i. *Congruent angry trials*. To determine if there were differences between groups in the facilitation bias for angry faces, a 2 (Anxiety Group: HTA vs. LTA) X 2 (Emotion: Angry vs. Neutral) ANOVA was conducted on congruent trial RTs. There was a marginally significant Anxiety Group X Emotion interaction, *F* (1, 76) = 3.22, *MSE* = 130.44, *p* = .08, $\eta_p^2$ = .041, such that the facilitation bias was marginally larger for the HTA group (6 ms) compared to the LTA group (-1 ms).

ii. *Incongruent angry trials*. To determine if there were differences between groups in the disengagement bias for angry faces, a 2 (Anxiety Group: HTA vs. LTA) X

2 (Emotion: Angry vs. Neutral) ANOVA was conducted on incongruent trial RTs. There was a significant Anxiety Group X Emotion interaction, $F$ (1, 76) = 4.38, $MSE$ = 981.89, $p$ = .04, $\eta_p^2$ = .055, such that the disengagement bias was larger for the HTA group (19 ms) than the LTA group (-2 ms).

iii. *Congruent happy trials*. A 2 (Anxiety Group: HTA vs. LTA) X 2 (Emotion: Happy vs. Neutral) ANOVA on RTs for congruent trials revealed no interaction, $F$ (1, 76) = 0.31, $MSE$ = 278.32, $p$ = .58.

iv. *Incongruent happy trials*. A 2 (Anxiety Group: HTA vs. LTA) X 2 (Emotion: Happy vs. Neutral) ANOVA on RTs for incongruent trials revealed a significant interaction, $F$ (1, 76) = 9.61, $MSE$ = 862.05, $p$ = .003, $\eta_p^2$ = .112. The HTA group had a larger disengagement bias for happy trials (14 ms) than the LTA group (-15 ms).

Next, planned t-tests were conducted comparing bias scores to 0 for each group and condition. Results are presented in Table 13. In the Calm condition, the HTA group (15 ms effect; $p$ = .02) and the LTA group (17 ms effect; $p$ = .07) showed difficulty disengaging from angry faces. In the Anxious condition, LTA participants showed negative overall and disengage biases for happy faces, which may represent avoidance (e.g., Koster et al., 2006a). HTA participants in the Anxious condition showed biases (facilitation, overall, and disengage biases) for angry faces, which is typically interpreted as biased attention toward threat (e.g., Mogg et al., 2008; Fox et al., 2002). HTA participants also showed difficulty disengaging from happy faces.

Correlations between the bias indices and the questionnaire data are presented in Table 15. In the Anxious condition, trait anxiety correlated with the overall bias and the disengage bias for happy faces but not angry faces. There were no other significant correlations.

#### 4.3.4.1 Emotional cueing reliability

Table 14 presents the reliabilities for the difference scores (congruency effects, overall biases, facilitation biases, and disengage biases) from the emotional cueing paradigm. In the overall sample, the congruency scores (which are not a measure of bias but rather measure differences in RTs between incongruent and congruent cues) had Spearman-Brown corrected reliability estimates of .56 for angry trials and .59 for happy trials. In contrast, the overall, facilitation, and disengage bias scores had very low reliability: Spearman-Brown corrected estimates for the biases were between 0 and .19 in the overall sample. There were no notable differences in reliability between the Anxious and Calm conditions.

### 4.3.5 Between Task Correlations

To determine whether bias scores on the two tasks were related, Pearson's correlations were computed between all bias scores on the dot probe and emotional cueing tasks. There were no significant correlations in the full sample ($r$s (152) < .15, $p$s > .05). When the sample was divided into the Anxious and Calm conditions, the dot probe happy bias and the emotional cueing happy facilitation bias were correlated in the Calm condition only, $r$ (75) = .23, $p$ = .047.

## 4.4 Discussion

Study 3 examined the psychometric properties of bias scores from classic dot probe and emotional cueing tasks in the context of a state anxiety manipulation. The ABT, reliability, and convergent validity results are discussed in turn below.

### 4.4.1 State Anxiety, Trait Anxiety, and Attentional Biases

4.4.1.1 Dot probe task

The current dot probe task was identical to Mogg and Bradley's (1999) task with the addition of a state anxiety manipulation. In the RT analysis, there were no main effects or interactions involving state anxiety condition. Unfortunately, the results failed to replicate Mogg and Bradley's (1999) finding of larger bias scores to angry faces for HTA participants compared with LTA participants. There was a trend whereby HTA participants showed marginally larger bias scores toward *happy* faces than LTA participants (see Bradley, Mogg, White, Groom, & De Bono, 1999, for a similar effect in Generalized Anxiety Disorder), but there were otherwise no group differences. The failure to replicate Mogg and Bradley's (1999) group difference findings may be due to the inclusion of the state anxiety manipulation or due to the relatively low reliability of dot probe bias scores.

When the biases for each group were compared to 0, there was a significant *negative bias* (-6 ms) from angry faces for the HTA group in the Anxious condition. This pattern was unexpected and is the opposite of what is typically found with HTA participants in the absence of a state anxiety manipulation (e.g., Mogg & Bradley, 1999). Notably, a similarly sized negative bias (-5 ms) was also found in the LTA group in the Anxious condition in Study 3. Although the effect did not reach significance in the LTA group, this pattern suggests that the negative biases are related to state anxiety rather than trait anxiety. Thus, the lab-induced state anxiety manipulation appears to have resulted in a bias away from threat in the dot probe task. This bias was not present in the Calm condition. This pattern is inconsistent with most previous dot probe research on the effects of lab-induced stressors, as other studies typically show larger biases toward threat (e.g., Mogg, Mathews, Bird, & Macgregor-Morris, 1990) in the high anxiety conditions.

Mathews and Mackintosh (1998), however, also found negative dot probe biases in their anxious state condition—but only for the LTA group. In their cognitive model, Mathews and Mackintosh propose that when state anxious participants are asked to perform a task, they make efforts to focus *on the task* and away from both their anxious thoughts and the anxious stimuli. This increased effort or task demand may result in the elimination or even the reversal of biases toward threat in high state anxious conditions, especially when the threat stimuli are only mildly threatening. It is unclear why this pattern was evident in the HTA group in the current study.

4.4.1.2 Emotional cueing task

In the Anxious condition, Mogg et al.'s (2008) Anxiety Group X Emotion X Congruency interaction was successfully replicated. Furthermore, their finding of a larger disengagement bias for angry faces for the HTA group than the LTA group was also replicated. Unlike in Mogg et al.'s (2008) study, the current HTA group also showed a marginally larger facilitation bias for angry faces and a larger disengagement bias for happy faces compared to the LTA group in the Anxious condition. These additional findings (facilitation ABTs; difficulty disengaging from happy faces) are likely a result of the state anxiety manipulation, whereas the disengagement bias with angry faces appears to be associated with trait anxiety generally (see Table 13).

As discussed earlier, RT difference scores are more reliable when effect sizes are large (Miller & Ulrich, 2013). The current emotional cueing results are particularly important because they allow for the examination of the reliability of ABT biases in a context where ABTs were actually observed.

### 4.4.2 Reliability

#### 4.4.2.1 Dot probe

Although the observed reliabilities for the dot probe task were higher than some previous estimates in the literature (e.g., Staugaard, 2009; Waechter et al., 2014), they remained unacceptably low and were generally not higher than the reliabilities from Study 1 and Study 2. The reliability estimates for angry trials were somewhat higher in the Calm condition than the Anxious condition, which suggests that the anxiety induction did *not* improve the reliability of dot probe bias scores. Overall, the results of Study 3 support the growing consensus that dot probe scores have low reliability.

#### 4.4.2.2 Emotional cueing

Study 3 is among the first studies to both *find* significant biases and to examine their reliability (see also Bar-Haim et al., 2010, Cooper et al., 2011). Generally, the reliability estimates for emotional cueing biases were very low (Spearman-Brown corrected estimates between 0 and .19 with the overall sample) and similar to the reliability estimates from Study 1 and Study 2. There were no clear changes in reliability across the Calm and Anxious conditions; again, the state anxiety induction did not consistently improve reliability. These results clearly indicate that merely finding significant biases does not guarantee these biases have adequate reliability.

We did replicate Enock et al.'s (2014) observation that the reliability of the congruency scores was higher than that of the bias scores; however, congruency scores are not a measure of bias in the emotional cueing paradigm as they simply measure the difference in RT between incongruent and congruent trials for each emotion.

### 4.4.3 Convergent Validity

The bias scores in Study 3 again had low convergent validity. There was only one significant correlation between the bias scores from the two tasks, and this correlation was small ($r$ = .23) and occurred only in the Calm condition. It is likely that the correlations between tasks were limited at least to some extent by the poor reliability of the bias scores or by other factors (as discussed in Miller & Ulrich, 2013). Alternatively, these tasks may be indexing different constructs, such as ABT at different time points. For example, attentional biases measured 250 ms after the onset of the emotional face (as in the emotional cueing paradigm used in Study 3) may be largely unrelated to attentional bias measured 500 ms after the onset of the emotional face (as in the dot probe paradigm used in Study 3). Notably, however, Study 1 and Study 2 used identical timing parameters for both the dot probe and emotional cueing tasks, and still failed to find consistent and substantial evidence of convergent validity.

### 4.4.4 Study 3 Summary

Study 3 assessed the reliability of dot probe and emotional cueing bias scores in high anxious and low anxious participants under conditions of high state anxiety to increase the likelihood of producing large and potentially reliable attentional biases. In the dot probe task, there were no significant differences in attention measures between the HTA and LTA groups. When the biases for each group and condition were compared to 0, the HTA group in the Anxious condition actually showed a bias *away* from threat faces. In the emotional cueing task, there were significant group differences: HTA participants showed larger facilitation and disengage biases to angry faces than LTA participants. When the biases for each group and condition were compared to 0, the Anxious and Calm HTA groups each showed difficulty disengaging from angry faces. In the Anxious condition, the HTA group also showed facilitated attention to angry faces and an overall bias for angry faces.

Despite the fact that there were significant attentional biases in Study 3, the reliability estimates for both dot probe and emotional cueing biases were unacceptably low. The anxiety induction did not substantially improve the reliability of the bias scores.

# Chapter 5
# General Discussion

## 5.1 Summary of Findings

Cognitive processing biases, including ABT, are seen as a critical (e.g., Beck & Clark, 1997) and often as a causal (e.g., Mathews & MacLeod, 2002) factor in the development and maintenance of anxiety disorders. Cognitive processing biases have inspired hundreds of empirical and theoretical papers (see Bar-Haim et al., 2007; Ouimet, Gawronski, & Dozois, 2009 for reviews), and are seen as an important target of intervention in anxiety disorders treatment (e.g., Cognitive Therapy: Beck & Clark, 1997; Cognitive Bias Modification: MacLeod & Mathews, 2012). Despite the clear theoretical importance of ABT, recent investigations have revealed that the most common RT measures of ABT have either poor or unknown reliability. As stated earlier, the use of measures with low reliability in ABT research (a) reduces power to detect correlations and group differences (e.g., Kopriva & Shaw, 1991); (b) contributes to inconsistencies in the literature (e.g., Cooper et al., 2011); and (c) impedes progress in our understanding of ABT and their association with anxiety disorders (see Van Bockstaele et al., 2014).

