The Rise and Fall of the Dot-Probe Task: Opportunities for Metascientific Learning

Benjamin T. Sharpe[1], Monika Halls[2], Thomas E. Gladwin[3]

Much of the extensive literature on spatial attentional bias is built on measurements using the dot-probe task. In recent years, concerns have been raised about the psychometric properties of bias scores derived from this task. The goal of the current paper is to look ahead and evaluate possible responses of the field to this situation from a metascientific perspective. Therefore, educated guesses are made on foreseeable but preventable future (repeats of) errors. We discuss, first, the issue of overreactions to the disappointing findings, especially in the context of the potential of a new generation of promising variations on the traditional dot-probe task; second, concerns with competition between tasks; and third, the misuse of rationales to direct research efforts. Alternative directions are suggested that may be more productive. We argue that more adequately exploring and testing methods and adjusting scientific strategies will be critical to avoiding suboptimal research and potentially failing to learn from mistakes. The current articulation of arguments and concerns may therefore be of use in discussions arising around future behavioural research into spatial attentional bias and more broadly in psychological science.

Psychological tasks play a central role in psychological science. These are relatively simple stimulus-response games that participants are asked to play, containing specific variations that affect cognition and behaviour. Tasks are designed so that effects, i.e., differences between conditions on observable measurements such as reaction time, can be interpreted in relation to a particular psychological construct of interest. Such effects are used either to test hypotheses directly or to create individual difference variables for use as building blocks in further analyses. Progress in scientific fields may thus heavily depend on the quality of tasks and the processes via which they evolve. The current paper considers developments around one such task: the dot-probe task.

The Dot-probe Task

When confronted with a salient stimulus - for instance, an angry face looking at you in a crowd, a spider crawling across your desk, or a tasty treat on a kitchen table - our

[1] Institute of Psychology, Business, and Human Sciences, University of Chichester, Chichester, United Kingdom

[2] Independent researcher

[3] Independent researcher; corresponding author: thomas.gladwin@gmail.com

attention is preferentially drawn to that particular stimulus over many other stimuli present in the same environment. Attentional biases refer to the ability of certain stimuli to draw or repel attention automatically. For over three decades (MacLeod et al., 1986), the dot-probe task has been the basis for a significant amount of research into attentional bias (Cisler & Koster, 2010; Field & Cox, 2008; Jiang & Vartanian, 2018; Starzomska, 2017). In the usual version of this well-known task (Figure 1), two task-irrelevant cues are presented, followed by a probe at one of their locations. One cue is drawn from a salient category hypothesized to cause an attentional bias; the other is drawn from a neutral control category. Faster (or slower) reaction times to probes presented at the location of salient cues would reflect an attentional bias towards (or away from) that category. The behavioural measure of spatial attentional bias provided by this task has been used to study potential underlying mechanisms of a wide range of disorders, e.g., anxiety (Bantin et al., 2016; B. P. Bradley et al., 1998; Koster et al., 2006), obesity (Vervoort et al., 2021), posttraumatic stress disorder (Bomyea et al., 2017; Wald et al., 2013), and addiction (Field et al., 2016; Townshend & Duka, 2001; Wiers et al., 2016).

Figure 1. *Illustration of the dot-probe task*
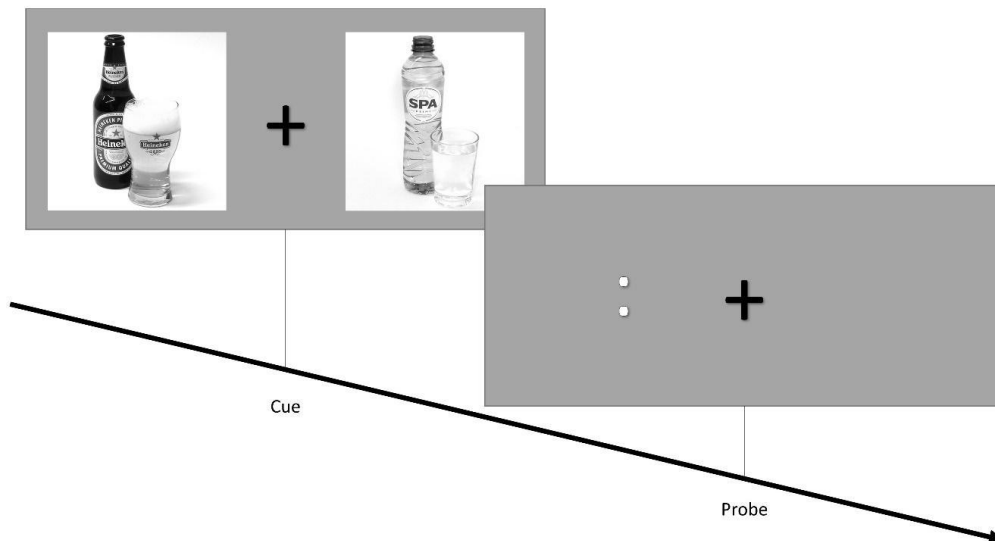
Note. The Figure shows an example of a trial from a dot-probe task. The cue consists of a salient (here, an alcoholic drink) stimulus and a relatively neutral control stimulus (here, a non-alcoholic drink). After an interval, in the order of hundreds of milliseconds to a second, a probe stimulus is presented requiring a speeded choice response. Examples of probes are a single dot or double dots, each requiring a different response. The critical manipulation of the task is whether the probe is presented at the location of the salient versus the control stimulus.

