RESEARCH ARTICLE

Extended testing with the dot-probe task increases test–retest reliability and validity

Jacob S. Aday[1,2] · Joshua M. Carlson[1]

Abstract

The dot-probe task is used to measure attentional biases toward threat. However, the test–retest reliability and validity of the task have been called into question. No studies to date have included an extended number of testing sessions or tailored the task to individual participants by incorporating self-relevant stimuli—doing so may improve reliability. Here, participants provided 10 words that caused them anxiety and 10 neutral words which were incorporated into a dot-probe task for 6 sessions. The test–retest reliability of their bias indices was stronger in bottom-target trials relative to top-target trials and stronger among later relative to earlier sessions. State and trait anxiety were moderately correlated with bias indices in later sessions, but not earlier sessions. Overall reaction time in each session was moderately correlated with state and trait anxiety. These results suggest that including extended testing may facilitate dot-probe task test–retest reliability and validity.

Keywords Attention bias · Dot-probe task · Test–retest reliability · Target location · Anxiety

Introduction

Threatening stimuli have been shown to automatically capture, hold, and direct attention (Salemink et al. 2007; Carlson and Reinke 2008; Carlson et al. 2009a, b, 2012; Becker 2010; Carlson and Mujica-Parodi 2014; Torrence et al. 2017; Carlson and Aday 2017). According to cognitive theories, this capture of attention is exaggerated in those with anxiety disorders (Bar-Haim 2010). This attention bias toward threat is thought to not just be a symptom of anxiety, but a causal factor leading to the development and maintenance of other anxious symptoms (MacLeod et al. 2002). Attentional biases toward threat are consistent with what many consider to be a cardinal symptom of anxiety—hypervigilance toward potential threats. While the underlying mechanisms involved in attentional bias are still heavily debated, there is a general

✉ Jacob S. Aday
aday1js@cmich.edu

[1] Northern Michigan University, Marquette, MI, USA

[2] Department of Psychology, Central Michigan University, 101 Sloan Hall, Mount Pleasant, MI, USA

consensus that it is a symptom and causal factor in some, though not all, cases of anxiety (Van Bockstaele et al. 2014).

The dot-probe task is one cognitive test researchers have developed in order to get a quantifiable measure of attention bias (Macleod et al. 1986). In the task, participants are seated at a computer and instructed to fixate on a central cue to begin each trial. Then, two stimuli (typically, one threat-related and one neutral) are briefly presented on opposing sides of the screen. After the stimuli disappear, a small dot (or some other target stimulus) appears in the location previously occupied by one of the stimuli. Faster reaction times when the target replaces a threatening stimulus (congruent trial) compared to a neutral stimulus (incongruent trial) are thought to be indicative of an attentional bias toward threat. That is, if the participant's attention is automatically biased toward the threatening stimulus, then targets appearing in that location should be found faster and targets appearing behind the neutral stimulus will take longer to find as they must disengage from the threatening stimulus on the other side of the screen. Therefore, it appears that there are two distinct aspects of attention in which attention biases seem to manifest—facilitated orienting toward threat and delayed disengagement from threat (Koster et al. 2004; Salemink et al. 2007; Carlson and Reinke 2008; Carlson and Mujica-Parodi 2014).

Although the dot-probe task has been extensively used in research settings as a psychometric instrument, its test–retest reliability has been called into question (Schmukle 2005; Staugaard 2009; Price et al. 2015). While several studies have reported low test–retest reliability (*r*'s 0–0.3), group data does seem to consistently show attention biases—suggesting that there are likely individual-level and state-level factors that influence whether participants will display a bias at any one time (Staugaard 2009). It may also be that these studies have not yielded high test–retest reliability because they have failed to consider the spatial location of the target in calculating their bias indices. Indeed, Price et al. (2015) found that when calculating top and bottom-target bias indexes separately, test–retest reliability was high for bottom-target trials, but not top-target trials. Based off these results, it would seem that calculating top and bottom-target bias indices separately for various types of threat stimuli may lead to higher test–retest reliability. Additionally, dot-probe task studies examining test–retest reliability have exclusively used only two sessions to date (Schmukle 2005; Staugaard 2009; Price et al. 2015). It is possible that the test–retest reliability may change as a function of extended testing. Yet, this possibility remains untested.

