Test-retest reliability of attention bias for food

Appetite 136 (2019) 86–92

Contents lists available at ScienceDirect

Appetite

journal homepage: www.elsevier.com/locate/appet

Test-retest reliability of attention bias for food: Robust eye-tracking and reaction time indices

Welmoed van Ens[a], Ulrike Schmidt[a], Iain C. Campbell[a], Anne Roefs[b], Jessica Werthmann[a,c,*]

[a] Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK
[b] Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, the Netherlands
[c] Clinical Psychology and Psychotherapy, Institute of Psychology, Albert-Ludwigs-Universität Freiburg, Germany

ARTICLE INFO

Keywords:
Attention bias
Eye-tracking
Reaction times
Visual dot probe task
Food

ABSTRACT

This study assessed internal reliability and test-retest reliability of attention bias scores for food derived from the dot probe task. A visual dot probe task with food and non-food pictures (presented for 3000 ms) was administered to 53 healthy women on two occasions. Attention bias scores for food were calculated based on manual response latencies (reaction time bias) and concurrent assessment of eye-movements (direction bias and dwell-time bias). Subjective hunger and blood glucose levels were measured on both testing occasions. Dietary restraint and eating disorder symptoms were assessed during the second session. Results showed that direction bias had poor internal and test-retest reliability. Dwell time bias had excellent internal and acceptable test-retest reliability. Reaction time bias had acceptable internal and good test-retest reliability. Exploratory correlational analyses found that hunger, blood glucose, dietary restraint and eating disorder symptoms were not consistently significantly correlated with indices of attention bias for food. Overall, these findings contradict previous studies that reported low reliability of attention bias indices derived from the visual dot probe task. The implications are that a longer stimulus presentation time (i.e. ≥ 3000 ms), the use of eye-tracking and the use of appetizing stimuli can yield reliable attention bias scores for food. However, the interpretation of dot-probe scores of attention bias for food based on a dot probe task with 3000 ms presentation time and the score's relationship to theoretically relevant constructs such as hunger, eating restraint and eating disorder symptoms, require further clarification.

Attention bias (AB) is a tendency to preferentially allocate attention to personally, motivationally, and emotionally relevant stimuli (Kuckertz & Amir, 2015). MacLeod, Mathews, and Tata (1986) were the first to develop a dot-probe task paradigm, which assessed AB. The dot-probe task has become a commonly used measure of AB (Kappenman, MacNamara, & Proudfit, 2013). However, it has also been deemed a problematic instrument (Rodebaugh et al., 2016). The internal and test-retest reliability of reaction time bias scores for threatening stimuli obtained using the dot-probe task are extremely poor[1] (Cooper et al., 2011; Schmukle, 2005; Staugaard, 2009). Poor reliability has been identified as one of the central methodological and theoretical challenges to the AB field (Doolan, Breslin, Hanna, & Gallagher, 2014; Rodebaugh et al., 2016; Schmukle, 2005).

AB in the dot-probe task is calculated by subtracting response times to congruent trials from those to incongruent trials (MacLeod et al., 1986). A reliable difference score can only be obtained if the two raw scores have exceptionally good reliability (Rodebaugh et al., 2016). In the case of AB, Waechter, Nelson, Wright, Hyatt, and Oakman (2014) suggested that individual characteristics such as processing speed may have a large influence on raw reaction times (Waechter et al., 2014). Moreover, the difference in reaction times to neutral and salient stimuli in critical trials within an individual is typically small, making it difficult to detect reliably. This may mean that individual differences in processing speed may be easier to detect than differences in attention processing of neutral versus salient stimuli (i.e. the bias itself). In addition, the reaction time based index of AB can only provide us with a "snap-shot view" of attention, as this indirect measure is based on presumed attentional focus at the end of the stimulus presentation, right before the probe is presented (Mogg, Bradley, Field, & De Houwer, 2003; Wald et al., 2011).

* Corresponding author. University of Freiburg, Institute of Psychology,Department of Clinical Psychology and Psychotherapy, Engelbergerstr. 41, 79106, Freiburg, Germany.

E-mail address: Werthmann@psychologie.uni-freiburg.de (J. Werthmann).

[1] The current paper reports all reliability estimates of < 0.70 as poor, > 0.70 as acceptable, > 0.80 as good and > 0.90 as excellent, which is consistent with the way reliability is reported in the work cited.

https://doi.org/10.1016/j.appet.2019.01.020
Received 18 November 2018; Received in revised form 20 January 2019; Accepted 22 January 2019
Available online 22 January 2019
0195-6663/

In the field of anxiety and addiction, several ways of improving the reliability of the dot-probe paradigm have been explored. It has been argued that the reliability of the dot probe task can be improved using eye-tracking (Christiansen, Mansfield, Duckworth, Field, & Jones, 2015; Waechter et al., 2014). Eye-tracking provides information on concurrent attentional focus during the stimulus presentation itself. It can therefore be regarded as a more direct measure of attention allocation. AB indices derived from eye-tracking measures are somewhat more reliable than traditional reaction time bias scores, but their reliability is still not always acceptable (Christiansen et al., 2015; Price et al., 2015). A careful selection of stimuli, such as personalized stimuli, which match the preferences of the participant, may further improve internal reliability (Christiansen et al., 2015). Another way to improve reliability is prolonging stimulus presentation times. Previous research suggested that the internal reliability of eye-tracking bias indices (i.e. dwell time) is good when recorded over an entire 5 s segment, but poor when assessed over intervals of 1 s or less (Waechter et al., 2014).

