



UvA-DARE (Digital Academic Repository)

Reliability and validity of measures of attentional bias towards threat in unselected student samples: seek, but will you find?

Van Bockstaele, B.; Lamens, L.; Salemink, E.; Wiers, R.W.; Bögels, S.M.; Nikolaou, K.

DOI

[10.1080/02699931.2019.1609423](https://doi.org/10.1080/02699931.2019.1609423)

Publication date

2020

Document Version

Final published version

Published in

Cognition & Emotion

License

CC BY-NC-ND

[Link to publication](#)

Citation for published version (APA):

Van Bockstaele, B., Lamens, L., Salemink, E., Wiers, R. W., Bögels, S. M., & Nikolaou, K. (2020). Reliability and validity of measures of attentional bias towards threat in unselected student samples: seek, but will you find? *Cognition & Emotion*, *34*(2), 217-228. <https://doi.org/10.1080/02699931.2019.1609423>

General rights




It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

Reliability and validity of measures of attentional bias towards threat in unselected student samples: seek, but will you find?

Bram Van Bockstaele ^{a,b}, Luuk Lamens^a, Elske Salemink ^{a,c}, Reinout W. Wiers ^a, Susan M. Bögels^{a,b} and Kyriaki Nikolaou^{a,d}

^aDepartment of Psychology, University of Amsterdam Amsterdam, The Netherlands; ^bResearch Institute of Child Development and Education, University of Amsterdam, Amsterdam, The Netherlands; ^cDepartment of Clinical Psychology, University of Utrecht Utrecht, The Netherlands; ^dBehavioural and Clinical Neuroscience, School of Psychology, University of Sussex, Brighton, UK

ABSTRACT

Although attentional bias (AB) is considered a key characteristic of anxiety problems, the psychometric properties of most AB measures are either problematic or unknown. We conducted two experiments in which we addressed the reliability, convergent validity, and concurrent validity of different AB measures in unselected student samples. In Experiment 1 ($N = 66$), the visual probe task and the emotional flanker task yielded unreliable estimates of AB. Both the relevant and irrelevant feature visual search task yielded better reliability estimates, yet AB scores did not correlate significantly with each other nor with self-reported social anxiety. In Experiment 2 ($N = 60$), we retained only the visual search tasks. The relevant feature visual search task was again highly reliable, but it did not correlate significantly with anxiety measures. The irrelevant feature visual search task yielded only small reliability estimates, yet one of the scores was significantly correlated with implicit (but not self-reported or physiological) measures of social anxiety. Together, our results advocate the use of variants of visual search tasks to measure AB and they underline the importance of fundamental psychometric testing in AB research.

ARTICLE HISTORY

Received 5 March 2018
Revised 12 March 2019
Accepted 25 March 2019

KEYWORDS


Reliability; validity;
attentional bias; social
anxiety

Attentional bias (AB) to threat is the preferential allocation of attention to threatening stimuli over non-threatening stimuli (Van Bockstaele et al., 2014). According to information processing theories of anxiety, AB is a core feature of anxiety problems and may even be causally involved in the aetiology or maintenance of anxiety disorders (e.g. Mogg & Bradley, 2016). Given this prominent role of AB in anxiety research, the paradigms measuring AB need to be accurate and reliable. The visual probe task (VPT; MacLeod, Mathews, & Tata, 1986) is probably the most often used AB paradigm. In this paradigm, one threatening and one non-threatening cue (typically a picture or a word) are presented on two different locations of the computer screen. Following the offset of the cues, participants respond to the location or the identity of an emotionally irrelevant

target stimulus (e.g. a letter) that appears on either of the two previously cued locations. AB to threat is inferred from faster reaction times (RTs) on threat-congruent trials (target appearing on the threatening cue location) compared to threat-incongruent trials (target appearing on the non-threatening cue location).

Despite its frequent use, the VPT is not beyond contention. Results of individual studies using the VPT often diverge and the pattern of correlations between individual AB scores and measures of anxiety is highly inconsistent (Van Bockstaele et al., 2014), and even clinical samples often have no significant AB (e.g. Mogg, Waters, & Bradley, 2017). A likely cause of these inconsistencies is the VPT's poor reliability. Schmukle (2005; see also Waechter, Nelson, Wright, Hyatt, & Oakman, 2014) assessed AB in an unselected sample of students using different versions of

CONTACT Bram Van Bockstaele  Bram.Vanbockstaele@gmail.com

 Supplemental data for this article can be accessed <https://doi.org/10.1080/02699931.2019.1609423>.

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

the VPT. He found simple split-half correlations of AB scores ranging between $-.16$ and $.19$, test-retest correlations ranging between $-.22$ and $.32$, and correlations with trait anxiety ranging between $-.13$ and $.26$. These findings led Schmukle to conclude that the VPT is an unreliable measure of AB in non-clinical samples.

A relatively well-known alternative for the VPT is the *relevant feature* visual search task (RFVST, e.g. Öhman, Flykt, & Esteves, 2001). In this task, participants are required to find a specific target stimulus in an array of distracting stimuli. By varying the threat value of targets and distractors, AB in this task is typically inferred from faster RTs on trials with a threatening target embedded within an array of non-threatening distractors compared to trials with a non-threatening target embedded within an array of threatening distractors. Although the psychometric properties of the RFVST have not yet been systematically assessed, Van Bockstaele, Saleminck, Bögels, and Wiers (2017) found relatively high split-half reliabilities of $.43$ and $.59$ for AB scores in an RFVST.

Dodd, Vogt, Turkileri, and Notebaert (2017) proposed a variant of the RFVST. In their *irrelevant feature* visual search task (IFVST), participants were required to find either a young or old face in an array of middle-aged faces. Crucially, on some trials, the target face displayed either a happy or an angry expression, while the distractors were always neutral. As participants were required to respond to the age-group of the target faces, the emotional valence of the target stimuli was task-irrelevant. While the authors did not report the reliability of the AB scores, they found a medium sized significant positive correlation between trait anxiety and AB in the IFVST, defined as the RT difference between trials with angry targets and trials with happy targets. In contrast, the relationship between trait anxiety and AB scores in an RFVST was not significant.

