Conscious robots may merit legal protections.

PHOTO: AP PHOTO/IRAM ASIM

LETTERS

Edited by **Jennifer Sills**

Conscious machines: Defining questions

In their Review "What is consciousness, and could machines have it?" (27 October 2017, p. 486), S. Dehaene *et al.* argue that the science of consciousness indicates that we are not on the verge of creating conscious machines. However, Dehaene *et al.* ask and answer the wrong questions.

To determine whether machines are conscious, we must ask whether they have subjective experiences: Do machines consciously perceive and sense colors, sounds, and smells? Do they feel emotions? Unfortunately, Dehaene *et al.* relegate this issue to the final paragraph of their Review, dismissing it as a philosophical question "beyond the scope of the present paper." Instead, they ask whether machines "mimic" consciousness by exhibiting the global availability of information (the ability to select, access, and report information) and metacognition (the capacity for self-monitoring and confidence estimation). Questions concerning to what extent machines have these capacities are interesting, but neither capacity is necessary or sufficient for subjective experience (*1*, *2*). Furthermore, Dehaene *et al.*'s emphasis on metacognition and global broadcasting presumes that the prefrontal cortex is the home of consciousness, which remains a matter of debate (*3*, *4*).

Finally, when arguing that machines are not yet conscious, Dehaene *et al.* highlight the "feedforward"—i.e., sequential—nature of information processing typical of these computing systems. Although current machines may not exhibit propagation of information into the broadcasting network hub or formal Bayesian metacognition, many artificial intelligence systems involve global projection of winning representations and have metacognitive capacity, such as confidence estimates. If these types of processes were strongly indicative of consciousness, we would have to admit that some machines are already conscious.

We agree with Dehaene *et al.* that to address the question of machine consciousness, we must start with a theory of human consciousness. Given current disagreement on that topic, we are all left to speculate whether machines will ever be conscious. A more pertinent question for the field might be: "What would constitute successful demonstration of artificial consciousness?"

Olivia Carter,[1] Jakob Hohwy,[2] Jeroen van Boxtel,[2] Victor Lamme,[3] Ned Block,[4] Christof Koch,[5] Naotsugu Tsuchiya[2]*
[1]University of Melbourne, Melbourne, VIC 3010, Australia. [2]Monash University, Mebourne, VIC 3800 Australia. [3]University of Amsterdam, 1018 XA Amsterdam, Netherlands. [4]New York University, New York, NY 10003, USA. [5]Allen Institute for Brain Science, Seattle, WA 98103, USA.
*Corresponding author.
Email: naotsugu.tsuchiya@monash.edu

REFERENCES

1. C. Koch *et al.*, *Nat. Rev. Neurosci.* **17**, 307 (2016).
2. N. Tsuchiya *et al.*, *Trends Cogn. Sci.* **19**, 757 (2015).
3. M. Boly *et al.*, *J. Neurosci.* **37**, 9603 (2017).
4. B. Odegaard, R. T. Knight, H. Lau, *J. Neurosci.* **37**, 9593 (2017).

10.1126/science.aar4163

Conscious machines: Robot rights

In their Review "What is consciousness, and could machines have it?" (27 October 2017, p. 486), S. Dehaene *et al.* suggest that machine consciousness, which would model human cognitive functions within a physical architecture other than the human brain, has not yet been achieved. Creating consciousness is still a goal for the future, but now is the time to consider the implications of conscious machines.

Developing machine capacities such as artificial intelligence (AI) and robots for human-robot interaction is of extreme technological value, but it raises moral and ethical questions. For example, one of the major sectors in robotics is the development of sexual robots (*1*). Although some researchers consider this phenomenon a gateway to the future acceptance of human-robot interactions, others see a danger for human society as robots modify the social representation of human behaviors (*2*, *3*).

The robotics field must wrestle with these questions: Is it ethical to create and continue to use robots with consciousness in the way we use robots that were originally designed for our needs? Do robots deserve to be protected? This debate could challenge the limits of human morality and polarize society's views on whether conscious robots are objects or living entities (*4*–*6*).

As we approach an era when conscious robots become part of daily life, it is important to start thinking about the current status of robots today. The purpose of granting legal status to robots is not only to prevent inappropriate human-robot interactions but also to recognize and formalize the role of robots in society, thus normalizing their existence (*7*).