Methodological Concerns

Relatively recent concerns with the reliability of the task, extensively discussed elsewhere (Christiansen et al., 2015; Jones et al., 2018; McNally, 2018; Rodebaugh et al., 2016; Van Bockstaele et al., 2020) are therefore concerning. We only briefly recap this literature. Very low reliability was found in a number of studies that assessed it (Ataya et al., 2012; Brown et al., 2014; Chapman et al., 2017; Dear et al., 2011; Kappenman et al., 2014; Waechter et al., 2014). In some cases, the reliability was close to zero, and certainly not near the level needed for detecting relationships involving individual differences with sufficient statistical power. Such statistical effects would include, for examples, correlations between bias scores and mental health-related variables - roughly speaking, if a measure cannot predict itself, how could it predict a different variable?

If there is a systemic problem with the reliability of measures, then even though there is presumably a subset of important and true findings in the literature, we don't know whether a particular finding is in that set. Further, such issues add to the plausibility that the literature is affected by publication biases and questionable research practices (J. M. Simmons et al., 2011). This renders it difficult to draw any conclusions from the literature as a whole and implies any such attempt will require statistical methods focused on critical assessment of the literature (Bartoš & Schimmack, 2020). Different authors could currently draw different conclusions on the overall state of the literature; however, the number of publications discussing such psychometric problems, with authors including prominent researchers, confirm that at least a significant part of the field has accepted that there is indeed a serious problem to be addressed. Similarly, as discussed further below, research groups are investing in developing alternative task versions, implying awareness of a need for improvement. That there is, in fact, a problem therefore seems undeniable. (Please see Appendix A for some additional points on assessing psychometric properties.)

The Current Paper

The aim of the current paper, rather than to provide another review of psychometric issues, is to consider what went wrong from a metascientific perspective and to consider the future of the field. That is: Perhaps it is worth making and communicating educated guesses to avoid something akin to the current situation re-occurring, and only at that point looking back with hindsight, when the damage has been done? From this perspective, we caution against three speculative but predictable reactions to this "psychometric crisis" of spatial attentional bias measurement and suggest constructive responses.

Potential reactions to the psychometric issues

Reaction 1: Rejection versus improvement

One potential reaction to the psychometric crisis would be to reject altogether effects in behavioural performance as a valid way to assess automatic processes, at least concerning spatial attentional bias. This would have quite serious implications. Studies could no longer be performed using the relatively basic equipment needed for performance-based measurement, which would strongly increase inequalities as researchers would differ in the resources needed to make such a shift. Additionally, we would have to accept as a stable situation that our associated understanding and theory is so inadequate that we cannot even consistently measure attentional biases behaviourally.

However, there is an alternative view: That the problem is about the details of the methods rather than the fundamental approach, and that the solution to the psychometric crisis could involve finding better task variants, study designs or analyses. This would involve a much stronger focus on, and appreciation of, basic research exploring and testing details of how task variations affect behavioural measures, as opposed to studies aiming to use those measures to answer complex, real-world, e.g., clinical, questions. The latter kind of studies may be attractive given their apparent potential for impact (Bayley & Phipps, 2019) but would be premature unless the implicit assumption is confirmed that the measures they use are sufficiently reliable and valid. We note a number of alternative versions of spatial attentional bias tasks that have arisen relatively recently to combat some of the concerns about reliability.

First, in the Odd-One Out Task (Heitmann et al., 2021), a matrix of stimuli is presented, and the task is to indicate whether one of the stimuli is an "odd one out" target, e.g., a tool in a matrix otherwise consisting of alcoholic drinks serving as distractors. The nature of the target and of the non-target stimuli varies over trials, and can be used to define various biases; e.g., it can be assessed whether reaction times are slowed when the distractors are alcoholic, or are reduced when the target is alcoholic. The reliability of the task has been improved to relatively good levels as it has been developed.

Second, the Predictive Visual Probe Task aims to assess an *anticipatory* form of attentional bias (Gladwin et al., 2021). The task presents visually neutral cues that predict the location of upcoming emotionally salient stimuli. Trials throughout the task are randomly determined to be "picture" trials in which only the salient and control stimuli appear, or "probe" trials in which only probe stimuli requiring a response appear. Performance is thus not affected by the actual presentation of the particular exemplars of stimuli on that trial. The anticipatory attentional bias evoked by the predictive cues has been found to have good reliability with optimized designs and to be related to anxiety. The task design may also reveal trial-to-trial effects (Gladwin & Figner, 2019) that are not measurable in the traditional dot-probe task (Maxwell et al., 2022).

