Forensic Science International: Synergy 2 (2020) 705–707

Contents lists available at ScienceDirect

Forensic Science International: Synergy

journal homepage: https://www.journals.elsevier.com/forensic-science-international-synergy/

Author response: No need for throwing stones – Wherever you live…

We welcome a constructive debate on the merits of rigorous and ecologically valid research on cognitive bias in forensic decisions. In our editorial we tried to highlight the need for a deeper understanding of cognitive bias by drawing on a recent literature review by Cooper and Meterko [1]. We acknowledge the important work Kukucka and other colleagues have made in their effort to identify and understand bias. However, we have reservations regarding his interpretation of our methodological criticisms and of our position regarding the research on bias.

In his refutation, Kukucka [2] argues that we misrepresented the methodological rigor of existing research on cognitive bias. In addition, he claims that we endorsed a problematic interpretation of bias in forensic decision making. We will respond to each point in turn.

1. Reactivity

In contrast to Kukucka's claim that only 1 of the 29 studies in Cooper and Meterko's [1] review show critical deficiencies, we would stand by our original assessment that a number of studies did not have adequate blinding procedures. Cooper and Meterko [1] repeatedly raised the issue of blinding in their review. On page 42 they write: "As noted previously, we identified two studies with critical deficiencies relating to participant blinding and to comparability of groups. Among the other studies, design elements fell short of what we considered 'ideal', either because important details were not reported or because specific standards were not met." They further state: "The extent of blinding also varied among the studies, with some studies designed in a way that prevented participants from knowing that a particular sample was a study sample, and other studies only blinding participants to the specific hypothesis or to group allocation." Their table in the appendix includes "evidence of a failure of participant blinding" and "evidence of failure of blinding, affecting participant behaviour or assessors' decision-making", suggesting that issues around blinding and/or reporting of blinding went beyond the single study acknowledged by Kukucka. Besides, it is always possible to raise standards. In medical or pharmaceutical research "double-blind procedures", where neither the participant nor the experimenter knows which condition has been administered, are seen as the "gold standard".

2. Randomization and generalizability

Kukucka correctly states that Cooper and Meterko [1] concluded that only one reviewed study showed critical deficiencies in randomization. However, in our editorial we wrote: "None of the 29 reviewed studies provided information about randomization of trials" which echoes the following statement by Cooper and Meterko [1]: "none of the studies provided information about the randomization procedures". We argue that information about randomization of trials and conditions would reassure readers that carry-over/order effects were avoided.

Our comment about differences between experimental and control groups refers to the lack of detailed demographic information. Kukucka states "As every scientist knows, participants must be randomly assigned to conditions in order to establish causality; in theory, this practice creates two groups that are, on average, identical apart from the experimental manipulation". This statement is surprising, because in the previous paragraph Kukucka argues that a lack of reporting demographics is hardly a "fatal flaw". Yet, how can we rest assured that experimental and control groups are equivalent when no demographic information about the groups is available? We stated accordingly: "Many of the reviewed studies did not address differences between experimental and control groups".

Kukucka further asserts that "There is simply no evidence that cognitive bias is moderated by a person's age, sex or race", citing a paper on demand characteristics that is not an empirical study on demographic differences [4]; and was published in 1962. This misses the point as we need to check whether randomization was successful. For smaller sample sizes stratified sampling or matched sampling rather than convenience sampling is often the preferred randomization strategy [5]. We stand by our original claim and suggest that research into forensic bias should make an effort to report the demographic variables of participants in different groups; particularly since reported sample sizes have been relatively small.

Kukucka tries to convince the reader that sample sizes have increased over time by performing a linear regression on 19 data points. Analysing this data is problematic, not only because the number of studies is low, but also because one outlier (a recent study with N = 192 deviates almost 4 SDs from the mean) drives the significant increase. Kukucka examined only total sample sizes, regardless of the number of groups in different research designs. He included studies from 2006-onwards but not earlier studies (e.g., two studies from 1987 and 1984). Nevertheless, if we repeat the analysis without the outlier then publication date is no longer a statistically significant predictor of total sample size ($b = 3.16$, $t = 1.73$, $p = .103$). His conclusion that "sample sizes have steadily increased over time" is therefore unconvincing, if not misleading.