Another task that has shown promise for the measurement of AB is the emotional flanker task (EFT). Developed in the context drinking behaviours, Nikolaou, Field, and Duka (2013) asked participants to respond to the direction of a central arrow while ignoring the directions of flanking arrows. The central and flanking arrows pointed either in the same direction (congruent trials) or in different directions (incongruent trials). Crucially, the arrows were superimposed on either alcohol-related (e.g. beer can) or neutral (e.g. office stationary) task-irrelevant background pictures. By subtracting the flanker effect (i.e. the RT difference between congruent and

incongruent trials) of trials with alcohol backgrounds from the flanker effect of trials with neutral backgrounds, they found a medium to large positive correlation between the resulting alcohol AB index and weekly alcohol consumption. The EFT has not been used previously in anxiety research, and the reliability of its AB index is unknown.

Many researchers have argued that the psychometric limitations of AB measures are one of the main challenges for AB research (e.g. Evans, Walukevich, Seager, & Britton, 2018; McNally, 2018; Rodebaugh et al., 2016; Van Bockstaele et al., 2014; Waechter & Stolz, 2015). The poor reliability of measures likely accounts for the wealth of diverging findings and non-replications, thus adding to the inconsistencies in the field. We set out to add some clarity by systematically assessing the psychometric properties of different AB measures in the context of anxiety. In Experiment 1, we assessed the reliability, the convergent validity, and the concurrent validity of the VPT, the RFVST, the IFVST, and the EFT. We anticipated poor reliability estimates for the VPT (Schmukle, 2005), relatively high reliability scores for the RFVST (Van Bockstaele et al., 2017), while for the IFVST and the EFT we had no *a priori* expectations. Provided adequate reliability, we expected the tasks to show convergent validity, demonstrated by significant positive correlations between the AB scores. Finally, we expected positive correlations between the AB indices and self-reported measures of social anxiety.

Experiment 1

Method

Participants

Sixty-six students from the University of Amsterdam (53 women, M age = 23.55, SD = 8.57) participated in the experiment in exchange for course credits or cash. A post-hoc power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) showed that a sample size of 67 was needed to detect medium-sized ($r = .30$) one-tailed correlations with a power of $.80$. In order to avoid the problem of range restriction in our correlational approach, we tested an unselected sample displaying a wide range of anxiety levels (see Data Preparation, Scoring, and Outliers).

Questionnaires

Fear of Negative Evaluation Scale. The brief Fear of Negative Evaluation Scale (FNES; Leary, 1983) consists

of 12 statements regarding social evaluation. Participants responded on 5-point Likert scales ranging from 0 to 4. Cronbach's alpha in the present experiment was .97.

Social Interaction Anxiety Scale. The Social Interaction Anxiety Scale (SIAS; Mattick & Clarke, 1998) assesses anxiety in different social contexts. It consists of 20 statements, scored on a 5-point scale ranging from 0 to 4. Cronbach's alpha in the present experiment was .91.

Social Phobia Scale. The Social Phobia Scale (SPS; Mattick & Clarke, 1998) consists of 20 statements concerning fear of being scrutinised during routine activities such as eating or drinking. Participants responded on 5-point scales ranging from 0 to 4. Cronbach's alpha in the present experiment was .89.

Materials

The stimuli used in the AB tasks were the same pictures of neutral and angry faces as the ones used by Dodd et al. (2017). These faces were selected from the FACES database (Ebner, Riediger, & Lindenberger, 2010). A validation study demonstrated that people can, on average, accurately infer both the age and emotional expression of the faces in this database (Ebner et al.). The same 96 pictures (angry and neutral facial expression from 24 male and 24 female actors with each gender subset consisting of 8 young, 8 middle-aged, and 8 old actors) were used in all tasks. For the practice blocks in the VPT, the IFVST, and the EFT, we selected an additional 20 neutral faces of mixed gender and age from the FACES database. For the practice blocks in the RFVST we selected 4 angry and 4 happy facial expressions from the Karolinska Directed Emotional Faces database (KDEF; Lundqvist, Flykt, & Öhman, 1998). Pictures were presented in greyscale and hair and ears were cropped using an oval template.

Visual probe task

Each trial started with a 500 ms presentation of a white fixation cross flanked by two 3.9×5.4 cm (visual angle = $3.72^\circ \times 5.15^\circ$)¹ white rectangles against a black background. The distance between the fixation cross and the centre of the white rectangles was 5.3 cm (5.06°). Next, the white rectangles were replaced by one angry and one neutral face picture of the same actor that remained on the screen for 500 ms until they were masked by the white rectangles. After 20 ms, the target appeared in the centre of one of the two white rectangles. The

target was a 0.4×0.4 cm ($0.29^\circ \times 0.29^\circ$) black dot, and participants were required to respond as fast and as accurately as possible to the target location by pressing the A- or the L-key on a standard QWERTY keyboard. The screen was erased upon responding and the next trial started 500 ms later.

Participants completed two identical test blocks, each consisting 48 congruent (target on the angry face location) and 48 incongruent (target on the neutral face location) trials. The location of the angry versus the neutral face was randomised across trials, as were the target location and the identities of the actors. All face pairs were presented equally often, and angry faces, neutral faces, and targets appeared equally often on the left and the right location. Prior to the test phase, participants completed a practice block consisting of 8 trials with only neutral faces and error feedback.

Relevant feature visual search task

Each trial started with the presentation of a black fixation cross on a white background. After 500 ms, 8 unique face pictures (3.9×5.4 cm; $3.72^\circ \times 5.15^\circ$) of varying ages and genders appeared on the screen in a 3×3 rectangular grid (15.5×22 cm; $14.72^\circ \times 20.78^\circ$) with the middle position always empty. All faces were presented equally often, and target faces appeared equally often in any of the eight possible locations. The task consisted of two test blocks that were presented in a counterbalanced order across participants. On each trial of the "find angry" block, consisting of 48 trials, the grid included 7 neutral faces and a single angry face, and participants were asked to click as fast and as accurately as possible on the angry face. The "find neutral" block also consisted of 48 trials. On each trial, the grid included 7 angry faces and a single neutral face, and participants were required to click as fast and as accurately as possible on the neutral face. As soon as a response was registered, the screen was erased and the next trial started 500 ms later. Each test block was preceded by an appropriate practice block consisting of 8 trials with error feedback.