Third, the dual probe task has been developed as an alternative to the dot-probe task (Grafton et al., 2021). In this task, two probes are presented, so briefly that participants can only attend and respond to one of them. Results of this task showed excellent reliability and sensitivity to experimental manipulations of attention. All these tasks may be seen as steps in the right direction, arguably representing the evolution of traditional methods. This is not to say that these approaches do not need confirmation of results or that they could not be improved or perhaps combined with each other. However, they do indicate the feasibility of the use of behavioural measures of the dot-probe-like spatial attentional bias.

Reaction 2: Dysfunctional task-level competition

The second suboptimal reaction relates to the way in which such novel methods might compete and evolve, as a specific case of the more general role of competition in science (Fang & Casadevall, 2015; Tiokhin et al., 2021). Given that a new generation of tasks assessing spatial attentional bias is a viable way forward for research into spatial attentional bias, it seems plausible that there will be a tendency for researchers and research groups to aim to produce the one dominant replacement for the traditional dot-probe task.

However, this would risk incentivizing fast but sloppy science, leading to premature selection of a new method based on limited and possibly biased data used to suggest some form of superiority. Further, it seems unlikely that any single task will hit on a truly optimal combination of features for all purposes, especially in relatively early phases of discovery and exploration. Finally, quick convergence on a winning task for use in further studies might well reduce observed heterogeneity, or unexplained variability beyond that expected by sampling error (Linden & Hönekopp, 2021). However, if there are in fact differences between variations in tasks that should measure the same construct, heterogeneity should be considered a valuable signal of a gap in understanding (Linden & Hönekopp, 2021). Such heterogeneity should, therefore, not be obscured by failing to sufficiently explore task variations.

In contrast, research could more positively focus on task features, regardless of the pedigree of researchers, rather than tasks as a whole. Such features could have varying pros and cons, be more or less suitable in different situations, and be more or less suited to combinations of features. Research that permits the time and resources to explore this "task space", rather than more rigidly comparing candidate tasks, seems essential to providing sufficient data to support sufficient understanding of attentional bias measurement for it to be justified as a method in more complex or applied, e.g., clinical, studies.

Reaction 3: Sensitivity to narratives

The third reaction regards the role of theoretical rationales to conclude that a given task or task feature is either an improvement or flawed a priori. For instance, the dot-probe task could be argued to be more interpretable or informative than the emotional Stroop

---

task (McNally, 2018). It was perhaps more immediately obvious that it is uncertain what could cause differences between trial types in the emotional Stroop task – but does that mean investment of resources in the former task will have been more justified and more productive than in the latter? (Please see also the note in Appendix B on Attention Bias Modification.)

While theoretical arguments are of course an essential component of hypothesis generation and designing, interpreting, and criticizing individual studies, the concern expressed here is that conceptual, narrative (whether verbally or formally/computationally stated) arguments may be given more weight than deserve and thereby potentially derail the direction of research investments. The risk of this inherently follows from the uncertainties around the theory underlying attentional bias. It is not even trivial to define what the "attention" of attentional bias is; it has been questioned whether it is a useful construct at all (Hommel et al., 2019). Early models of attention concerned fundamental, relatively well-defined information processing bottlenecks (Broadbent, 1958; Norman, 1968; Posner & Petersen, 1990; Schneider & Shiffrin, 1977; Treisman, 1964). However, in linking attention to emotion, via theoretical frameworks such as motivated attention (Lang, 1995), motivational activation (M. M. Bradley, 2009), or dual process models (Strack & Deutsch, 2004), a more complex research question arises, namely whether the emotional meaning of stimuli to the individual affects cognitive processing. While there are certainly interesting and potentially valuable studies and models that focus on differentiating aspects of attentional bias (Koster et al., 2006), meaningful scientific theoretical development must be built on sufficiently reliable and precise empirical support, which is of course exactly what has been questioned.

Further, despite their continuing use, the dual systems/dual processes models that form the usual theoretical underpinnings of attentional bias research have been fundamentally criticized, for instance regarding conceptual precision (Keren, 2013; Keren & Schul, 2009). A lack of empirically supported and detailed theory has implications for how valid it is to use theoretical rationales to evaluate methods. Arguments made to decide between tasks can become somewhat arbitrary – which and whose theoretical constructs and considerations will the field pay attention to? And which features of tasks will be considered relevant versus ignored? Implications that, we suggest, follow from this situation are (1) that caution is necessary towards narrative rationales for using one task or another, or steering fields of research, as arguments may rest on only a very limited aspect of potentially relevant factors, all of which are uncertain; (2) that where there are variations in tasks that preclude any one task providing a definitive interpretation, to consider whether they can provide convergent evidence, e.g., by demonstrating effects that are consistent over different threats to interpretations; (3) that empirical evidence of psychometric quality should always be prioritized over merely conceptual (dis)advantages; and (4) that alternative

interpretations of effects, drawing on different hypothetical influences, should be seen as an important and positive driver of research progress.