The dot-probe task is also used to study AB for food. Heightened AB for food has been linked to disordered eating behavior, dietary restraint, and obesity, but these findings are inconsistent (Brooks, Prince, Stahl, Campbell, & Treasure, 2011; Castellanos et al., 2009; Doolan et al., 2014; Field et al., 2016; Giel et al., 2011; Giel et al., 2011; Werthmann, Jansen, & Roefs, 2014, 2016). Findings on the relationship between AB for food and hunger are also mixed (Castellanos et al., 2009; Loeber, Grosshans, Herpertz, Kiefer, & Herpertz, 2013; Mogg, Bradley, Hyare, & Lee, 1998; Papies, Stroebe, & Aarts, 2008; Werthmann, Jansen, & Roefs, 2016). It is not known whether bias scores for food suffer from a similar lack of reliability as those for threat and addiction related stimuli, as this has not been assessed. Insufficient reliability could explain the inconsistent results in the field, because instruments that produce unreliable scores are likely to yield false positives and chance findings (Rodebaugh et al., 2016; Schmukle, 2005). To determine whether modifications to the paradigm are needed, it is therefore critical to first establish whether AB scores for food also suffer from a lack of reliability.

1. Current study

The primary aim of the current study was to assess the internal reliability and the test-retest reliability of AB scores for pictures of food based on reaction time bias and two eye-tracking indices (i.e. direction bias and dwell time bias). Direction bias and dwell time bias were chosen because they are the most commonly used eye-tracking indices in attention bias research, representing early orientation of attention and maintained attention, respectively (Doolan et al., 2014; Field et al., 2016; Werthmann et al., 2014). We hypothesized that dwell-time bias, which is based on the total amount of time participants fixated on food versus non-food stimuli, would yield acceptable internal reliability. This hypothesis was based on research, which demonstrated that prolonged stimulus presentation time and the use of eye-tracking improve internal reliability (Christiansen et al., 2015; Waechter et al., 2014). On the other hand, direction bias scores and reaction time bias scores were hypothesized to be internally unreliable, in line with previous research (e.g., Waechter et al., 2014). We also hypothesized that reaction time bias scores would not reach acceptable levels of test-retest reliability, in line with previous research (e.g. Schmukle, 2005). As test-retest reliability of eye-tracking bias scores has, to our knowledge, not yet been assessed, we aimed to extend previous research by exploring these indices of reliability in the current study.

Moreover, in an attempt to evaluate the impact of calorie content on the reliability of AB, we aimed to explore if reliability indices would differ for high-versus low calorie food stimuli. We also aimed to explore the correlation of potentially theoretically relevant constructs, such as hunger, blood glucose levels, dietary restraint, and eating disorder symptoms with attention bias indices. Additionally, intercorrelations between the bias-indices were also assessed. Considering that reaction-time bias and direction bias are thought to provide only a snap-shot view of attention (Mogg et al., 2003) and both have been shown to be unreliable in previous research, no correlation of these indices with dwell-time bias was expected.

2. Materials and methods

2.1. Sample

Participants were recruited through online research advertisements at King's College London and through social media. Eligibility was confirmed through an online screening questionnaire that assessed BMI (body mass index), mental and physical health, psychiatric history and medication use. Like most studies on AB for food, we only invited females with a healthy BMI between 18.5 and 25 kg/m^2 to participate to obtain a homogenous sample (for example Werthmann et al., 2013). Individuals with severe psychological or medical problems were excluded, as was anyone with significant eating disorder symptoms. This study was approved by the King's College London Research Ethics, Psychiatry, Nursing and Midwifery Research Ethics Subcommittee.

2.2. Procedure

Once eligibility had been verified, participants were invited to attend two sessions of approximately 30 min each. The sessions were scheduled on two separate days within a period of fourteen days. On average, sessions were two and a half days apart, with a range of one to sixteen days. Participants were requested to refrain from eating or drinking anything besides water 2 h prior to testing to reduce variability in hunger and satiety. Providing participants with eating instructions is standard practice in studies of AB for food (Castellanos et al., 2009; Mogg et al., 1998).

Informed consent and demographic information were obtained at the start of the first session. AB, hunger, and blood glucose (for details see section 2.3 and 2.4) were assessed at both sessions. At the end of the second session, participants also completed measures of dietary restraint and eating disorder symptoms, and their weight and height were measured.

2.3. Attention bias paradigm

2.3.1. Dot-probe task

AB was measured using the dot-probe task (see Fig. 1). The EyeLink 1000 (SR Research Ltd, 2005) was used to record eye-movements. The task was built using Experiment Builder software and data were extracted and pre-processed using Data Viewer (SR Research Ltd, 2005). At each trial of the dot-probe task, a fixation cross was displayed in the centre of the screen until participants had fixated their gaze on it for 100 ms. The cross was then replaced by two images displayed for 3000 ms on either side of the screen. Immediately after these stimuli disappeared, a star appeared in the previous location of one of the stimuli. Participants were instructed to indicate the location of the probe as quickly as they could by pressing a corresponding arrow key. The probe disappeared immediately after they responded and the next trial started automatically.