Nicolas Spatola* and Karolina Urbanska
Université Clermont Auvergne, CNRS, LAPSCO, F-63000 Clermont-Ferrand, France.
*Corresponding author. Email: nicolas.spatola@uca.fr

REFERENCES

1. M. Scheutz, T. Arnold, "Intimacy, bonding, and sex robots: Examining empirical results and exploring ethical ramifications" (2017); https://hrilab.tufts.edu/publications/scheutz2017intimacy.pdf.
2. J. Robertson, *Body Soc.* **16**, 1 (2010).
3. M. Coeckelbergh, *Int. J. Soc Robot.* **1**, 217 (2009).
4. P. Lin *et al.*, Eds., *Robot Ethics* (MIT Press, 2012).
5. A. L. Peláez, D. Kyriakou, *Technol. Forecast. Soc. Change* **75**, 1176 (2008).
6. K. Richardson, *Comp. Soc.* **45**, 290 (2016).
7. I. Yeoman, M. Mars, *Futures* **44**, 365 (2012).

10.1126/science.aar5059

Response

Any discussion of machine consciousness should start with empirical evidence, and our Review primarily consists of an empirical look at how nonconscious and conscious processing differ in humans [see also (*1*, *2*)]. In contrast to Carter *et al.*'s interpretation of our conclusions, we suggest that conscious subjective states are in fact on the verge of becoming implementable in machines and that two computational ingredients (global information sharing and self-monitoring), if jointly and correctly implemented, may provide machines with conscious subjectivity.

Carter *et al.* claim that "neither capacity is necessary or sufficient for subjective experience," but that is begging the question. It is possible that the subjective experiences of humans are simply information-bearing representations with the same specific properties we used for those of machines: being globally available and therefore reportable, and entering into dedicated self-monitoring processes capable of evaluation and criticism. Empirical evidence suggests that whenever human subjective experience is impaired, such as in psychosis (*3*) or blindsight (*4*), aspects of both of these processing functions are also disrupted. Thus, our working hypothesis is that subjective experience comes down to nothing but a combination of specific forms of processing (including reality monitoring as a crucial component).

It has been suggested that failing to recognize that machines may have subjective states reflects a lack of imagination (*5*). In fact, many if not all visual illusions, in which subjective perception diverges from objective reality, arise from efficient computing, such as applying a Bayesian prior to noisy inputs (*6*) or taking efficient shortcuts in otherwise intractable computations (*7*). Thus, such subjective percepts would arise in any efficient machine.

To move forward, Carter *et al.* point out that it would be advantageous if we could agree on the criteria that demonstrate consciousness. But it is precisely because we lack such a consensus that we think it is best to start with empirical evidence based on the known features of the human brain. Rapid progress in mapping the human brain mechanisms of consciousness has indeed revealed an important contribution of the prefrontal cortex [e.g., (*8*–*10*)].

When machines share enough features with conscious human brain processing, we should be prepared to accept the possibility that they are conscious. Of course, even if a machine shared those features and reported having subjective experiences, Carter *et al.* could still deny that it experienced anything at all. But such a solipsist position also applies to humans: By such a standard, we likewise cannot prove that other human beings are conscious. This position, and the associated insistence on "qualia" and the "hard problem" of consciousness, are unproductive (*11*). In the future, denying machines any form of subjectivity, when it is caused by computations similar to those that constitute core ingredients of consciousness in the human brain, may become as contentious as denying it to other human beings or to nonhuman animals with brain architectures similar to ours.

We therefore agree with Spatola and Urbanska that the predictable emergence of conscious machines calls for an immediate consideration of its societal consequences. The potential benefits should not be neglected: A powerful sentient artificial intelligence (AI) may collaborate with humans in addressing major issues such as energy management, ecology, or care in an aging society. The risks, however, are equally real and include job loss, concentration of power in a few hands, a military arms race, and social upheaval as humans and AI increasingly compete for the same societal roles. Mitigating these disorders will require a major international effort, and we can only heed here the conclusion of a recent academic statement on the power and limits of AI (*12*): "Just like crash tests for transportation, the passing

of ethical and safety tests, evaluating, for instance, social impact or racial prejudice, could become a prerequisite to the release of [artificial intelligence] software."