Irrelevant feature visual search task

The general appearance and trial timing of the IFVST was identical to the RFVST. However, on each trial, 7 of the 8 presented faces were of middle-aged people with a neutral facial expression, while the single target face was either of a young or an old person. Participants were required to click as fast and as accurately

as possible on the non-middle-aged face. The target face was equally often young or old and was presented equally often in each of the 8 possible positions. Crucially, on one third of the trials the target face had an angry facial expression, while on the remaining two thirds of the trials it had a neutral facial expression. The target was neutral in the majority of the trials to deter participants from adopting an emotion-driven search strategy. The task consisted of 4 identical test blocks of 48 trials each. Prior to the test phase, participants completed a brief practice block consisting of 8 trials containing only neutral faces.

Emotional flanker task

Each trial started with the presentation of a white fixation cross on a black background. After 800 ms, the fixation cross was replaced by a stimulus display consisting of an arrow configuration superimposed on a background picture. The arrow configuration consisted of 5 arrows with the central and flanking arrows either pointing in the same direction (congruent trials: “>>>>>” or “<<<<<”) or in opposite directions (incongruent trials: “>><<>>” or “<<>><<”). Participants were asked to respond as quickly and as accurately as possible to the direction of the middle arrow by pressing the A- or the L-key on a standard QWERTY keyboard. Trials ended as soon as a response was registered or if a response deadline of 1750ms had passed. The inter-trial interval varied randomly between 350, 500, and 650 ms.

The arrow configurations were superimposed on background pictures (10.6 × 15.0 cm; 10.10° × 14.25°), consisting of an angry facial expression, a neutral facial expression, or a control white rectangle. Each type of background was presented equally often. Facial expressions varied randomly across age and gender, and each face was presented equally often. Participants completed two blocks of 144 trials each, with as many congruent as incongruent trials. The trials with white rectangle backgrounds were inserted to reduce the effects of habituation to the faces. Prior to the first test block, participants completed a practice block of 12 trials with error feedback (6 congruent, 6 incongruent). Eight of the practice trials contained neutral face backgrounds and 4 trials contained the white rectangle background.

Procedure

The experiment was programmed and presented using Inquisit 4 (2014). Participants were informed about the general nature of the tasks before providing

written informed consent. The experiment was conducted in a quiet lab with a maximum of four participants at the same time. Participants first completed the questionnaires in the order described above. The order of the AB tasks was counterbalanced across participants. After completing the tasks, participants were debriefed and reimbursed. The entire procedure lasted 60 minutes and the study was approved by the ethical committee of the University of Amsterdam.

Results

Data Preparation, scoring, and outliers

The same overall outlier analysis was used for the four AB tasks. First, we removed practice trials and we calculated error percentages. Next, we removed errors (VPT = 2.57%, RFVST = 8.22%, IFVST = 22.39%, EFT = 6.63%) and trials with outlying RTs (VPT = 6.19%, RFVST = 7.41%, IFVST = 6.96%, EFT = 5.24%) using the median absolute deviation procedure described by Leys, Ley, Klein, Bernard, and Licata (2013), with the moderately conservative threshold of 2.5.² From the remaining data, we calculated AB scores for each task separately so that higher scores reflect a stronger AB towards angry faces. Thus, for the VPT, AB scores were calculated by subtracting average RTs on congruent trials from average RTs on incongruent trials. For the RFVST, we subtracted the average RT in the “find angry” block from the average RT in the “find neutral” block. For the IFVST, we subtracted the average RTs on angry face target trials from the average RT on neutral face target trials (irrespective of target age). For the EFT, we subtracted the congruency effect of the neutral face trials from the congruency effect of angry face trials. Finally, indicating a lack of motivation or misunderstanding the task instructions, we set AB scores of specific tasks missing if participants scored at chance level or below on any trial type in a task. This led to the removal of the IFVST data of 6 participants.³

Stable reliability indices for the four tasks were calculated using a Monte Carlo simulation process similar to the one used by Enock, Hofmann, and McNally (2014). For each task and each participant, the algorithm first randomly split the data in two halves and calculated individual AB scores for each half. Next, the correlation between these two AB scores was calculated. This process was repeated for 2000 iterations, and our final reliability estimate is the average of those 2000 split-half correlations. In addition, because the

full tasks were twice as long as the split-half tasks, we corrected these split-half correlations for test length using the Spearman-Brown prophecy formula.

Reflecting our recruitment of unselected students, our sample showed a wide range of social anxiety levels (FNES: $M = 18.55$, $SD = 12.24$, range = 0–48; SIAS: $M = 19.24$, $SD = 12.23$, range = 1–56; SPS: $M = 13.42$, $SD = 10.14$, range = 0–48). Scores on the questionnaires were highly correlated (r s between .62 and .82). We created a single all-encompassing social anxiety index by standardising the scores from each of the three questionnaires and computing the mean of these three standardised values. These averaged standardised scores as well as three of the AB scores were not normally distributed. Hence, all validity estimates are based on one-tailed Spearman correlation coefficients.

Reliability, convergent validity, and concurrent validity of Attentional bias measures

The results of our key analyses on the reliability and validity of the AB measures are presented in Table 1. The diagonal of this table shows that the reliability estimates of the AB scores in the VPT and EFT were very low. The Spearman-Brown corrected reliability of the AB score in the RFVST could be considered acceptable when dealing with psychological constructs (Kline, 1999), but the reliability of the AB score in the IFVST was poor. All of the convergent validity correlations were small to medium (all ρ s between .025 and .312) and not significant after

correcting for multiple comparisons, suggesting that the four AB estimates had little in common. Finally, none of the AB estimates correlated significantly with self-reported social anxiety, questioning the linear relation between AB and anxiety.

Post-Hoc group comparisons

Comparing AB scores between high and low socially anxious participants, we used both median and tertile splits on our social anxiety index. Independent samples t -tests revealed no significant differences in any of the AB scores between high and low anxious participants, all t s < 1.76 , all p s $> .08$. Separate one-way ANOVAs comparing AB scores between high, medium, and low anxious groups also revealed no significant group differences after correcting for multiple comparisons, all F s < 3.26 , all p s $> .04$. In sum, we found no evidence in either of the tasks for a larger AB towards threat in high versus low socially anxious individuals.

Discussion

The RFVST and IFVST showed acceptable and poor internal consistency, respectively, and the EFT and the VPT showed very little internal consistency. The poor reliability of the VPT is in line with previous reports (Schmukle, 2005). Given their poor reliabilities, it is also not surprising that the AB scores of the VPT and EFT did not correlate significantly with other measures of AB or with social anxiety. Given that the

Table 1. Descriptive statistics, mean reliability estimates, convergent validity, and concurrent validity of the attentional bias measures from Experiment 1.

