Commentary on: J., Cesario (forthcoming). What can experimental studies of bias tell us about real-world group disparities? *Behavioral and Brain Sciences.*
An edited version of this commentary will be published in *Behavioral and Brain Sciences.*

External validity of social psychological experiments is a concern, but these models are useful

Youri L. Mora[1]*, Olivier Klein[1], Christophe Leys[1], Annique Smeding[2]

We agree that external validity of social psychological experiments is a concern, we disagree these models are useless. Experiments, reconsidered from a situated cognition perspective and non-linearly combined with other methods (qualitative, simulations) allow grasping decision dynamics beyond bias outcomes. Dynamic (vs. discrete) insights regarding these processes are key to understand missing forces and bias in real-world social groups.

In this commentary, we aim at extending Cesario's critique on the "use of experimental social psychology to explain real-world group disparities". While we agree that focusing on average bias may impair external validity in significant ways, we disagree with the lesson Cesario draws from the use of experimental paradigms: Using such paradigms should be a tool of last resort to explain real-world disparities. Indeed, situated cognition experiments have already demonstrated their usefulness in shedding light upon the decision dynamics involved in bias, and not only on the *presence* of bias, which is Cesario's focus. We will develop why drawing Cesario's lesson would amount to throwing the baby out with the bathwater and will propose alternative solutions to his.

First, experimental paradigms in social psychology serve as models of the real world and as such (i) are by definition a simplification (parsimony in modelling); but (ii) are still useful to understand the psychological processes (Smaldino, 2017) that drive behaviour in the "real world". This concern is not new and is aptly illustrated by George Box's aphorism (1979, p. 202) « All models are wrong, but some are useful ». What Cesario (2021) is essentially saying is that current models in experimental social psychology are wrong and not even useful to explain real-world disparities. The three flaws accurately identified by Cesario are actually three types of missing variables – moderators – whose absence is involved in deterioration of external validity. One can reformulate this criticism as suggesting that the studied effect sizes are smaller than the smallest effect size of interest when more ecological variables – absent in most experimental models – are taken into account.

[1] Université libre de Bruxelles
[2] Univ. Savoie Mont Blanc
* Corresponding author can be reached at youri.mora@ulb.be.

Appraising the problem through this lens leads us to disagree with Cesario's idea that the « research is *fundamentally* flawed ». Still, the question remains: How to design experiments that yield more robust and meaningful effect sizes while accounting for these missing variables and which are applicable to real-life situations? The three identified flaws raise questions whose answers can help incrementally elaborate models that include crucial moderators.

Second, and related, while demonstrating average bias in "the general population" (often undergraduate, non-expert psychology students) with indirect measures like the Implicit Association Test (IAT) may fall short in accounting for real-world group disparities, using such paradigms to understand differential processes in real-world social groups has proven valuable. We were surprised that Cesario's section on implicit bias and STEM-related IAT did not refer to lines of research grounded in situated social cognition (Smith & Semin, 2007) or Freeman's work on social perception (Freeman, 2014; Freeman et al., 2016). For instance, using a mouse-tracking adapted gender-math IAT, completed by female and male STEM and non-STEM majors, research (Smeding et al., 2016) has shown meaningful group differences in decision-making dynamics (i.e. attractions) and their early emergence in time (around 300 ms). Millisecond differences – or deviations in mouse trajectories while decision-making is unfolding – may thus represent one of those real-world forces that characterise real-world group disparities. These can be measured with experimental paradigms. Also, Cesario frames the interest about bias in decision-making in the minds of "gatekeepers" discriminating against ambiguous candidates. However, candidates themselves will be the first depleting link in the decision-making chain if they are biased by held stereotypes (e.g. Shapiro & Williams, 2012) or because of imperfect inferences drawn from observed regularities (Kutzner & Fiedler, 2017) ; leading attrition to occur way before formal selection by potential gatekeepers. By comparing engineering and humanities female students, Study 3 in Smeding et al. (2016) has shown that self-congruency trumps the role of stereotype-congruency in a "Math vs. Language" IAT. Self-congruency would here be categorised by Cesario (2021) as a "missing force". This moderator could not have been identified with a convenience sample. However, it can still be studied through an IAT paradigm, reconsidered from a situated social cognition approach. To explain underrepresentation of women in STEM majors, "men and women differ[ing] in their interest" (Cesario, 2021, p. 8) can be considered as a mere demographic difference across groups, discarding the role of decision bias. However, the key role of self-congruency shows that the underrepresentation phenomenon can still be explained by decision bias (in the minds of candidates), which an IAT can meaningfully investigate while providing leads to implement change (e.g. impact on self-domain related associations). Still, such investigation requires participants who are involved in the actual field of application and not convenience samples, as called by Cesario. Besides, mouse-tracking-based paradigms, anchored in dynamical systems theory (which has nurtured many real-life applications spanning from physics to cognitive science, Krpan, 2017) represent a user-friendly tool (Rivollier et al., 2020) that allows understanding (i) continuity and (nonlinear) competition in decision-making (beyond discrete judgments) and (ii) the influence of social category triggers specifically when people are ambiguous on a relevant real-world characteristic (Freeman, 2014; Freeman et al., 2016).