Three studies were conducted to examine the reliability and convergent validity of dot probe and emotional cueing bias scores. The results are clear and consistent throughout the three studies: bias scores from dot probe and emotional cueing tasks have unacceptably low reliability. In fact, for all three studies, all of the corrected reliability estimates were below .60. The data from the convergent validity analyses were also relatively clear in that the correlations between dot probe and emotional cueing scores were low (maximum of .34) and sometimes even negative (e.g., for happy faces in Study 2). Thus, the psychometric properties of dot probe and emotional cueing bias scores appear to be quite poor.

These studies successfully clarify the impact of a number of methodological factors on the reliability of these ABT scores. First, these studies were the first to examine the reliability of dot probe and emotional cueing bias scores specifically in HTA and LTA individuals, as opposed to unselected samples (e.g., Schmukle, 2005; Staugaard, 2009). It was hypothesized that selecting extreme groups in terms of trait anxiety would result in more variability in ABT true scores, which should increase reliability. Unfortunately, the selection of HTA and LTA participants did not substantially improve the reliability of the ABT measures compared with previous estimates (e.g., Staugaard, 2009).

Study 1 and Study 2 were also the first to investigate the reliability of ABT measures using fear faces. Unfortunately, fear faces do not appear to result in higher reliability than other emotional faces used previously in the ABT reliability literature (e.g., angry faces: Staugaard, 2009; disgust faces: Waechter et al., 2014).

Finally, Study 3 examined the effects of a state anxiety manipulation on the reliability of ABT scores. It was thought that high levels of state anxiety might increase ABT effect sizes, which might improve reliability (e.g., Miller & Ulrich, 2013). The state anxiety manipulation had the intended effect on ABT scores and effect sizes, in that there were significant biases in the Anxious condition that did not occur in the Calm condition (see Table 13); however, the reliability of these biases remained low. In sum, none of the factors investigated here (e.g., selecting HTA and LTA participants, adding fear faces, using classic dot probe and emotional cueing tasks; including a state anxiety manipulation; etc.) resulted in adequate reliability for dot probe and emotional cueing bias scores.

## 5.2 Research Implications

In Study 3, an effort was made to optimize the size of the effects through the use of classic replicable procedures, adequate sample sizes, and the addition of a state anxiety induction. Although these efforts resulted in significant attentional biases, the reliability of these biases remained below acceptable levels. Continued research to improve the reliability of ABT measures is critical for understanding ABT and its role in the development and maintenance of anxiety and anxiety disorders. If researchers continue to use RT tasks to index ABT in the future, they should assess reliability of the resulting ABT scores. This will aid in interpretations of results and contribute to the understanding of the factors (e.g., effect sizes, number of trials, methodological differences, etc.) that impact reliability. Continued research on the reliability of ABT measures will be essential in the discovery / development of more reliable ABT measures.

Are there any factors in particular that might be promising in developing more psychometrically-sound ABT tasks? Number of trials may be important, in that Miller and Ulrich (2013) showed that the number of trials has a strong effect on the reliability of difference scores. In Study 1 and Study 2 of this dissertation, the dot probe and emotional cueing tasks both had 32 trials per participant and condition. In Study 3, the dot probe task had 32 trials per participant and condition; the emotional cueing task had 48 (congruent conditions) and 16 (incongruent conditions) trials per condition. All tasks took participants 15 to 20 minutes to complete. Because the observed reliability estimates for ABT scores were so low, however, increasing the number of trials actually shows relatively little promise in terms of reaching acceptable levels of reliability. This can be easily illustrated by using the Spearman-Brown prophecy formula to examine corrected reliabilities at different test lengths. For example, if each task used in Study 1 had four times as many trials per condition (and therefore each took participants roughly a full *hour* to complete), only three of the eighteen bias scores would reach a reliability above .60. It is impractical to simply increase the number of trials in these tasks until acceptable levels of reliability are attained.

Miller and Ulrich (2013) also demonstrated that reliabilities are higher when the difference scores are taken from *opposing task processes,* rather than from *unrelated* or *common task processes.* Using Miller and Ulrich's (2013) definition, it appears that some of the bias scores resulting from the dot probe and emotional cueing tasks result from opposing task processes and some result from unrelated task processes. Specifically, in the dot probe task, the overall dot probe bias scores result from opposing task processes because increasing attention to threat theoretically results in faster RTs in the congruent condition and slower RTs in the incongruent condition. The facilitation and disengage indices, in contrast, result from unrelated task processes. In the emotional cueing task, the congruency effects (which are not a measure of bias) result from opposing task processes. Because the overall bias score is the difference between two congruency effects, it represents the difference between two opposing processes. Finally, the facilitation and disengagement scores in the emotional cueing task result from unrelated task processes. Because they result from opposing processes, the overall dot probe bias scores and the emotional cueing congruency scores are more likely to show good reliability compared to the other indices (Miller & Ulrich, 2013). Study 3, in particular, shows some limited support for this pattern. That said, it is clear that the overall dot probe bias scores have very poor reliability despite the fact that they result from opposing task processes. Future research might focus on developing new ABT tasks that use opposing task processes, in particular, in the hope that these bias scores might show better reliability.

Ultimately, it may be time to look beyond response time measures in ABT research. A recent study suggests that eye tracking indices of ABT have adequate to good reliability after the first 1500 ms of image display (Waechter et al., 2014). For researchers who are interested in overt attention or late processes in attention to threat (e.g., avoidance), eye tracking may provide a good alternative to traditional measures like dot probe or emotional cueing bias scores.

In addition, a recent study by Britton et al. (2013; Experiment 2) demonstrated that children had stable neural activation in the ventrolateral prefrontal cortex and amygdala on two separate occasions when they completed a dot probe task. Twenty-one typically-developing children completed the dot probe tasks while in an MRI scanner on two separate visits. Intra-class correlation (ICC) analyses on the fMRI data were used to measure the reliability of neural activation across the two visits. For the contrasts representing ABT (e.g., fear incongruent versus fear congruent trials), there was significant stability (ICCs > .70) across the two visits in the ventrolateral prefrontal cortex and amygdala. Despite the consistency in their neural activation, the children's RT bias scores across the two sessions were not reliable. These results are very new, but they do suggest that neural activation may provide more stable measures of ABT than RTs and provide an interesting avenue for future research.

Finally, these results have some implications for the development and refinement of cognitive models of attention in anxiety. The amount of measurement error inherent in both dot probe and emotional cueing bias scores likely contributes to the inconsistent results observed in the ABT literature. In turn, these inconsistent results impair researchers' understanding of ABT and make it more difficult to develop and successfully test models of cognitive and attentional processing in anxiety. The fact that so many excellent and detailed models (e.g., Bar-Haim et al., 2007; Eysenck, 1997; Mogg & Bradley, 1998; Ouimet et al, 2009; Williams et al., 1997) do exist in this area is a testament to researchers' tenacity and the impressive and sustained level of interest in this research over the past three decades. To advance our understanding of ABT and to more easily test and improve current theoretical models, more psychometrically-sound measures of ABT must be developed.

### 5.3 Clinical Implications

A number of researchers suggest that ABT or changes to ABT may be a useful predictor of important clinical variables on an individual level (e.g., of therapy outcome; Reinecke, Waldenmaier, Cooper, & Harmer, 2013; Legerstee, Tulen, Dierckx, Treffers, Verhulst, & Utens, 2010; Waters, Mogg, & Bradley, 2012). Given the poor reliability of dot probe and emotional cueing bias scores, however, these measures may not allow us to conclude anything of interest about an individual. In effect, these bias scores may index largely measurement error. In clinical settings, then, these measures are unlikely to be a useful predictor of individual progress or recovery in clients who struggle with anxiety.

These results of the current studies also have implications for our understanding of Attentional Bias Modification (ABM), which is a new experimental intervention for anxiety disorders with its roots in the ABT literature (MacLeod, Rutherford, Campbell, Ebsworthy, & Holker, 2002). As described in Chapter 1, typical ABM training tasks are similar to dot probe tasks except that the targets appear preferentially in the location previously occupied by the non-threat (e.g., neutral or positive/happy) stimulus. The goal of this intervention is to modify ABT by training participants to attend to non-threat stimuli. Because ABT is thought to play a causal role in anxiety disorders, the objective is to ultimately alleviate individuals' anxiety problems by modifying ABT.

To date, seven randomized control trials of ABM with clinically-anxious participants have been conducted (Amir, Beard, Burns & Bomyea, 2009; Amir, Beard, Taylor et al., 2009; Boettcher et al., 2013; Carlbring et al., 2012; Heeren, Reese, McNally, & Philippot, 2012; Neubauer, von Auer, Murray, Petermann, Helbig-Lang, & Gerlach, 2013; Schmidt, Richey, Buckner & Timpano, 2009). In one of the first RCTs, Amir, Beard, Burns and Bomyea (2009) randomly assigned participants with Generalized Anxiety Disorder (GAD) to complete 8 sessions of either ABM training or a control task. In both conditions, participants completed a dot probe-like task with two words (one threat word, one neutral word) on the screen. In the ABM condition, the target always appeared in the location of the neutral word. In the control condition, the target was equally likely to appear in the location of either word. Amir, Beard, Burns and Bomyea (2009) found large reductions in anxiety symptoms in the ABM condition, such that 50% of this group no longer met diagnostic criteria for GAD after the 8 sessions of training (compared to 13% in the control condition). Participants in the other six RCTs mentioned above also reported reductions in anxiety symptoms after ABM, although in three studies the ABM group did not differ from the control group post-treatment (Boettcher et al., 2013; Carlbring et al., 2012; Neubauer et al., 2013; all of which used internet-delivered ABM).

Despite the generally positive results of the published RCTs, studies on ABM using subclinical populations have been mixed. Some studies show significant effects (e.g., Li, Tan, Qian, & Liu, 2008; Macleod et al., 2002) whereas others show no change from pre-treatment to post-treatment (Julian, Beard, Schmidt, Powers, & Smits; 2012) or no differences between the ABM group and the control group (Enock et al., 2014). Even with these mixed results, though, the most recent ABM meta-analysis (Mogoase, David, & Koster, 2014) was able to conclude that ABM has a small but significant effect ($g$ = .26) on anxiety symptoms.

Thus, ABM is a novel intervention with at least some promise for ameliorating anxiety disorder symptoms. In order to improve ABM interventions and optimize symptom change, there has been a recent push to better understand the active ingredients and mechanisms that underlie post-ABM changes in symptoms (Beard, 2011; Enock et al., 2014; Mogoase et al., 2014). A leading theory is that ABM works by changing an individual's attentional biases, which in turn leads to symptom improvement (e.g., Amir, Beard, Taylor et al., 2009). Unfortunately, a lack of reliable and valid ABT measures makes it difficult to determine if ABM actually works in this way. Mogoase et al. (2014) conducted a mediational analysis to test this proposed mechanism and found no significant mediation. The mediation model may have failed because ABT is not the true mechanism of change in ABM studies. Alternatively, the low reliability of ABT measures may have reduced statistical power to detect significant effects in the mediation analysis. Developing more reliable ABT measures will be a critical step in understanding and improving ABM and other novel treatments that target attentional biases to alleviate anxiety symptoms.