	VPT	RFVST	IFVST	EFT	Mean standardised social anxiety
Mean (SD)	0 (6)	388 (411)	233 (300)	2 (19)	0.00 (0.90)
VPT	–.082 (.104) [–0.179]	.025 ($p = .420$)	.047 ($p = .360$)	.046 ($p = .355$)	.272 ($p = .014$)
RFVST		.557 (.081) [.715]	.312 ($p = .008$)	.250 ($p = .021$)	–.014 ($p = .455$)
IFVST			.306 (.097) [.469]	.131 ($p = .159$)	.018 ($p = .444$)
EFT				–.100 (.098) [–.222]	.077 ($p = .269$)

Notes: VPT = Visual Probe Task, RFVST = Relevant Feature Visual Search Task, IFVST = Irrelevant Feature Visual Search Task, EFT = Emotional Flanker Task. The first row contains the mean scores of each variable and their corresponding standard deviations. Split half reliability estimates, corresponding standard deviations (between round brackets), and Spearman-Brown corrected reliability estimates [between square brackets] are shown in the correlation diagonal. Convergent validity estimates, shown above the reliability diagonal, and concurrent validity estimates, shown in the outer right column, are based on one-tailed Spearman correlation coefficients, with p -values between round brackets. No validity correlations are significant after correction for multiple comparisons using the procedure described by Benjamini and Hochberg (1995) with the false discovery rate set at 5%.

RFVST and IFVST AB estimates showed better internal consistency, the nonsignificant convergent validity correlation (after correcting for multiple comparisons) suggests that these two search tasks measure different constructs. The RFVST measures AB in both a top-down and a bottom-up manner, as participants' task is to find a specific emotion while ignoring other emotions. In the IFVST, AB is operationalised in a bottom-up manner, as the emotional expressions of the faces should be ignored in order to focus on the relevant age dimension. Furthermore, AB in the RFVST was defined as the difference in RT between finding neutral faces *in angry arrays* and finding angry faces in neutral arrays, while in the IFVST, AB was defined as the difference in RTs between finding neutral faces *in neutral arrays* and finding angry faces in neutral arrays. This different operationalisation may have contributed to the lack of correlation between the two search tasks. More worrisome is the lack of significant correlations between the visual search AB scores and social anxiety, suggesting that there is no linear relation between AB and self-reported anxiety.

However, the data are subject to several limitations: The IFVST proved quite difficult, old faces are typically perceived as more negative and young faces as more positive (Dodd et al., 2017; Ebner, 2008), emotions of older faces are less reliably identified (Ebner et al., 2010), and we only assessed social anxiety through self-report. Therefore, we conducted a second experiment in which we aimed to replicate and extend our findings. Given their poor reliabilities, we omitted the VPT and the EFT. Attempting to make the IFVST easier and avoiding the possible confound between face age and valence, we changed the irrelevant feature to gender. Because neutral faces may have been perceived as ambivalent, we changed the neutral faces to happy faces (see also Dodd et al., 2017). Finally, to assess social anxiety in a more comprehensive manner, we also included physiological and implicit measures of social anxiety. We expected to replicate the reliability estimates of both visual search tasks, and we expected to find significant positive correlations between AB scores and social anxiety measures.

Experiment 2

Method

Participants

Sixty unselected students (44 women, M age = 26.35, SD = 10.73) participated in exchange for course

credits or cash. A post-hoc power analysis showed that a sample size of 67 would have been needed to detect medium-sized ($r = .30$) one-tailed correlations with a power of .80.

Questionnaires

Fear of Negative Evaluation Scale. Self-reported social anxiety was assessed using the FNES. Cronbach's alpha was .97.

Materials

We retained the young and middle-aged actors from Experiment 1, and we replaced the old actors by 4 young (2 men and 2 women) and 4 middle-aged (2 men and 2 women) actors from the FACES database. These additional stimuli were selected arbitrarily and not based on a systematic analysis of picture ratings. The same actors were selected once with an angry expression and once with a happy expression, and they were used in both the IFVST and the RFVST. For the practice blocks, we selected 2 men and 2 women, both once angry and once happy, from the KDEF. Pictures were presented in greyscale and hair and ears were cropped using an oval template.

Relevant feature visual search task

The RFVST was identical to the one that we used in Experiment 1, apart from using happy rather than neutral faces. In addition, to reduce between-subject variations in AB due to specific task demands, we no longer counterbalanced the block order: Participants were first instructed to find an *angry* face in a *happy* crowd and next to find a *happy* face in an *angry* crowd.

Irrelevant feature visual search task

The general appearance of the IFVST was identical to the one used in Experiment 1. In six alternating blocks of 48 trials each, participants were instructed to either find a male face in an array of female faces or a female face in an array of male faces. Each block contained 16 trials with an angry target in an angry array, 16 trials with a happy target in a happy array, 8 trials with an angry target in a happy array, and 8 trials with a happy target in an angry array. Trials were presented in a random order, and targets appeared equally often on each of the 8 possible locations. Half of the participants started with a block with female face target, while the other half started with a block with male target faces. The first block of each task (find female versus find male) was preceded by an 8-trial practice block.

Social anxiety identity implicit association test

Assessing implicit associations between self-concept and anxiety, we included a social anxiety identity Implicit Association Test (IAT: Greenwald, McGhee, & Schwartz, 1998). In this task, participants were required to sort words, presented in the centre of the screen, as quickly as possible using the left or right arrow keys of the keyboard. All words were presented in black on a white background. The inter-trial interval was 350 ms. The relevant response labels (ANXIOUS, CALM, SELF, and NOT SELF; see below) were presented in the top left and top right corners of the screen and remained on the screen for the entire duration of each block. Implicit social anxiety is inferred from differences in RTs between blocks where self is paired with anxious and not-self with calm versus blocks where self is paired with calm and not-self with anxious.