Conclusion

In conclusion, there is a risk of various unproductive reactions to problematic psychometric findings regarding the dot-probe task. There certainly appears to be reason to look back critically on previous research practices, in line with more general concerns with psychological research (John et al., 2012; J. M. Simmons et al., 2011); however, this results at worst in a disappointing lack of usable results, not evidence against hypotheses concerning attentional bias altogether, or against the potential for behavioural study using implicit measures. There are promising avenues of research that could validly be pursued at this point. While the current paper focuses on the tasks used to generate data, such avenues also include innovations in data analysis, such as the analysis of attentional bias variability (Iacoviello et al., 2014; Zvielli et al., 2014). Nonetheless, to avoid potentially repeating history, future work should align itself with practices and attitudes in support of methods and ethics that tend to lead to scientifically valid productivity. Some ways in which alignment could be expressed in the current context are listed below.

(1) Spending sufficient time exploring task features and qualities, focused on solid science (Frith, 2020) and avoiding questionable practices such as hyping (Martin, 2016); this implies incentivizing accurate and ethical behaviour around the acknowledgement of previous work on the development of our fundamental research tools.

(2) Requiring a sufficient empirical basis before the use of task-based measures in costly studies; such use requires it to be true that those measures are reliable and valid, and therefore the replication and extension work necessary for this must be valued.

(3) Acknowledging the risks of potentially premature uniformization of methods, rather than only the possible benefits of uniformization assuming sufficient psychometric qualities.

(4) Applying and advocating caution and scepticism about narrative theoretical rationales used to promote or suppress measures or lines of research.

(5) Accepting that strategic simplifications may need to be accepted; e.g., methodologically, by starting out with the assumption of uniform populations to find robust overall patterns of effects, or theoretically, by allowing that it is not always necessary or optimal to go beyond a general, ecological concept of attention as a naturally unified process of selection for action (Balkenius & Hulth, 1999).

The intention here is to raise awareness and explicitly identify choices. Which choices will in fact be made in practice will depend on a multitude of factors; indeed, while beyond the scope of the current paper, we briefly discuss in Appendix C how the case

of the dot-probe task could tie into broader questioning of scientific systems (Tijdink et al., 2021). We hope the articulation of the risks and suggestions raised in the current paper will be of use for researchers interested in implicit measures, as well as broader metascientific issues. There may be unexpected benefits, beyond improvements within the relatively constrained interests of implicit measures, that arise from the current disappointing situation, if it could lead to better understanding of what features of the academic ecosystem tend to lead to good and bad outcomes for scientific progress.

Gardener Comments

Dan James:
This paper is a cogent examination of the dot-probe paradigm and sustains a consistently high level of engagement throughout. What especially recommends this paper is its coordination of the perceived successes and weaknesses of the dot-probe construct, within a broader epistemological context that raises questions about what constitutes evidence not just for Psychology but for Science as a whole.

This added philosophical dimension makes this paper relevant for researchers of any discipline who may be faced with a similar conundrum as detailed in the paper - that as part of the research process, logically coherent constructs, or the selection of experimental setups and the resulting statistical analysis employed, cannot avoid generalising or simplifying heuristically, and in doing so will always run the risk of biased or false models.

Revision comments:
In an otherwise comprehensive account, this paper could potentially benefit from greater attention to a specific and often neglected problem in psychology – heterogeneity (Linden & Hönekopp 2021). Whilst the problem of individual variance is acknowledged in the paper, I feel a more explicit account of heterogeneity could conceivably form part of the authors discussion concerning a critical analysis of research practices that they highlight as a possible way forward to address some identified problematic psychometric findings of the dot-probe task.

As the authors explain, the essential practicality of the dot-probe construct is its ability to measure attentional bias as a proxy for potential underlying mechanisms or disorders. However, heterogeneity suggests such measurements and their purported behavioural association, should be qualified, especially if the effect sizes are small, because to do otherwise would be to make an assumption that the dot-probe task applies equally well to a large portion, if not all, of the population. In contrast, and to minimise the problem of heterogeneity, would better specified and smaller subsections or target populations produce better measurement and predictive success?

Clearly, as the paper explains, in any meta-analysis, measuring individual variance is an important factor in determining the theoretical scope of a psychometric assessment and the dot-probe paradigm is no exception to this.

But there is another sense in which a consideration of heterogeneity is important as part of a methodological principle, that is not mentioned in the paper, and that is to avoid the so-called 'paradox of convergence' where there is a disconnect between an accumulation or convergence of evidence in favour of a particular theory and yet at the same time an inability to be sure that such theory can accurately be applied to any individual (Davis-Stober C, Regenwetter M.,2019).

Recommendation:
The interaction of heterogeneity with an accumulation of evidence, means it's not hard to foresee a crisis of confidence questioning the theoretical rationale for any psychometric construct, in particular the dot-probe paradigm, the precise subject of the paper under review and the reason why I suggest a greater emphasis on heterogeneity would only be of benefit to an already fascinating and well-researched paper that is highly recommended for publication.

Heidi Zamzow:
This piece is very well-written and referenced, with a strong and coherent argument. The contribution to the field is clearly demonstrated, not only with respect to methods using implicit measures specifically, but to a systematic critique of methodology more broadly. In particular, I appreciated the discussion on potential lines of future research. Overall, I found the paper to be highly relevant, and I strongly endorse its publication.

