

What makes a theory of consciousness unscientific?

Authors	Klincewicz,Michat; Cheng,Tony; Schmitz,Michael; Sebastián,Miguel Ángel et al
Published in	Nature Neuroscience
DOI	10.1038/s41593-025-01881-x
Publication Date	2025-04
Link	https://research.tilburguniversity.edu/en/publications/509e11bc-ded5-4d07-ab90-bdfe9c8fbf91
Citation	Klincewicz, M, Cheng, T, Schmitz, M, Sebastián, M Á, Snyder, J S, Arnold, D H, Baxter, M G, Bekinschtein, T A, Bengio, Y, Bisley, J W, Browning, J, Buonomano, D, Carmel, D, Carrasco, M, Carruthers, P, Carter, O, Chang, D H F, Charest, I, Cherkaoui, M, Cleeremans, A, Cohen, M A, Corlett, P R, Christoff, K, Cumming, S, Cushing, C A, de Gelder, B, Brigard, F D, Dennett, D C, Dijkstra, N, Doerig, A, D ...
Download Date	2026-04-13 08:16:33
Rights	<p>General rights</p> <p>Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.</p> <ul style="list-style-type: none"> - Users may download and print one copy of any publication from the public portal for the purpose of private study or research. - You may not further distribute the material or use it for any profit-making activity or commercial gain - You may freely distribute the URL identifying the publication in the public portal" <p>Take down policy</p> <p>If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.</p>

What makes a theory of consciousness unscientific?

Corresponding authors:

Michał Klincewicz¹, Tony Cheng², Michael Schmitz³, Miguel Ángel Sebastián⁴, Joel S. Snyder⁵

- 1: Department of Cognitive Science and Artificial Intelligence, Tilburg University, the Netherlands;
Institute of Cognitive Science, Jagiellonian University, Poland
Email: m.w.klincewicz@tilburguniversity.edu
ORCID: 0000-0003-2354-197X
- 2: Waseda Institute for Advanced Study, Waseda University, Tokyo
Email: h.cheng.12@ucl.ac.uk
ORCID: 0000-0002-2288-0964
- 3: Institut für Philosophie, Universität Wien, Austria
Email: Michael.Schmitz@univie.ac.at
ORCID: 0000-0002-3898-1276
- 4: National Autonomous University of México, México
Email: msebastian@gmail.com
ORCID: 0000-0003-1890-8319
- 5: Department of Psychology, University of Nevada, Las Vegas, Nevada, USA
Email: joel.snyder@unlv.edu
ORCID: 0000-0002-5565-3063

Contributing Authors:

IIT concerned, Derek H. Arnold¹, Mark G. Baxter², Tristan A. Bekinschtein³, Yoshua Bengio^{4,5}, James W. Bisley^{6,7}, Jacob Browning⁸, Dean Buonomano^{6,7,9}, David Carmel¹⁰, Marisa Carrasco^{11,12}, Peter Carruthers¹³, Olivia Carter¹⁴, Dorita H. F. Chang¹⁵, Ian Charest¹⁶, Mouslim Cherkaoui⁷, Axel Cleeremans¹⁷, Michael A. Cohen^{18,19}, Philip R. Corlett^{20,21,22}, Kalina Christoff²³, Sam Cumming²⁴, Cody A. Cushing⁷, Beatrice de Gelder²⁵, Felipe De Brigard^{26,27,28,29}, Daniel C. Dennett³⁰, Nadine Dijkstra³¹, Adrien Doerig^{32,33}, Paul E. Dux¹, Stephen M. Fleming^{34,31}, Keith Frankish³⁵, Chris D. Frith^{31,36}, Sarah Garfinkel³⁷, Melvyn A. Goodale^{38,39}, Jacqueline Gottlieb^{40,41}, Jake Hanson⁴², Ran R. Hassin^{43,44}, Michael H. Herzog⁴⁵, Cecilia Heyes^{46,47}, Po-Jang Hsieh⁴⁸, Shao-Min Hung⁴⁹, Robert Kentridge⁵⁰, Tomas Knapen^{51,52}, Nikos Konstantinou⁵³, Konrad Kording^{54,55,56,57}, Timo L. Kvamme⁵⁸, Sze Chai Kwok^{29,59}, Renzo C. Lanfranco⁶⁰, Hakwan Lau^{61,62,63}, Joseph LeDoux^{64,65,66}, Alan L. F. Lee⁶⁷, Camilo Libedinsky⁶⁸, Matthew D. Lieberman⁷, Ying-Tung Lin⁶⁹, Ka-Yuet Liu^{61,70}, Maro G. Machizawa^{71,72,73}, Janet Metcalfe⁷⁴, Matthias Michel⁷⁵, Kenneth D. Miller^{76,77,78,79,41}, Partha P. Mitra⁸⁰, Dean Mobbs^{81,82}, Robert M. Mok⁸³, Jorge Morales^{84,85}, Myrto Mylopoulos⁸⁶, Brian Odegaard⁸⁷, Charles C.-F. Or⁸⁸, Adrian M. Owen^{38,39,89}, David Pereplyotchik⁹⁰, Franco Pestilli⁹¹, Megan A. K. Peters^{92,93}

Ian Phillips⁹⁴, Rosanne L. Rademaker⁹⁵, Dobromir Rahnev⁹⁶, Geraint Rees³⁷, Dario Ringach^{7,6}, Adina Roskies⁹⁷, Daniela Schiller^{98,99}, Aaron Schurger^{100,101}, D. Samuel Schwarzkopf¹⁰², Ryan B. Scott¹⁰³, Aaron R. Seitz^{84,104,105}, Joshua Shepherd^{106,107}, Juha Silvanto¹⁰⁸, Heleen A. Slagter^{109,110}, Barry Smith³⁷, Guillermo Solovey¹¹¹, David Soto¹¹², Hugo Spiers¹¹³, Timo Stein^{114,115}, Vincent Taschereau-Dumouchel^{116,117}, Frank Tong¹¹⁸, Peter U. Tse¹¹⁹, Jonas Vibell¹²⁰, Sebastian Watzl¹²¹, Taylor Webb⁷, Josh Weisberg¹²², Thalia Wheatley^{119,123}, Michał Wierchoń¹²⁴, Martijn E. Wokke¹²⁵, Karen Yan⁶⁷