In the first block, consisting of 20 trials, participants practiced the attribute category by sorting words referring to anxiety (Dutch translations of “afraid”, “nervous”, “ashamed”, “criticized”, and “insecure”) or calmness (Dutch translations of “calm”, “relaxed”, “accepted”, “carefree”, and “secure”). Each word was presented twice. In the second block, also consisting of 20 trials with each word presented twice, participants practiced the target category by sorting words referring to themselves (Dutch translations of “I”, “me”, “my”, “myself”, and “own”) or not-self (Dutch translations of “themselves”, “they”, “their”, “others”, and “you”). In the third block, participants practiced the combinations of self + anxious and not-self + calm. Each word was presented once, and the block consisted of 20 trials. The fourth block was identical to the third block, but consisted of 40 trials with each word presented twice. In the fifth block, the target category keys were switched: Participants practiced in 20 trials, with each word presented twice, to respond to words referring to self and not-self using the opposite response mapping as in the second block. In the sixth block, consisting of 20 trials with each word presented once, participants practiced the combinations of self + calm and not-self + anxious. Finally, the seventh block was identical to the sixth, but consisted of 40 trials with each word presented twice.

Social stress task

Prior to the social stress task, heart rate electrodes were attached. Participants were asked to avoid movements. All instructions appeared on the computer

screen. During the first 2 minutes, we only presented a fixation cross. The second of these 2 minutes was used as a baseline measure for heart rate and heart rate variability (HRV). After 2 minutes, a text appeared on the screen, informing participants that they had to perform a 5-minute speech about the (dis-)advantages of abortion, which would be videotaped and evaluated, and for which they had 2 minutes to prepare. There was a videorecorder in the room, clearly visible for participants. Participants were not allowed to take notes, and they were reminded to move as little as possible to avoid artefacts in the heart rate measurement. A 2-minute countdown clock started running as soon as the text appeared and was shown underneath the text. After 2 minutes, the text was replaced by another text informing participants that, based on their participant number, they did not have to give the speech. Instead, they were asked to remain seated for another 2 minutes without moving, and that after these 2 minutes the experiment ended.

ECG was measured using a custom made portable amplifier with a 1G Ω input resistance and a bandwidth of 0.1 Hz (6 dB/oct) to 250 Hz (24 dB/oct) containing a National Instruments NI-USB6210 A/D converter to digitise the analogue data at a rate of 1000 S/s. We used disposable pre-gelled Ag/AgCl 3M Red Dot electrodes to measure ECG in LEAD-II configuration.

Procedure

The entire procedure was implemented using Inquisit 4 (2014). Participants were informed about the nature of the stimuli and tasks before providing written informed consent. The experiment was conducted in a sound-proof cubicle, and only one participant was tested at a time. Participants first completed the FNES, followed by the RFVST, the IFVST, and the IAT, in this fixed order. After the IAT, the experimenter briefly entered the lab to attach the heart rate devices, after which the social stress task started as described above. Upon completion of the experiment, participants were debriefed and reimbursed. The entire procedure lasted for 60 minutes and was approved by the ethical committee of the University of Amsterdam.

Results

Data Preparation, outliers, and scoring

We used the same criteria to remove outliers as in Experiment 1. In brief, we removed practice trials,

errors (RFVST = 0.66%, IFVST = 9.45%), and trials with outlying RTs (RFVST = 5.70%, IFVST = 6.63%) using the same procedure as in Experiment 1.⁴ For the RFVST, AB was calculated as in Experiment 1, subtracting the average RT in the “find angry” block from the average RT in the “find happy” block. For the IFVST, we calculated 2 separate AB scores: First, mirroring the IFVST AB scores of Experiment 1, we subtracted average RTs to find an angry target in a happy array from the average RT to find a happy target in a happy array. Second, reflecting the operationalisation of AB in the RFVST, we subtracted the average RT to find angry targets in a happy array from the average RT to find happy targets in an angry crowd. We calculated the reliability of each of these AB indices using the same procedure as described for Experiment 1.

For the IAT, we calculated the D600 score (Greenwald, Nosek, & Banaji, 2003). The D600 includes RTs in blocks 3 and 6, error latencies are given a 600 ms penalty, and latencies are corrected for individual variability. Positive scores indicated a stronger implicit association of the self with anxiety, while negative scores indicated a stronger implicit association of the self with calmness. The odd-even split-half Spearman correlation of the D600 score was large, $\rho = .722$, $p < .001$.

Vsrp98 (2011) was used to detect R-tops from the ECG recording and to calculate heart rate and HRV (RMSSD, the root mean square of successive differences in inter-beat-intervals) allowing a maximum difference of $\pm 33\%$ in successive IBI length for HRV. ECGs were visually inspected and areas with poor signal and/or movement artefacts were manually removed prior to scoring. Data from two participants were not recorded and were set to missing. The data from three additional participants were set to missing because the signal was lost or became very noisy during the baseline or shortly after the baseline measurement, leaving not enough data to measure heart rate during the crucial stress and recovery phases. Finally, for one participant whose signal was lost shortly after starting the recovery phase, we retained the data of the stress phase and set the data of the recovery phase to missing. Heart rate variables were calculated for five 1-minute windows: The baseline minute, the first and second stress minute, and the first and second recovery minute. Illustrating the internal consistency of the heart rate measurements, Spearman correlation coefficients between the first and second minute of each phase were large and significant (heart rate stress phase: $\rho = .90$;

heart rate recovery phase: $\rho = .93$; HRV stress phase: $\rho = .81$; HRV recovery phase: $\rho = .94$; all $ps < .001$). For the concurrent validity analyses, we calculated two change scores: A stress score by subtracting the baseline minute from the mean of the two stress minutes and a recovery score by subtracting the mean of the two recovery minutes from the mean of the two stress minutes.

As in Experiment 1, our recruitment of unselected students resulted in a wide range of social anxiety levels (FNES: $M = 19.03$, $SD = 12.31$, range = 0–48). The scores on the FNES, IFVST, and all the heart rate scores were not normally distributed. Consequently, we used Spearman correlation coefficients for all validity estimates.

Reliability, convergent validity, and concurrent validity of Attentional bias measures

The results of our key analyses on the reliability and validity of the different AB measures are presented in Table 2. As in Experiment 1, the reliability of the RFVST was high, yet after correcting for multiple comparisons, it did not correlate significantly with AB scores of the IFVST or with any of the social anxiety measures. In the IFVST, when AB was operationalised in a similar way to Experiment 1 (happy target in happy crowd minus angry target in happy crowd), the reliability score was near zero, thus failing to replicate the findings of Experiment 1. Consequently, the validity indices were small and non-significant after correcting for multiple comparisons. When IFVST AB was operationalised as in the RFVST (happy target in angry crowd minus angry target in happy crowd), the reliability score was still poor, it was significantly associated with the IAT scores, but after correcting for multiple comparisons not with other social anxiety measures.

