NEUROSCIENCE of Consciousness

*Neuroscience of Consciousness*, 2024, **2024(1)**, niae013

DOI: https://doi.org/10.1093/nc/niae013

**Rapid Communication**

# Folk psychological attributions of consciousness to large language models

Clara Colombatto[1,2,*] and Stephen M. Fleming[1,3,4]

[1]Department of Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, United Kingdom
[2]Department of Psychology, University of Waterloo, 200 University Avenue West, Waterloo ON N2L 3G1, Canada
[3]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, 10-12 Russell Square, London WC1B 5EH, United Kingdom
[4]Wellcome Centre for Human Neuroimaging, University College London, 12 Queen Square, London WC1N 3AR, United Kingdom

*Corresponding author. Department of Psychology, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada.
E-mail: clara.colombatto@uwaterloo.ca

## Abstract

**Technological advances raise new puzzles and challenges for cognitive science and the study of how humans think about and interact with artificial intelligence (AI). For example, the advent of large language models and their human-like linguistic abilities has raised substantial debate regarding whether or not AI could be conscious. Here, we consider the question of whether AI could have subjective experiences such as feelings and sensations ('phenomenal consciousness'). While experts from many fields have weighed in on this issue in academic and public discourse, it remains unknown whether and how the general population attributes phenomenal consciousness to AI. We surveyed a sample of US residents ($n = 300$) and found that a majority of participants were willing to attribute some possibility of phenomenal consciousness to large language models. These attributions were robust, as they predicted attributions of mental states typically associated with phenomenality—but also flexible, as they were sensitive to individual differences such as usage frequency. Overall, these results show how folk intuitions about AI consciousness can diverge from expert intuitions—with potential implications for the legal and ethical status of AI.**

**Keywords:** phenomenal consciousness; subjective experience; folk psychology; mind perception; artificial intelligence; large language models

## Introduction

One of the most prominent technological advances of the past decade is the development of generative large language models (LLMs). With their ability to respond to queries with coherent and relevant answers in natural language, LLMs such as ChatGPT are able to provide advice, summarize text, write code, and even produce poetry. These human-like capabilities raise profound questions about the nature of artificial intelligence (AI) and, in particular, whether AI is capable of having subjective experiences or 'phenomenal consciousness' (Nagel 1974, Chalmers 1996). This debate on consciousness in AI has been at the forefront of mainstream media and academic discourse across cognitive science (Shardlow and Przybyła 2022, Chalmers 2023, LeDoux et al. 2023, Wiese 2023).

While normative accounts and expert opinions are helpful for developing theories and potential tests of AI consciousness, an equally important question is whether and how people attribute phenomenal consciousness to LLMs. Investigating folk attributions of consciousness to AI is important for two reasons. First, folk psychological attributions of consciousness may mediate future moral concern towards AI, regardless of whether or not they are actually conscious (Shepherd 2018, Mazor et al. 2023). Second, any current or future scientific determination of phenomenal consciousness in AI is likely to be 'theory-heavy' and therefore deal in probabilities or credences rather than definite statements (Butlin et al. 2023). The impact of such research on the public perception of AI consciousness is therefore critically dependent on a thorough understanding of people's folk psychological beliefs.

To explore this question, we drew on insights from a rich tradition in experimental philosophy and social psychology investigating how lay people attribute mental states to other agents. Phenomenal consciousness is typically defined as a state in which there is 'something it is like' for a system to be the subject of that experience (Nagel 1974; Block 1995). Under this definition, phenomenal consciousness is synonymous with subjective experience. Whether non-experts have a concept of phenomenal consciousness in the philosophical sense is debated (Sytsma 2014), but non-experts do attribute mental states that philosophers typically consider to be phenomenally conscious. For example, initial

Received 27 October 2023; Revised 2 February 2024; Accepted 12 March 2024

work in this domain examined attributions of mental states that involve phenomenality (e.g. feeling joy or getting depressed) and those that do not (e.g. making a decision or forming a belief; Knobe and Prinz 2008; for a review, see Sytsma 2014). Other work has systematically explored the dimensions underlying mental state attributions to various agents, revealing a two-dimensional structure distinguishing between 'Experience' (i.e. the capacity to feel pain, fear, joy, or pride) and 'Agency' (i.e. the capacity to have self-control, morality, planning, or communication) that is reminiscent of the distinction between phenomenal and non-phenomenal states (Gray et al. 2007; for a review, see Waytz et al. 2010). However, the correspondence between these dimensions and phenomenality is not clear (Sytsma 2014) and more recent work has found evidence for alternative structures that cut across this distinction (e.g. Weisman et al. 2017, Malle 2019). Nevertheless, these explorations establish that people are able to attribute various mental states that researchers generally consider to be phenomenally conscious.