### 5.4 Limitations

There are, of course, a number of limitations to these studies. First, the participants consisted of a convenience sample of undergraduate students selected for high and low trait anxiety. This sample was chosen because the majority of the ABT literature is made up of similar samples (e.g., Mogg & Bradley, 1999; Mogg et al., 2008; Nelson et al., 2014), and because a meta-analysis (Bar-Haim et al., 2007) found no differences in ABT between clinically-diagnosed participants and non-clinical participants with high self-reported anxiety. When designing the study, it therefore seemed more important to maximize the sample size and thus increase power to detect ABT than to recruit a (inevitably) much smaller number of clinically-diagnosed anxious participants. There is no compelling evidence that the psychometric properties of ABT would be different in a clinically-diagnosed sample; however, it is a least *possible* that these results might not generalize to individuals suffering from clinically-significant anxiety disorders.

Second, the current reliability results may apply only to the specific dot probe and emotional cueing tasks employed in this study. Research on response time tasks has shown that reliability can vary depending on the experimental methodology (e.g., as demonstrated by Borgmann et al., 2007, using the Simon effect); thus, different methodologies (e.g., different display and timing parameters, identification versus localization versus detection tasks, etc.) may result in greater reliability. Notably, however, ABT scores from the six unique tasks examined in the current studies *all* had low reliability, and these results converge with other investigations in the literature using a variety of methodologies (e.g., Brown et al., 2014; Enock et al., 2014; Schmukle, 2005; Staugaard, 2009).

Third, it should be noted that the current reliability results apply only to dot probe and emotional cueing bias scores, and do not necessarily reflect the reliability of ABT measures in general. Similarly, the impact of the timing changes and the state anxiety manipulation on reliability may not generalize to other ABT tasks (e.g., the emotional Stroop task, the visual search task, attentional blink tasks, and eye tracking tasks). Currently, relatively little is known about the specific factors that impact the reliability of bias measures from these tasks.

Finally, it is important to note that "attention" is generally understood to be a cognitive operation that results in the preferential selection of information regardless of modality (e.g., Weierich et al., 2008); yet, the current series of studies assessed only measures of *visual/spatial* attention to threat. The true clinical phenomenology of ABT is likely complex and may include biased attention to auditory cues (e.g., Kimble, Fleming, Bandy, & Zambetti, 2010), internal sensations (Hayward, Ahmad, & Wardle, 2000), and thoughts (e.g., Beck & Clark, 1997). Future research should investigate the phenomenology of attentional processes in anxiety generally, including the psychometric properties of ABT in different modalities.

### 5.5 Future Directions

There are many possible avenues for future research in this area. As discussed earlier, it is important that researchers using ABT tasks consistently evaluate and disseminate the psychometric properties of their measures. This will provide key information about the factors (e.g., task, effect sizes, number of trials, methodological differences, etc.) that impact reliability. Because so little is known about the reliability of ABT measures derived from attentional blink and visual search tasks, these two tasks may be particularly important areas for future research. If researchers want to reach an understanding of how attention operates in anxiety, a sustained scientific effort to understand the psychometric properties of the tools used to measure ABT is absolutely essential (see Van Bockstaele et al., 2014, for a similar argument).

Interestingly, of the existing ABT measures with known psychometric properties, *eye-tracking* measures appear to be the most promising. Waechter et al. (2014) found adequate to good reliability for eye tracking bias scores in a free viewing paradigm after the first 1500 ms of image display (Waechter et al., 2014); in addition, the poor reliability of the bias measures prior to the first 1500 ms appeared to be due to participants' consistent tendency to look at the top image first, regardless of the image content. Reducing this tendency to look at the top image (e.g., through altering display parameters) may result in adequate reliability for measures of early attention to threat. A study testing this hypothesis is currently in progress (Waechter, Nelson, & Purdon, in preparation).

A particular advantage of eye-tracking technology is that it can shed light on how attentional biases operate in more "real world" situations (e.g., Smilek, Birmingham, Cameron, Bischof, & Kingstone, 2006; Mühlberger, Wieser, & Pauli, 2008), such as when completing tasks or interpreting complex scenes. It remains to be seen whether eye tracking bias measures remain reliable in these more complex contexts. If they do, eye tracking measures of ABT may potentially lead to exciting insight into how attention to threat impacts anxiety and its treatment in people's daily lives.

### 5.6 Conclusions

Three studies were conducted to better understand the psychometric properties of dot probe and emotional cueing bias scores in high anxious and low anxious participants. Even under arguably ideal conditions (e.g., Study 3), the psychometric properties of these measures were poor.

These results underscore the importance of developing more reliable and clinically-useful ABT measures. As long as researchers continue to use measures with low reliability (which in fact index more measurement error than ABT), progress in understanding the role of ABT in anxiety and its treatment will be limited. Further research, with a focus on developing new tools for measuring ABT or improving the reliability of existing measures, is critical.

## Appendix A:
## Mood Induction Procedure

<u>Verbal Script:</u>

Before we begin the visual task I am going to ask you to get into a mood that makes you as **[anxious, or calm]** as you feel comfortable. You can do this by thinking about an event in your life where you felt especially **[same mood word].** I know that this may not be the easiest thing to do, but it is very important for our research.

I've done this a few times myself so I'll tell you a few things about it. I found that since I was the one asking myself to become **[same mood word]** by thinking about events in my own life, I was very much in control of the mood. I could intensify, lessen, and later even end the mood quite easily by changing my thoughts.

I'll begin by turning on some music that people usually find helpful for getting into mood that makes them feel **[same mood word].** While you are listening to the music, please think about a particular event from your past where you were especially (same mood word).

While you are listening to the music I'd like you to relive this event. When I did this, I thought about the time **[I was just about to enter an important job interview (anxious condition), or I was sitting outside the cottage on vacation (calm condition)].**

It is important to remember that the more detail you can re-create in your mind about the event, the more intensely you'll re-live the same feelings.

But I also want to reassure you that I will take time at the end of the session to make sure you are feeling good again before you leave today. Remember that the goal is to feel as **[same mood word]** as possible for this short period of time. I know this may not be easy, but are you willing to try?

I'll leave you alone now with the music and your thoughts. Please relax in the chair while you think about these events. I'll be coming back in 5 minutes and will ask you to rate how you are feeling on the mood grid. Because I want you to keep the feeling of **[anxiety or calm]** as long as possible, I won’t be talking with you more than necessary through the rest of the experiment.

Music Selection

Anxious Mood Induction Music:

- Symphony No. 8 in C Minor: 4. Finale (Bruckner)
- Mars, the Bringer of War (Holst)
- Grosse Fugue B Flat Major, Op. 133 (Beethoven)

Calm Mood Induction Music:

- Rodeo: No.2 Corral Nocturne (Copland)
- Appalachian Spring (Copland)
- Carnival of the Animals: The Swan (Sant-Saens)
- Venus, the Bringer of Peace (Holst)