In the following we assess the two studies in Kukucka and Kassin [6] to illustrate some of our methodological concerns. We acknowledge, however, that different studies warrant different criticisms.

3. Sampling

We stated that Kukucka and Kassin [6] did not report

https://doi.org/10.1016/j.fsisyn.2020.03.008

demographic information about their participants beyond their status as psychology undergraduate students. Obviously, this was an error as we mixed up information about the pilot study with Study 1 and we apologise for this mistake. Nevertheless, Kukucka and Kassin [6] did not provide demographics for the between-subjects groups. As a consequence, there is little reassurance that their random assignments produced comparable groups. Study 1 used a "snowball sample" using social media whereas Study 2 employed representative sampling from the general population. In Study 2, the sampling from the general population may be representative for jurors but is unlikely to generalize to handwriting examiners. Although the group sizes seem sufficiently large, no power analysis was conducted to determine optimal participant numbers. Especially, in Study 2 there was unexplained variability of responses at Time 1 (before any contextual information had been shown) across the three groups, indicating considerable group differences. Statistically significant differences between Time 1 (before any contextual information had been shown) and Time 2 (confession-present) were observed for similarity and matching, but the non-significant difference in the similarity variable at Time 2 between the control, confession absent group and the confession-present group remains unexplained.

4. Multiple testing

Three dependent variables were collected together with two additional confidence ratings. Participants rated in succession "similarity of handwritings", "possible match" and "guilt". This makes each participant "judge and jury", possibly inflating correspondences between similarity ratings and subsequent judgments on possible matches and the guilt of the suspect. Using three dependent variables also increases the chances of finding statistically significant effects through multiple testing [7]. No adjustments of significance levels are reported (e.g., Bonferroni, Newman-Keuls).

5. Accuracy and ecological validity

In Study 1, two pairs of handwritings with low similarity and two pairs with high similarity were selected, based on ratings of 15 pairs in the pilot study. Throughout the paper the authors do not take into account that the four selected handwriting pairs were from different persons; whereas the only pair that was authored by the same person was not included. When making decisions, forensic examiners can make four types of responses to targets (hits, misses) and foils (correct rejections, false alarms). Without any responses to targets (hit, misses) it is difficult to evaluate decision performance of participants because only one type of error and correct response is observed [8]. The same argument applies to Study 2. In conclusion, these shortcomings cast doubts on the interpretation and ecological validity of the results.

On the second point - the alleged misinterpretation of cognitive bias - we refer the reader to our recent reply to a letter by Thompson (see Curley et al. [9,10], and Thompson [11]). In this reply we acknowledge that bias, regardless of the interpretation by individuals or organisations, does exist in forensic decision making. We point out that task-irrelevant contextual information is difficult to define, especially in terms of probative value in multi-phase forensic decisions, and that we do not endorse the use of task-irrelevant contextual information in forensic decisions. We acknowledge the problems associated with cognitive bias, but do not necessarily see it as a "scourge" that can and needs to be eradicated in every instance. On the contrary, we believe that certain biases which have no or a positive impact on accuracy should be explored in more detail. We believe that a perception of bias that is too narrow (e.g., bias can only be negative) comes from a lack of appreciation surrounding bounded rationality and decision making literature in general (Gigerenzer & Brighton [12]).

We also note that Kukucka did not comment on our statement that "Several results sections did not include basic statistical information such as effect size, measures of variability and/or inferential test statistics; some *worryingly* even conducted inappropriate statistical tests" [9]. Indeed, in his letter, Kukucka treats every methodological issue separately from each other, as if, for example, issues with randomization should not be considered alongside participant blinding or sample sizes. In reality, these issues interact with and conflate one another and consequently should be treated together. We stand by our conclusion that these problems are significant and are worth considering when interpreting evidence of cognitive bias in forensic decision making.