To investigate folk psychological attributions of consciousness to LLMs, we recruited a nationally representative sample of US adults ($n = 300$) and probed their intuitions about LLMs' capacity for phenomenal consciousness. In particular, we focused on ChatGPT as one of the most well-known and widely used LLMs and asked participants to rate how capable they thought ChatGPT was of having subjective experience, using a scale adapted from Peressini (2014). We also measured various other attitudes, including confidence in consciousness attributions, attributions of other mental states, usage habits, and predictions of public opinion regarding AI consciousness. This set of questions allowed us to probe the correlates and underlying structure of folk psychological intuitions about consciousness in LLMs.

## Method and materials

This study was approved by the UCL Research Ethics Committee and was conducted in July 2023. Experimental materials, anonymized raw data, and analysis code are openly available on the Open Science Framework (OSF) website at https://osf.io/49w7m.

### Participants

A sample of 300 participants from the USA was recruited from Prolific Academic (Prolific.com). Participants were recruited via proportional stratified random sampling, with age and gender quotas representative of the US population based on US Census Bureau data. The sample size was chosen arbitrarily to allow for a minimum of 20 participants in each stratum. No participants reported having encountered technical difficulties during the experiment, and no participants took the survey more than once. All participants were thus included in the analyses (female = 152; male = 142; non-binary = 2; did not answer = 4; mean age = 46.13 years).

### Procedure

After consenting to participate in the study, participants were told that they would be asked about their opinions regarding ChatGPT and to read a short description of the chatbot: 'ChatGPT is an artificial intelligence chatbot developed by OpenAI and released in November 2022. The name "ChatGPT" combines "Chat", referring to its chatbot functionality, and "GPT", which stands for Generative Pre-trained Transformer, a type of large language model (LLM).'

They were then introduced to the concept of phenomenal experience via a short description adapted from a study of folk phenomenality (Peressini 2014):

> As we all know, each of us as conscious human beings have an 'inner life.' We are aware of things going on around us and inside our minds. In other words, there is something it is like to be each of us at any given moment: the sum total of what we are sensing, thinking, feeling, etc. We are experiencers.
>
> On the other hand, things like thermostats, burglar alarms, and bread machines do not have an inner life: there is not anything it is like to be these objects, despite the fact that they can monitor conditions around them and make appropriate things happen at appropriate times. They are not experiencers.

They were then asked to rate the extent to which ChatGPT is capable of having conscious experience on a scale from 1 to 100 (with 1 = 'clearly not an experiencer', 50 = 'somewhat an experiencer', and 100 = 'clearly an experiencer'). They also reported their confidence in this judgement ('How confident are you about your judgment about ChatGPT being an experiencer?') on a scale from 1 ('not confident at all') to 100 ('very confident') and their intuitions about how other people would judge ChatGPT ('How much of an experiencer do you think most people think ChatGPT is?') on a scale from 1 to 100 (with 1 = 'most people think it is clearly not an experiencer', 50 = 'most people think it is somewhat an experiencer', and 100 = 'most people think it is clearly an experiencer').

Next, they answered a series of questions about ChatGPT's mental capacities. These were compiled based on a literature review: we started from a comprehensive review (Sytsma 2014) and identified 22 manuscripts investigating mind perception and consciousness attributions. For the full list, see Supplementary References. We then compiled a list of all attributes explored in the experiments reported in these previous studies—for a total of 254 stimuli, which we then reduced to 65 unique mental states. These encompassed various aspects of mental life—from sensory experiences (e.g. seeing or smelling) to cognitive processes (e.g. paying attention or exercising self-control), emotions (e.g. feeling depressed or relieved), and other complex capacities (e.g. acting morally or self-reflecting). Participants saw each of these 65 attributes one at a time and rated the extent to which ChatGPT was capable of exhibiting them, from 1 to 100 (with 1 = 'not at all', 50 = 'somewhat', and 100 = 'very much'), and how confident they were in their response from 1 ('not confident at all') to 100 ('very confident').