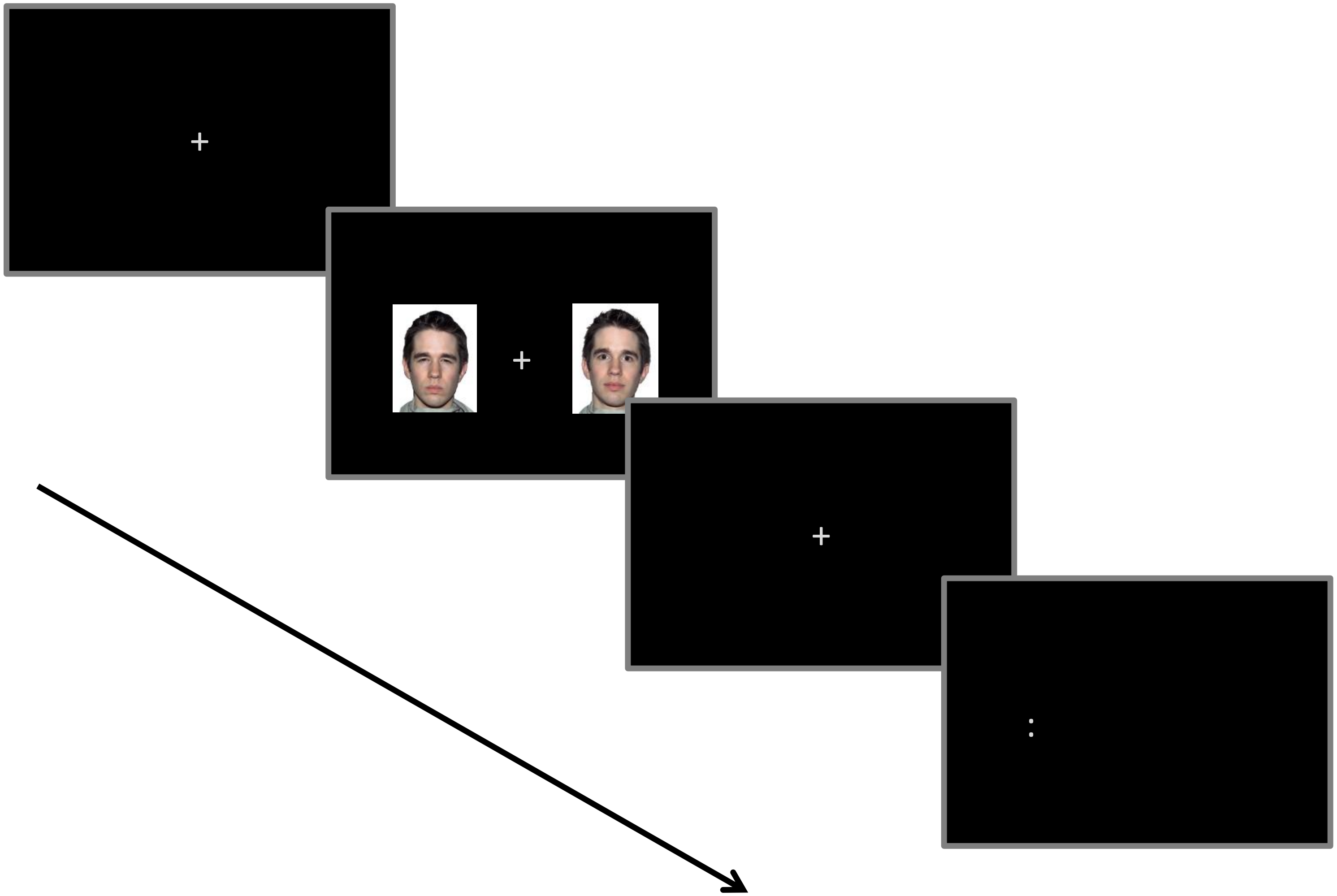

Figure 1. Dot probe task

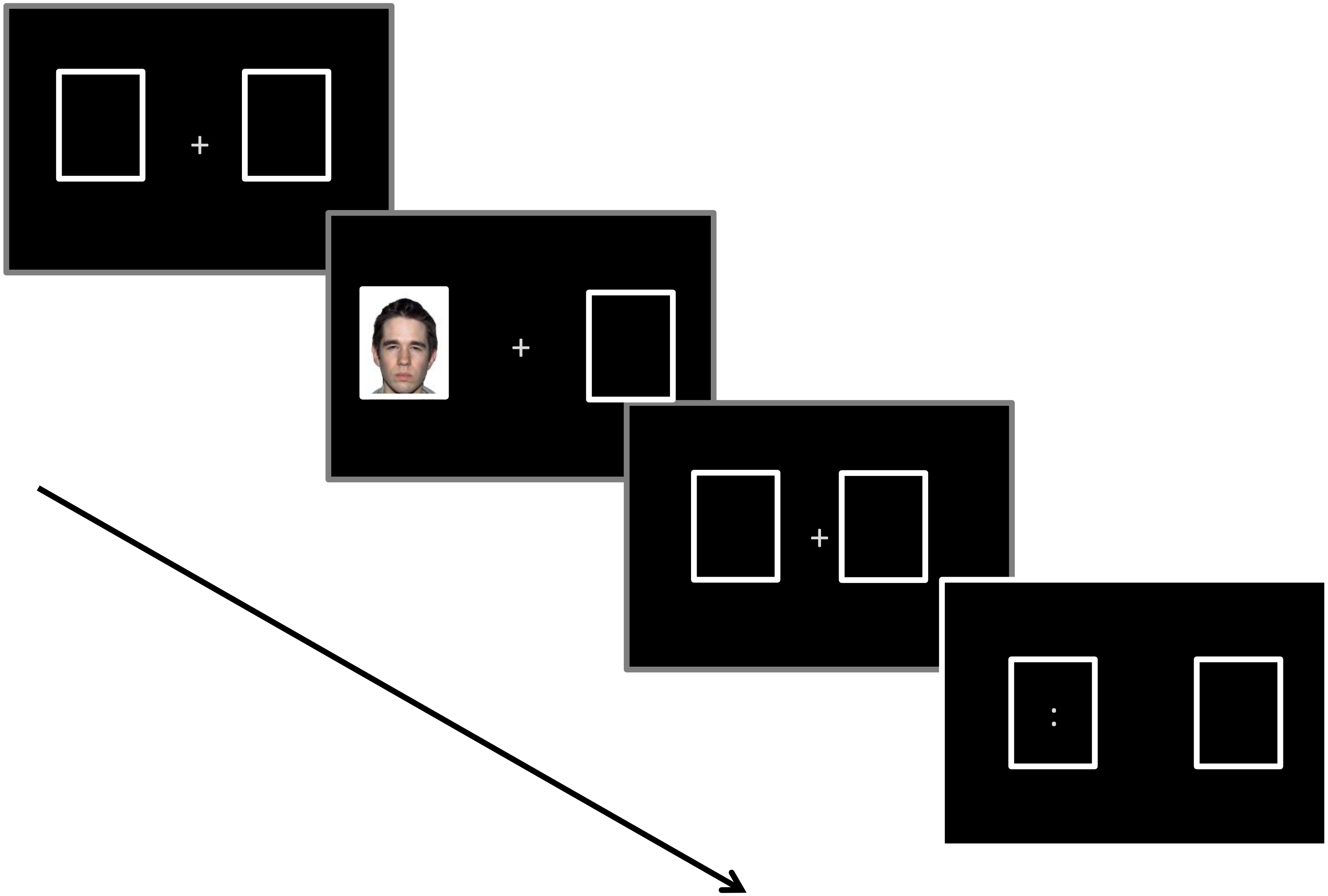

Figure 2. Emotional cueing task


Figure 3. Sample emotional face images

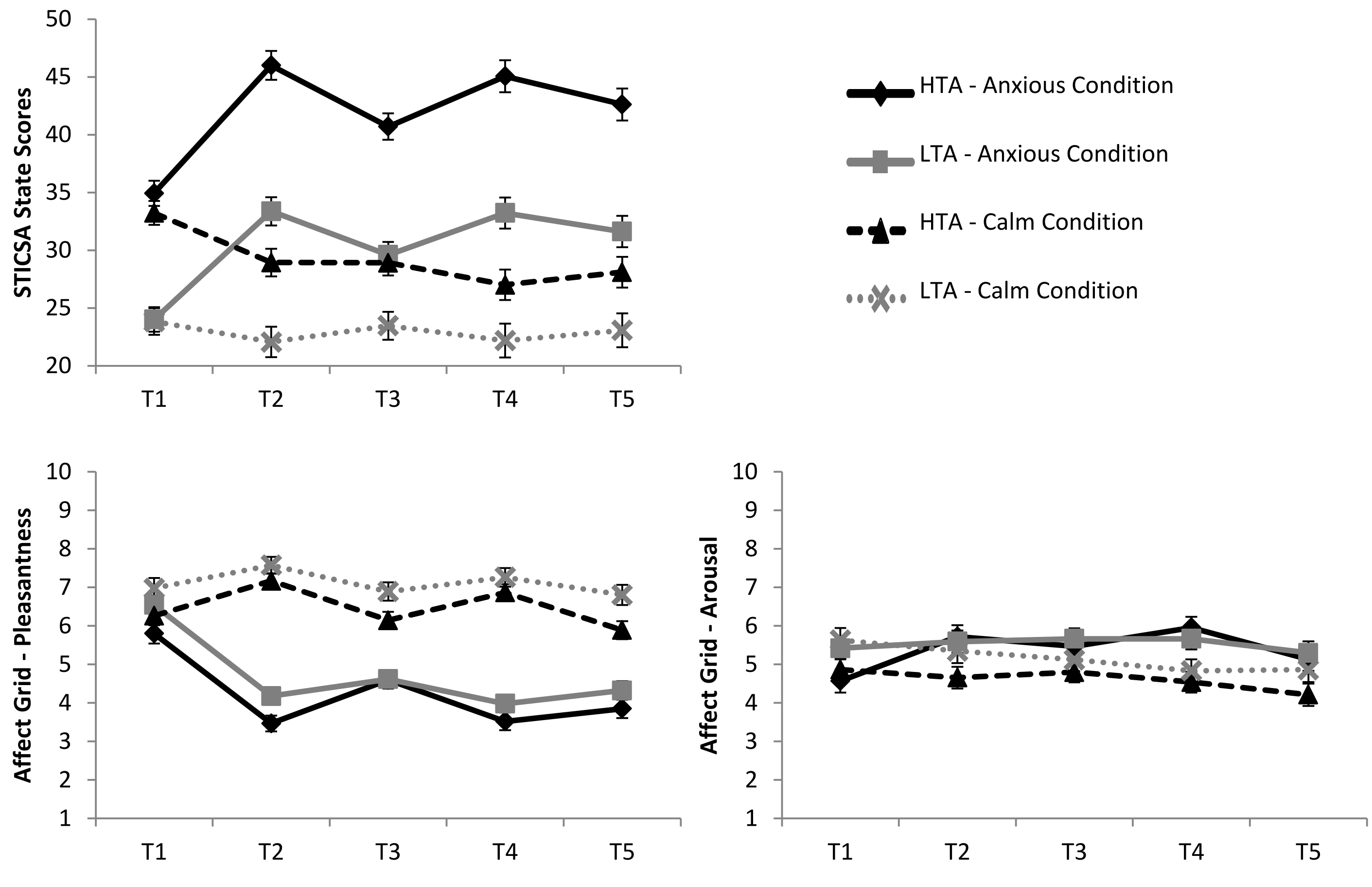


Figure 4. Study 3 mood ratings by group and observation time

**Table 1**

Study 1 Means and Standard Deviations for Demographic and Questionnaire Data by Trait Anxiety Group

| | Low Trait Anxiety Group | | High Trait Anxiety Group | | | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | *t* | *p* |
| Age | 20.45 | (3.05) | 19.27 | (1.81) | 2.902 | .004 |
| Sex (1=Female) | 0.70 | (0.46) | 0.75 | (0.43) | -.794 | .429 |
| STICSA-Trait | 29.28 | (4.84) | 48.66 | (5.72) | -22.868 | .000 |
| STICSA-State | 27.54 | (5.08) | 39.93 | (9.25) | -10.336 | .000 |
| DASS-Depression | 5.32 | (4.83) | 17.05 | (9.11) | -10.018 | .000 |
| DASS-Anxiety | 3.62 | (3.91) | 14.83 | (8.63) | -10.407 | .000 |
| DASS-Stress | 8.72 | (6.90) | 21.82 | (8.76) | -10.357 | .000 |
| SDS | 10.47 | (3.74) | 7.81 | (3.61) | 4.523 | .000 |

SD = Standard Deviation; STICSA=State-Trait Inventory of Cognitive and Somatic Anxiety; DASS=Depression Anxiety Stress Scales; SDS=Social Desirability Scale

**Table 2**

Study 1 Response Times for Dot Probe Task and Emotional Cueing Task by Group

| | Dot Probe Task | | | Emotional Cueing Task | |
|---|---|---|---|---|---|
| | **Incongruent** | **Congruent** | **Neutral-Neutral** | **Incongruent** | **Congruent** |
| **LTA** | | | | | |
| Angry | 552 | 553 | | 505 | 504 |
| Fear | 552 | 553 | | 504 | 502 |
| Happy | 552 | 547 | | 499 | 501 |
| Neutral | | | 547 | 505 | 503 |
| **HTA** | | | | | |
| Angry | 535 | 532 | | 492 | 491 |
| Fear | 535 | 533 | | 493 | 493 |
| Happy | 540 | 536 | | 494 | 491 |
| Neutral | | | 537 | 495 | 490 |

**Table 3**

Study 1 Attentional Bias Scores by Group and Condition

| | Dot Probe Task | | | | Emotional Cueing Task | | | |
|---|---|---|---|---|---|---|---|---|
| | Bias Score (ms) | *t* | *df* | *p* | Bias Score (ms) | *t* | *df* | *p* |
| **LTA Group** | | | | | | | | |
| Angry Overall Bias | -1.71 | -.30 | 77 | .77 | -1.70 | -.38 | 76 | .71 |
| Angry Facilitation Bias | -6.67 | -.98 | 77 | .33 | -1.71 | -.49 | 76 | .63 |
| Angry Disengage Bias | 4.96 | 1.12 | 77 | .26 | .01 | .00 | 76 | 1.00 |
| Fear Overall Bias | -1.41 | -.23 | 77 | .82 | .01 | .00 | 76 | 1.00 |
| Fear Facilitation Bias | -6.55 | -1.86 | 77 | .07 | .23 | .06 | 76 | .95 |
| Fear Disengage Bias | 5.14 | 1.07 | 77 | .29 | -.22 | -.07 | 76 | .95 |
| Happy Overall Bias | 5.08 | 1.12 | 77 | .27 | -3.48 | -.74 | 76 | .46 |
| Happy Facilitation Bias | -.38 | -.12 | 77 | .91 | 1.61 | .50 | 76 | .62 |
| Happy Disengage Bias | 5.46 | 1.27 | 77 | .21 | -5.09 | -1.52 | 76 | .13 |
| **HTA Group** | | | | | | | | |
| Angry Overall Bias | 2.47 | .72 | 75 | .47 | -4.96 | -1.02 | 73 | .31 |
| Angry Facilitation Bias | 4.38 | 1.44 | 75 | .15 | -1.57 | -.55 | 73 | .59 |
| Angry Disengage Bias | -1.91 | -.74 | 75 | .46 | -3.39 | -.99 | 73 | .33 |
| Fear Overall Bias | 2.12 | .66 | 75 | .51 | -5.32 | -.84 | 73 | .40 |
| Fear Facilitation Bias | 3.59 | 1.12 | 75 | .27 | -3.53 | -.77 | 73 | .44 |
| Fear Disengage Bias | -1.47 | -.64 | 75 | .53 | -1.80 | -.55 | 73 | .58 |
| Happy Overall Bias | 3.50 | .94 | 75 | .35 | -2.18 | -.52 | 73 | .61 |
| Happy Facilitation Bias | .13 | .05 | 75 | .96 | -1.49 | -.49 | 73 | .62 |
| Happy Disengage Bias | 3.37 | 1.21 | 75 | .23 | -.69 | -.22 | 73 | .83 |

