PSYCHOLOGY

Theory as adversarial collaboration

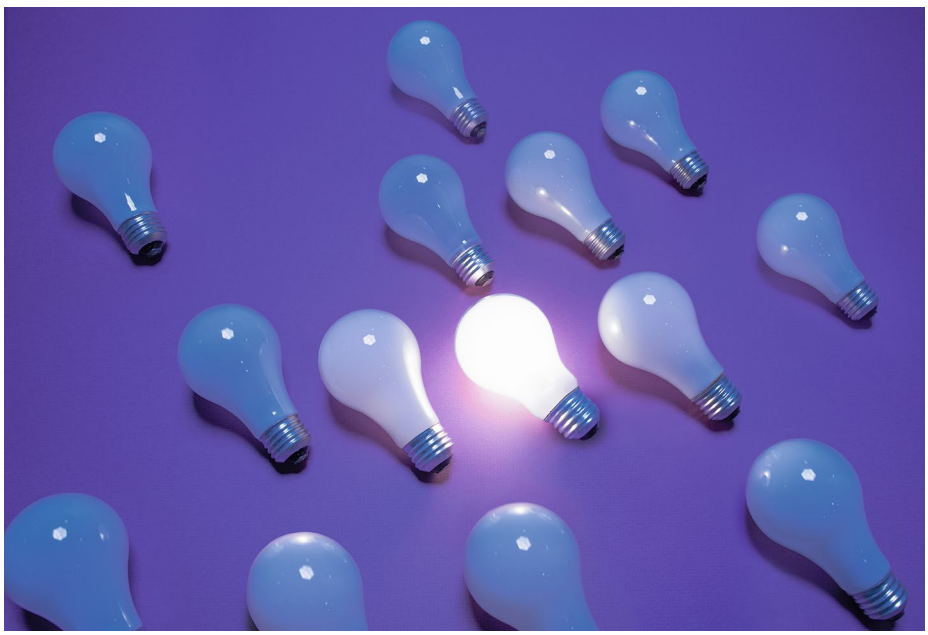
Developing theories by designing experiments that are aimed at falsifying them is a core endeavour in empirical sciences. By analysing 365 articles dedicated to the study of consciousness, Yaron et al.'s study[1] shows that there is almost no dialogue between the four main theories of this elusive phenomenon and gives us an interactive database with which to probe the literature.

Axel Cleeremans

When reasoning on the basis of hypotheses (or more generally, beliefs), we tend to seek confirmatory rather than contradictory evidence — exhibiting the well-known confirmation bias[2]. This manner of reasoning about how to establish truth, however, flies in the face of the core tenets of the scientific method. Popper's falsification principle[3] is at the core of the scientific endeavour: theories are only true as long as they are not proven false, and falsifying the predictions of a theory is what scientists should strive to achieve.

This verificationist posture applies not only to lay people, but also to scientists themselves. Negative or discomfirmatory findings are becoming less and less frequent in published studies, regardless of their domain[4]. This problem is further compounded, or perhaps amplified, by the fact that articles that report null results are much harder to publish than articles that offer significant results. The replication crisis in psychology stems in large part from the fact that failures to replicate, important as they are in contributing to reducing the publication bias, are often rejected by academic editors and reviewers alike. The adoption of statistical methods that make it possible to reason on the basis of null effects, and the emergence of preregistered studies that are published regardless of their outcomes, are both recent developments that are vital in addressing the publication bias and hence correcting the published record.

Yaron et al.'s article[1] clearly documents these trends by examining 365 articles (selected out of 6,938) published between 2001 and 2019 that either mention one of four main theories of consciousness in their title, abstract or keywords, or that cite a key paper from the authors of a given theory. The problem of understanding how the biological activity of the brain produces our mental states is now widely taken to constitute one of the most challenging scientific questions of the twenty-first century. And yet, the science of consciousness is but a nascent field of enquiry. There is continuing debate[5] not only about which methods are most appropriate to detect consciousness, but also about the very mechanisms that produce it. Four broad theories currently dominate the field: global neuronal workspace theory[6], integrated information theory[7], recurrent processing theory[8] and higher-order thought theories[9]. Although one may occasionally find tantalizing overlap between their core assumptions, the four theories are generally taken to be incompatible with each other. What do we know about each, and which have received the most empirical support?



PM Images / DigitalVision / Getty.

It is in answering this simple question that Yaron et al.'s analysis[1] already proves valuable, as it makes it possible to sense the field quantitatively for the first time. For instance, one realizes that global neuronal workspace theory is by far the theory that has received the most attention from experimenters; by contrast, higher-order thought theories remain surprisingly marginal despite their conceptual weight. But Yaron et al.'s study[1] goes much further than simply tallying support. Its most remarkable result is perhaps the finding that the vast majority of the included studies mention only one theory: they are simply meant to be confirmatory. Further, only one third of such studies were explicitly aimed at testing the predictions of a theory; the rest appeal to post hoc interpretation or mention a theory only in passing.

