Review of: "Why are there different versions of the COM-B model diagram?"

William Lawless[1]

1 Paine College

Potential competing interests: No potential competing interests to declare.

Review: West & Michie's Why are there different versions of the COM-B model diagram? Qeios ID: AWANWG

William F. Lawless

I have completed my review of West & Michie's (2023) paper on the different versions of their COM-B model. While interesting, the model as presented in their paper is unable to stand on its own in their very short paper. It is likely to be unfair of me to critique this short version of their model, but it is more unfair for the authors to publish their model when it is incomplete.

Autonomous human-machine teams and systems

First, before my critique, I review our program of research so that readers have sufficient context to appreciate my perspective. Our goal is to craft the algorithms deployed with Artificial Intelligence (AI) in open systems by machines and humans to structure human-machine teams to be able to reach their maximum performance autonomously. Unlike the laboratory, in open systems, teams face complexity, uncertainty and conflict. Complexity in this new domain is primarily the environment and opponents, which are affected by uncertainty and conflict. In our research, we contrast the individual and interdependenceapproaches to teams. The individual approach focuses on building teams and systems by aggregating the best available information for individuals, their thoughts, behaviors and skills. Its theoretical background is characterized chiefly by one-to-one relations between mind and body (derived from Hume and Spinoza).

In Spinoza's view, he did not believe that the human being is a union of two substances(Nadler, 2022). Instead, Spinoza believed that the human mind and its human body are two different expressions, one under thought and another under extension, of one and the same thing, the human person. Thus, his believing that a causal interaction between a human's mind and body, the so-called mind-body problem of dualism, does not come about. Similarly, Hume's (1975) copy principle holds that there is one–to–one correspondence between ideas and reality.

But, in practice, summing these individual mental and physical attributes results in several degrees of freedom; however, the lack of practicality is characterized by the many researchers who have invested in this approach for almost a century yet finding few effects that can be generalized to human-machine interactions; by the replication crisis of today; and by its disembodied nature. In contrast, our approach is based on the interdependence found in every interaction between living organisms. It allows us to theorize about the bistability of mind and body, but it proposes a measurement problem and

non-factorable information. Bistability addresses team structure and performance; the measurement problem solves the replication crisis; and the non-factorable information created by a reduction in the degrees of freedom among teammates matches findings by the National Academies of Science (e.g., Cooke & Hilton, 2015). For example, in 2021, the Academy reported that the "performance of a team is not decomposable to, or an aggregation of, individual performances" (Endsley, 2021, p. 11). In our research, we review the science of teams and a focus on human-machine team research in the laboratory versus in the open; justifications for rejecting traditional social science while supporting our approach; the quantum-like mathematics involved; a review of results from our model versus the open field; a discussion of the results; and the path forward.

Review: Why are there different versions of the COM-B model diagram?

Now, we review what little exists of the COM-B model presented by West and Michie (2023). The authors did not provide statistical support, nor attempts by others of its validation. Online, however, there is a validation attempt reported (Howlett et al., 2021): "The COM-B strongly predicted sitting behaviour (27% variance explained), with Capability, Opportunity, and habit strength as key drivers. The TPB [theory of planned behavior] explained a large amount of variance (23%) in sitting behaviour, with intention and habit strength as key drivers." But not discussed is whether "sitting behaviour" is self-reported or measured independently. A different article, however, republished in PubMed with the same authors and title, stated "Self-reported sitting behaviour."

Self-reported results are not noteworthy, but they allow me to address the issue directly. We report on three examples, and a fourth based on a retraction. First, based on self-reported data, self-esteem was regarded as one of the American Psychology Association's "elements in the highest levels of human functioning." From Bednar and Peterson (1995):

Although, relatively little is known about self-esteem, it is generally considered to be a highly favorable personal attribute, the consequences of which are assumed to be as robust as they are desirable. Books and chapters on mental hygiene and personality development consistently portray self-esteem as one of the premier elements in the highest levels of human functioning . . . Its general importance to a full spectrum of effective human behaviors remains virtually uncontested. We are not aware of a single article in the psychological literature that has identified or discussed any undesirable consequences that are assumed to be a result of realistic and healthy levels of personal self-regard.