Dr. Payal B. Joshi:
The article takes a dive in a difficult premise and attempts to provide an explanation on the psychometric crisis. Authors have critically discussed and presented their arguments in a clear and engaging manner. Overall, the article is a decent attempt to dot-probe tasks on ABM and metascience literature.

Declarations

Conflict of interest statement: This research was not supported by any grant. The authors have no competing interests to declare that are relevant to the content of this article.

Ethical Statement and Informed Consent: The current paper did not involve any human or animal data collection and did not involve ethical approval or informed consent.

Data availability statement: Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

Ataya, A. F., Adams, S., Mullings, E., Cooper, R. M., Attwood, A. S., & Munafò, M. R. (2012). Internal reliability of measures of substance-related cognitive bias. *Drug and Alcohol Dependence*, *121*(1–2), 148–151. https://doi.org/10.1016/j.drugalcdep.2011.08.023

Balkenius, C., & Hulth, N. (1999). Attention as selection-for-action: A scheme for active perception. *1999 Third European Workshop on Advanced Mobile Robots (Eurobot'99). Proceedings (Cat. No.99EX355)*, 113–119. https://doi.org/10.1109/EURBOT.1999.827629

Bantin, T., Stevens, S., Gerlach, A. L., & Hermann, C. (2016). What does the facial dot-probe task tell us about attentional processes in social anxiety? A systematic review. *Journal of Behavior Therapy and Experimental Psychiatry*, *50*. https://doi.org/10.1016/j.jbtep.2015.04.009

Bartoš, F., & Schimmack, U. (2020). *Z-curve. 2.0: Estimating replication rates and discovery rates*.

Bayley, J. E., & Phipps, D. (2019). Building the concept of research impact literacy. *Evidence & Policy: A Journal of Research, Debate and Practice*, *15*(4), 597–606. https://doi.org/10.1332/174426417X15034894876108

Bomyea, J., Johnson, A., & Lang, A. J. (2017). Information Processing in PTSD: Evidence for Biased Attentional, Interpretation, and Memory Processes. *Psychopathology Review*, *a4*(3), 218–243. https://doi.org/10.5127/pr.037214

Bradley, B. P., Mogg, K., Falla, S. J., & Hamilton, L. R. (1998). Attentional Bias for Threatening Facial Expressions in Anxiety: Manipulation of Stimulus Duration. *Cognition & Emotion*, *12*(6), 737–753. https://doi.org/10.1080/026999398379411

Bradley, M. M. (2009). Natural selective attention: Orienting and emotion. *Psychophysiology*, *46*(1), 1–11. https://doi.org/10.1111/j.1469-8986.2008.00702.x

Broadbent, D. E. (1958). The effects of noise on behaviour. In *Perception and communication* (pp. 81–107). Pergamon Press. https://doi.org/10.1037/10037-005

Brown, H. M., Eley, T. C., Broeren, S., MacLeod, C. M., Rinck, M., Hadwin, J. A., & Lester, K. J. (2014). Psychometric properties of reaction time based experimental paradigms measuring anxiety-related information-processing biases in children. *Journal of Anxiety Disorders*, *28*(1), 97–107. https://doi.org/10.1016/j.janxdis.2013.11.004

Chapman, A., Devue, C., & Grimshaw, G. M. (2017). Fleeting reliability in the dot-probe task. *Psychological Research*. https://doi.org/10.1007/s00426-017-0947-6

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. https://doi.org/10.1016/j.drugalcdep.2015.07.672

Cisler, J. M., & Koster, E. H. W. (2010). Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. *Clinical Psychology Review*, *30*(2), 203–216. https://doi.org/10.1016/j.cpr.2009.11.003

Cristea, I. A., Kok, R. N., & Cuijpers, P. (2015). Efficacy of cognitive bias modification interventions in anxiety and depression: Meta-analysis. *The British Journal of Psychiatry*, *206*(1), 7–16. https://doi.org/10.1192/bjp.bp.114.146761

Cristea, I. A., Kok, R. N., & Cuijpers, P. (2016). The effectiveness of cognitive bias modification interventions for substance addictions: A meta-analysis. *PloS One*, *11*(9), e0162226. https://doi.org/10.1371/journal.pone.0162226

De Schryver, M., Hughes, S., Rosseel, Y., & De Houwer, J. (2016). Unreliable Yet Still Replicable: A Comment on LeBel and Paunonen (2011). *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.02039

Dear, B. F., Sharpe, L., Nicholas, M. K., & Refshauge, K. (2011). The psychometric properties of the dot-probe paradigm when used in pain-related attentional bias research. *The Journal of Pain : Official Journal of the American Pain Society*, *12*(12), 1247–1254. https://doi.org/10.1016/j.jpain.2011.07.003

Dehue, T. (1995). *Changing the rules: Psychology in the Netherlands, 1900–1985* (pp. ix, 204). Cambridge University Press.