Post-Hoc group comparisons and supplementary analyses

Independent samples *t*-tests revealed no AB differences between FNES-based median split high and low anxious participants, all $ts < 1.99$, all $ps > .05$. Separate one-way ANOVAs comparing high, medium, and low anxious groups based on FNES scores also revealed no significant group differences on any of the AB indices, all $Fs < 2.89$, all $ps > .06$. In sum, we found no evidence for a larger AB towards angry faces in high versus low anxious individuals.

Validating the different measures of social anxiety, FNES scores were significantly correlated with

Table 2. Descriptive Statistics, Mean Reliability Estimates, Convergent Validity, and Concurrent Validity of the Visual Search Attentional Bias Measures from Experiment 2.

	RFVST	IFVST AB1	IFVST AB2	FNES	D600	Δ BPM stress	Δ BPM recovery	Δ HRV stress	Δ HRV recovery
Mean (<i>SD</i>)	−284 (214)	−22 (228)	−323 (304)	19.03 (12.31)	−0.31 (0.38)	7.23 (8.43)	−7.08 (8.10)	−10.10 (19.32)	8.15 (16.62)
RFVST	.650 (.079) [.788]	−.054 (<i>p</i> = .341)	.235 (<i>p</i> = .035)	−.029 (<i>p</i> = .413)	.118 (<i>p</i> = .185)	−.067 (<i>p</i> = .313)	.054 (<i>p</i> = .348)	−.143 (<i>p</i> = .149)	.103 (<i>p</i> = .229)
IFVST AB1		.046 (.112) [.088]	.558 (<i>p</i> = .000)*	−.107 (<i>p</i> = .209)	.223 (<i>p</i> = .043)	−.069 (<i>p</i> = .309)	.116 (<i>p</i> = .202)	−.097 (<i>p</i> = .240)	.051 (<i>p</i> = .358)
IFVST AB2			.209 (.106) [.346]	.229 (<i>p</i> = .039)	.346 (<i>p</i> = .003)*	.195 (<i>p</i> = .077)	−.155 (<i>p</i> = .132)	−.268 (<i>p</i> = .024)	.201 (<i>p</i> = .072)

Notes: RFVST = Relevant Feature Visual Search Task, IFVST = Irrelevant Feature Visual Search Task, AB1 = happy target in happy crowd minus angry target in happy crowd, AB2 = happy target in angry crowd minus angry target in happy crowd, FNES = Fear of Negative Evaluation Scale, D600 = Implicit Association Scores as described by Greenwald et al. (2003), Δ BPM stress = change in heart rate from baseline to stress induction, Δ BPM recovery = change in heart rate from stress induction to recovery, Δ HRV stress = change in heart rate variability from baseline to stress induction, Δ HRV recovery = change in heart rate variability from stress induction to recovery. The first row contains the mean scores of each variable and their corresponding standard deviations. Split half reliability estimates, corresponding standard deviations (between round brackets), and Spearman-Brown corrected reliability estimates [between square brackets] are shown in the correlation diagonal. Validity estimates are based on one-tailed Spearman correlation coefficients, with *p*-values between round brackets. Validity correlations marked with asterisks are significant after correction for multiple comparisons using the procedure described by Benjamini and Hochberg (1995) with the false discovery rate set at 5%.

changes in heart rate from baseline to stress ($p = .32$, $p < .01$) and with changes in heart rate and changes in HRV from stress to recovery ($p = -.38$, $p < .005$ and $p = .32$, $p < .01$, respectively) but not with changes in HRV from baseline to stress ($p = -.14$, $p = .15$). All the heart rate measurements were strongly interrelated, with the absolute values of all $ps > .43$, all $ps < .001$. Finally, IAT scores were not significantly correlated with any of the other measures, all $ps < .22$, all $ps > .05$. Comparing high and low socially anxious subgroups based on median split scores on the FNES, high anxious participants showed a stronger increase in heart rate following the stress induction ($M = 10.13$, $SD = 8.39$) and a stronger decrease in heart rate during recovery ($M = -9.94$, $SD = 8.50$) than low anxious participants ($M = 4.44$, $SD = 7.62$ and $M = -4.42$, $SD = 6.82$, respectively), both $ts > 2.63$, both $ps < .05$. On the IAT, there were no significant differences between high ($M = -0.23$, $SD = 0.39$) and low socially anxious participants ($M = -0.40$, $SD = 0.36$), $t < 1.70$, $p > .09$.

Discussion

In Experiment 2, we wanted to replicate the promising reliability indices of the visual search tasks and we wanted to assess social anxiety beyond self-report. We replicated the strong reliability of the RFVST, but RFVST AB scores were, again, not related to social anxiety measures. For the IFVST, we found that

internal consistency was near zero when the AB score was calculated in a similar way as in Experiment 1. Operationalising IFVST AB scores as the difference between happy targets in angry crowds and angry targets in happy crowds still gave a poor reliability estimate, this score correlated significantly with the IAT-score, but (after correcting for multiple comparisons) not with self-reported or physiological measures of social anxiety. We found no significant correlations between AB scores in the RFVST and the IFVST, suggesting that these tasks measure different constructs.

General discussion

In the present experiments, we assessed the reliability, convergent validity, and concurrent validity of different AB measures. Neither the VPT nor the EFT yielded reliable estimates of AB. The visual search tasks proved more promising, with consistently high reliability indices in the RFVST and less consistent and smaller reliability indices for the IFVST. In Experiment 2, the IFVST correlated with an implicit measure of social anxiety in the expected direction, while the RFVST did not correlate significantly with any of the social anxiety measures.

The poor reliability of the VPT does not come as a surprise. The results of our first experiment only confirm the low reliabilities also found by Schmukle (2005) and Waechter et al. (2014). Nevertheless,

given the surge of interest in Attentional Bias Modification (ABM) research, in which changes in AB are experimentally induced to influence responsiveness to stress, our finding is still relevant and timely. Many ABM studies rely on a modified version of the VPT to induce changes in AB or have used the VPT to assess these changes in AB (for reviews, see Mogg et al., 2017; Van Bockstaele et al., 2014). Similar to the AB literature, the overall pattern of results in ABM studies is marked by inconsistencies. Most studies in which the manipulation of AB proved successful also yielded significant changes in emotion-related outcomes (Clarke, Notebaert, & MacLeod, 2014). Our results indicate that ABM research can benefit from relying less on the VPT to assess changes in AB, as these estimates of AB are highly unreliable and thus inaccurate, at least as measures of individual differences.