Another factor that could increase reliability and validity is incorporating self-relevant stimuli. Given that anxiety is thought to stem from biasing attention toward perceived threats in one's day-to-day life, it would seem that incorporating those same stimuli into the task may yield a more ecologically valid measure of anxiety. Anxiety is incredibly heterogenous as evidenced by the wide variety of disorders classified under "Anxiety Disorders" (e.g., specific phobia, social phobia, generalized anxiety disorder; American Psychological Association 2013) as well as the diverse patterns in symptoms and disease trajectories (Nandi et al. 2009). Given this heterogeneity, anxiety is often accompanied by person-specific fears toward stimuli that generally should not induce anxiety, at least at that time. As such, there is likely a wide variety in the type of stimuli that induce anxiety and experimenter-generated stimuli may not be threat-relevant or salient enough to reliably capture attention. Indeed, researchers have found that individuals with subtypes of anxiety, such as those with social anxiety or specific phobias, show attention biases toward stimuli associated with their disorder (i.e., faces and spiders, respectively; Gilboa-Schechtman et al. 1999; Mogg and Bradley 2006). This suggests that attention biases may be more pronounced when using stimuli tailored toward individuals' specific fears.

To our knowledge, other than Price et al. (2015), no studies have looked at the test–retest reliability of the dot-probe task based on the spatial location of the target. Based on Price et al.'s (2015) promising findings, it seems that doing so may increase reliability. Additionally, it is currently unclear how the test–retest reliability of the task changes across extended testing since previous studies have only utilized two sessions. Lastly, it would seem that incorporating self-relevant stimuli into the task would increase reliability and validity as well given that these stimuli reliably induce anxiety in the real world. To address these gaps in the literature, participants completed a self-relevant variant of the dot-probe task across six testing sessions.

Methods

Participants

Twenty-five unselected participants aged between 18 and 30 (female = 21, M_{age} = 20.27) were recruited from Northern Michigan University to participate in the experiment. Participants completed the Spielberger State–Trait Anxiety Inventory which yields measures of trait (STAI-T) and state (STAI-S) anxiety (Spielberger et al. 1970). Using these scores, participants were median split into a low (*N* = 12; STAI-T *M* = 35.50, SD = 4.40; STAI-S *M* = 35.92, SD = 6.14) and high (*N* = 13; STAI-T *M* = 50.31, SD = 7.12; STAI-S *M* = 48.46, SD = 6.39) trait anxiety group during data analysis in order to look at individual differences in trait anxiety. Given that Knight et al. (1983), reported a mean STAI-T score of 36.10 (SD = 7.86) in an ages 18–29 unselected sample, our low-anxiety group's STAI-T scores are comparable to a healthy population and our high-anxiety group scored almost 2 standard deviations higher. These participants were the control group in a larger study looking at attention bias modification (ABM) with self-relevant stimuli. As part of the ABM study, participants came to complete a dot-probe task utilizing emotional faces while near-infrared spectroscopy (NIRS) measures were recorded from the prefrontal cortex during the first and last session.

Procedure

Before data collection, all participants provided informed consent and the study was approved by the Northern Michigan University Institutional Review Board (Project # HS16-723). After collecting informed consent and basic demographic information, the participants were asked to give a list of the 10 things that cause them the most anxiety. Participants then came to the lab on six different days within a 2–3 week span to complete the self-relevant dot-probe task. Participants were only allowed to complete one session per day and all sessions needed to be completed within 3 weeks. There were no other limits on the number of days between sessions. Participants completed the STAI-S and STAI-T during the first and last sessions.