Evans, T. M., Bira, L., Gastelum, J. B., Weiss, L. T., & Vanderford, N. L. (2018). Evidence for a mental health crisis in graduate education. *Nature Biotechnology 2018 36:3*, *36*(3), 282–284. https://doi.org/10.1038/nbt.4089

Fang, F. C., & Casadevall, A. (2015). Competitive Science: Is Competition Ruining Science? *Infection and Immunity*. https://doi.org/10.1128/IAI.02939-14

Field, M., & Cox, W. M. (2008). Attentional bias in addictive behaviors: A review of its development, causes, and consequences. *Drug and Alcohol Dependence*, *97*(1–2), 1–20. https://doi.org/10.1016/j.drugalcdep.2008.03.030

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

767–780. https://doi.org/10.1037/hea0000405

Fodor, L. A., Georgescu, R., Cuijpers, P., Szamoskozi, Ş., David, D., Furukawa, T. A., & Cristea, I. A. (2020). Efficacy of cognitive bias modification interventions in anxiety and depressive disorders: A systematic review and network meta-analysis. *The Lancet Psychiatry*, *7*(6), 506–514. https://doi.org/10.1016/S2215-0366(20)30130-9

Frith, U. (2020). Fast Lane to Slow Science. *Trends in Cognitive Sciences*, *24*(1), 1–2. https://doi.org/10.1016/j.tics.2019.10.007

Gladwin, T. E. (2017). Negative effects of an alternating-bias training aimed at attentional flexibility: A single session study. *Health Psychology and Behavioral Medicine*, *5*(1), 41–56. https://doi.org/10.1080/21642850.2016.1266634

Gladwin, T. E. (2018). Educating students and future researchers about academic misconduct and questionable collaboration practices. *International Journal for Educational Integrity*, *14*(1), 10. https://doi.org/10.1007/s40979-018-0034-9

Gladwin, T. E., & Figner, B. (2019). Trial-to-trial carryover effects on spatial attentional bias. *Acta Psychologica*, *196*, 51–55. https://doi.org/10.1016/J.ACTPSY.2019.04.006

Gladwin, T. E., Halls, M., & Vink, M. (2021). Experimental control of conflict in a predictive visual probe task: Highly reliable bias scores related to anxiety. *Acta Psychologica*, *218*, 103357. https://doi.org/10.1016/j.actpsy.2021.103357

Grafton, B., Teng, S., & MacLeod, C. M. (2021). Two probes and better than one: Development of a psychometrically reliable variant of the attentional probe task. *Behaviour Research and Therapy*, *138*, 103805. https://doi.org/10.1016/j.brat.2021.103805

Heitmann, J., Jonker, N. C., & de Jong, P. J. (2021). A Promising Candidate to Reliably Index Attentional Bias Toward Alcohol Cues–An Adapted Odd-One-Out Visual Search Task. *Frontiers in Psychology*, *12*, 102. https://doi.org/10.3389/fpsyg.2021.630461

Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J.-H., & Welsh, T. N. (2019). No one knows what attention is. *Attention, Perception, & Psychophysics*, *81*(7), 2288–2303. https://doi.org/10.3758/s13414-019-01846-w

Huistra, P., & Paul, H. (2022). Systemic Explanations of Scientific Misconduct: Provoked by Spectacular Cases of Norm Violation? *Journal of Academic Ethics*, *20*(1), 51–65. https://doi.org/10.1007/s10805-020-09389-8

Iacoviello, B. M., Wu, G., Abend, R., Murrough, J. W., Feder, A., Fruchter, E., Levinstein, Y., Wald, I., Bailey, C. R., Pine, D. S., Neumeister, A., Bar-Haim, Y., & Charney, D. S. (2014). Attention bias variability and symptoms of posttraumatic stress disorder. *Journal of Traumatic Stress*, 27(2), 232–239. https://doi.org/10.1002/jts.21899

Jiang, M. Y. w., & Vartanian, L. R. (2018). A review of existing measures of attentional biases in body image and eating disorders research. *Australian Journal of Psychology*, 70(1), 3–17. https://doi.org/10.1111/ajpy.12161

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*, 23(5), 524–532. https://doi.org/10.1177/0956797611430953

Jones, A., Christiansen, P., & Field, M. (2018). Failed attempts to improve the reliability of the alcohol visual probe task following empirical recommendations. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors*, 32(8), 922–932. https://doi.org/10.1037/adb0000414

Kappenman, E. S., Farrens, J. L., Luck, S. J., & Proudfit, G. H. (2014). Behavioral and ERP measures of attentional bias to threat in the dot-probe task: Poor reliability and lack of correlation with anxiety. *Frontiers in Psychology*, 5(DEC), 1368. https://doi.org/10.3389/fpsyg.2014.01368

Keren, G. (2013). A tale of two systems: A scientific advance or a theoretical stone soup? Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science*, 8(3), 257–262. https://doi.org/10.1177/1745691613483474

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, 4(6), 533–550. https://doi.org/10.1111/j.1745-6924.2009.01164.x

Koster, E. H. W., Crombez, G., Verschuere, B., Van Damme, S., & Wiersema, J. R. (2006). Components of attentional bias to threat in high trait anxiety: Facilitated engagement, impaired disengagement, and attentional avoidance. *Behaviour Research and Therapy*, 44(12), 1757–1771. https://doi.org/10.1016/j.brat.2005.12.011

Kruijt, A.-W., Parsons, S., & Fox, E. (2019). A meta-analysis of bias at baseline in RCTs of attention bias modification: No evidence for dot-probe bias towards threat in clinical anxiety and PTSD. *Journal of Abnormal Psychology*, 128(6), 563–573. https://doi.org/10.1037/abn0000406

Lang, P. J. (1995). The emotion probe. Studies of motivation and attention. *The American Psychologist*, 50(5), 372–385.