**Table 4**

Study 1 Correlations between Bias Scores and Questionnaire Data

| | **STICSA Trait** | **STICSA State** | **DASS Depression** | **SDS** |
|---|---|---|---|---|
| **Dot Probe** | | | | |
| Angry Overall Bias | -.01 | .01 | -.09 | **.22*** |
| Angry Facilitation Bias | .04 | .07 | .02 | .08 |
| Angry Disengage Bias | -.07 | -.08 | -.16 | **.17*** |
| Fear Overall Bias | .06 | -.05 | .09 | .00 |
| Fear Facilitation Bias | .11 | .02 | .11 | -.03 |
| Fear Disengage Bias | -.01 | -.09 | .02 | .02 |
| Happy Overall Bias | -.11 | -.07 | -.05 | -.01 |
| Happy Facilitation Bias | -.08 | -.04 | -.08 | -.14 |
| Happy Disengage Bias | -.05 | -.05 | .01 | .11 |
| **Emotional Cueing** | | | | |
| Angry Overall Bias | -.03 | **-.17*** | -.10 | .11 |
| Angry Facilitation Bias | .01 | -.11 | -.08 | .08 |
| Angry Disengage Bias | -.05 | -.13 | -.06 | .08 |
| Fear Overall Bias | -.04 | -.07 | -.08 | .08 |
| Fear Facilitation Bias | -.09 | -.10 | -.13 | .10 |
| Fear Disengage Bias | .04 | .00 | .03 | .01 |
| Happy Overall Bias | .01 | -.04 | .03 | .06 |
| Happy Facilitation Bias | -.06 | -.10 | -.09 | .07 |
| Happy Disengage Bias | .07 | .04 | .13 | .01 |

STICSA=State-Trait Inventory of Cognitive and Somatic Anxiety; DASS=Depression Anxiety Stress Scales; SDS= Social Desirability Scale; * $p < .05$

**Table 5**

Study 1 Reliability of Dot Probe and Emotional Cueing Scores

| | Dot Probe Task | | | Emotional Cueing Task | | |
|---|---|---|---|---|---|---|
| | Permutation Reliability | 95% Confidence Interval | Spearman-Brown Corrected Reliability | Permutation Reliability | 95% Confidence Interval | Spearman-Brown Corrected Reliability |
| **Angry Trials** | | | | | | |
| Congruency Score | - | - | - | .09 | -.07 to .25 | .17 |
| Bias Score | .09 | -.07 to .24 | .16 | .00 | -.16 to .16 | .00 |
| Facilitation Score | .33 | .18 to .46 | .50 | .02 | -.14 to .18 | .03 |
| Disengage Score | .11 | -.05 to .26 | .20 | -.01 | -.17 to .15 | - |
| **Fear Trials** | | | | | | |
| Congruency Score | - | - | - | .28 | .13 to .42 | .44 |
| Bias Score | .28 | .13 to .42 | .44 | .18 | .02 to .33 | .30 |
| Facilitation Score | .06 | -.10 to .22 | .11 | .28 | .13 to .42 | .43 |
| Disengage Score | .15 | 0 to .30 | .26 | .05 | -.11 to .21 | .09 |
| **Happy Trials** | | | | | | |
| Congruency Score | - | - | - | .16 | 0 to .31 | .28 |
| Bias Score | .05 | -.11 to .21 | .10 | .02 | -.14 to .18 | .03 |
| Facilitation Score | -.08 | -.24 to .08 | - | .00 | -.16 to .16 | .00 |
| Disengage Score | .06 | -.10 to .22 | .12 | .03 | -.13 to .19 | .06 |

**Table 6**

Study 2 Means and Standard Deviations for Demographic and Questionnaire Data by Trait Anxiety Group

| | Low Trait Anxiety Group | | High Trait Anxiety Group | | | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | *t* | *p* |
| Age | 20.79 | (2.96) | 20.11 | (3.75) | 1.042 | .300 |
| Sex (1=Female) | 0.78 | (0.42) | 0.78 | (0.42) | -.050 | .960 |
| STICSA-Trait | 29.02 | (4.33) | 47.98 | (8.66) | -14.492 | .000 |
| STICSA-State | 25.57 | (4.43) | 37.40 | (11.52) | -7.098 | .000 |
| DASS-Depression | 5.19 | (4.85) | 17.13 | (9.98) | -7.967 | .000 |
| DASS-Anxiety | 3.59 | (4.44) | 14.51 | (8.67) | -8.298 | .000 |
| DASS-Stress | 7.81 | (6.48) | 21.53 | (10.07) | -8.471 | .000 |
| SDS | 10.83 | (3.88) | 7.25 | (3.52) | 5.040 | .000 |

SD = Standard Deviation; STICSA=State-Trait Inventory of Cognitive and Somatic Anxiety; DASS=Depression Anxiety Stress Scales; SDS= Social Desirability Scale

**Table 7**

Study 2 Response Times for Dot Probe Task and Emotional Cueing Task by Group

| | Dot Probe Task | | | Emotional Cueing Task | |
|---|---|---|---|---|---|
| | **Incongruent** | **Congruent** | **Neutral-Neutral** | **Incongruent** | **Congruent** |
| **LTA** | | | | | |
| Angry | 529 | 524 | | 531 | 509 |
| Fear | 529 | 522 | | 533 | 507 |
| Happy | 522 | 515 | | 523 | 512 |
| Neutral | | | 517 | 523 | 500 |
| **HTA** | | | | | |
| Angry | 514 | 519 | | 506 | 497 |
| Fear | 515 | 519 | | 512 | 497 |
| Happy | 513 | 515 | | 514 | 501 |
| Neutral | | | 516 | 511 | 497 |

**Table 8**

Study 2 Attentional Bias Scores by Group and Condition

| | Dot Probe Task | | | | | Emotional Cueing Task | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Bias Score (ms) | *t* | *df* | *p* | *d* | Bias Score (ms) | *t* | *df* | *p* | *d* |
| **<u>LTA Group</u>** | | | | | | | | | | |
| Angry Overall Bias | 4.87 | .91 | 52 | .37 | | -1.06 | -.15 | 51 | .88 | |
| Angry Facilitation Bias | -6.45 | -1.76 | 52 | .08 | | -9.48 | -2.31 | 51 | .02* | -.32 |
| Angry Disengage Bias | 11.32 | 1.94 | 52 | .06 | | 8.42 | 1.59 | 51 | .12 | |
| Fear Overall Bias | 7.40 | 1.84 | 52 | .07 | | 3.08 | .44 | 51 | .66 | |
| Fear Facilitation Bias | -4.15 | -.97 | 52 | .34 | | -7.40 | -1.95 | 51 | .06 | |
| Fear Disengage Bias | 11.55 | 2.65 | 52 | .01* | .37 | 10.48 | 2.06 | 51 | .04* | .29 |
| Happy Overall Bias | 7.04 | 1.76 | 52 | .08 | | -12.33 | -1.16 | 51 | .25 | |
| Happy Facilitation Bias | 2.77 | .63 | 52 | .53 | | -12.33 | -2.91 | 51 | .01* | -.41 |
| Happy Disengage Bias | 4.26 | 1.01 | 52 | .32 | | .00 | .00 | 51 | 1.00 | |
| **<u>HTA Group</u>** | | | | | | | | | | |
| Angry Overall Bias | -5.05 | -1.14 | 54 | .26 | | -3.83 | -.63 | 53 | .53 | |
| Angry Facilitation Bias | -2.95 | -.65 | 54 | .52 | | .59 | .13 | 53 | .89 | |
| Angry Disengage Bias | -2.11 | -.42 | 54 | .68 | | -4.43 | -1.05 | 53 | .30 | |
| Fear Overall Bias | -4.33 | -.92 | 54 | .36 | | 1.11 | .16 | 53 | .87 | |
| Fear Facilitation Bias | -3.00 | -.87 | 54 | .39 | | .13 | .03 | 53 | .97 | |
| Fear Disengage Bias | -1.33 | -.30 | 54 | .76 | | .98 | .21 | 53 | .84 | |
| Happy Overall Bias | -1.65 | -.40 | 54 | .69 | | .26 | .04 | 53 | .97 | |
| Happy Facilitation Bias | 1.82 | .42 | 54 | .67 | | -3.22 | -.67 | 53 | .51 | |
| Happy Disengage Bias | -3.47 | -.79 | 54 | .43 | | 3.48 | .71 | 53 | .48 | |

**Table 9**

Study 2 Correlations between Bias Scores and Questionnaire Data

| | STICSA Trait | STICSA State | DASS Depression | SDS |
|---|---|---|---|---|
| **Dot Probe Scores** | | | | |
| Angry Overall Bias | -.17 | -.12 | -.10 | .08 |
| Angry Facilitation Bias | .03 | -.11 | -.03 | -.04 |
| Angry Disengage Bias | -.18 | -.02 | -.07 | .11 |
| Fear Overall Bias | -.12 | -.04 | -.06 | .07 |
| Fear Facilitation Bias | .03 | .01 | -.02 | .04 |
| Fear Disengage Bias | -.14 | -.04 | -.04 | .03 |
| Happy Overall Bias | -.10 | -.06 | -.03 | .15 |
| Happy Facilitation Bias | .06 | -.06 | -.01 | .04 |
| Happy Disengage Bias | -.16 | .00 | -.02 | .11 |
| **Emotional Cueing Scores** | | | | |
| Angry Overall Bias | .01 | .11 | .02 | .09 |
| Angry Facilitation Bias | .12 | .11 | .13 | -.06 |
| Angry Disengage Bias | -.10 | .05 | -.09 | .18 |
| Fear Overall Bias | -.01 | .02 | -.03 | .05 |
| Fear Facilitation Bias | .00 | -.01 | -.03 | .03 |
| Fear Disengage Bias | -.01 | .03 | -.02 | .04 |
| Happy Overall Bias | .10 | .03 | .07 | -.01 |
| Happy Facilitation Bias | .09 | .05 | .06 | .01 |
| Happy Disengage Bias | .07 | .01 | .05 | -.02 |