Finally, experiments – as one of the methods available to psychologists – in combination not solely with in-depth (qualitative) field research (as suggested by Cesario), but also computational modelling have the potential to provide insight into real-world human behaviour, including group disparities. In Smeding et al.

(2016), results for simulated social groups and real-world social groups were compared. While the former provided proof of concept regarding the (hypothesised) psychological processes, the latter sustained their real-world validity. Both contributed to a finer-grained understanding of sex differences in STEM engagement which, admittedly, seemed to be less related to average stereotypic bias than to differential associations related to the self. But an experimental paradigm like the IAT happened to be of paramount importance in such findings. Mixed-methods (including qualitative, experimental, correlational, observational, computational, but also emerging real-world data-driven approaches such as machine learning) would all greatly benefit the study of bias and group disparities in social psychology. However, their use in a research program is certainly nonlinear and more dynamic than the fixed sequence depicted by Cesario in his suggested new/rehashed approach.

Conflict of interest and funding statements

Authors declare no conflict of interest. This work was supported by the Fund for Scientific Research (F.R.S.–FNRS) under Grant No. 40000042.

References

Box, G. E. P. (1979). Robustness in the strategy of scientific model building. *Robustness in Statistics*, 201–236. Elsevier. https://doi.org/10.1016/B978-0-12-438150-6.50018-2

Cesario, J. (2021). What can experimental studies of bias tell us about real-world group disparities? *Behavioral and Brain Sciences*, 1-80. https://doi.org/10.1017/S0140525X21000017

Freeman, J. B. (2014). Abrupt category shifts during real-time person perception. *Psychonomic Bulletin & Review*, 21, 85–92. https://doi.org/10.3758/s13423-013-0470-8

Freeman, J., Pauker, K., & Sanchez, D. (2016). A perceptual pathway to bias. *Psychological Science*, *27*(4), 502–517. https://doi.org/10.1177/0956797615627418

Krpan, D. (2017). Behavioral priming 2.0: Enter a dynamical systems perspective. *Frontiers in Psychology*, *8*, 1204. https://doi.org/10.3389/fpsyg.2017.01204

Kutzner, F., & Fiedler, K. (2017). Stereotypes as pseudocontingencies. *European Review of Social Psychology*, *28*(1), 1–49. https://doi.org/10.1080/10463283.2016.1260238

Rivollier, G., Quinton, J.-C., Gonthier, C., & Smeding, A. (2020). Looking with the (computer) mouse: How to unveil problem-solving strategies in matrix reasoning without eye-tracking. *Behavior Research Methods*. https://doi.org/10.3758/s13428-020-01484-3

Shapiro, J. R., & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. *Sex Roles*, 66(3-4), 175–183. https://doi.org/10.1007/s11199-011-0051-0

Smaldino, P. E. (2017). Models are stupid, and we need more of them. R. R. Vallacher, S. J. Read, & A. Nowak (Eds.), *Computational Social Psychology* (1st ed.), 311–331. Routledge. https://doi.org/10.4324/9781315173726-14

Smeding, A., Quinton, J.-C., Lauer, K., Barca, L., & Pezzulo, G. (2016). Tracking and simulating dynamics of implicit stereotypes: A situated social cognition perspective. *Journal of Personality and Social Psychology*, *111*(6), 817–834. https://doi.org/10.1037/pspa0000063

Smith, E. R., & Semin, G. R. (2007). Situated social cognition. *Current Directions in Psychological Science*, *16*(3), 132-135. https://doi.org/10.1111/j.1467-8721.2007.00490.x