Stimuli

Participants each provided a list of the 10 things that cause them the most anxiety. For the full list of anxiety-related and neutral words see the Supplementary Online Materials. Each answer was limited to one 3–11 letter word because of logistical limitations with the task design. That is, participants may not have time to read words longer than 11 letters in 500 ms; additionally, an affective database for phrases would be needed to match a neutral phrase with their anxiety-provoking phrase. If they had trouble coming up with a list of 10 things that cause them anxiety, they were prompted to think of things they find threatening or worrying. After compiling their list of anxiety/threat-related words, they were asked to choose a corresponding word of equal length from a neutral word list. They were told to select words that were "emotionally neutral" to themselves. The neutral list was generated from the Affective Norms for English Words List (Bradley and Lang 1999); words in this database are rated from 0 to 10 (negative to positive). Words rated between 4 and 6 (i.e., within 1 standard deviation of a "neutral" rating) were compiled for our neutral word list. We had participants choose their words from a list rather than generating them themselves because (1) it can be difficult to freely come up with a word of a certain length, which becomes more difficult when it must specifically be an emotionally neutral word and (2) we did not want them to implicitly generate neutral words that were somewhat related to their threat-related words.

Dot-probe task

After a research assistant programmed their words into the task using E-Prime2 (Psychology Software Tools, Pittsburgh, PA), participants were seated approximately 59 cm from the 60 Hz 16″ computer screen and the research assistant read aloud the instructions. Each trial began with a white plus sign (1.74° of the visual angle) centered on a black screen for 1000 ms. Then, two white lowercase words (depending on word length each word occupied 2.93°–9.52° × 1.48° of the visual angle) were briefly presented in Courier New 55 point font: one on the top and one on the bottom half of the screen (5.14° from the midline). One word was a stimulus they previously reported as causing them anxiety and one was a neutral word of equal length. After a short period of presentation (500 ms; Salemink et al. 2007), the words disappeared and were replaced with a small dot (0.4° of the visual angle) on one-half of the screen. The participants' task was to locate this dot by pressing "1" on a Chronos response box for top-sided targets and "2" for bottom-sided targets with their right hand. An intertrial interval of 1000 ms occurred after the participant's response. The target appeared an equal number of times behind threat-related (i.e., congruent trials) and neutral (i.e., incongruent

trials) words and targets appeared equally on top and bottom. Faster responding on congruent compared to incongruent trials is thought to be indicative of an attentional bias. Each of the 6 sessions was divided into 4 80-trial blocks (i.e., 1920 trials total).

STAI

The STAI-S yields a measure of state anxiety—how anxious one *currently* feels, while the STAI-T yields a measure of trait anxiety—how anxious one *generally* feels. The questionnaire consists of 40 questions, half of which are relevant to the STAI-S and STAI-T scales, respectively. Consistent with previous research (Quek et al. 2004; Julian 2011), Cronbach's alpha for the 20 STAI-S and 20 STAI-T items were 0.89 and 0.94, respectively.

Data analysis

Incorrect, premature (i.e., < 150 ms), and delayed (i.e., > 750 ms) responses were excluded from all analyses. After filtering, 95.63% of the data were available for further analyses. Participants who did not finish all six sessions were excluded from all analyses ($N = 5$). A mixed methods analysis of variance (ANOVA) was used to test the between-subjects effect of anxiety group (low vs. high) and the within-subjects effects of target location (top vs. bottom), trial type (congruent vs. incongruent), and session (1–6). When appropriate, Bonferroni-corrected post hoc pairwise comparisons were performed. To examine test–retest reliability, Pearson correlations were run between the attention bias scores from sessions 1–6. Attention bias (AB) scores were calculated by subtracting the mean RT on congruent trials from the mean RT on incongruent trials. Pearson correlations were also used to examine the relationship between anxiety and attention bias to examine validity.

Results

Using SPSS Statistics Software, a $2 \times 2 \times 2 \times 6$ ANOVA was utilized to examine two levels of anxiety group (low and high trait anxiety), two levels of target location (top and bottom), 2 levels of trial type (congruent and incongruent), and six levels of session (1–6) on RT (statistics for main effects and interactions can be found in Table 1). The main effect of session was significant such that RT in session 1 ($M = 410.91$ ms, SE = 9.72) was significantly slower than every other session (M's = 373.65-379.58 ms, SE's = 9.34-10.60; $p < 0.01$), while none of the other sessions were significantly different from one another ($p > 0.05$). The main effect of target location was significant such that top-target trials ($M = 376.94$ ms, SE = 9.27) were faster