The stimuli consisted of 30 pairs of images carefully matched for color and complexity, with twenty critical pairs and ten filler pairs (Blechert, Meule, Busch, & Ohla, 2014).[2] Critical pairs consisted of one non-food and one food item, with 10 food items of high caloric density and 10 food items of low caloric density. The ten filler pairs contained

[2] Pictures were used either from previous research (e.g. Werthmann et al., 2011) or from the food-pics database (Blechert, Meule, Busch, & Ohla. 2014), namely pictures 0070, 0350, 0229, 0332, 0267, 0198, 0197, 0413, 0199, 0386, 0466, 0531 and 1276 were used.

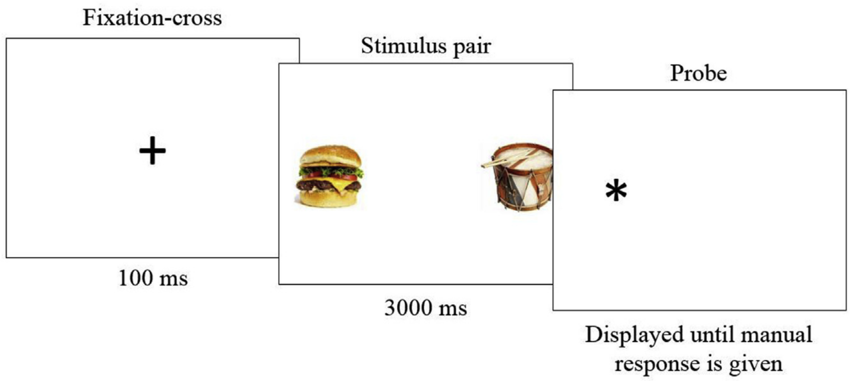

Fig. 1. Example of a congruent critical trial in the dot-probe task. Fixation cross, stimulus pair and probe not to scale.

non-food images. All pairs were presented four times resulting in a total of 120 trials. The trials were presented in two blocks of 60 separated by a short break. One block contained images of high-calorie foods and the other contained images of low-calorie foods. The position of the probe was counterbalanced. The order of the blocks was randomized for each participant. Participants completed a practice round of 16 trials before starting the task.

2.3.2. Attention bias

Three bias indices were computed based on the dot probe task in the current study following standard procedures (Christiansen et al., 2015; Werthmann et al., 2015). Only data from critical trials were used. A direction bias score and gaze dwell time bias score were calculated based on eye-tracking data, and a reaction time bias score was calculated based on response latencies.

Direction bias reflects a bias in the initial orientation of attention towards food or non-food stimuli (Werthmann et al., 2015). The direction is determined by whether the first fixation occurred on a food or non-food stimulus (Werthmann et al., 2015). A fixation was defined as a period lasting at least 100 ms in which no saccades or blinks occurred (Werthmann et al., 2015). Direction bias is calculated as the percentage of the total initial fixations on either food or non-food cues. A score above 50% indicates a direction bias towards food, whereas a score below 50% indicates a direction bias away from food.

To calculate the dwell-time bias, the total amount of time that participants fixated their gaze on food versus non-food cues was extracted from the eye-tracking data (Werthmann et al., 2015). The total fixation durations were averaged across all critical trials for food and non-food images. The average gaze dwell time for non-food images was then subtracted from the average gaze dwell time for food images to obtain dwell time bias. A positive score indicates an AB towards food, and a negative score indicates AB away from food. Because the overall dwell time bias is measured across the 3000 ms presentation of the stimuli, this score reflects the maintenance of visual attention on food cues (Werthmann et al., 2015).

The eye-tracking data showed that eight of the participants were “starers” during both sessions, meaning that they did not make any eye-movements in more than half of all critical trials (Bradley, Mogg, Wright, & Field, 2003). These participants were removed from the analysis as it was assumed that no valid dwell-time or direction bias score could be calculated from their eye-tracking data.[3]

Reaction time data were trimmed by removing reaction times faster than 200 ms, slower than 2000 ms and subsequently those more than three standard deviations above the individual mean, following standard procedures (Christiansen et al., 2015). Based on this procedure, 2.24% of data were discarded in session one and 2.37% of data were discarded in session two. Reaction time bias was then calculated as a difference score by subtracting the sum of all reaction times for the congruent trials from the incongruent trials. Accordingly, positive scores indicate AB towards food and negative scores indicate AB away from food.

All bias indices were calculated as an overall measure (across all critical trials) per session to evaluate test-retest reliability. Additionally, for each session and each bias index, a separate bias measure for critical trials with low calorie food items and for high calorie food items was calculated. To assess internal reliability, item bias scores were also computed following the procedure used by Christiansen et al. (2015): each bias measure (direction bias, dwell time bias, reaction time bias) was calculated separately, as described above, for each critical picture pair within each session, thus yielding 20 bias scores for each bias measure per session.

2.4. Materials

Subjective hunger was measured using a self-report question asking participants to rate their hunger on a 7-point Likert scale, with 0 being not hungry at all and 6 being extremely hungry.

Blood glucose was measured as a physiological marker of hunger and recorded using the Accu-Chek Aviva Blood Glucose Meter (Roche Diabetes Care). Participants were instructed to obtain their own glucose readings from a small drop of blood drawn from the finger using a prick-pen system. Blood glucose concentration was recorded in mmol per liter.