The second important finding is that the authors could actually predict which theory a study supports using machine learning: a random-forest classifier could achieve 80% accuracy in establishing which theory was supported by the study based exclusively on its methods. This finding reveals consistent

bias in methodological choices. For example, studies about global neuronal workspace theory tend to use high-level stimuli such as words and numbers whereas studies supporting recurrent processing theory[8] tend to use low-level stimuli such Gabor patches[10]. As striking as it is, the finding that one can predict theoretical orientation purely on the basis of methodology is not necessarily indicative of pervasive bias. Rather, it is an indication that the different theories do not always target the same aspect of consciousness, which in turn demonstrates the need for more crosstalk: a mature theory of consciousness should explain all of the data.

The most notable finding of Yaron and colleagues[1], however, concerns the heterogeneity of the spatial localizations of the cerebral regions identified by the four contender theories as being critically involved in subtending consciousness: when the brain imaging findings of the entire set of included studies are combined, the whole brain essentially lights up! This is rather disheartening and unhelpfully reinforces the perception that consciousness science is hopelessly confused. However, the fact that the different theories document different regions as subtending the contrast between conscious and unconscious processing also opens up the possibility of formulating distinct predictions that could be subject to empirical investigation.

Yaron et al.'s quantitative approach to meta-analysis is both novel and timely. Machine-readable hypothesis testing, in which theoretical predictions and empirical data can be processed algorithmically, will probably constitute an important development in the conduct of community-driven empirical research. Congruently, the authors have made the outcomes of their analysis publically available in the form of a interactive database that makes it possible for anyone to explore and probe the set of included studies — thus encouraging precisely the dialogue that is necessary to advance the field.

How do we move ahead, and go beyond the uneasy stasis that now characterizes consciousness research? Designing experiments that are specifically aimed at invalidating your theory requires courage, and possibility. However, when your standing in the field depends on the extent to which your theory is supported, courage may be hard to come by. This has led many authors to be content designing experiments to reinforce the (sometimes vague (often so out of necessity)) predictions of their theory, rather than to disprove them.

One remarkable initiative developed to counter this tendency towards confirmation is the Templeton World Charity Foundation's programme 'Accelerating research on consciousness'[11], which aims to foster 'adversarial collaboration' between theory leaders. The core aspect of such adversarial collaborations is to bring scientific opponents together and invite them to collaboratively design a critical experiment that they both agree has the power to falsify one of the theoretical positions at hand. The experiment is then carried out by independent teams and leveraging the full spectrum of good practices when conducting empirical research: preregistration, open data and replication. Different initiatives of this sort are currently underway, one[12] of which aims to compare the contrasting predictions of global neuronal workspace theory[6] and of integrated information theory[7]. The first predicts involvement of the prefrontal cortex in conscious processing, whereas the second assumes the implication of a posterior cortical 'hot zone' — clearly distinct predictions that are amenable to an empirical test.

Theory development, which is nascent when it comes to our understanding of consciousness, requires constant dialogue with empirical data, lest it be doomed to lose all explanatory power. What Yaron et al.'s remarkable study[1] shows, however, is that theory development also requires dialogue with competing theories. This dialogue should be buttressed by genuine collaboration and by a collective, concerted effort to advance knowledge for the benefit of all: Yaron et al.'s study is a first step in that direction. ❐

Axel Cleeremans ✉

Centre for Research in Cognition & Neurosciences, ULB Neuroscience Institute, Université libre de Bruxelles, Brussels, Belgium.

✉*e-mail: axcleer@ulb.ac.be*

Published online: 21 February 2022

https://doi.org/10.1038/s41562-021-01285-4

References

1. Yaron, I. et al. *Nat. Hum. Behav.*, https://doi.org/10.1038/s41562-021-01284-5 (2022).
2. Wason, P. C. *Q. J. Exp. Psychol.* **20**, 273–281 (1968).
3. Popper, K. R. *The Logic of Scientific Discovery* (Routledge, 2002) (original work published 1959).
4. Fanelli, D. *Scientometrics* **90**, 891–904 (2012).
5. Michel, M. et al. *Nat. Hum. Behav.* **3**, 104–107 (2019).
6. Mashour, G. A., Roelfsema, P., Changeux, J. P. & Dehaene, S. *Neuron* **105**, 776–798 (2020).
7. Koch, C., Massimini, M., Boly, M. & Tononi, G. *Nat. Rev. Neurosci.* **17**, 307–321 (2016).
8. Supèr, H., Spekreijse, H. & Lamme, V. A. *Nat. Neurosci.* **4**, 304–310 (2001).
9. Lau, H. & Rosenthal, D. *Trends Cogn. Sci.* **15**, 365–373 (2011).
10. Windey, B., Gevers, W. & Cleeremans, A. *Cognition* **129**, 404–409 (2013).
11. Templeton World. Accelerating research on consciousness. *templetonworldcharity.org*, https://go.nature.com/3gQOu5U (Templeton World Charity Foundation, 2021).
12. Melloni, L., Mudrik, L., Pitts, M. & Koch, C. *Science* **372**, 911–912 (2021).

Competing interests

The author declares no competing interests.