Stanislas Dehaene,[1,2]* Hakwan Lau,[3,4] Sid Kouider[5]

[1]Chair of Experimental Cognitive Psychology, Collège de France, 75005 Paris, France. [2]Cognitive Neuroimaging Unit, Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), INSERM, Université Paris-Sud, Université Paris-Saclay, NeuroSpin Center, 91191 Gif/Yvette, France. [3]Department of Psychology and Brain Research Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA. [4]Department of Psychology, University of Hong Kong, Hong Kong. [5]Brain and Consciousness Group (École Normale Supérieure, École des Hautes Études en Sciences Sociales, CNRS), Département d'Études Cognitives, École Normale Supérieure–Paris Sciences et Lettres Research University, Paris, France.
*Corresponding author.
Email: stanislas.dehaene@cea.fr

REFERENCES

1. S. Kouider, S. Dehaene, *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **362**, 857 (2007).
2. S. Dehaene, *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts* (Penguin Books, Reprint edition, 2014).
3. P. C. Fletcher, C. D. Frith, *Nat. Rev. Neurosci.* **10**, 48 (2009).
4. Y. Ko, H. Lau, *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **367**, 1401 (2012).
5. D. Dennett, *Cognition* **79**, 221 (2001).
6. Y. Weiss *et al.*, *Nat. Neurosci.* **5**, 598 (2002).
7. T. D. Ullman, E. Spelke, P. Battaglia, J. B. Tenenbaum, *Trends Cogn. Sci.* 10.1016/j.tics.2017.05.012 (2017).
8. M. Wang, D. Arteaga, B. J. He, *Proc. Natl. Acad. Sci. U.S.A.* **110**, E3350 (2013).
9. T. I. Panagiotaropoulos *et al.*, *Neuron* **74**, 924 (2012).
10. B. Odegaard, R. T. Knight, H. Lau, *J. Neurosci. Off. J. Soc. Neurosci.* **37**, 9593 (2017).
11. D. C. Dennett, in *Consciousness in Modern Science*, A. Marcel, E. Bisiach, Eds. (Oxford University Press, 1988), pp. 42–77.
12. Pontifical Academy of Sciences, "Final statement of the workshop: Power and limits of artificial intelligence" (2016); www.pas.va/content/accademia/en/events/2016/intelligence/statement.html.

10.1126/science.aar8639

TECHNICAL COMMENT ABSTRACTS

Comment on "Precipitation drives global variation in natural selection"

Isla H. Myers-Smith and Judith H. Myers

Siepielski *et al*. (Reports, 3 March 2017, p. 959) claim that "precipitation drives global variation in natural selection." This conclusion is based on a meta-analysis of the relationship between climate variables and natural selection measured in wild populations of invertebrates, plants, and vertebrates. Three aspects of this analysis cause concern: (i) lack of within-year climate variables, (ii) low and variable estimates of covariance relationships across taxa, and (iii) a lack of mechanistic explanations for the patterns observed; association is not causation.
Full text: dx.doi.org/10.1126/science.aan5028

Response to Comment on "Precipitation drives global variation in natural selection"

Adam M. Siepielski, Michael B. Morrissey, Mathieu Buoro, Stephanie M. Carlson, Christina M. Caruso, Sonya M. Clegg, Tim Coulson, Joseph DiBattista, Kiyoko M. Gotanda, Clinton D. Francis, Joe Hereford, Joel G. Kingsolver, Kate E. Augustine, Loeske E. B. Kruuk, Ryan A. Martin, Ben C. Sheldon, Nina Sletvold, Erik I. Svensson, Michael J. Wade, Andrew D. C. MacColl

The Comment by Myers-Smith and Myers focuses on three main points: (i) the lack of a mechanistic explanation for climate-selection relationships; (ii) the appropriateness of the climate data used in our analysis; and (iii) our focus on estimating climate-selection relationships across (rather than within) taxonomic groups. We address these critiques in our response.
Full text: dx.doi.org/10.1126/science.aan5760

Conscious machines: Defining questions

Olivia Carter, Jakob Hohwy, Jeroen van Boxtel, Victor Lamme, Ned Block, Christof Koch and Naotsugu Tsuchiya

Science **359** (6374), 400.
DOI: 10.1126/science.aar4163