Table 1 Main effects and interactions

Effect	F	df	P	eta2
Session*	14.58	5, 115	< 0.0001	0.39
Target location*	6.63	1, 23	0.02	0.22
Trial type	3.79	1, 23	0.06	0.14
Group	3.10	1, 23	0.10	0.12
Trial type × group	3.13	1, 23	0.09	0.12
Location × group	2.21	1, 23	0.15	0.09
Session × group	0.97	5, 115	0.44	0.04
Session × trial type	0.99	5, 115	0.43	0.04
Session × target location	2.03	5, 115	0.08	0.08
Trial type × target location	0.35	1, 23	0.56	0.02
Trial type × group × target location	0.04	1, 23	0.84	0.00
Target location × session × group	1.32	1, 23	0.26	0.04
Session × trial type × target location	0.88	5, 115	0.50	0.04
Group × session × trial type*	2.31	5, 115	0.04	0.09
Trial type × session × target location × group	1.45	5, 115	0.21	0.06

*Effect is significant at $p < 0.05$

Table 2 Correlations between anxiety, attention bias (AB), and reaction time (RT)

	AB 1	AB 2	AB 3	AB 4	AB 5	AB 6
STAI-S	0.05	0.20	0.09	0.55***	0.40**	0.32*
STAI-T	0.03	0.22	0.06	0.47**	0.33*	0.29*
	RT 1	RT 2	RT 3	RT 4	RT 5	RT 6
STAI-S	0.39**	0.40**	0.38**	0.40**	0.42**	0.36*
STAI-T	0.52***	0.53***	0.54***	0.52***	0.53***	0.51***

*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$

than bottom-target trials ($M = 386.01$ ms, SE = 9.37). The main effect of trial type did not reach significance although congruent trials ($M = 380.43$ ms, SE = 9.04) tended to be faster than incongruent trials ($M = 382.52$ ms, SE = 9.29). Importantly, the anxiety group × session × trial type interaction was significant such that the high-anxiety anxiety group responded faster on congruent relative to incongruent trials in sessions 4 ($M_{diff} = 5.64$, SE = 2.33, $p < 0.05$), 5 ($M_{diff} = 5.67$, SE = 2.78, $p = 0.05$), and 6 ($M_{diff} = 9.24$, SE = 3.50, $p < 0.05$; i.e., those high in anxiety showed an attention bias in later sessions). Consistent with the three-way interaction reported above, AB scores in later sessions (r's = 0.29–0.47, p's = 0.01–0.09) also generally correlated better with trait anxiety than did earlier sessions (r's = 0.03–0.22, $p > 0.05$; see Table 2). The same pattern was also observed for state anxiety (earlier: r's = 0.05–0.20, $p > 0.05$; later: r's = 0.32–0.55, p's = 0.004–0.08). State (r's = 0.36–0.42, p's = 0.03–0.06) and trait (r's = 0.51–0.54, p's < 0.05) anxiety also moderately correlated with overall RT in each session (Table 2). Finally, state and trait anxiety were highly correlated with each other ($r = 0.88$, $p < 0.001$).

AB scores from sessions 1–6 were subjected to one-tailed Pearson correlations to evaluate test–retest reliability. In general, later sessions (4–6) had greater test–retest reliability (r's = 0.41–0.59, $p < 0.05$; mean $r = 0.49$) with one another than did earlier sessions (1–3; r's = − 0.24–0.24, $p > 0.05$; mean $r = 0.05$; Table 3).[1] To test the effect of target location on test–retest reliability, separate AB scores were calculated for top and bottom-target trials. As shown in Table 3, bottom-target trials had greater test–retest reliability than top-target trials. To confirm this general pattern, a paired samples t test was run between the reliability correlation coefficients calculated on top and bottom-target trials separately (Table 3). The effect of target location was significant, $t(14) = − 2.90$, $p < 0.05$, such that bottom-target trials had greater test–retest reliability (mean $r = 0.33$, SD = 0.21) than top-target trials (mean $r = 0.07$, SD = 0.36).

---

[1] This effect was not related to changes in variance given that the overall standard deviation did not change across sessions, $F(5, 23) = 2.22$, $p > 0.05$.

