**Letter to the Editors of *Psychological Science*: Considerations When Investigating Metacognitive Sensitivity: Regarding Miller et al. (2023)**

Building on evidence that synthetic, AI-generated face images are indistinguishable from real face photographs (e.g., Nightingale & Farid, 2022), a recent study by Miller and colleagues (2023) found that these synthetic (specifically White) faces were judged as human more often than actual faces. The researchers also considered participants’ insights into their detection errors, demonstrating a positive association between performance (cognitive sensitivity, *d’*) and insight (metacognitive sensitivity, meta-*d’*). This was interpreted as good (poor) performers showing good (poor) insight. However, some important issues regarding the application of this approach and interpretation of these results were not considered.

In order to fit signal detection theoretic models and therefore estimate metacognitive sensitivity, we must assume that stimulus strength or task difficulty is held roughly constant. That way, fluctuations in accuracy and confidence are the result of noise internal to the participant rather than external differences in signal strength (Fleming & Lau, 2014). Put simply, trial difficulties should be approximately equal. However, the 200 trials used by Miller and colleagues to test real/synthetic face discrimination showed substantial variation in difficulty (accuracy *M* = 0.41, *SD* = 0.18, range = 0.10-0.92; based on the original data of Nightingale & Farid, 2022). In addition, these 200 trials were divided into two tests based on image gender, with participants only completing one test. As such, alongside the variation in trial difficulties, half of the participants completed an entirely different set of trials to the other half, making a combined analysis of insight inappropriate.

Further, participants provided a “human” or “AI” response on each trial (similar to responding “yes” versus “no” in a simple detection task). In these circumstances, we might expect asymmetries in metacognitive sensitivity for these two response options (e.g., Fleming & Dolan, 2010) and so response-conditional meta-*d’* estimation is the more suitable approach, i.e., fitting models to estimate meta-*d’* for each type of response separately.

Finally, even if the task and modelling approach were suitable, the analytical focus was on the correlation between *d’* and meta-*d’* (*r* = .48), with the interpretation that lower insight was associated with poorer performance. However, meta-*d’* indexes the information accessed by metacognition and so we should expect a strong, positive relationship with *d’* if metacognitive performance is based on (and, indeed, constrained by) the same information as cognitive performance (Maniscalco & Lau, 2012). Instead, it would be more informative to consider the association between performance and the quality of metacognitive processing (termed ‘metacognitive efficiency’, given by meta-*d’*/*d’*). Analysing Miller and colleagues’ data produced a negligible correlation between these measures (*r* = .07, or .05 after removing extreme, misleading values; see McIntosh et al., 2022) and so failed to support the idea that the quality of metacognitive processing was worse among poor performers after accounting for differences in the quality of information available to metacognition.

To conclude, the use of meta-*d’* may allow researchers to tackle important questions involving insight into face perception tasks. However, the application and interpretation of this approach require careful consideration.

Robin S. S. Kramer

*School of Psychology, University of Lincoln*

[rkramer@lincoln.ac.uk](mailto:rkramer@lincoln.ac.uk)

**References**

Fleming, S. M., & Dolan, R. J. (2010). Effects of loss aversion on post-decision wagering: Implications for measures of awareness. *Consciousness and Cognition, 19*(1), 352-363.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience, 8*, 443.

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition, 21*(1), 422-430.

McIntosh, R. D., Moore, A. B., Liu, Y., & Della Sala, S. (2022). Skill and self-knowledge: Empirical refutation of the dual-burden account of the Dunning–Kruger effect. *Royal Society Open Science, 9*(12), 191727.

Miller, E. J., Steward, B. A., Witkower, Z., Sutherland, C. A., Krumhuber, E. G., & Dawel, A. (2023). AI hyperrealism: Why AI faces are perceived as more real than human ones. *Psychological Science*. Advance online publication.

Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences, 119*(8), e2120481119.