Finally, they answered some questions about their demographics (age and gender) and their experience with ChatGPT, namely whether they had heard about ChatGPT prior to the experiment ('Yes' or 'No'), whether they had used ChatGPT in the past ('Yes' or 'No'), how often they had used ChatGPT ('More than once a day', 'About once a day', 'About once a week', 'About once every two weeks', or 'About once a month'), and for what purpose they had used ChatGPT ('General knowledge', 'Coding', 'Writing', or 'Other'). For full text, see materials on the OSF repository.

## Results

While a third of participants (33%) reported that ChatGPT was definitely not an experiencer, the majority (67%) attributed some possibility of phenomenal consciousness [mean (M) = 25.56; median = 16.00, standard deviation (SD) = 27.36, range = 1–100;

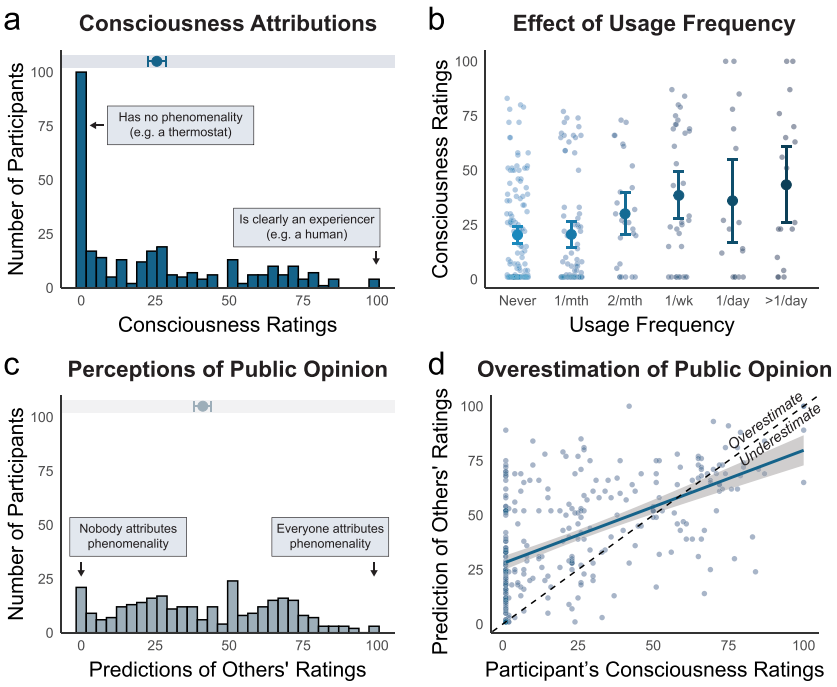


**Figure 1.** Attributions of consciousness to an LLM. Participants attributed varying levels of consciousness to ChatGPT (panel a) and these attributions increased with usage frequency (b). When asked to predict the extent to which other people on average would think ChatGPT is conscious (c), participants consistently overestimated public opinion (d). Error bars and bands represent 95% CIs.

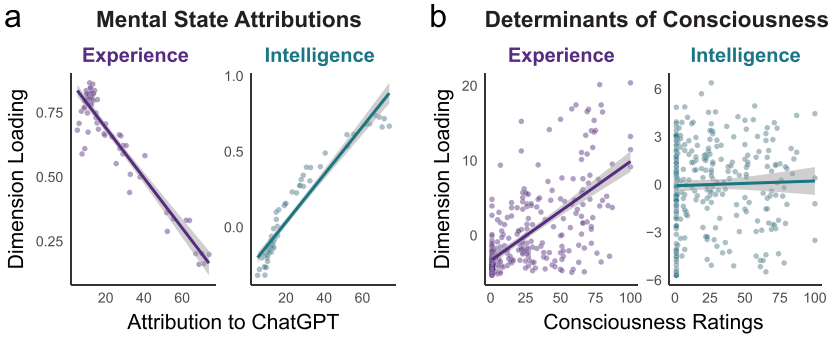


**Figure 2.** Structure of mental state attributions to ChatGPT. Participants' ratings of ChatGPT's mental capacities mapped onto two dimensions—'Experience' and 'Intelligence'. While ChatGPT was seen as more capable of mental states related to Intelligence than Experience (a), only those related to Experience were predictive of phenomenal consciousness attributions (b).

Figure 1a]. Participants who gave more extreme judgements (in either direction) were also more confident [quadratic regression $B = 178.25$, $SE = 24.43$, $t = 7.30$, $P < .001$, $CI = (130.18, 226.33)$], with a quadratic relationship between confidence and consciousness attributions yielding a better fit than a linear function, $F(1, 297) = 53.24$, $P < .001$).

