

Review of: "Metacognitive Agents for Ethical Decision Support: Conceptual Model and Research Roadmap"

Robert West¹

¹ University College London, University of London

Potential competing interests: No potential competing interests to declare.

This is an interesting paper that addresses an important issue. The gap between attitudes and behaviour and latterly between intentions and behaviour has been the subject of a huge amount of research over the decades in the psychology, and the idea of seeking to close that gap through decision support tools seems like a good thing to attempt to do.

Like the other reviewers, I got a bit lost in the details and, although I think the material is there, perhaps the structure could be altered to make it easier to follow. Though I recognise sometimes readers need to put the effort in to grasp complexity when it is in the nature of the subject matter.

My main substantive comments are:

1. The intention-action gap is not the same as the value action gap. Intention and value are different psychological entities. I think that here the emphasis is on the discrepancy between values as generalised appraisals and behaviour. This could easily be mediated by the people concerned NOT intending to behaviour in accordance with their values because they have other priorities. In that case there would not be an intention-action gap. This is important because it determines what any kind of decision support should do.
2. Secondly, the Human Behaviour Change Project is about to publish an ontology of 'mechanisms of action' for behaviour change interventions that builds on the 'mental functioning' ontology that was developed by Janna Hastings. This may be of help to the author in being precise about the entities in the logic model underpinning the proposals. Check out 'bciosearch.org' and Google 'Human Behaviour Change Project'. We are running webinars on all this at the moment and putting them up on YouTube (<https://www.eventbrite.co.uk/cc/human-behaviour-change-project-webinar-series-1684489>).
3. Thirdly, I would like to see, where possible, the article to use Qeios' facility to link to definitions of constructs as this is an area where the words typically mean different things to different people.
4. Finally, I wonder whether the COM-B model^[1] might help to structure thinking and modelling in this area. As a general model of behaviour it seems to me to be very applicable to what this paper is trying to do and could help to bring out the logic behind the proposals.

References

1. [^] [Robert West, Susan Michie. \(2020\). *A brief introduction to the COM-B Model of behaviour and the PRIME Theory of motivation*. Qeios. doi:10.32388/WW04E6.](#)