Table 3 Attention bias (AB) test–retest reliability correlation coefficients

Session	1	2	3	4	5	6
1						
Overall AB	1	0.24	0.15	0.19	− 0.01	0.09
Top AB	1	0.32	− 0.28	0.13	− 0.23	− 0.11
Bottom AB	1	0.02	**0.34***	0.04	0.01	0.15
2						
Overall AB	0.24	1	− 0.24	0.03	0.17	**0.56****
Top AB	0.32	1	− 0.01	0.05	0.11	0.21
Bottom AB	0.02	1	0.30	**0.35***	**0.53****	**0.52****
3						
Overall AB	0.15	− 0.24	1	0.16	0.20	0.15
Top AB	− 0.28	− 0.01	1	**0.41***	0.30	0.13
Bottom AB	**0.34***	0.30	1	**0.39***	**0.42***	0.31
4						
Overall AB	0.19	0.03	0.16	1	**0.41***	**0.45***
Top AB	0.13	0.05	**0.41***	1	**0.35***	**0.39***
Bottom AB	0.04	**0.35***	**0.39***	1	**0.71****	**0.46****
5						
Overall AB	− 0.01	0.17	0.20	**0.41***	1	**0.59****
Top AB	− 0.23	0.11	0.30	**0.35***	1	0.24
Bottom AB	0.01	**0.53****	**0.42***	**0.71****	1	**0.54****
6						
Overall AB	0.09	**0.56****	0.15	**0.45***	**0.59****	1
Top AB	− 0.11	0.21	0.13	**0.39***	0.24	1
Bottom AB	0.15	**0.52****	0.31	**0.46****	**0.54****	1

Bold values indicate $p < 0.05$

$**p < 0.01$; $*p < 0.05$

Discussion

To summarize, all participants responded marginally, but not significantly, faster on congruent compared to incongruent trials. However, those high in trait anxiety responded faster on congruent compared to incongruent trials in later, but not earlier, sessions (i.e., they displayed an attention bias in the later sessions). The test–retest reliability of the task was considerably higher in later sessions relative to earlier sessions. Accordingly, attention bias scores in later sessions were also much more strongly correlated with anxiety than earlier sessions. Overall reaction time in each session was also moderately correlated with state and trait anxiety. Top-target trials were overall faster than bottom-target trials; however, bottom-target bias indices had stronger test–retest reliability than top-target indices.

The test–retest reliability of the self-relevant dot-probe task was higher during later sessions than earlier sessions. This finding may explain why many of the studies assessing the reliability of the dot-probe task, which have previously only utilized one or two sessions, have reported strikingly low test–retest reliability coefficients (Schmukle 2005; Staugaard 2009; Price et al. 2015). It seems that biases are not

reliably shown until participants have fairly extensive experience with the task (at least when self-relevant word stimuli are used). Similar to how attention bias indexes were not reliable until later sessions, bias scores did not correlate with anxiety until later sessions—suggesting that the validity of the task also increases with extended testing. Importantly, this effect suggests that participants did not habituate to the repeated usage of the same self-relevant anxiety stimuli across extended testing sessions. This is consistent with previous findings of sustained (i.e., not habituated) attentional bias scores across repeated sessions (Weber et al. 2016). A recent review has noted that in *most* dot-probe task studies, self-reported anxiety does not correlate with attention biases (Mogg et al. 2017), greatly compromising the tasks' utility as a psychometric instrument. Our results suggest that perhaps most dot-probe studies have failed to find relationships with anxiety because they have exclusively included only one or two sessions. Including more sessions/trials appears to reduce noise in the data related to task familiarity and boosts reliability. Extended testing with the addition of self-relevant stimuli seems to yield a more valid measure of attentional bias that is related to anxiety. However, based solely off these results, it is unclear if the effect is due to

the inclusion of extended testing and self-relevant threat stimuli or whether the effect generalizes to extended testing and any threat-related stimulus. Future research probing a direct comparison between experimenter-generated and self-relevant stimuli across extended testing would appease this limitation.