Fundamentally however, our data present a cause for concern. While the RfVST yielded reliable AB scores, these scores were unrelated to measures of social anxiety. This poor validity of the RfVST is in line with previous studies (e.g. Dodd et al., 2017) and limits the applicability of the RfVST. Thus far, the IFVST seems to be a more valid measure of AB (see also Dodd et al.). However, in our present study, only one operationalisation of AB in the IFVST in Experiment 2 was correlated with an implicit measure of social anxiety and the reliability index of this AB score was poor. As such, the evidence for the validity and reliability of the IFVST remains limited and more research is needed to establish the best conditions to measure AB using this paradigm.

One remarkable finding concerns the apparent dissociation in the results of different anxiety measures: While an AB score in the IFVST correlated significantly with social anxiety as measured in the IAT, the same score did not correlate significantly with self-reported or physiological measures of anxiety. While implicit, physiological, and self-reported measures of anxiety can be expected to show substantial overlap because they are assumed to share the same underlying construct (i.e. social anxiety), this overlap is often far from perfect (e.g. Van Bockstaele et al., 2011). As mentioned in the supplementary analyses of Experiment 2, the correlations between the different social anxiety outcomes varied greatly in size and significance. Given this limited overlap in the outcome measures, it is not surprising that the pattern of correlations between attentional bias

measures and social anxiety measures differed from measure to measure.

A recent line of research has suggested to change the operationalisation of AB in the VPT, either in terms of RT variability (e.g. Iacoviello et al., 2014) or in terms of trial-level bias scores, in which different bias estimates are calculated from trial-by-trial comparisons (e.g. Zvielli, Bernstein, & Koster, 2015). These new indices of AB have been argued to be more reliable (Caudek, Ceccarini, & Sica, 2017; Rodebaugh et al., 2016), but this claim is also subject to debate (Kruijt, Field, & Fox, 2016). Although a full analysis of these new indices is beyond the scope of our article, the AB variability index (Iacoviello et al.) for the VPT data of Experiment 1 yielded a medium-sized simple split-half correlation ($\rho = .370$, $p = .001$) but it did not correlate significantly with social anxiety ($\rho = .031$, $p = .403$) or with any of the other AB scores (after correction for multiple comparisons), all p s between $-.156$ and $.223$, all p s between $.044$ and $.407$. It is also unclear whether these new indices are restricted to the VPT, or whether they can also be used in other tasks or in blocked test formats, like the RfVST. We therefore agree that these operationalizations hold promise and may move the field forward, but more research is needed to fully understand their potential and limitations.

Several limitations are worth mentioning. We only assessed AB in the social anxiety domain, using facial stimuli. It is possible that the psychometric properties of AB measures in other psychopathological domains using corresponding stimulus materials do not mirror our findings. In addition, AB measures may show higher reliability in more homogeneous anxious samples (but see Waechter et al., 2014). Finally, we limited ourselves to RT-based tasks. A recent model of RT-measures has raised fundamental concerns about the reliabilities and correlations of RT-based difference scores, like the AB scores that we used (Miller & Ulrich, 2013). According to this model, reliabilities of difference scores are compromised by large positive correlations between RTs on the different trial types (see also McNally, 2018). In the data of our present experiments, these correlations between the components of the difference scores were all larger than $.80$. Miller and Ulrich argued that in such cases, thousands of trials per condition may be required to reach acceptable levels of reliability. As for correlations involving RT difference scores (i.e. our validity estimates), Miller and Ulrich showed that these correlations are strongly affected by so many

different parameters that these correlations (or lack thereof) should only be interpreted with extreme caution (e.g. see also Waechter & Stolz, 2015). While findings of small or near-zero correlations may reflect true dissociations between two variables (e.g. AB for negative faces in the RFVST is *not* related to social anxiety), they may just as well reflect variations in other parameters that – unfortunately – cannot be estimated directly and thus cannot be controlled for. As such, the model of Miller and Ulrich poses a major challenge for RT-based AB research. One solution could be to focus on other outcome measures of AB, like eye-movements (e.g. Waechter et al., 2014) or electrophysiological measures (e.g. Wieser, Hambach, & Weymar, 2018). However, as these scores are also typically difference scores, they may be subject to the same limitations as RT-based scores. Alternatively, a regression-based approach deriving AB from multivariate residual scores (Evans et al., 2018) has been argued to counter some of the problems raised by the model of Miller and Ulrich, but even when using this alternative approach, the reliability of the VPT remained unacceptably low.

Keeping these limitations in mind, we found in two experiments that while the RFVST yielded reliable estimates of AB, these estimates did not correlate with measures of social anxiety. The IFVST yielded overall lower and less stable reliability estimates, and only one of the resulting AB scores correlated significantly with an implicit but not with explicit or physiological measures of social anxiety.

Notes

1. Visual angles were calculated using a viewing distance of 60 cm.
2. The general pattern of results remained the same using different outlier exclusion criteria, including Tukey's fences with $k=1.5$ and $k=3$, and the procedure based on $M \pm 3SDs$ described by Van Bockstaele et al. (2017). Full results with these different outlier analyses are provided in online supplementary Tables S1-S3.
3. Inclusion of these participants did again not change the general pattern of results. Full results of the entire sample with different outlier analyses are provided in online supplementary Tables S4-S7.
4. The general pattern of results remained the same using different outlier exclusion criteria, including Tukey's fences with $k=1.5$ and $k=3$, and the procedure based on $M \pm 3SDs$ described by Van Bockstaele et al. (2017). Full results with these different outlier analyses are provided in online supplementary Tables S8-S10. No participants scored at or below chance level on any of the trial types in both tasks.

Acknowledgements

Bram Van Bockstaele is a postdoctoral researcher of the Research Priority Area Yield of the University of Amsterdam.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Bram Van Bockstaele  <http://orcid.org/0000-0002-5860-6449>

Elske Saleminck  <http://orcid.org/0000-0002-6885-2598>

Reinout W. Wiers  <http://orcid.org/0000-0002-4312-9766>

Data availability statement

The original data that were used in Experiment 1 and Experiment 2 are accessible in the following Open Science Framework data deposit: osf.io/uzjk5 (doi: 10.17605/OSF.IO/UZJK5).