Linden, A. H., & Hönekopp, J. (2021). Heterogeneity of research results: A new perspective from which to assess and promote progress in psychological science. *Perspectives on Psychological Science*, *16*, 358–376. https://doi.org/10.1177/1745691620964193

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

Martin, B. (2016). Plagiarism, misrepresentation, and exploitation by established professionals: Power and tactics. In *Handbook of Academic Integrity* (pp. 913–927). Springer Singapore. https://doi.org/10.1007/978-981-287-098-8_75

Maxwell, J. W., Fang, L., & Carlson, J. M. (2022). Do Carryover Effects Influence Attentional Bias to Threat in the Dot-Probe Task? *Journal of Trial & Error*, *2*(1). https://doi.org/10.36850/e9

McNally, R. J. (2018). Attentional bias for threat: Crisis or opportunity? *Clinical Psychology Review*, *69*, 4–13. https://doi.org/10.1016/J.CPR.2018.05.005Siri

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*, 216770261769635. https://doi.org/10.1177/2167702617696359

Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review*, *75*(6), 522–536. https://doi.org/10.1037/h0026699

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42. https://doi.org/10.1146/annurev.ne.13.030190.000325

Puls, S., & Rothermund, K. (2018). Attending to emotional expressions: No evidence for automatic capture in the dot-probe task. *Cognition and Emotion*, *32*(3), 450–463. https://doi.org/10.1080/02699931.2017.1314932

Rodebaugh, T. L., Scullin, R. B., Langer, J. K., Dixon, D. J., Huppert, J. D., Bernstein, A., Zvielli, A., & Lenze, E. J. (2016). Unreliability as a threat to understanding psychopathology: The cautionary tale of attentional bias. *Journal of Abnormal Psychology*, *125*(6), 840–851. https://doi.org/10.1037/abn0000184

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66. https://doi.org/10.1037/0033-295X.84.1.1

Simmons, J. M., Minamimoto, T., Murray, E. A., Richmond, B. J., Bouret, S., Kennerley, S. W., & Wallis, J. D. (2011). Reward-Related Neuronal Activity During Go-Nogo

Task Performance in Primate Orbitofrontal Cortex Reward-Related Neuronal Activity During Go-Nogo Task Performance in Primate Orbitofrontal Cortex. *Journal of Neurophysiology*, 1864–1876.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Starzomska, M. (2017). Applications of the dot probe task in attentional bias research in eating disorders: A review. *Psicológica*, *38*(2), 283–346.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220–247. https://doi.org/10.1207/s15327957pspr0803

Tiggemann, M., & Kemps, E. (2020). Is Sham Training Still Training? An Alternative Control Group for Attentional Bias Modification. *Frontiers in Psychology*, *11*. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.583518

Tijdink, J. K., Horbach, S. P. J. M., Nuijten, M. B., & O'Neill, G. (2021). Towards a Research Agenda for Promoting Responsible Research Practices. *Journal of Empirical Research on Human Research Ethics*, *16*(4), 450–460. https://doi.org/10.1177/15562646211018916

Tiokhin, L., Yan, M., & Morgan, T. J. H. (2021). Competition for priority harms the reliability of science, but reforms can help. *Nature Human Behaviour*, *5*(7), 857–867. https://doi.org/10.1038/s41562-020-01040-1

Townshend, J. M., & Duka, T. (2001). Attentional bias associated with alcohol cues: Differences between heavy and occasional social drinkers. *Psychopharmacology*, *157*(1), 67–74. https://doi.org/10.1007/s002130100764

Treisman, A. M. (1964). Selective attention in man. *British Medical Bulletin*, *20*(1), 12–16.

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

Vervoort, L., Braun, M., De Schryver, M., Naets, T., Koster, E. H. W., & Braet, C. (2021). A Pictorial Dot Probe Task to Assess Food-Related Attentional Bias in Youth With and Without Obesity: Overview of Indices and Evaluation of Their Reliability. *Frontiers in Psychology*, *12*, 561. https://doi.org/10.3389/fpsyg.2021.644512

Waechter, S., Nelson, A. L., Wright, C., Hyatt, A., & Oakman, J. (2014). Measuring attentional bias to threat: Reliability of dot probe and eye movement indices. *Cognitive Therapy and Research*, *38*(3), 313–333. https://doi.org/10.1007/s10608-013-9588-2