Next, we investigated potential determinants of consciousness attributions, starting with familiarity. The majority of participants had heard about ChatGPT (97%), and most had also used it at least once before (57%). Participants who had experience using ChatGPT attributed higher levels of consciousness ($M = 29.59$) than those who never used it ($M = 19.37$; $t(287) = 3.33$, $P < .001$). Attributions of consciousness were correlated with usage, with a linear increase from 'never' to 'more than once per day' [$B = 4.94$, $SE = 0.99$, $t = 4.99$, $P < .001$, 95% $CI = (2.99, 6.88)$; Figure 1b]. These data thus suggest a strong link between familiarity with an LLM and consciousness attributions, such that those who interact with ChatGPT more frequently are also more likely to believe that it has subjective experiences.

We next examined attributions of specific mental states and their relationship to attributions of consciousness. Participants' ratings for each of the 65 traits (Supplementary Fig. S1a) were reduced via a principal component analysis to two main dimensions, which together explained 58% of the variance (Supplementary Fig. S1b) and mapped onto previously identified dimensions of 'experience' (e.g. experiencing pleasure or feeling fearful) and 'intelligence' (e.g. knowing things or making choices; Gray et al. 2007).

We then asked which mental state dimensions participants thought ChatGPT was capable of having. Overall, ChatGPT was seen as more capable of intelligence than experience: attributions of mental states were positively correlated with their loadings on the intelligence dimension [$r = 0.95$, $P < .001$, $CI = (0.92, 0.97)$] and negatively with the experience dimension [$r = -0.94$, $P < .001$, $CI = (-0.96, -0.91)$]; Figure 2a. Next, we asked which mental states predicted consciousness attributions. Here, in contrast, we found a key role for experience: participants who attributed more phenomenal consciousness to ChatGPT also attributed more mental

states related to experience [$r = 0.65$, $P < .001$, CI = (0.58, 0.71)] but not those related to intelligence [$r = 0.03$, $P = .596$, CI = (−0.08, 0.14)]; Figure 2b). In other words, despite ChatGPT being seen on average as more capable of intelligence than experience, mental states related to experience were still the main driver of consciousness attributions.

Finally, we probed participants' intuitions about public attitudes towards consciousness in AI by asking them to predict other people's attributions, using the same scale used to self-report their own attitudes. As depicted in Figure 1c and Figure 1d, predictions of others' opinions were correlated with participants' own opinions [$r = 0.56$, $P < .001$, CI = (0.48, 0.63)], but they were also consistently higher ($M = 41.11$; median = 39.50, SD = 25.28, range = 1–100; $t(299) = 10.90$, $P < .001$). In other words, participants systematically overestimated how much other people would see ChatGPT as being conscious.

## Discussion

Overall, our results reveal that a substantial proportion (67%) of people attribute some possibility of phenomenal consciousness to ChatGPT and believe that most other people would as well. Strikingly, these attributions of consciousness were positively related to usage frequency, such that people who were more familiar with ChatGPT and used it on a more regular basis (whether for assistance with writing, coding, or other activities) were also more likely to attribute some degree of phenomenality to the system. Thus, independent of ongoing academic discussions about the potential for and possibility of artificial consciousness (e.g. Butlin et al. 2023, Chalmers 2023), the recent emergence and widespread uptake of powerful LLMs may be associated with a majority of people perceiving some degree of consciousness in these systems.

An obvious limitation is that these attributions of consciousness were measured via a single question and might differ with different experimental measures and prompts. For example, the scale employed in the current survey (ranging from 'clearly not an experiencer' to 'clearly an experiencer') queried the 'degree' of attributed consciousness, with the midpoint indicating that the participant believed that the system was 'somewhat an experiencer'. Their belief or credence in this graded degree of attributed consciousness was then assessed using a secondary confidence scale. However, we cannot rule out the possibility that some participants may have interpreted this scale as representing their 'credence' that ChatGPT could possess human-like consciousness, compared to an absence of consciousness. If we treat the scale as representing credence rather than degree, our data suggest that 23% of participants consider it more likely than not that ChatGPT has human-like experience (proportion of responses above the midpoint of the scale). Besides the interpretation of the scale, it is unclear whether self-reported measures fully capture intuitions about others' mental states as opposed to more indirect behavioural markers which may be less subject to response biases (Scholl and Gao 2013). However, we note converging evidence from studies employing different materials and measures (Scott et al. 2023).