The Restraint Scale (RS) measures dietary restraint (Herman & Polivy, 1980). It consists of ten items about dieting concerns and weight fluctuations rated on a four to five-point Likert scale. Scores for the instrument are reliable in healthy weight participants (Ruderman, 1983).

The Eating Disorder Examination Questionnaire Short (EDE-QS) was used to assess eating disorder symptoms (Gideon et al., 2016). This is a shortened version of the self-report EDE-Q (Fairburn & Beglin, 1994). It contains 12 items that assess symptoms over the last seven days and are rated on a four-point Likert scale. The EDE-QS has excellent reliability, good validity, and high correlations with the scores from the original questionnaire (Fairburn & Beglin, 1994; Gideon et al., 2016).

2.5. Analysis

To test internal reliability, we computed Cronbach's alpha for each bias index, following the procedures described by Christiansen et al. (2015). To evaluate test-retest reliability, Pearson's correlation was calculated between the AB scores obtained at the first and second session.

[3] It is worth noting that removing the starers from the sample did not alter our findings, except for the mean dwell-time bias score, which was slightly lower without the starers.

3. Results

3.1. Sample characteristics

Sixty-one healthy females participated in the study. Eight participants were identified as "starers" (Bradley et al., 2003) based on their eye-tracking data (see description above) and removed from further analyses, resulting in a final sample of 53 participants. One participant missed a blood glucose measurement, but her data were retained for the analyses. The average age of the participants was 26 years (*SD* 7.37). Although all participants reported a weight within the healthy range for their height at the screening, the BMIs recorded in the study ranged from 16.8 to 26.4 kg/m^2 with an average of 21.7 (*SD* 2.14). Two of the participants had a BMI lower than 18.5, four had a BMI between 25 and 26 and two had a BMI slightly above 26.[4] The average score on the Restraint Scale was 9.77 (*SD* 3.48) with a range of 2–19. Similar mean scores of 10.27 and 10.69 have been reported for healthy weight participants in previous research (Bohrer, Forbush, & Hunt, 2015). Only two participants in the current study scored higher than 15 on dietary restraint, which has been used as a cutoff score for high dietary restraint in previous research (Rieger et al., 1998). Furthermore, the average score on the EDE-QS was 2.53 (*SD* 2.21) with a range of 0–10. In previous research, the average score on the EDE-QS was 5.0 for healthy controls and 17.5 for individuals with an eating disorder (Gideon et al., 2016). The sample characteristics suggest that the screening excluded individuals with clinically significant disordered eating or high dietary restraint from participating in the current study. Additional descriptive statistics on subjective hunger and blood glucose can be found in Table 1.

Participants mostly reported adhering to the instructions not to eat or drink during the 2 h prior to the study sessions. One participant reported eating and drinking during the 2 h prior to the first session and three reported drinking something other than water. Prior to the second session, one participant had something to drink and another had something to eat.

To determine whether the sample displayed a bias towards food, a one sample *t*-test was conducted with 50% as the test value for direction bias and 0 ms as the test value for dwell time bias and reaction time bias. The sample showed a small significant direction bias towards food at the second session ($t(52) = 2.08$, $p = .042$, $d = 0.29$) but not at the first session ($t(52) = 1.84$, $p = .072$, $d = 0.25$). There was a small significant dwell time bias towards food at session one (($t(52) = 2.71$, $p = .009$, $d = 0.37$) and a moderate significant dwell time bias at session two ($t(52) = 4.10$, $p < .001$, $d = 0.56$). There appeared to be a substantial difference in the mean dwell time bias score at session one and session two, which prompted us to perform a post-hoc *t*-test. This test confirmed that the bias-score increased significantly ($t(52) = 2.20$, $p = .032$, $d = 0.19$). Finally, the sample showed a small reaction time bias for food at the second session ($t(52) = 2.31$, $p = .025$, $d = 0.32$), but not at the first ($t(52) = 1.61$, $p = .114$, $d = 0.22$).

3.2. Reliability

Direction bias. Neither the overall direction bias score, nor the low and high caloric subscores achieved acceptable internal reliability at either session. Although there was significant test-retest reliability for high caloric foods, it was not of an acceptable level (see Table 2 for reliability estimates of attention bias scores).

Dwell time bias. Dwell-time bias scores had excellent internal reliability at both sessions and acceptable test-retest reliability. The internal reliability of the low and high caloric subscores was good to excellent and the test-retest reliability was acceptable.

Reaction time bias. The internal reliability of the reaction time bias score was acceptable at the first session and poor at the second session. The test-retest reliability was good. At the first session, the low caloric subscore had good internal reliability and the high caloric subscore had poor internal reliability. However, at the second session, only the high caloric subscore had acceptable internal reliability and the internal reliability of the low caloric subscore was extremely poor. Neither subscale had acceptable test-retest reliability.

Table 1
Self-reported hunger and blood glucose in session one and session two.

	Session 1				Session 2			
	Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
Hunger	3.08	1.45	0	5	3.21	1.69	0	6
Blood glucose	4.9	0.6	3.8	7.5	5.0	0.6	3.1	6.3

Note. Hunger was rated from 0 (not hungry at all) to 6 (extremely hungry). Blood glucose concentration was recorded in mmol per liter.