Wald, I., Degnan, K. A., Gorodetsky, E., Charney, D. S., Fox, N. A., Fruchter, E., Goldman, D., Lubin, G., Pine, D. S., & Bar-Haim, Y. (2013). Attention to threats and combat-related posttraumatic stress symptoms: Prospective associations and moderation by the serotonin transporter gene. *JAMA Psychiatry*, *70*(4), 401–408. https://doi.org/10.1001/2013.jamapsychiatry.188

Wiers, C. E., Gladwin, T. E., Ludwig, V. U., Gröpper, S., Stuke, H., Gawron, C. K., Wiers, R. W., Walter, H., & Bermpohl, F. (2016). Comparing three cognitive biases for alcohol cues in alcohol dependence. *Alcohol and Alcoholism*, *1*(7), 1–7. https://doi.org/10.1093/alcalc/agw063

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), e104025. https://doi.org/10.1371/journal.pone.0104025

Appendices

A. Additional notes on assessing psychometric properties of the dot-probe task

We briefly note that it is essential that the reliability of the bias scores is assessed correctly. If the interest is in a *bias*, the relevant reliability is that of the *difference score* between reaction times on different conditions measuring that bias. For instance, consider the reaction times to probes at threat locations, and reaction times to probes at control locations. The reliability of the simple reaction times per condition separately, or the reliability of the reaction time averaged over both conditions, do not assess the reliability of the bias. In contrast, the reliability of simple or averaged reaction times will trivially reflect systematic variability due to participants being generally fast or slow. Taking the difference score per participant and assessing its reliability is necessary to assess the reliability of the bias.

Further, we emphasize that low reliability, while critical to studies involving individual differences, does not impact the replicability or size of within-subject effects (De Schryver et al., 2016). To consistently measure a within-subject effect, any between-subject variability in the effect would ideally be low, while the reliability of an individual difference measure benefits from high between-subject variability of individuals' true scores (reducing between-subject error, of course, benefits both aims). However, there are also concerns even with the robustness of such relatively simple, overall effects as well (Puls & Rothermund, 2018). For instance, in a re-analysis of pre-test data originally used for intervention studies, there was no significant bias towards threat in clinically anxious individuals, where it would have been expected

(Kruijt et al., 2019). This was an exceptional finding as the primary aim of these studies involved testing effects over time of interventions, rather than biases at pre-test, mitigating the most likely effects of questionable research practices (John et al., 2012; J. M. Simmons et al., 2011).

B. From the dot-probe task to Attention Bias Modification (ABM)

We briefly note that issues involving assessment tasks may also affect the course of research on derived training paradigms such as ABM. In ABM, the dot-probe task is converted to a training intervention. This is done by manipulating the probability of the location of the probe stimulus, in order to train attention towards versus away from particular stimulus categories. For example, in a task with alcoholic and non-alcoholic beverages, the probe would be presented at the non-alcoholic location, in hopes of reducing an assumed attentional bias towards alcohol. While the current article focuses on the assessment form of the dot-probe task, results of directional ABM versus control conditions have arguably been similarly disappointing (Cristea et al., 2015, 2016; Fodor et al., 2020; Mogg et al., 2017). The rationale of the dot-probe task suggests specifically directional training manipulations - "towards" versus "away". The control training condition would then be random, with probe location being unrelated to cue locations, or trained in the opposite direction as desirable from a clinical perspective. An issue raised as a possible reason for weak effects of ABM is control conditions are, in fact, still serving a training purpose (Tiggemann & Kemps, 2020); one possibility is that participants are learning that the cues are irrelevant (Gladwin, 2017). Similarly, despite the directionality, training attention "away" from a given stimulus category still makes cue locations task-relevant and hence maintains, or induces, the salience of the cues. Interestingly, these concerns would seem to have been more immediately evident from the perspective of the emotional Stroop task. It seems likely that the focus of ABM, if it had been linked primarily to that task, would have shifted towards training aimed at the reinterpretation of salient stimuli as irrelevant. It is of course unknown whether this would have produced more clinically relevant results.

C. The dot-probe task as a crowbar into scientific structure

That is: it may seem that methodological debate around the dot-probe task merely concerns a rather localized issue of suboptimal scientific behaviours around one particular implicit measure. However, it may be useful to consider a connection to concerning scientific events that are more "spectacular" (Huistra & Paul, 2022), such as the Stapel affair and continued subsequent (inter)nationally publicized cases of PIs committing fraud or other unethical breaches. At least some academics have considered systemic explanations for these events, rather than merely issues related to the individual (Huistra & Paul, 2022). Less conspicuous problems that exist within business as usual - such as the current topic, but also questionable research practices in general (J. P. Simmons et al., 2011), academic exploitation (Gladwin, 2018; Martin, 2016), and mental health crises in PhD students (Evans et al., 2018) - could similarly warrant questioning of scientific systems. It seems at least conceivable that there are

common underlying causes involving the way in which science is currently organized. A detailed analysis of the history of the dot-probe task could therefore be of interest from a philosophy of science perspective. How do methods and ideas spread, become acquired, and suppressed, and how do such processes relate to the distribution, acquisition and maintenance of institutional power? Or, put differently, what are the assumed and perhaps unscrutinised organizing structures and "rules of the game" (Dehue, 1995) and what are their consequences?