References

- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 289–300.
- Caudek, C., Ceccarini, F., & Sica, C. (2017). Facial expression movement enhances the measurement of temporal dynamics of attentional bias in the dot-probe task. *Behaviour Research and Therapy*, 95, 58–70. doi:10.1016/j.brat.2017.05.003
- Clarke, P. J. F., Notebaert, L., & MacLeod, C. (2014). Absence of evidence or evidence of absence: Reflecting on therapeutic implementations of attentional bias modification. *BMC Psychiatry*, 14, 8. doi:10.1186/1471-244X-14-8
- Dodd, H. F., Vogt, J., Turkileri, N., & Notebaert, L. (2017). Task relevance of emotional information affects anxiety-linked attention bias in visual search. *Biological Psychology*, 122, 13–20. doi:10.1016/j.biopsycho.2016.01.017
- Ebner, N. C. (2008). Age of face matters: Age-group differences in ratings of young and old faces. *Behavior Research Methods*, 40, 130–136. doi:10.3758/BRM.40.1.130
- Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES - A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, 42, 351–362. doi:10.3758/BRM.42.1.351
- 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, 200–216. doi:10.1007/s10608-014-9606-z
- Evans, T. C., Walukevich, K. A., Seager, I., & Britton, J. C. (2018). A psychometric comparison of anxiety-relevant attention measures. *Anxiety, Stress, and Coping*, 31, 539–554. doi:10.1080/10615806.2018.1489536

- 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. doi:10.3758/BF03193146
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464–1480. doi:10.1037/0022-3514.74.6.1464
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85, 197–216. doi:10.1037/0022-3514.85.2.197
- Iacoviello, B. M., Wu, G., Abend, R., Murrough, J. W., Feder, A., Fruchter, E., ... Charney, D. S. (2014). Attention bias variability and symptoms of posttraumatic stress disorder. *Journal of Traumatic Stress*, 27, 232–239. doi:10.1002/jts.21899
- Inquisit 4. (2014). [Computer software]. Retrieved from https://www.millisecond.com..
- Kline, P. (1999). *The handbook of psychological testing* (2nd ed.). London: Routledge.
- Kruijt, A.-W., Field, A. P., & Fox, E. (2016). Capturing dynamics of biased attention: Are new attention variability measures the way forward? *PLoS ONE*, 11, e0166600. doi:10.1371/journal.pone.0166600
- Leary, M. R. (1983). A brief version of the Fear of Negative Evaluation Scale. *Personality and Social Psychology Bulletin*, 9, 371–375. doi:10.1177/0146167283093007
- Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49, 764–766. doi:10.1016/j.jesp.2013.03.013
- Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces - KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet.
- MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology*, 95, 15–20. doi:10.1037/0021-843X.95.1.15
- Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*, 36, 455–470. doi:10.1016/S0005-7967(97)10031-6
- McNally, R. J. (2018). Attentional bias for threat: Crisis or opportunity? *Clinical Psychology Review*, 69, 4–13.
- Miller, J., & Ulrich, R. (2013). Mental chronometry and individual differences: Modeling reliabilities and correlations of reaction time means and effect sizes. *Psychonomic Bulletin & Review*, 20, 819–858. doi:10.3758/s13423-013-0404-5
- Mogg, K., & Bradley, B. P. (2016). Anxiety and attention to threat: Cognitive mechanisms and treatment with attention bias modification. *Behaviour Research and Therapy*, 87, 76–108. doi:10.1016/j.brat.2016.08.001
- Mogg, K., Waters, A. M., & Bradley, B. P. (2017). Attention bias modification (ABM): Review of effects of multisession ABM training on anxiety and threat-related attention in high-anxious individuals. *Clinical Psychological Science*, 5, 698–717. doi:10.1177/2167702617696359
- Nikolaou, K., Field, M., & Duka, T. (2013). Alcohol-related cues reduce cognitive control in social drinkers. *Behavioural Pharmacology*, 24, 29–36. doi:10.1097/FBP.0b013e32835cf458
- Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General*, 130, 466–478. doi:10.1037/0096-3445.130.3.466
- Rodebaugh, T. L., Scullin, R. B., Langer, J. K., Dixon, D. J., Huppert, J. D., Bernstein, A., ... Lenze, E. J. (2016). Unreliability as a threat to understanding psychopathology: The cautionary tale of attentional bias. *Journal of Abnormal Psychology*, 125, 840–851. doi:10.1037/abn0000184
- Schmukle, S. C. (2005). Unreliability of the dot probe task. *European Journal of Personality*, 19, 595–605. doi:10.1002/per.554
- Van Bockstaele, B., Salemink, E., Bögels, S. M., & Wiers, R. W. (2017). Limited generalisation of changes in attentional bias following attentional bias modification with the visual probe task. *Cognition and Emotion*, 31, 369–376. doi:10.1080/02699931.2015.1092418
- Van Bockstaele, B., Verschuere, B., Koster, E. H. W., 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, 166–174. doi:10.1016/j.ijpsycho.2010.10.003
- Van Bockstaele, B., Verschuere, B., Tibboel, H., De Houwer, J., Crombez, G., & Koster, E. H. W. (2014). A review of current evidence for the causal impact of attentional bias on fear and anxiety. *Psychological Bulletin*, 140, 682–721. doi:10.1037/a0034834
- Vsrrp98 (v10.4). (2011). [Computer software]. Amsterdam, The Netherlands: University of Amsterdam.
- Waechter, S., Nelson, A. L., Wright, C., Hyatt, A., & Oakman, J. (2014). Measuring attentional biases to threat: Reliability of dot probe and eye movement indices. *Cognitive Therapy and Research*, 38, 313–333. doi:10.1007/s10608-013-9588-2
- Waechter, S., & Stolz, J. A. (2015). Trait anxiety, state anxiety, and attentional bias to threat: Assessing the psychometric properties of response time measures. *Cognitive Therapy and Research*, 39, 441–458. doi:10.1007/s10608-015-9670-z
- Wieser, M. J., Hambach, A., & Weymar, M. (2018). Neurophysiological correlates of attentional bias for emotional faces in socially anxious individuals: Evidence from a visual search task and N2pc. *Biological Psychology*, 132, 192–201. doi:10.1016/j.biopsycho.2018.01.004
- Zvielli, A., Bernstein, A., & Koster, E. H. (2015). Temporal dynamics of attentional bias. *Clinical Psychological Science*, 3, 772–788. doi:10.1177/2167702614551572