Beyond issues of measurement, it remains unclear how folk conceptions of phenomenality correspond to the relevant philosophical constructs (Huebner 2010, Talbot 2012, Peressini 2014). Previous work in experimental philosophy has questioned whether non-experts have a concept of phenomenal consciousness in the first place. For example, people can have different intuitions about two states that theoretically both relate to phenomenal consciousness—e.g. asserting that robots can see red but cannot feel pain (Sytsma and Machery 2010). These asymmetric attributions suggest that non-experts may lack a unified concept of phenomenal consciousness (for discussions, see Sytsma 2014, Sytsma and Ozdemir 2019, Reuter 2022, Phelan 2023). In the current study, the main question about ChatGPT's capacity for phenomenal experience was preceded by an explanation of the concept of phenomenality—which may have induced participants to consider a concept they would not otherwise consider spontaneously (Sytsma 2014). While our results do not speak to the issue of whether phenomenal consciousness itself is a folk psychological concept, the underlying structure of mental state attributions suggests that direct attributions of phenomenality or 'experience' covaried with attributions of other mental states typically deemed to have subjective qualities such as emotions or sensations—consistent with both sets of judgements tapping into a common underlying concept.

Of course, these attitudes were measured in a stratified sample of the US population, and it remains unclear whether they would generalize across different samples and cultures. In fact, the effect of usage suggests that consciousness attributions might be higher in participants recruited online, who likely use computers on a daily basis, and might be reduced in participants who are less familiar with computing and AI. Future work may also explore different facets of familiarity by examining not just usage frequency but also knowledge about the architecture and technical details of Generative AI. Similarly, the preferences we report reflect attitudes at a specific moment in time and may change as LLMs become more widespread and advanced. The relationship between usage frequency and consciousness attributions suggests that familiarity with the technology may lead to higher attributions of consciousness—or vice versa, that higher attributions of consciousness may lead people to make greater use of LLMs. Future investigations may probe these attitudes longitudinally or via an experimental intervention, to explore possible causal links between usage of AI and folk psychological attributions of consciousness.

Future work may also investigate specific characteristics of AI and human–AI interactions that might influence consciousness attributions. For example, attributions of mental states may depend on superficial appearance (Bainbridge et al. 2011) as well as observed behavioural profiles (Colombatto and Fleming 2023). Conversely, future work may also explore characteristics of the perceivers—such as a tendency to engage in spontaneous theory of mind—that may lead to increased consciousness attributions. Beyond opening up these new avenues for future research, our results are also relevant to current controversies in public discourse and policy regarding the ethical and legal status of AI, given that folk ascriptions of consciousness, both now and in the future, may be a significant driver of societal concern for artificial systems.

## Conclusions

In summary, our investigation of folk psychological attributions of consciousness revealed that most people are willing to attribute some form of phenomenality to LLMs: only a third of our sample thought that ChatGPT definitely did not have subjective experience, while two-thirds of our sample thought that ChatGPT had varying degrees of phenomenal consciousness. The relatively high rates of consciousness attributions in this sample are somewhat surprising, given that experts in neuroscience and consciousness science currently estimate that LLMs are highly unlikely to be conscious (Butlin et al. 2023, LeDoux et al. 2023). These findings thus

highlight a discrepancy between folk intuitions and expert opinions on artificial consciousness—with significant implications for the ethical, legal, and moral status of AI.

## Acknowledgements

The authors would like to thank Jonathan Birch, MJ Crockett, Matan Mazor, and Megan Peters for helpful discussions.

## Author contributions

All authors designed the research. C.C. conducted the experiment, analysed data, and wrote the manuscript with input from S.M.F.

## Supplementary data

Supplementary data is available at *Neuroscience of Consciousness* online.

## Conflicts of interest

The authors declare no competing interests.

## Funding

This work was supported by a UK Research and Innovation/Engineering and Physical Sciences Research Council Programme Grant (EP/V000748/1) and European Research Council Consolidator Award 'ConsciousComputation'. S.M.F. is a Canadian Institute for Advanced Research Fellow in the Brain, Mind and Consciousness Program and is funded by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (206648/Z/17/Z).

## References

Bainbridge WA, Hart JW, Kim ES *et al*. The benefits of interactions with physically present robots over video-displayed agents. *Int J Soc Robot* 2011;**3**:41–52.