In 2005, however, self-esteem was found to be invalid regarding actual (not self-reported) academic and work performance (Baumeister et al., 2005).

Second, a team led by Tetlock (Blanton et al., 2009) found that implicit racism was invalid based on replications; despite this finding, over the past two decades, extraordinary expense, time and effort has been consumed in multiple, repeated attempts to reduce implicit bias but with results that are "dispiriting" (Paluck et al., 2021). Further, the U.S. National Institutes of Health (2021, 9/27) held a public workshop: "Is Implicit Bias Training Effective?" NIH concluded that "scant scientific data suggests these interventions prevent unintended discrimination and create more inclusive workplaces in the long term."

Third, emboldened by the success with the invalidation found by his team on implicit racism, Tetlock published a book with

Gardiner in 2015 on how to predict social outcomes. Their thesis was that politics and human affairs are not inscrutable, but a bit like weather forecasting, where short-term predictions are possible and reasonably accurate: "forecasting ... is a skill that can be cultivated." They collected superforecasters from around the world, trained them, started a website (since redacted), and made their very first two super forecasts, that Brexit would not occur, and Trump would not become president. Both occurred in 2016, but these failures have not deterred Tetlock who has struggled to labor on (e.g., see the interview by Klein, 2021).

Fourth, under the tutelage of Bazerman, a leading researcher in ethics among social scientists, his student, Ariely, published the honesty scale in 2012 in the Proceedings of the National Academy of Sciences (Shu et al., 2012). However, that article was retracted in 2021 after the discovery that the scale was based on fabricated data (Berenbaum, 2021). In sum, what can we conclude about a self-reported scale like COM-B?

Conclusions

The first three failures have led to, or continue, the replication crises in social science (Nosek, 2015). While these failures are extraordinary, and while social science, mired in its replication dysfunction, soldiers on, for my team's research, the larger issue of generalization is overlooked or ignored. Thus, none of this, COM-B included, is of value for the development or the performance of autonomous human-machine teams and systems. A more generous conclusion is that we not only need to solve the theory of autonomous human-machine teams and systems, but we also need to know what causes models like self-esteem, implicit racism, or honesty to fail, and yet still be pursued? We speculate that the failure of our first three cases discussed above can be explained by self-reported beliefs or logics that are disembodied which means that they are unlikely to connect to reality, COM-B included. Why pursued? The fourth case deals with deception; our theory explains that its pursuit is instrumental. For example, in the case of implicit racism, there is a large amount of money involved in attempting to reduce its effects even if the evidence is non-existent (Singal, 2017). We can only guess why other models are pursued. Instead, we hope that the pursuit is to build models that can be generalized to human-machine teams.

References:

Baumeister, R. F., Campbell, J.D., Krueger, J.I., & Vohs, K.D. (2005, January). Exploding the self-esteem myth.*Scientific American, 292(1):* 84-91*; retrieved 1/10/2012* from https://www.uvm.edu/~wgibson/PDF/Self-Esteem%20Myth.pdf

Bednar, R.L. & Peterson, S.R. (1995), *Self-esteem Paradoxes and innovations in clinical practice, 2nd edition* Washington, DC: American Psychological Association (APA).

Berenbaum, M.R. (2021), Retraction for Shu et al., Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end, PNAS, 118 (38) e2115397118; https://doi.org/10.1073/pnas.2115397118

Blanton, Hart, Klick, J., Mitchell, G., Jaccard, J., Mellers, B. & Tetlock, P.E. (2009), Strong Claims and Weak Evidence: Reassessing the Predictive Validity of the IAT, *Journal of Applied Psychology, 94*(3): 567–582.

Cooke, N.J. & Hilton, M.L. (Eds.) (2015), Enhancing the Effectiveness of Team Science. National Research Council. Washington (DC): National Academies Press.