6. Summary

We believe that research on cognitive bias needs to move on from illustrating the existence of bias in different examples to exploring and quantifying the underlying causes and effects in more ecologically valid settings (maybe inspiration can be drawn from the seminal work of Itiel Dror in relation to creating more ecologically valid studies [13]: How strong is the effect of contextual information for different examiners and probes in various scenarios? Where to draw the line between task-relevant and task-irrelevant contextual information? When are participants representative of a population of forensic examiners and when is the selection and administration of stimuli/cases ecologically valid? Further, more research into jurors and how they integrate forensic information is needed - particularly when they are aware that forensic examiners used task-irrelevant information in their decisions.

These questions suggest renewed research efforts. It seems obvious that simply manipulating contextual information (present, absent) in different forensic domains (fingerprint, dental, footprint, tools, DNA) and observing responses in diverse samples (practitioners, trainees, naïve participants) is insufficient to understand the causes and effects of cognitive bias. New studies may specify random factors that contribute to overall variability (participants, stimuli/cases) using improved measures, designs and analyses in order to quantify individual as well as stimulus-specific bias. Crossed/mixed-effect designs and hierarchical models with more ecologically valid material and samples may be a way forward.

Of course, it is easy to criticise limitations and shortcomings of previous studies rather than to conduct new research. However, we don't see this as "throwing stones" as Kukucka puts it. Instead, by engaging in constructive debate, building on past research, and suggesting new avenues we are trying to encourage and facilitate scientific progress.

Declaration of competing interest

The authors have no conflict of interest to declare.

References

[1] G.S. Cooper, V. Meterko, Cognitive bias research in forensic science: a systematic review, Forensic Sci. Int. (2019), https://doi.org/10.1016/j.forsciint.2019.01.016.
[2] J. Kukucka, People Who Live in Ivory Towers Shouldn't Throw Stones: a Refutation of Curley et al, Forensic Sci. Int.: Syn, 2020.
[4] M.T. Orne, On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications, Am. Psychol. 17 (1962) 776–783.
[5] S. Singh, Advanced sampling theory with applications, in: How Michael ŝelected Ămy, vol. 1, Kluwer Ac. Pub (Springer, Dordrecht, 2003.
[6] J. Kukucka, S.M. Kassin, Do confessions taint perceptions of handwriting

evidence? An empirical test of the forensic confirmation bias, Law Hum. Behav. 38 (3) (2014) 256, https://doi.org/10.1037/lhb0000066.
[7] P. Dattalo, Analysis of Multiple Dependent Variables. New York, NY: Oxford University Press.
[8] T.D. Wickens, Elementary Signal Detection Theory, Oxford University Press, Oxford, 2001.
[9] L.J. Curley, J. Munro, M. Lages, R. MacLean, J. Murray, Assessing cognitive bias in forensic decisions: a review and outlook, J. Forensic Sci. (2020), https://doi.org/10.1111/15564029.14220.
[10] L.J. Curley, J. Munro, M. Lages, R. MacLean, J. Murray, Authors' Response: is the definition of task-irrelevant contextual information black and white? J. Forensic Sci. 2020 (2020) https://doi.org/10.1111/1556-4029.14273.
[11] W.C. Thompson, Commentary on: Curley Lj, J. Munro, M. Lages, R. MacLean, J. Murray, Assessing cognitive bias in forensic decisions: a review and outlook, J. Forensic Sci. (2020), https://doi.org/10.1111/15564029.14220.
[12] G. Gigerenzer, H. Brighton, Homo heuristicus: why biased minds make better inferences, Top. Cognit. Sci. 1 (1) (2009) 107–143, https://doi.org/10.1111/j.1756-8765.2008.01006.x.
[13] I. E. Dror, D. Charlton, A.E. Péron, Contextual Information Renders Experts Vulnerable to making erroneous identifications. Forensic Sci. Int., 156(1), 74-78, 10.1016/j.forsciint.2005.10.017

Lee J. Curley*
Faculty of Arts and Social Sciences, School of Psychology and Counselling, The Open University, Milton Keynes, UK

James Munro
School of Applied Sciences, Edinburgh Napier University, Edinburgh, UK

Martin Lages
College of Science and Engineering, The School of Psychology, The University of Glasgow, Glasgow, UK

* Corresponding author. Faculty of Arts and Social Sciences, School of Psychology, the Open University, Milton Keynes, UK.
E-mail address: Lee.Curley@open.ac.uk (L.J. Curley).

Available online 1 April 2020