3.3. Inter-correlations attentional bias indices

The reaction time bias, direction bias and dwell time bias scores were all significantly correlated to each other except for direction bias and reaction time bias at the second session. After applying Bonferroni correction for multiple testing, reaction time bias and dwell time bias scores remained significantly correlated at session 1 and session 2. However, direction bias only correlated with dwell time bias at session 2. The correlations were moderate and ranged from 0.417 between reaction time bias and dwell time bias in session 2, and 0.650 between dwell time bias and reaction time bias at session one (see Table 3).

3.4. Correlational analyses

Although the current study aimed to standardize levels of hunger, it was noted that subjective hunger and blood glucose levels still showed some variability (see Table 1). Exploratory correlational analyses between these and other theoretically potentially relevant states and traits and AB were conducted. Hunger and blood glucose were moderately correlated at both sessions. Dietary restraint and EDE-QS were also moderately correlated. These correlations remained significant at the adjusted *alpha*-level of 0.004 after applying a Bonferroni correction for multiple testing (see Tables 3 and 4).

There were significant correlations between dwell time bias and EDE-QS at session one and two, dwell time bias and hunger at session two, and direction bias and EDE-QS at session two. Out of these, only the correlation between EDE-QS and dwell-time bias at session one remained significant after the multiple testing correction was applied and alpha adjusted to 0.004 (see Table 3).

4. Discussion

The aim of the current study was to evaluate the internal and test-retest reliability of eye-tracking and reaction time indices of AB obtained using a pictorial version of the dot-probe task. We also explored the relationships between the different indices of AB and their association with hunger, blood-glucose, dietary restraint and eating disorder symptoms. Overall, the current sample showed a significant dwell time bias towards food at both sessions. On the direction bias and reaction time bias indices, participants only showed a bias towards food during the second session.

[4] To ensure that BMI was not a confounder, analyses were also conducted after removing the underweight and overweight participants from the sample. The results were not significantly different for the healthy weight sample, therefore all participants were retained in the current analysis.

Table 2
Descriptives and reliability of attention bias indices.

Type of food stimuli	Session 1		Session 2		Test-retest
	M (SD)	*α*	*M (SD)*	*α*	*r*
Direction bias					
All foods (overall score)	51.24 (4.90)	.056	51.48 (5.16)	.234	.246
High caloric density	53.25 (8.64)	.414	53.59 (7.56)	.185	.297*
Low caloric density	49.15 (6.25)	-.074	49.40 (6.19)	.105	.005
Dwell time bias					
All foods (overall score)	134.58 (361.59)	.921	204.21 (362.32)	.923	.798**
High caloric density	167.56 (462.10)	.911	238.61 (438.55)	.898	.739**
Low caloric density	101.60 (324.38)	.824	169.81 (323.25)	.815	.757**
Reaction time bias					
All foods (overall score)	9.23 (41.84)	.793	7.49 (23.66)	.615	.835**
High caloric density	10.06 (33.24)	.357	8.48 (38.45)	.701	.611**
Low caloric density	7.77 (59.09)	.844	6.68 (18.08)	-.233	.361**

Note. α = Cronbach's α; r = Pearson's r. *p < .050, **p < .010.

Table 3
Correlations session one.

	Hunger	Blood glucose	Dietary restraint	EDE-QS	Direction bias	Dwell time bias
Hunger						
Blood glucose	-.494**					
Dietary restraint	.030	-.069				
EDE-QS	.047	.109	.574**			
Direction bias	.204	-.138	.108	.156		
Dwell time bias	.100	.098	.036	.416**	.359	
Reaction time bias	.142	-.080	.086	.193	.273	.650**

Note. **p < .004 (Bonferroni adjusted significance level).

Table 4
Correlations session two.

	Hunger	Blood glucose	Dietary restraint	EDE-QS	Direction bias	Dwell time bias
Hunger						
Blood glucose	-.318					
Dietary restraint	-.067	-.034				
EDE-QS	.150	.093	.574**			
Direction bias	.103	-.027	.125	.288		
Dwell time bias	.342	-.204	-.044	.331	.538**	
Reaction time bias	.100	-.029	-.024	.079	.263	.417**

Note. **p < .004 (Bonferroni adjusted significance level).

4.1. Reliability

The internal and test-retest reliability of direction bias was poor, providing additional support for the notion that the direction of the first fixation is not a coherent and stable measure of AB (Waechter et al., 2014). One reason for this may be that participants have a tendency to initially direct their gaze to the left of a display rather than towards any specific stimulus type (Durgin et al., 2008; Guo et al., 2012). It is possible this tendency overshadowed early attentional preference for certain stimuli. Alternatively, direction bias may have poor reliability as it relies on a "snap shot" view of attention allocation, which may need more samples (thus more trials) to achieve reliable proportions.

In contrast, dwell-time bias scores showed excellent internal reliability and acceptable test-retest reliability. In agreement with previous research, dwell time bias scores were more reliable than reaction time bias scores. Reaction time bias scores showed poor to acceptable internal reliability and good test-retest reliability. However, it is of note that the observed reliability of both indices in our study exceeds that reported previously (Christiansen et al., 2015; Schmukle, 2005; Waechter et al., 2014). One potential explanation for this could be the use of a relatively long presentation time of 3000 ms, as Waechter et al. (2014) previously demonstrated that a prolonged presentation time can improve reliability.