The test–retest reliability of the task also seems to be influenced in part by target location. Bias indices calculated solely from bottom-target trials had overall stronger test–retest reliability coefficients than top-target indices. This effect is consistent with the only other study to our knowledge that has examined separate top and bottom-target test–retest reliability coefficients (Price et al. 2015). Given that fearful eyes automatically capture attention (Carlson and Reinke 2014; Carlson et al. 2016) and these researchers presented faces in a vertical orientation, it could be argued that the stronger test–retest reliability they found on bottom-target trials was driven by the fearful eyes being closer to the central fixation cue when placed on bottom. However, our results suggest that this is not the case given that participants showed a more reliable bias on the bottom when words were utilized. Clearly, more research is needed to elucidate how the spatial location of the target interacts with bias indices and reliability. In particular, many dot-probe task studies present competing threat versus non-threat stimulus pairs and subsequent targets on the horizontal axis (rather than the vertical axis). Several such studies have reported lateralized attention bias effects (Mogg and Bradley 1999, 2002; Fox 2002; Carlson et al. 2009b), but it is unclear whether reliability estimates differ depending on dot location in the left–right variant of the task. Our results provide evidence that the stronger test–retest reliability of bottom-target attention bias indices is still shown when using word stimuli and thus, not driven by emotional eyes being spatially closer to the fixation point.

In addition to location-specific reliability effects, participants also displayed location-specific reaction time effects by responding faster on top-target relative to bottom-target trials. This pattern is consistent with previous studies, which utilized eye-tracking measures of attention bias (Waechter et al. 2014; Price et al. 2015). Waechter et al. (2014) found that participants overwhelmingly made their initial fixation on the top stimulus rather than the bottom stimulus regardless of its emotional content. It is possible that this "look up" bias may be related to the strategies we typically use when scanning computer screens and books in Western cultures (i.e., typically read from left-to-right and top-to-bottom). Given that Waechter and colleagues used emotional faces in their task, the use of word stimuli in the current study may have accentuated this "look up" bias even more. Consistent with this cultural-visuospatial hypothesis, it has been shown across multiple studies that attention biases are more pronounced when targets are

presented on the left than on the right (i.e., suggesting a "look left" bias; Mogg and Bradley 1999, 2002; Fox 2002; Carlson et al. 2009b). Price et al. (2015) argue that this "look up" bias may be a factor driving the increased test–retest reliability on bottom-target trials. Since some argue that attention biases primarily index disengagement (e.g., Koster et al. 2004), it would make sense that biases would be most reliably pronounced when participants must disengage from the top stimulus (which they almost always orient toward first).

State and trait anxiety were moderately correlated with overall reaction time within each session, such that those higher in anxiety had slower reaction times. This is consistent with the results of Mogg et al. (2008) as well as Zvielli et al. (2015), both of whom found a general slowing effect in high-anxiety participants. Given that this effect was present in early sessions and remained relatively consistent across sessions, it seems unlikely that the extended testing aspect of this study is driving this effect. Perhaps when presented with stimuli that induce anxiety, those with higher levels of anxiety spend more cognitive resources evaluating and thinking about those stimuli or are more distracted by them. This would slow reaction times regardless of trial type as we observed. In other words, while overall RT does not directly measure attention bias toward threat, it does seem to index some aspects of cognition related to anxiety based on these results.

While this study puts forth several corroborating and novel results, there are limitations that are important to keep in mind. The first limitation regards our use of an unselected sample. It is possible that not all of our results will generalize to anxious individuals. However, previous dot-probe research has utilized non-clinical samples extensively (Carlson and Reinke 2008; Carlson et al. 2012, 2016; Carlson and Reinke 2014; Carlson 2016) and more pertinently, non-clinical samples have been used to evaluate the reliability of the task (Schmukle 2005; Staugaard 2009; Waechter et al. 2014). Furthermore, Waechter et al. (2014) found that reliability measures do not differ between clinical and non-clinical samples, suggesting that is possible to accurately estimate task reliability with an unselected group. Another limitation is that research suggests that the STAI has poor discriminant validity with measures of depression (Julian 2011) and could be considered a general measure of negative affect. The self-relevant anxiety words in our study differed from neutral words on valence, but also familiarity as the anxiety-related words were self-generated and the neutral words were chosen from an affective word list. We did not include a group with experimenter-generated words or participant-generated neutral words. This would allow us to definitively conclude the extent to which self-relevant stimuli per se affect the reliability and validity of the task. Regardless, the current results provide important exploratory

information regarding the reliability of bias indices and how they relate to anxiety across extended testing.