**Table 10**

Study 2 Reliability of Dot Probe and Emotional Cueing Scores

| | Dot Probe Task | | | Emotional Cueing Task | | |
|---|---|---|---|---|---|---|
| | Permutation Reliability | 95% Confidence Interval | Spearman-Brown Corrected Reliability | Permutation Reliability | 95% Confidence Interval | Spearman-Brown Corrected Reliability |
| **Angry Trials** | | | | | | |
| Congruency Score | - | - | - | .25 | .06 to .42 | .40 |
| Bias Score | .07 | -.12 to .26 | .13 | .07 | -.12 to .26 | .13 |
| Facilitation Score | .05 | -.14 to .24 | .10 | .08 | -.11 to .27 | .15 |
| Disengage Score | .36 | .18 to .51 | .53 | .09 | -.10 to .28 | .16 |
| **Fear Trials** | | | | | | |
| Congruency Score | - | - | - | .31 | .13 to .47 | .47 |
| Bias Score | .01 | -.18 to .20 | .01 | .06 | -.13 to .25 | .12 |
| Facilitation Score | .02 | -.17 to .21 | .04 | .02 | -.17 to .21 | .03 |
| Disengage Score | .17 | -.02 to .35 | .29 | .09 | -.10 to .28 | .17 |
| **Happy Trials** | | | | | | |
| Congruency Score | - | - | - | .24 | .05 to .41 | .39 |
| Bias Score | -.04 | -.23 to .15 | - | .30 | .11 to .46 | .46 |
| Facilitation Score | .19 | 0 to .37 | .32 | .14 | -.05 to .32 | .24 |
| Disengage Score | .15 | -.04 to .33 | .26 | .41 | .24 to .56 | .58 |

**Table 11**

Study 3 Means and Standard Deviations for Demographic and Questionnaire Data by Trait Anxiety Group

| | Low Trait Anxiety Group | | High Trait Anxiety Group | | | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | *t* | *p* |
| Age | 20.34 | (1.78) | 19.82 | (1.88) | 1.73 | .086 |
| Sex (1=Female) | 0.76 | (0.43) | 0.77 | (0.42) | -0.08 | .940 |
| STICSA-Trait | 26.71 | (3.42) | 44.80 | (8.05) | -18.62 | .000 |
| STICSA-State | 23.92 | (2.75) | 34.03 | (9.05) | -9.65 | .000 |
| DASS-Depression | 3.97 | (5.4) | 13.87 | (9.77) | -7.96 | .000 |
| DASS-Anxiety | 2.58 | (3.81) | 13.14 | (8.00) | -10.71 | .000 |
| DASS-Stress | 6.21 | (6.18) | 17.70 | (9.50) | -9.08 | .000 |
| SDS | 11.95 | (3.82) | 8.77 | (3.35) | 5.54 | .000 |

SD = Standard Deviation; STICSA=State-Trait Inventory of Cognitive and Somatic Anxiety; DASS=Depression Anxiety Stress Scales; SDS= Social Desirability Scale

**Table 12**

Study 3 Response Times for Dot Probe Task and Emotional Cueing Task by Group and Condition

| | Dot Probe Task | | Emotional Cueing Task | |
|---|---|---|---|---|
| | **Incongruent** | **Congruent** | **Incongruent** | **Congruent** |
| LTA – Calm Condition | | | | |
| Angry Trials | 400 | 399 | 560 | 477 |
| Happy Trials | 396 | 401 | 548 | 482 |
| Neutral Trials | - | - | 543 | 476 |
| LTA - Anxious Condition | | | | |
| Angry Trials | 405 | 410 | 545 | 472 |
| Happy Trials | 405 | 410 | 531 | 478 |
| Neutral Trials | - | - | 546 | 471 |
| HTA - Calm Condition | | | | |
| Angry Trials | 415 | 418 | 552 | 461 |
| Happy Trials | 414 | 413 | 537 | 464 |
| Neutral Trials | - | - | 536 | 459 |
| HTA - Anxious Condition | | | | |
| Angry Trials | 408 | 414 | 546 | 454 |
| Happy Trials | 411 | 411 | 541 | 464 |
| Neutral Trials | - | - | 527 | 460 |

**Table 13**

Study 3 Attentional Bias Scores by Group and Condition

| | Dot Probe Task | | | | | Emotional Cueing Task | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Bias Score (ms) | *t* | *df* | *p* | *d* | Bias Score (ms) | *t* | *df* | *p* | *d* |
| **LTA - Calm Condition** | | | | | | | | | | |
| Angry Overall Bias | 1.40 | 0.42 | 34 | .68 | | 16.54 | 1.66 | 34 | .11 | |
| Angry Facilitation Bias | | | | | | -0.49 | -0.18 | 34 | .86 | |
| Angry Disengage Bias | | | | | | 17.03 | 1.89 | 34 | .07 | |
| Happy Overall Bias | -5.06 | -1.53 | 34 | .13 | | -0.66 | -0.08 | 34 | .94 | |
| Happy Facilitation Bias | | | | | | -5.54 | -1.79 | 34 | .08 | |
| Happy Disengage Bias | | | | | | 4.89 | 0.60 | 34 | .55 | |
| **LTA - Anxious Condition** | | | | | | | | | | |
| Angry Overall Bias | -5.08 | -1.31 | 39 | .20 | | -2.45 | -0.36 | 39 | .72 | |
| Angry Facilitation Bias | | | | | | -0.80 | -0.30 | 39 | .77 | |
| Angry Disengage Bias | | | | | | -1.65 | -0.27 | 39 | .79 | |
| Happy Overall Bias | -5.58 | -1.80 | 39 | .08 | | -22.00 | -2.82 | 39 | .01 | -.45 |
| Happy Facilitation Bias | | | | | | -6.95 | -1.61 | 39 | .12 | |
| Happy Disengage Bias | | | | | | -15.05 | -2.31 | 39 | .03 | -.37 |
| **HTA - Calm Condition** | | | | | | | | | | |
| Angry Overall Bias | -3.33 | -0.66 | 42 | .51 | | 13.88 | 1.86 | 39 | .07 | |
| Angry Facilitation Bias | | | | | | -1.38 | -0.41 | 39 | .68 | |
| Angry Disengage Bias | | | | | | 15.25 | 2.51 | 39 | .02 | .40 |
| Happy Overall Bias | 1.56 | 0.43 | 42 | .67 | | -4.48 | -0.72 | 39 | .48 | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Happy Facilitation Bias | | | | | | -5.05 | -1.24 | 39 | .22 | |
| Happy Disengage Bias | | | | | | 0.58 | 0.11 | 39 | .91 | |
| **HTA - Anxious Condition** | | | | | | | | | | |
| Angry Overall Bias | -5.92 | -2.33 | 38 | .03 | -.37 | 25.13 | 2.98 | 37 | .01 | .48 |
| Angry Facilitation Bias | | | | | | 5.76 | 2.32 | 37 | .03 | .38 |
| Angry Disengage Bias | | | | | | 19.37 | 2.43 | 37 | .02 | .39 |
| Happy Overall Bias | -0.13 | -0.04 | 38 | .97 | | 10.13 | 1.23 | 37 | .23 | |
| Happy Facilitation Bias | | | | | | -3.97 | -1.30 | 37 | .20 | |
| Happy Disengage Bias | | | | | | 14.11 | 2.08 | 37 | .04 | .34 |

**Table 14**

Study 3 Reliability of Dot Probe and Emotional Cueing Scores

| | Angry Trials | | | Happy Trials | | |
|---|---|---|---|---|---|---|
| **Difference Score** | Permutation Reliability | 95% Confidence Interval | Spearman-Brown Corrected Reliability | Permutation Reliability | 95% Confidence Interval | Spearman-Brown Corrected Reliability |
| Dot Probe Bias | | | | | | |
| Overall Sample | .28 | .13 to .39 | .44 | .09 | -.07 to .22 | .17 |
| Calm Condition | .42 | .22 to .59 | .59 | .17 | -.05 to .38 | .29 |
| Anxious Condition | .08 | -.14 to .30 | .15 | .04 | -.18 to .26 | .09 |
| Emotional Cueing Congruency | | | | | | |
| Overall Sample | .39 | .25 to .52 | .56 | .42 | .28 to .54 | .59 |
| Calm Condition | .42 | .21 to .59 | .59 | .47 | .27 to .63 | .64 |
| Anxious Condition | .39 | .18 to .56 | .56 | .40 | .20 to .57 | .57 |
| Emotional Cueing Bias | | | | | | |
| Overall Sample | .05 | -.11 to .21 | .10 | .07 | -.09 to .23 | .12 |
| Calm Condition | .06 | -.17 to .28 | .11 | .06 | -.17 to .28 | .11 |
| Anxious Condition | .05 | -.17 to .27 | .10 | .08 | -.15 to .30 | .15 |
| Emotional Cueing Facilitation | | | | | | |
| Overall Sample | -.15 | -.30 to .01 | - | .10 | -.06 to .25 | .19 |
| Calm Condition | -.12 | -.34 to .11 | - | .07 | -.16 to .29 | .13 |
| Anxious Condition | -.17 | -.38 to .05 | - | .14 | -.09 to .35 | .24 |
| Emotional Cueing Disengage | | | | | | |
| Overall Sample | .07 | -.09 to .23 | .13 | .06 | -.10 to .22 | .11 |
| Calm Condition | .08 | -.15 to .30 | .15 | .06 | -.17 to .28 | .11 |
| Anxious Condition | .09 | -.14 to .31 | .16 | .07 | -.15 to .29 | .13 |

**Table 15**

Study 3 Correlations between Emotional Cueing Biases and Questionnaire Data

| | STICSA Trait | DASS Depression | SDS |
|---|---|---|---|
| **Overall Sample (N=153)** | | | |
| Angry Overall Bias | .063 | .074 | .030 |
| Angry Facilitation Bias | .056 | -.015 | .104 |
| Angry Disengage Bias | .049 | .088 | -.007 |
| Happy Overall Bias | .131 | .045 | -.043 |
| Happy Facilitation Bias | .028 | .001 | .056 |
| Happy Disengage Bias | .134 | .051 | -.079 |
| **Calm Condition (n=75)** | | | |
| Angry Overall Bias | -.011 | .089 | .106 |
| Angry Facilitation Bias | .043 | .003 | .157 |
| Angry Disengage Bias | -.030 | .101 | .057 |
| Happy Overall Bias | -.035 | -.033 | .002 |
| Happy Facilitation Bias | -.012 | -.014 | .074 |
| Happy Disengage Bias | -.030 | -.028 | -.038 |
| **Anxious Condition (n=78)** | | | |
| Angry Overall Bias | .140 | .049 | -.038 |
| Angry Facilitation Bias | .077 | -.016 | .038 |
| Angry Disengage Bias | .126 | .060 | -.055 |
| Happy Overall Bias | **.272*** | .108 | -.072 |
| Happy Facilitation Bias | .067 | .014 | .042 |
| Happy Disengage Bias | **.289*** | .122 | -.109 |

92

## References


Amir, N., Beard, C., Burns, M., & Bomyea, J. (2009). Attention modification program in individuals with Generalized Anxiety Disorder. *Journal of Abnormal Psychology, 118*(1), 28.