Stimuli in the dot-probe task are typically presented for a shorter time (e.g. Schmukle, 2005), presumably to detect a stimulus-driven, so called "bottom-up", bias. Whereas "top-down" visual attention is driven by a more time-consuming goal directed selection of relevant stimuli, bottom-up attention is driven by the rapid detection of salient stimuli (Buschman & Miller, 2007). Hence, limiting the presentation time of images in the dot probe task has been assumed to limit the degree to which participants were able to manipulate their scores. Although our results appear to provide further evidence in favor of a prolonged stimulus presentation time in the dot-probe task, the adjustment of stimulus presentation times may have substantial impact on the interpretation of the observed AB, which may then reflect unintentional processes to a significantly lesser extent than originally intended.

Additionally, the careful selection of appetizing stimuli used in our study may have contributed to the good reliability. Personally relevant and ecologically valid stimuli elicit more robust scores (Caudek, Ceccarini, & Sica, 2017; Christiansen et al., 2015). The use of less effective stimuli may have contributed to the poor reliability bias scores in previous work. If the stimuli used are too diverse or do not match an individual's preference, they are unlikely to elicit a strong bias consistently. More attention to the choice of stimuli is needed to improve the psychometric properties of the dot-probe task. To test whether appetizing stimuli can improve reliability, participants in future research could be asked to rate how appetizing they found the images of food. No clear pattern emerged with regards to the reliability of AB for foods that are low versus high in caloric density. This may, at least in part, be due to the lower number of trials each of these bias indices was based on compared to the overall bias score (incorporating both high- and low-calorie critical trials). Accordingly, further research is needed to establish if the use of high-versus low-calorie pictures influences reliability measures of the dot-probe task.

Although the dwell-time scores achieved acceptable test-retest reliability, we did observe a significant increase in bias scores from session one to session two. The acceptable test-retest reliability of the dwell time scores suggests that the scores are reliable enough to be used in studies examining group differences. However, studies investigating a change in AB over time within individuals will need to consider that the score may increase with repeated administrations of the task. This limitation can be overcome by including a control-condition. Moreover, the observed increase in dwell time bias scores from session one to session two needs further replication and clarification.

4.2. Exploratory analyses

The reaction time bias and dwell time bias were moderately correlated with each other. However, direction bias only correlated with dwell time bias at session two. The finding that the direct measures of overt attention (eye-tracking) correlated with the indirect measure of attention (reaction times), provides support for the assumption that AB based on reaction times in the dot-probe task truly reflects biased visual attention.

A post-hoc analysis revealed no strong relationships between AB and theoretically potentially relevant constructs. The only significant correlation in this respect was between EDE-QS scores and dwell time at session one. The remaining relationships between attention bias and hunger, blood glucose levels, dietary restraint and eating disorder symptoms were not significant, or did not survive a multiple testing correction. This is in line with research that found no consistent pattern of associations between hunger, dietary restraint, disordered eating and AB for food (e.g., Field et al., 2016). It is also in line with research showing that AB scores based on the dot probe paradigm fail to consistently relate to theoretically important constructs in the field of addiction and anxiety.

The lack of significant relationships between AB and restrained or disordered eating could be attributed to some limitations in our study. Research has suggested that findings may be inconsistent because AB for food can be attenuated by priming and mindset manipulations (Papies et al., 2008; Werthmann et al., 2016). Individuals can also show AB both towards and away from different threatening stimuli within a single assessment (Zvielli, Bernstein, & Koster, 2014). Controlling for these factors and examining fluctuations within the session was beyond the scope of the present study. In future research, the indices derived from trial level bias scores such as variability (Liu, Roefs, Werthmann, & Nederkoorn, 2019; Price et al., 2015; Rodebaugh et al., 2016; Zvielli, Bernstein, & Koster, 2015) might inform the relationship between restrained eating and AB given the previous evidence of the instability of their AB. Moreover, the current sample consisted only of healthy participants with low rates of disordered eating and restrained eating. It is possible that AB for food is most strongly related to pathological mechanisms (in the way we would expect AB to alcohol-related stimuli to be higher among individuals who score high on addiction). If this is the case, our sample may have been too homogenous in terms of food-related pathology to detect this association. That is, restriction-of-range may have contributed to not finding stronger relationships.

Despite standardized eating instructions, the sample showed fluctuations and variability in subjective hunger and blood glucose levels. Previous authors have also observed this (Mogg et al., 1998). Standard practice eating instructions appear to be ineffective at standardizing subjective hunger and future research should explore alternative approaches, such as asking participants to eat until they feel satiated. In the current sample, no association between hunger and AB was found. However, our study was not designed to maximize variability in hunger. Moreover, we only measured two aspects of hunger (subjective hunger and blood level glucose). As hunger is not a unidimensional concept, future research could consider other relevant aspects such as gastrointestinal signals (Schultes et al., 2016). As it stands, the inconsistent pattern of associations between hunger and AB for food remains in need of clarification.

4.3. Summary and conclusions

The current study examined the reliability of different AB indices for food and explored their relationships to theoretically relevant constructs. Results showed that direction bias had poor internal and test-retest reliability. Dwell time bias had excellent internal and acceptable test-retest reliability. Reaction time bias had acceptable internal and good test-retest reliability. Thus, results of our study indicate that dwell time and reaction time indices derived from the dot-probe task can provide reliable measures of attention bias for food.