Despite these limitations, this paper does present novel findings and corroborating results with previous research. First, this study corroborates the Price et al. (2015) finding that bottom-target bias indices are generally more reliable than top-target indices. This is the first study of our knowledge to replicate this finding. We also present novel evidence that test–retest reliability with the dot-probe task improves after extended testing. Importantly, bias scores also did not correlate with anxiety until later sessions—indicating that reliability and validity are improved after extended testing. This is the first study to our knowledge to utilize self-relevant stimuli in the dot-probe task. However, further research is needed to elucidate its effectiveness in the task relative to experimenter-generated stimuli. A final confirmatory finding is that overall reaction time moderately correlated with trait anxiety in each session. Altogether, it is clear that there is much more work that needs to be done to refine the dot-probe task as a valid and reliable psychometric instrument. Nevertheless, the findings presented here indicate that the task can be manipulated (e.g., by considering extended testing, target location, and potentially self-relevant stimuli) to optimize its psychometric properties.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest.

References

American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders (DSM-5). American Psychiatric Pub, Washington

Bar-Haim Y (2010) Research review: attention bias modification (ABM): a novel treatment for anxiety disorders. J Child Psychol Psychiatry 51(8):859–870. https://doi.org/10.1111/j.1469-7610.2010.02251.x

Becker MW (2010) The effectiveness of a gaze cue depends on the facial expression of emotion: evidence from simultaneous competing cues. Atten Percept Psychophys 72(7):1814–1824. https://doi.org/10.3758/app.72.7.1814

Bradley MM, Lang PJ (1999) Affective norms for English words (ANEW): instruction manual and affective ratings. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida

Carlson JM (2016) Facilitated orienting underlies fearful face-enhanced gaze cueing of spatial location. Cogent Psychol 3(1):1147120

Carlson JM, Aday JS (2017) In the presence of conflicting gaze cues, fearful expression and eye-size guide attention. Cogn Emot. https://doi.org/10.1080/02699931.2017.1391065

Carlson JM, Mujica-Parodi LR (2014) Facilitated Attentional orienting and delayed disengagement to conscious and nonconscious fearful faces. J Nonverbal Behav 39(1):69–77. https://doi.org/10.1007/s10919-014-0185-1

Carlson JM, Reinke KS (2008) Masked fearful faces modulate the orienting of covert spatial attention. Emotion 8(4):522–529. https://doi.org/10.1037/a0012653

Carlson JM, Reinke KS (2014) Attending to the fear in your eyes: facilitated orienting and delayed disengagement. Cogn Emot 28(8):1398–1406

Carlson JM, Fee AL, Reinke KS (2009a) Backward masked snakes and guns modulate spatial attention. Evol Psychol 7(4):147470490900700. https://doi.org/10.1177/147470490900700404

Carlson JM, Reinke KS, Habib R (2009b) A left amygdala mediated network for rapid orienting to masked fearful faces. Neuropsychologia 47(5):1386–1389. https://doi.org/10.1016/j.neuropsychologia.2009.01.026

Carlson JM, Beacher F, Reinke KS, Habib R, Harmon-Jones E, Mujica-Parodi LR, Hajcak G (2012) Nonconscious attention bias to threat is correlated with anterior cingulate cortex gray matter volume: a voxel-based morphometry result and replication. NeuroImage 59(2):1713–1718. https://doi.org/10.1016/j.neuroimage.2011.09.040

Carlson JM, Torrence RD, Vander Hyde MR (2016) Beware the eyes behind the mask: the capture and hold of selective attention by backward masked fearful eyes. Motiv Emot 40(3):498–505. https://doi.org/10.1007/s11031-016-9542-1

Fox E (2002) Processing emotional facial expressions: the role of anxiety and awareness. Cogn Affect Behav Neurosci 2(1):52–63. https://doi.org/10.3758/cabn.2.1.52

Gilboa-Schechtman E, Foa EB, Amir N (1999) Attentional biases for facial expressions in social phobia: the face-in-the-crowd paradigm. Cogn Emot 13(3):305–318

Julian LJ (2011) Measures of anxiety: state-trait anxiety inventory (STAI), beck anxiety inventory (BAI), and hospital anxiety and depression scale-anxiety (HADS-A). Arthritis Care Res 63(S11):S467–S472