Amir, N., Beard, C., Taylor, C. T., Klumpp, H., Elias, J., Burns, M., & Chen, X. (2009). Attention training in individuals with Generalized Social Phobia: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *77*(5), 961.

Antony, M. M, Beiling, P. J., Cox, B. J, Enns, M. W., Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment, 10,* 176-181.

Bar-Haim, Y., Holoshitz, Y., Eldar, S., Frenkel, T. I., Muller, D., Charney, D. S., Pine, D. S., et al. (2010). Life-threatening danger and suppression of attention bias to threat. *The American Journal of Psychiatry, 167*(6), 694–698.

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.

Beard, C. (2011). Cognitive bias modification for anxiety: current evidence and future directions. *Expert Review of Neurotherapeutics*, *11*(2), 299-311.

Beck, A. T., & Clark, D. A. (1997). An information processing model of anxiety: Automatic and strategic processes. *Behaviour Research and Therapy*, *35*(1), 49-58.

Boettcher, J., Leek, L., Matson, L., Holmes, E. A., Browning, M., MacLeod, C. ... &

Carlbring, P. (2013). Internet-based attention bias modification for social anxiety: a randomised controlled comparison of training towards negative and training towards positive cues. *PloS One*, *8*(9), e71760.

Bögels, S. M., & Mansell, W. (2004). Attention processes in the maintenance and treatment of social phobia: Hypervigilance, avoidance and self-focused attention. *Clinical Psychology Review*, *24,* 827-856.

Borgmann, K. W., Risko, E. F., Stolz, J. A., & Besner, D. (2007). Simon says: Reliability and the role of working memory and attentional control in the Simon task. *Psychonomic Bulletin & Review*, *14*(2), 313-319.

Bradley, B. P., Mogg, K., Millar, N., Bonham-Carter, C., Fergusson, E., Jenkins, J., & Parr, M. (1997). Attentional biases for emotional faces. *Cognition & Emotion*, *11*(1), 25-42.

Bradley, B. P., Mogg, K., Falla, S. J., & Hamilton, L. R. (1998). Attentional bias for threatening facial expressions in anxiety: Manipulation of stimulus duration. *Cognition & Emotion*, *12*(6), 737–753.

Bradley, B. P., Mogg, K., & Millar, N. H. (2000). Covert and overt orienting of attention to emotional faces in anxiety. *Cognition & Emotion*, *14*(6), 789–808.

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

Britton, J. C., Bar-Haim, Y., Clementi, M. A., Sankin, L. S., Chen, G., Shechner, T. … & Pine, D. S. (2013). Training-associated changes and stability of attention bias in youth:

Implications for Attention Bias Modification Treatment for pediatric anxiety. *Developmental Cognitive Neuroscience, 4,* 52-64.

Brosschot, J., Deruiter, C., & Kindt, M. (1999). Processing bias in anxious subjects and repressors, measured by emotional Stroop interference and attentional allocation. *Personality and Individual Differences*, *26*(5), 777–793.

Broomfield, N. M., & Turpin, G. (2005). Covert and overt attention in trait anxiety: a cognitive psychophysiological analysis. *Biological Psychology*, *68*(3), 179-200.

Brown, H. M., Eley, T. C., Broeren, S., MacLeod, C., Rinck, M., Hadwin, J. A., & Lester, K. J. (2014). Psychometric properties of reaction time based experimental paradigms measuring anxiety-related information-processing biases in children. *Journal of Anxiety Disorders*, *28*(1), 97-107.

Carlbring, P., Apelstrand, M., Sehlin, H., Amir, N., Rousseau, A., Hofmann, S. G., & Andersson, G. (2012). Internet-delivered attention bias modification training in individuals with social anxiety disorder: A double blind randomized controlled trial. *BMC Psychiatry*, *12*(1), 66.

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.

Cooper, R. M., Bailey, J. E., Diaper, A., Stirland, R., Renton, L. E., Benton, C. P. … & Munafò, M. R. (2011). Effects of 7.5% CO2 inhalation on allocation of spatial attention to facial cues of emotional expression. *Cognition & Emotion, 25(4),* 626-638.

Cooper, R. M., & Langton, S. R. H. (2006). Attentional bias to angry faces using the dot probe task? It depends when you look for it. *Behaviour Research and Therapy*, *44*(9), 1321–9.

Cooper, R. M., Rowe, A. C., Penton-Voak, I. S., & Ludwig, C. (2009). No reliable effects of emotional facial expression, adult attachment orientation, or anxiety on the allocation of visual attention in the spatial cueing paradigm. *Journal of Research in Personality*, *43*(4), 643–652.

Dalgleish, T., Taghavi, R., Neshat-Doost, H., Moradi, A., Canterbury, R., & Yule, W. (2003). Patterns of processing bias for emotional information across clinical disorders: A comparison of attention, memory, and prospective cognition in children and adolescents with depression, generalized anxiety, and posttraumatic stress disorder. *Journal of Clinical Child and Adolescent Psychology, 32*, 10–21.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, 39,* 1–38.

Egloff, B., & Hock, M. (2003). Assessing attention allocation toward threat-related stimuli: A comparison of the emotional Stroop task and the attentional probe task. *Personality and Individual Differences, 35*(2), 475-483.

Eich, E., Ng, J. T. W., Macaulay, D., Percy, A. D., & Grebneva, I. (2007). Combining music with thought to change mood. In J. A. Coan & J. J. B. Allen (Eds.), *The Handbook of Emotion Elicitation and Assessment* (pp. 124–136). Oxford University Press.

Eide, P., Kemp, A., Silberstein, R. B., Nathan, P. J., & Stough, C. (2002). Test-retest reliability of the emotional Stroop task: Examining the paradox of measurement change. *The Journal of Psychology*, *136*(5), 514-520.

Enock, P. M., Hofmann, S. G., & McNally, R. J. (2014). Attention bias modification training via smartphone to reduce social anxiety: A randomized, controlled multi-session experiment. *Cognitive Therapy and Research, 38*(2), 200-216.

Eysenck, M. W. (1997). *Anxiety and cognition: A unified theory*. Hove, UK: Psychology Press.

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

Fischer, D. G., & Fick, C. (1993). Measuring social desirability: Short forms of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurement, 53*, 417-424.

Fox, E. (1993). Allocation of visual attention and anxiety. *Cognition & Emotion, 7,* 207-215.

Fox, E, Russo, R., Bowles, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety? *Journal of Experimental Psychology: General*, *130*(4), 681–700.

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

Fox, E, Russo, R., & Dutton, K. (2002). Attentional bias for threat : Evidence for delayed disengagement from emotional faces. *Cognition & Emotion*, *16*(3), 1–22.

Gotlib, I. H., Kasch, K. L., Traill, S., Joormann, J., Arnow, B. A., & Johnson, S. L. (2004). Coherence and specificity of information-processing biases in depression and social phobia. *Journal of Abnormal Psychology*, *113*(3), 386–398.

Grös, D. F., Simms, L. J., & Antony, M. M. (2010). Psychometric properties of the state-trait inventory for cognitive and somatic anxiety (STICSA) in friendship dyads. *Behavior Therapy*, *41*(3), 277–284.

Grös, D. F., Antony, M. M., Simms, L. J., & McCabe, R. E. (2007). Psychometric properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA): Comparison to the State-Trait Anxiety Inventory (STAI). *Psychological Assessment*, *19*(4), 369–81.

Hayward, P., Ahmad, T., & Wardle, J. (2000). Attention to bodily sensations: A test of the cognitive-attentional model of panic. *Depression and Anxiety*, *12*(4), 203-208.

Heeren, A., Reese, H. E., McNally, R. J., & Philippot, P. (2012). Attention training toward and away from threat in social phobia: effects on subjective, behavioral, and physiological measures of anxiety. *Behaviour Research and Therapy*, *50*(1), 30-39.

Hernandez, S., Vander Wal, J. S., & Spring, B. (2003). A negative mood induction procedure with efficacy across repeated administrations in women. *Journal of Psychopathology and Behavioral Assessment*, *25*(1), 49–55.

Horwitz, S. K., & Horwitz, I. B. (2012). Small is beautiful: Implications of reliability and statistical power for testing the efficacy of HR interventions. *Human Resource Management*, *51*(1), 143-160.

Jefferies, L. N., Smilek, D., Eich, E., & Enns, J. T. (2008). Emotional valence and arousal interact in attentional control. *Psychological Science*, *19*(3), 290–295.

Julian, K., Beard, C., Schmidt, N. B., Powers, M. B., & Smits, J. A. (2012). Attention training to reduce attention bias and social stressor reactivity: An attempt to replicate and extend previous findings. *Behaviour Research and Therapy*, *50*(5), 350-358.

Kimble, M. O., Fleming, K., Bandy, C., & Zambetti, A. (2010). Attention to novel and target stimuli in trauma survivors. *Psychiatry Research*, *178*(3), 501–506.

Kopriva, R. J., & Shaw, D. G. (1991). Power estimates: The effect of dependent variable reliability on the power of one-factor ANOVAs. *Educational and Psychological Measurement*, *51,* 585-595.

Koster, E.H., 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,* 1183-1192.

Koster, E. H.W., Leyman, L., De Raedt, R., & Crombez, G. (2006). Cueing of visual attention by emotional facial expressions: The influence of individual differences in anxiety and depression. *Personality and Individual Differences*, *41*(2), 329–339.

Koster, E.H.W, Crombez, G., Verschuere, B., Van Damme, S., & Wiersema, J. R. (2006). Components of attentional bias to threat in high trait anxiety: Facilitated engagement, impaired disengagement, and attentional avoidance. *Behaviour Research and Therapy*, *44*(12), 1757–71.

Koster, E. H., Verschuere, B., Crombez, G., & Van Damme, S. (2005). Time-course of attention for threatening pictures in high and low trait anxiety. *Behaviour Research and Therapy*, *43*(8), 1087-1098.

Legerstee, J. S., Tulen, J. H. M., Kallen, V. L., Dieleman, G. C., Treffers, P. D. a, Verhulst, F. C., & Utens, E. M. W. J. (2009). Threat-related selective attention predicts treatment success in childhood anxiety disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *48*(2), 196–205.

Li, S., Tan, J., Qian, M., & Liu, X. (2008). Continual training of attentional bias in social anxiety. *Behaviour Research and Therapy*, *46*(8), 905-912.

Lovibond, P. F, & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33,* 335-342.

MacLeod, J.W., Lawrence, M.A., McConnell, M.M., Eskes, G.A., Klein, R.M., & Shore, D.I. (2010). Appraising the ANT: Psychometric and theoretical considerations of the Attention Network Test. *Neuropsychology, 24,* 637–651.

MacLeod, C., & Mathews, A. (1988). Anxiety and the allocation of attention to threat. *The Quarterly Journal of Experimental Psychology Section A*, *40*(4), 653–670.