These findings are in contrast to previous studies that reported low reliability of AB indices derived from the visual dot probe task. Previous studies achieving high reliability have done so through substantial changes to the dot-probe paradigm (e.g. Christiansen et al., 2015; Zvielli et al., 2015), but our data suggest that it might be possible to achieve reliable scores with only slight modifications to the paradigm currently in use. Our results indicate that the combination of the use of dwell-time scores, longer stimulus presentation times and carefully selected appetizing stimuli produced the most reliable scores, whereas direction bias scores were the least reliable. However, even the reliable bias indices in the current study lacked robust relationships with theoretically relevant constructs, such as hunger, blood glucose and dietary restraint. Moreover, the adjustment of stimulus presentation time (3000 ms in the current study) may have substantial impact on the conceptualization of attention bias in terms of capturing versus holding of attention. Therefore, the interpretation of the AB indices as derived from the dot-probe task may require further clarification and possibly theoretical revision. In an attempt to further advance the study of food-related attention bias, we urge future studies to incorporate reliability assessments of their attention bias indices.

Declarations of interest

None.

Acknowledgement

Jessica Werthmann was supported financially by a Rubicon Grant (446-13-011) awarded by the Netherlands Organization for Scientific Research and by a NARSAD Young Investigator award (25269) from the Brain & Behavior Research Foundation. Ulrike Schmidt receives salary support from the National Institute for Health Research (NIHR) Biomedical Research Centre for Mental Health at South London and Maudsley NHS Foundation Trust and is supported by an NIHR Senior Investigator Award. Anne Roefs is supported by a VIDI-grant (452.16.007) of the Netherlands Organization for Scientific Research. We would like to thank Pia Lauffer and Clara Rockstroh for their help in recruiting and testing participants for this study.

References

Blechert, J., Meule, A., Busch, N. A., & Ohla, K. (2014). Food-pics: An image database for experimental research on eating and appetite. *Frontiers in Psychology, 5*, 617.
Bohrer, B. K., Forbush, K. T., & Hunt, T. K. (2015). Are common measures of dietary

restraint and disinhibited eating reliable and valid in obese persons? *Appetite, 87*, 344–351.

Bradley, B. P., Mogg, K., Wright, T., & Field, M. (2003). Attentional bias in drug dependence: Vigilance for cigarette-related cues in smokers. *Psychology of Addictive Behaviors, 17*(1), 66–72.

Brooks, S., Prince, A., Stahl, D., Campbell, I. C., & Treasure, J. (2011). A systematic review and meta-analysis of cognitive bias to food stimuli in people with disordered eating behaviour. *Clinical Psychology Review, 31*(1), 37–51.

Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science, 315*(5820), 1860–1862.

Castellanos, E. H., Charboneau, E., Dietrich, M. S., Park, S., Bradley, B. P., Mogg, K., et al. (2009). Obese adults have visual attention bias for food cue images: Evidence for altered reward system function. *International Journal of Obesity, 33*(9), 1063–1073 (2005).

Caudek, C., Ceccarini, F., & Sica, C. (2017). Facial expression movement enhances the measurement of temporal dynamics of attentional bias in the dot-probe task. *Behavior Research and Therapy, 95*, 58–70.

Christiansen, P., Mansfield, R., Duckworth, J., Field, M., & Jones, A. (2015). Internal reliability of the alcohol-related visual probe task is increased by utilising personalised stimuli and eye-tracking. *Drug and Alcohol Dependence, 155*, 170–174.

Cooper, R. M., Bailey, J. E., Diaper, A., Stirland, R., Renton, L. E., Benton, C. P., et al. (2011). Effects of 7.5% CO(2) inhalation on allocation of spatial attention to facial cues of emotional expression. *Cognition & Emotion, 25*(4), 626–638.

Doolan, K. J., Breslin, G., Hanna, D., & Gallagher, A. M. (2014). Attentional bias to food-related visual cues: Is there a role in obesity? *Proceedings of the Nutrition Society, 74*.

Durgin, F. H., Doyle, E., & Egan, L. (2008). Upper-left gaze bias reveals competing search strategies in a reverse Stroop task. *Acta Psychologica, 127*(2), 428–448.

Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders, 16*(4), 363–370.

Field, M., Werthmann, J., Franken, I. H. A., Hofmann, W., Hogarth, L., & Roefs, A. (2016). The role of attentional bias in obesity and addiction. *Health Psychology, 35*(8), 767–780.

Gideon, N., Hawkes, N., Mond, J., Saunders, R., Tchanturia, K., & Serpell, L. (2016). Development and psychometric validation of the EDE-QS, a 12 item short form of the eating disorder examination questionnaire (EDE-Q). *PLoS One, 11*(5), 1–19.

Giel, K. E., Friederich, H. C., Teufel, M., Hautzinger, M., Enck, P., & Zipfel, S. (2011). Attentional processing of food pictures in individuals with anorexia nervosa - an eye-tracking study. *Biological Psychiatry, 69*(7), 661–667.

Giel, K. E., Teufel, M., Friederich, H. C., Hautzinger, M., Enck, P., & Zipfel, S. (2011). Processing of pictorial food stimuli in patients with eating disorders-A systematic review. *International Journal of Eating Disorders, 44*(2), 105–117.

Guo, K., Smith, C., Powell, K., & Nicholls, K. (2012). Consistent left gaze bias in processing different facial cues. *Psychological Research, 76*(3), 263–269.

Herman, C. P., & Polivy, J. (1980). Restrained eating. *Obesity,* 208–225.