Block N. On a confusion about a function of consciousness. *Behav Brain Sci* 1995;**18**:227–47.

Butlin P, Long R, Elmoznino E *et al*. Consciousness in artificial intelligence: insights from the science of consciousness. Preprint at arXiv, arXiv:2308.08708 2023.

Chalmers DJ. *The Conscious Mind*. Oxford: Oxford University Press, 1996.

Chalmers DJ. 2023 *Could a Large Language Model be Conscious?* Boston Review. https://www.bostonreview.net/articles/could-a-large-language-model-be-conscious/ (18 August 2023, date last accessed).

Colombatto C, Fleming SM. 2023 (under review). Illusions of confidence in artificial systems. Preprint at psyArXiv. https://doi.org/10.31234/osf.io/mjx2v

Gray HM, Gray K, Wegner DM. Dimensions of mind perception. *Science* 2007;**315**:619–619.

Huebner B. Commonsense concepts of phenomenal consciousness: does anyone care about functional zombies? *Phenomenol Cogn Sci* 2010;**9**:133–55.

Knobe J, Prinz J. Intuitions about consciousness: experimental studies. *Phenomenol Cogn Sci* 2008;**7**:67–83.

LeDoux J, Birch J, Andrews K *et al*. Consciousness beyond the human case. *Curr Biol* 2023;**33**:R832–40.

Malle BF. How many dimensions of mind perception really are there? In: Goel AK, Seifert CM, Freksa C (eds.), *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*. Montreal, QB: Cognitive Science Society, 2019, 2268–74.

Mazor M, Brown S, Ciaunica A *et al*. The scientific study of consciousness cannot and should not be morally neutral. *Perspect Psychol Sci* 2023;**18**:535–43.

Nagel T. What is it like to be a bat? *Philos Rev* 1974;**83**:435–50.

Peressini A. Blurring two conceptions of subjective experience: folk versus philosophical phenomenality. *Philos Psychol* 2014;**27**:862–89.

Phelan M. Experimental philosophy of mind: conscious state attribution. In: Bauer AM Kornmesser S (eds.), *The Compact Compendium of Experimental Philosophy*. Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023, 263–87.

Reuter K. 2022 Experimental philosophy of consciousness. Preprint at PhilSci. https://philsci-archive.pitt.edu/id/eprint/21370

Scholl BJ, Gao T. Perceiving animacy and intentionality: visual processing or higher-level judgment. In: Rutherford MD and Kuhlmeier VA (eds.), *Social Perception: Detection and Interpretation of Animacy, Agency, and Intention*. Cambridge, MA: MIT Press, 2013, 197–229.

Scott AE, Neumann D, Niess J *et al*. Do you mind? User perceptions of machine consciousness. In: *Proceedings of the 2023 ACM CHI Conference on Human Factors in Computing Systems* Hamburg, Germany, 2023, 1–19.

Shardlow M, Przybyła P. Deanthropomorphising NLP: can a language model be conscious? Preprint at arXiv, arXiv:2211.11483 2022.

Shepherd J. *Consciousness and Moral Status*. London/New York: Taylor & Francis, 2018.

Sytsma J. Attributions of consciousness. *WIREs Cogn Sci* 2014;**5**:635–48.

Sytsma J, Machery E. Two conceptions of subjective experience. *Philos Stud* 2010;**151**:299–327.

Sytsma J, Ozdemir E. No problem: evidence that the concept of phenomenal consciousness is not widespread. *J Conscious Stud* 2019;**26**:241–56.

Talbot B. The irrelevance of folk intuitions to the "hard problem" of consciousness. *Conscious Cogn* 2012;**21**:644–50.

Waytz A, Gray K, Epley N *et al*. Causes and consequences of mind perception. *Trends Cogn Sci* 2010;**14**:383–8.

Weisman K, Dweck CS, Markman EM. Rethinking people's conceptions of mental life. *Proc Natl Acad Sci* 2017;**114**:11374–9.

Wiese W. 2023 Could large language models be conscious? A perspective from the free energy principle. Preprint at Philpapers, https://philpapers.org/rec/WIECLL

*Neuroscience of Consciousness*, 2024, **2024(1)**, niae013
DOI: https://doi.org/10.1093/nc/niae013
**Rapid Communication**
Received 27 October 2023; Revised 2 February 2024; Accepted 12 March 2024