MacLeod, C., & Mathews, A. (2012). Cognitive bias modification approaches to anxiety. *Annual Review of Clinical Psychology*, *8*, 189-217.

MacLeod, C., Rutherford, E., Campbell, L., Ebsworthy, G., & Holker, L. (2002). Selective attention and emotional vulnerability: assessing the causal basis of their association through the experimental manipulation of attentional bias. *Journal of Abnormal Psychology*, *111*(1), 107.

Maloney, E. A., Risko, E. F., Preston, F., Ansari, D., & Fugelsang, J. (2010). Challenging the reliability and validity of cognitive measures: The case of the numerical distance effect. *Acta Psychologica*, *134*(2), 154-161.

Mathews, A. (1990) . Why worry? The cognitive function of anxiety. *Behaviour Research and Therapy*, *28*, 455-468.

Mathews, A., & Mackintosh, B. (1998). A cognitive model of selective processing in anxiety. *Cognitive Therapy and Research, 22*, 539-560.

Mathews, A., & MacLeod, C. (2002). Induced processing biases have causal effects on anxiety. *Cognition & Emotion*, *16,* 331–354.

McNally, R. J., Amir, N., & Lipke, H. J. (1996). Subliminal processing of threat cues in posttraumatic stress disorder? *Journal of Anxiety Disorders, 10*(2), 115-128.

Miller, J. (1991). Reaction time analysis with outlier exclusion: Bias varies with sample size. *The Quarterly Journal of Experimental Psychology*, *43*(4), 907-912.

Miller J. & Ulrich, R. (2013). Mental chronometry and individual differences: Modeling reliabilities and correlations of reaction time means and effect sizes. *Psychonomic Bulletin and Review, 20 (5),* 819-858.

Mogg, K, Bradley, B. P., & Hallowell, N. (1994). Attentional bias to threat: roles of trait anxiety, stressful events, and awareness. *Quarterly Journal of Experimental Psychology*, *47*(4), 841–864.

Mogg, K., & Bradley, B. P. (1998). A cognitive-motivational analysis of anxiety. *Behaviour Research and Therapy*, *36*(9), 809-848.

Mogg, K, & Bradley, B. P. (1999). Some methodological issues in assessing attentional biases for threatening faces in anxiety : A replication study using a modified version of the probe detection task. *Behaviour Research and Therapy*, *37*, 595-604.

Mogg, K., Bradley, B. P., De Bono, J., & Painter, M. (1997). Time course of attentional bias for threat information in non-clinical anxiety. *Behaviour Research and Therapy*, *35*(4), 297-303.

Mogg, K, Bradley, B. P., Dixon, C., Fisher, S., Twelftree, H., & McWilliams, A. (2000). Trait anxiety, defensiveness and selective processing of threat: an investigation using two measures of attentional bias. *Personality and Individual Differences*, *28*(6), 1063–1077.

Mogg, K, Bradley, B. P., & Hallowell, N. (1994). Attentional bias to threat: roles of trait anxiety, stressful events, and awareness. *Quarterly Journal of Experimental Psychology, 47(4),* 841–864.

Mogg, K., Bradley, B. P., Williams, R., & Mathews, A. (1993). Subliminal processing of emotional information in anxiety and depression. *Journal of Abnormal Psychology, 102*(2), 304.

Mogg, K, Holmes, A., Garner, M., & Bradley, B. P. (2008). Effects of threat cues on attentional shifting, disengagement and response slowing in anxious individuals. *Behaviour Research and Therapy*, *46*(5), 656–667.

Mogg, K., Mathews, A., Bird, C., & Macgregor-Morris, R. (1990). Effects of stress and anxiety on the processing of threat stimuli. *Journal of Personality and Social Psychology, 59(6),* 1230.

Mogg, K., Mathews, A., & Weinman, J. (1987). Memory bias in clinical anxiety. *Journal of Abnormal Psychology,* 96(2), 94.

Mogoaşe, C., David, D., & Koster, E. H. (2014). Clinical Efficacy of Attentional Bias Modification Procedures: An Updated Meta-Analysis. *Journal of Clinical Psychology*.

Mühlberger, A., Wieser, M. J., & Pauli, P. (2008). Visual attention during virtual social situations depends on social anxiety. *CyberPsychology & Behavior*, *11*(4), 425-430.

Müller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention:

time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(2), 315–330.

Murphy, K., & Davidshofer, C. (2005). Psychological testing: Principles and applications (6th Ed). Upper Saddle River, NJ: Prentice Hall.

Nelson, A. L., Purdon, C., Quigley, L., Carriere, J., & Smilek, D. (2014). Distinguishing the roles of trait and state anxiety on the nature of anxiety-related attentional biases to threat using a free viewing eye movement paradigm. *Cognition & Emotion*, (ahead-of-print), 1-23.

Neubauer, K., von Auer, M., Murray, E., Petermann, F., Helbig-Lang, S., & Gerlach, A. L. (2013). Internet-delivered attention modification training as a treatment for social phobia: A randomized controlled trial. *Behaviour Research and Therapy*, *51*(2), 87-97.

Oehlberg, K. a, Revelle, W., & Mineka, S. (2012). Time-course of attention to negative stimuli: Negative affectivity, anxiety, or dysphoria? *Emotion*, *12*(5), 943–59.

Ouimet, A. J., Gawronski, B., & Dozois, D. J. (2009). Cognitive vulnerability to anxiety: A review and an integrative model. *Clinical Psychology Review*, *29*(6), 459-470.

Phillips, J. P., & Giancola, P. R. (2008). Experimentally induced anxiety attenuates alcohol-related aggression in men. *Experimental and Clinical Psychopharmacology*, *16*(1), 43–56.

Quigley, L., Nelson, A.L., Carriere, J., Smilek, D., & Purdon C. (2012). The effects of trait and state anxiety on attention to emotional images: an eye-tracking study. *Cognition & Emotion*, 26, 1390-411.

R Core Development Team. (2005). R: A Language and Environment for Statistical

Computing. R Foundation for Statistical Computing.

Rapee, R.M., & Heimberg, R.G. (1997). A cognitive-behavioral model of anxiety in social phobia. *Behaviour Research and Therapy*, *35*, 741–756.

Ree, M. J., French, D., MacLeod, C., & Locke, V. (2008). Distinguishing cognitive and somatic dimensions of state and trait anxiety: Development and validation of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA). *Behavioural and Cognitive Psychotherapy*, *36*(03), 313-332.

Reinecke, A., Waldenmaier, L., Cooper, M. J., & Harmer, C. J. (2013). Changes in automatic threat processing precede and predict clinical changes with exposure-based cognitive-behavior therapy for panic disorder. *Biological Psychiatry*, *73*(11), 1064-1070.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, *57*(3), 493–502.

Santesso, D. L., Meuret, A. E., Hofmann, S. G., Mueller, E. M., Ratner, K. G., Roesch, E. B., & Pizzagalli, D. A. (2008). Electrophysiological correlates of spatial orienting towards angry faces: A source localization study. *Neuropsychologia*, *46*(5), 1338–48.

Sattler, J.M. (2008). The Assessment of Children: Cognitive Foundations, Fifth Edition. San Diego: Jerome M. Sattler, Publisher, Inc.

Sayette, M. A., Martin, C. S., Perrott, M. A., Wertz, J. M., & Hufford, M. R. (2001). A test of the appraisal-disruption model of alcohol and stress. *Journal of Studies on Alcohol*, *62*, 247–256.

Schmidt, N. B., Richey, J. A., Buckner, J. D., & Timpano, K. R. (2009). Attention training for generalized social anxiety disorder. *Journal of Abnormal Psychology*, *118*(1), 5.

Schmukle, S. C. (2005). Unreliability of the dot probe task. *European Journal of Personality*, *19*(7), 595–605.

Smilek, D., Birmingham, E., Cameron, D., Bischof, W., & Kingstone, A. (2006). Cognitive ethology and exploring attention in real-world scenes. *Brain Research, 1080(1),* 101-119.

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

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.

Stolz, J. A., Besner, D., & Carr, T. H. (2005). Implications of measures of reliability for theories of priming: Activity in semantic memory is inherently noisy and uncoordinated. *Visual Cognition, 12,* 284-336.

Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology, 28*, 191-193.

Todd, R. M., Cunningham, W. A., Anderson, A. K., & Thompson, E. (2012). Affect-biased attention as emotion regulation. *Trends in Cognitive Sciences*, *16*(7), 1–8.

Tottenham, N., Tanaka, J., Leon, A., et al. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Research, 168,* 242-249.

Van Bockstaele, B., Verschuere, B., Koster, E. H., Tibboel, H., De Houwer, J., & Crombez, G. (2011). Differential predictive power of self-report and implicit measures on behavioural and physiological fear responses to spiders. *International Journal of Psychophysiology*, *79*(2), 166-174.

Van Bockstaele, B., Verschuere, B., Tibboel, H., De Houwer, J., Crombez, G., & Koster, E. H. (2014). A review of current evidence for the causal impact of attentional bias on fear and anxiety. *Psychological Bulletin*, 140(3), 682-721.

Van Damme, S., Crombez, G., Hermans, D., Koster, E. H., & Eccleston, C. (2006). The role of extinction and reinstatement in attentional bias to threat: A conditioning approach. *Behaviour Research and Therapy, 44(11),* 1555-1563.

Van Der Does, W. (2002). Different types of experimentally induced sad mood? *Behavior Therapy*, *33*(4), 551–561.

Van Selst, M. & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology*, *47*(3), 631-650.

Waechter, S., Nelson, A.L., Wright, C., Hyatt, A., & Oakman, J. (2014). Attentional bias to threat in social anxiety: Psychometric properties of dot probe and eye-tracking measures. *Cognitive Therapy and Research, 38*, 313-333.

Waechter, S., Stolz, J. A., & Besner, D. (2010). Visual word recognition: On the reliability of repetition priming. *Visual Cognition*, *18*(4), 537-558.

Waters, A. M., Mogg, K., Bradley, B. P., & Pine, D. S. (2012). Attention bias for angry faces in children with social phobia. *Journal of Experimental Psychopathology*, *2*, 475-489.

Weierich, M. R., Treat, T. A., & Hollingworth, A. (2008). Theories and measurement of visual attentional processing in anxiety. *Cognition & Emotion*, *22*(6), 985-1018.

Williams, J. M. G., Watts, F. N., MacLeod, C., & Mathews, A. (1988). *Cognitive psychology and the emotional disorders*. New York: Wiley.

Williams, J. M. G., Watts, F. N., MacLeod, C., & Mathews, A. (1997). *Cognitive sychology* *and the emotional disorders* (2nd ed.). New York: Wiley.

Yiend, J., & Mathews, A. (2001). Anxiety and attention to threatening pictures. *The Quarterly Journal of Experimental Psychology A*, *54*(3), 665–681.

York, D., Borkovec, T. D., Vasey, M., & Stern, R. (1987). Effects of worry and somatic anxiety induction on thoughts, emotion and physiological activity. *Behaviour Research and Therapy*, *25*(6), 523–526.