Endsley, M.R., et al. (2021). Human-AI Teaming: State-of-the-Art and Research Needs. The National Academies of Sciences-Engineering-Medicine. Washington, DC: National Academies Press. Retrieved 12/27/2021 from https://www.nap.edu/catalog/26355/human-ai-teaming-state-of-the-art-and-research-needs

Howlett, N., Schulz, J., Trivedi, D., Troop, N. & Chater, A. (2021). Determinants of weekly sitting time: construct validation of an initial COM-B model and comparison of its predictive validity with the Theory of Planned Behaviour, Psychology & Health, 36(1):96-114. doi: 10.1080/08870446.2020.1763994; republished same author list and title in PubMed, PMID: 32403945.

Hume, D. (1975). *A Treatise of Human Nature*, edited by L. A. Selby-Bigge, 2nd ed. revised by P. H. Nidditch, Oxford: Clarendon Press.

Klein, E. (2021, 12/3), “Predicting the future is possible. These ‘Superforecasters’ know how. The psychologist Philip Tetlock on the art and science of prediction,” New York Times, retrieved 12/3/2021 from https://www.nytimes.com/2021/12/03/opinion/ezra-klein-podcast-philip-tetlock.html?action=click&module=Well&pgtype=Homepage&section=OpEd%20Columnists

Lawless, W.F. (2022c), Interdependent Autonomous Human-Machine Systems: The Complementarity of Fitness, Vulnerability & Evolution, Entropy, 24(9):1308. doi: 10.3390/e24091308.

Lawless, W.F.; Sofge, Donald A.; Lofaro, Daniel; & Mittu, Ranjeev (2023), Editorial: Interdisciplinary Approaches to the Structure and Performance of Interdependent Autonomous Human Machine Teams and Systems, Frontiers in Physics, eBook, retrieved 3/1/2023 from https://www.frontiersin.org/articles/10.3389/fphy.2023.1150796/full

Nadler, S. (2022), "Baruch Spinoza", The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.), retrieved 4/15/2023 from https://plato.stanford.edu/archives/sum2022/entries/spinoza/

NIH (2021). U.S. National Institutes of Health, Scientific Workforce Diversity Seminar Series (SWDSS) Seminar Proceedings: “Is Implicit Bias Training Effective?” retrieved 3/15/2023 from https://diversity.nih.gov/sites/coswd/files/images/NIH_COSWD_SWDSS_Implicit_Bias_Proceedings_508.pdf

Nosek, B. (2015), Estimating the reproducibility of psychological science, Science, 349 (6251): 943; retrieved 2/24/2023 from https://www.science.org/doi/10.1126/science.aac4716

Paluck, E.L., Porat, R., Clark, C.S. & Green, D.P. (2021), Prejudice Reduction: Progress and Challenges, Annual Review of Psychology, 72:533–60; retrieved 1/23/2023 from https://www.annualreviews.org/doi/pdf/10.1146/annurev-psych-071620-030619

Shu, L.L., Mazar, N., Gino, F., Ariely, D. & Bazerman, M.H. (2021), Signing at the beginning makes ethics salient and

decreases dishonest self-reports in comparison to signing at the end. Edited by Daniel Kahneman (2012), PNAS, 109(38):15197-15200, https://doi.org/10.1073/pnas.1209746109; Retracted SEPTEMBER 13, 2021 by the editor, PNAS.

Singal, J. (2017), “Psychology’s Favorite Tool for Measuring Racism Isn’t Up to the Job Almost two decades after its introduction, the implicit association test has failed to deliver on its lofty promises,” *The New York’s The Cut*, 2017, https://www.thecut.com/2017/01/psychologys-racism-measuring-tool-isnt-up-to-the-job.html (Accessed 11 January 2017).

Tetlock, P.E. & Gardner, D. (2015), Superforecasting: The Art and Science of Prediction, Crown.

West, R. & Michie, S. (2023), Why are there different versions of the COM-B model diagram? Qeios, https://doi.org/10.32388/AWANWG

;