# Review of: "What is it like to be an AI bat?"

Yutao Yue

Potential competing interests: No potential competing interests to declare.

This paper is a good explanation of the difficult and debating status of the topic of AC. The philosophical aspect of qualia and the explanation gap is discussed. The dependences of consciousness on its neural correlates, on autopoietic features, on system complexity, and on its behavioral functionalities are analyzed. All these give readers a clear picture of what consciousness is and how different people interpret it. This is a well-written paper with good references and adequate concrete numbers (especially compared to lots of other papers on a similar topic) to support the contents.

The main opinions and points that the authors want to stress in the “Problems and Solutions” section, however, could be better organized, clearer, and more powerful. The title does not give a clear idea of what the main point of the paper is either. The authors suggest a pragmatic view of the AC problem. This can be further clarified by trying to answer questions such as, what are (and what are not) the key characteristics of consciousness that should be considered while pragmatically dealing with the AI-human relationship? What might be a concrete way to evaluate the status of consciousness of an AI system? Where might the borderline of consciousness for AI systems to be a threat to humans be? What kinds and forms of damage could an “over-conscious” AI system do to humans and society? What might be the ways to prevent AI from being “over-conscious” (or from making damage)?

I personally agree with the authors' opinion of putting the (in many cases, meaningless) debates aside and focusing on what the actual impacts are and what can be done in this fast AI era. In my personal opinion, consciousness is a certain advanced aspect among all kinds of characteristics of various intelligent systems. There is nothing mysterious about it. It can be decomposed into a few key components, such as self-awareness, self-defining intention, and subjective (and thus variational) internal representations and judgements of the same objective input, etc. I believe consciousness can eventually be mathematically formulized with a high degree of consensus among researchers from different scientific disciplines and will be equally applicable to biological (natural) and artificial intelligent systems.