Knight RG, Waal-Manning HJ, Spears GF (1983) Some norms and reliability data for the state-trait anxiety inventory and the Zung self-rating depression scale. Br J Clin Psychol 22(4):245–249

Koster EH, Crombez G, Verschuere B, De Houwer J (2004) Selective attention to threat in the dot probe paradigm: differentiating vigilance and difficulty to disengage. Behav Res Ther 42(10):1183–1192. https://doi.org/10.1016/j.brat.2003.08.001

Macleod C, Mathews A, Tata P (1986) Attentional bias in emotional disorders. J Abnorm Psychol 95(1):15–20

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. J Abnorm Psychol 111(1):107–123. https://doi.org/10.1037//0021-843x.111.1.107

Mogg K, Bradley BP (1999) Orienting of attention to threatening facial expressions presented under conditions of restricted awareness. Cogn Emot 13(6):713–740. https://doi.org/10.1080/026999399379050

Mogg K, Bradley BP (2002) Selective orienting of attention to masked threat faces in social anxiety. Behav Res Ther 40(12):1403–1414. https://doi.org/10.1016/s0005-7967(02)00017-7

Mogg K, Bradley BP (2006) Time course of attentional bias for fear-relevant pictures in spider-fearful individuals. Behav Res Ther 44(9):1241–1250

Mogg K, Holmes A, Garner M, Bradley BP (2008) Effects of threat cues on attentional shifting, disengagement and response slowing in anxious individuals. Behav Res Ther 46(5):656–667

Mogg K, Waters AM, Bradley BP (2017) Attention bias modification (ABM): review of effects of multisession ABM training on anxiety and threat-related attention in high-anxious individuals. Clin Psychol Sci. https://doi.org/10.1177/2167702617696359

Nandi A, Beard JR, Galea S (2009) Epidemiologic heterogeneity of common mood and anxiety disorders over the lifecourse in the general population: a systematic review. BMC Psychiatry 9(1):31

Price RB, Kuckertz JM, Siegle GJ, Ladouceur CD, Silk JS, Ryan ND, Dahl RE, Amir N (2015) Empirical recommendations for improving the stability of the dot-probe task in clinical research. Psychol Assess 27(2):365–376. https://doi.org/10.1037/pas0000036

Quek KF, Low WY, Razack AH, Loh CS, Chua CB (2004) Reliability and validity of the Spielberger state-trait anxiety inventory (STAI) among urological patients: a Malaysian study. Med J Malays 59(2):258–267

Salemink E, Van den Hout MA, Kindt M (2007) Selective attention and threat: quick orienting versus slow disengagement and two versions of the dot probe task. Behav Res Ther 45(3):607–615. https://doi.org/10.1016/j.brat.2006.04.004

Schmukle SC (2005) Unreliability of the dot probe task. Eur J Pers 19(7):595–605. https://doi.org/10.1002/per.554

Spielberger CD, Gorsuch RL, Lushene RE (1970) Manual for the state-trait anxiety inventory (self-evaluation questionnaire). Consulating Psychology Press, Palo Alto

Staugaard SR (2009) Reliability of two versions of the dot-probe task using photographic faces. Psychol Sci Q 51(3):339–350

Torrence RD, Wylie E, Carlson JM (2017) The time-course for the capture and hold of visuospatial attention by fearful and happy faces. J Nonverbal Behav 41(2):139–153. https://doi.org/10.1007/s10919-016-0247-7

Van Bockstaele B, Verschuere B, Tibboel H, De Houwer J, Crombez G, Koster EH (2014) A review of current evidence for the causal impact of attentional bias on fear and anxiety. Psychol Bull 140(3):682

Waechter S, Nelson AL, Wright C, Hyatt A, Oakman J (2014) Measuring attentional bias to threat: reliability of dot probe and eye movement indices. Cogn Ther Res 38(3):313–333

Weber MA, Morrow KA, Rizer WS, Kangas KJ, Carlson JM (2016) Sustained, not habituated, activity in the human amygdala: a pilot fMRI dot-probe study of attentional bias to fearful faces. Cogent Psychol 3(1):1259881

Zvielli A, Bernstein A, Koster EH (2015) Temporal dynamics of attentional bias. Clin Psychol Sci 3(5):772–788