Kappenman, E. S., MacNamara, A., & Proudfit, G. H. (2013). Electrocortical evidence for rapid allocation of attention to threat in the dot-probe task. *Social Cognitive and Affective Neuroscience, 10*(4), 577–583.

Kuckertz, J. M., & Amir, N. (2015). Attention bias modification for anxiety and phobias: Current status and future directions. *Current Psychiatry Reports, 17*(2).

Liu, Y., Roefs, A., Werthmann, J., & Nederkoorn, C. (2019). Dynamics of attentional bias for food in adults, children, and restrained eaters. *Appetite, 135*, 86–92. https://doi.org/10.1016/j.appet.2019.01.

Loeber, S., Grosshans, M., Herpertz, S., Kiefer, F., & Herpertz, S. C. (2013). Hunger modulates behavioral disinhibition and attention allocation to food-associated cues in normal-weight controls. *Appetite, 71*, 32–39.

MacLeod, C., Mathews, A. M., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology, 95*(1), 15–20.

Mogg, K., Bradley, B. P., Field, M., & De Houwer, J. (2003). Eye movements to smoking-related pictures in smokers: Relationship between attentional biases and implicit and explicit measures of stimulus valence. *Addiction, 98*(6), 825–836.

Mogg, K., Bradley, B. P., Hyare, H., & Lee, S. (1998). Selective attention to food-related stimuli in hunger: Are attentional biases specific to emotional and psychopathological states, or are they also found in normal drive states? *Behaviour Research and Therapy, 36*(2), 227–237.

Papies, E. K., Stroebe, W., & Aarts, H. (2008). The allure of forbidden food: On the role of attention in self-regulation. *Journal of Experimental Social Psychology, 44*(5), 1283–1292.

Price, R. B., Kuckertz, J. M., Siegle, G. J., Ladouceur, C. D., Silk, J. S., Ryan, N. D., et al. (2015). Empirical recommendations for improving the stability of the dot-probe task in clinical research. *Psychological Assessment, 27*(2), 365–376.

Rieger, E., Schotte, D. E., Touyz, S. W., Beumont, P. J. V., Griffiths, R., & Russell, J. (1998). Attentional biases in eating disorders: A visual probe detection procedure. *International Journal of Eating Disorders, 23*(2), 199–205.

Rodebaugh, T. L., Scullin, R. B., Langer, J. K., Dixon, D. J., Huppert, J. D., Bernstein, A., et al. (2016). Unreliability as a threat to understanding psychopathology: The cautionary tale of attentional bias. *Journal of Abnormal Psychology, 125*(6), 840–851.

Ruderman, A. J. (1983). The restraint scale: A psychometric investigation. *Behaviour Research and Therapy, 21*(3), 253–258.

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

Schultes, B., Panknin, A. K., Hallschmid, M., Jauch-Chara, K., Wilms, B., de Courbière, F., et al. (2016). Glycemic increase induced by intravenous glucose infusion fails to affect hunger, appetite, or satiety following breakfast in healthy men. *Appetite, 105*, 562–566.

SR Research Ltd (2005). *Eyelink 1000 [eye-tracking system]*.

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

Waechter, S., Nelson, A. L., Wright, C., Hyatt, A., & Oakman, J. (2014). Measuring attentional bias to threat: Reliability of dot probe and eye movement indices. *Cognitive Therapy and Research, 38*(3), 313–333.

Wald, I., Shechner, T., Bitton, S., Holoshitz, Y., Charney, D. S., Muller, D., et al. (2011). Attention bias away from threat during life threatening danger predicts PTSD symptoms at one-year follow-up. *Depression and Anxiety, 28*(5), 406–411.

Werthmann, J., Jansen, A., & Roefs, A. (2014). Worry or craving? A selective review of evidence for food-related attention biases in obese individuals, eating-disorder patients, restrained eaters and healthy samples. *Proceedings of the Nutrition Society,* 1–16.

Werthmann, J., Jansen, A., & Roefs, A. (2016). Make up your mind about food: A healthy mindset attenuates attention for high-calorie food in restrained eaters. *Appetite, 105*, 53–59.

Werthmann, J., Jansen, A., Vreugdenhil, A. C. E., Nederkoorn, C., Schyns, G., & Roefs, A. (2015). Food through the child's eye: An eye-tracking study on attentional bias for food in healthy-weight children and children with obesity. *Health Psychology, 34*(12), 1123–1132.

Werthmann, J., Roefs, A., Nederkoorn, C., Mogg, K., Bradley, B. P., & Jansen, A. (2011). Can(not) take my eyes off it: Attention bias for food in overweight participants. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association, 30*(5), 561–569. https://doi.org/10.1037/a0024291.

Werthmann, J., Roefs, A., Nederkoorn, C., Mogg, K., Bradley, B. P., & Jansen, A. (2013). Attention bias for food is independent of restraint in healthy weight individuals-An eye tracking study. *Eating Behaviors, 14*(3), 397–400.

Zvielli, A., Bernstein, A., & Koster, E. H. W. (2014). Dynamics of attentional bias to threat in anxious adults: Bias towards and/or away? *PLoS One, 9*(8).

Zvielli, A., Bernstein, A., & Koster, E. H. W. (2015). Temporal dynamics of attentional bias. *Clinical Psychological Science,* (5), 772–788.