- 1: School of Psychology, The University of Queensland, St Lucia, Queensland, Australia
- 2: Section on Comparative Medicine, Department of Pathology, Wake Forest University School of Medicine, Winston-Salem, North Carolina, USA
- 3: Consciousness and Cognition Lab, Department of Psychology, University of Cambridge, UK
- 4: Department of Computer Science and Operations Research Department, Université de Montréal
- 5: Montreal Institute for Learning Algorithms
- 6: Department of Neurobiology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA
- 7: Department of Psychology, University of California Los Angeles, Los Angeles, CA, USA
- 8: Computer Science Department, New York University, New York, NY, USA
- 9: Integrative Center for Learning and Memory, University of California Los Angeles, Los Angeles, CA, USA
- 10: School of Psychology, Victoria University of Wellington, New Zealand
- 11: Department of Psychology, New York University, New York, NY, USA
- 12: Center for Neural Science, New York University, New York, NY, USA
- 13: Department of Philosophy, University of Maryland, College Park, MD 20742, USA
- 14: Melbourne School of Psychological Sciences, University of Melbourne, Parkville, VIC, Australia
- 15: Department of Psychology, The University of Hong Kong, Hong Kong
- 16: Cerebrum, Département de psychologie, Université de Montréal, Montréal, Québec, Canada
- 17: Consciousness, Cognition and Computation Group (CO3), Center for Research in Cognition and Neuroscience (CRCN), ULB Neuroscience Institute (UNI), Université Libre de Bruxelles, Brussels, Belgium
- 18: Department of Psychology and Program in Neuroscience, Amherst College, USA
- 19: McGovern Institute of Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, USA
- 20: Department of Psychiatry, Yale University, New Haven, CT, USA
- 21: Department of Psychology, Yale University, New Haven, CT, USA
- 22: Wu Tsai Institute, Yale University, New Haven, CT, USA
- 23: Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada
- 24: Department of Philosophy, University of California Los Angeles, Los Angeles, CA, USA
- 25: Cognitive Neuroscience/Brain and emotion lab, Oxfordlaan 55, Maastricht University, The Netherlands
- 26: Department of Philosophy, Duke University, USA
- 27: Department of Psychology and Neuroscience, Duke University, USA
- 28: Center for Cognitive Neuroscience, Duke University, USA
- 29: Duke Institute for Brain Sciences, Duke University, USA
- 30: Center for Cognitive Studies, Tufts University, Medford, MA, USA

- 31: Department of Imaging Neuroscience, University College London, 12 Queen Square, London WC1N 3AR, UK
- 32: Institute of Cognitive Science, University of Osnabrück, Germany
- 33: Department of Psychology and Education, Freie Universität Berlin, Germany
- 34: Department of Experimental Psychology, University College London, London, UK
- 35: Department of Philosophy, University of Sheffield, Sheffield, UK
- 36: Institute of Philosophy, School of Advanced Study, University of London, Senate House, Malet Street, London WC1E 7HU, UK
- 37: Institute of Cognitive Neuroscience, University College London, London, UK
- 38: Western Institute for Neuroscience, University of Western Ontario, London, Ontario N6A 5B7, CANADA
- 39: Department of Psychology, University of Western Ontario, London, Ontario N6A 5C2, CANADA
- 40: Department of Neuroscience, The Kavli Institute for Brain Science, Columbia University, New York, NY, USA
- 41: Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA
- 42: Association for Mathematical Consciousness Science, Munich Center for Mathematical Philosophy, Munich, BY, Germany
- 43: Psychology Department, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel
- 44: The Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel
- 45: Laboratory of Psychophysics, École Polytechnique, Fédérale de Lausanne (EPFL), Switzerland
- 46: All Souls College, University of Oxford, Oxford OX1 4AL, UK
- 47: Department of Experimental Psychology, University of Oxford, Oxford OX2 6GG, UK
- 48: Department of Psychology, National Taiwan University, Taiwan
- 49: Waseda Institute for Advanced Study, Waseda University, Tokyo, Japan
- 50: Department of Psychology, University of Durham, Durham DH1 3LE, UK
- 51: Spinoza Centre for Neuroimaging & Netherlands Institute for Neuroscience, Royal Dutch Academy of Arts and Sciences, Amsterdam, the Netherlands
- 52: Department of Behavioral and Movement Sciences, Vrije Universiteit Amsterdam, the Netherlands
- 53: Department of Rehabilitation Sciences, Cyprus University of Technology, Cyprus
- 54: Penn Integrates Knowledge (PIK) program, University of Pennsylvania, USA
- 55: Department of Bioengineering, University of Pennsylvania, USA
- 56: Department of Neuroscience, University of Pennsylvania, USA
- 57: Learning in Machines and Brains program, Canadian Institute for Advanced Research, Canada
- 58: Center of Functionally Integrative Neuroscience (CFIN), Department of Clinical Medicine, Aarhus University, Denmark
- 59: Phylo-Cognition Laboratory, Division of Natural and Applied Sciences, Data Science Research Center, Duke Kunshan University, Kunshan, Jiangsu, China
- 60: Department of Neuroscience, Karolinska Institute, Solnavägen 9, 171 65 Stockholm, Sweden
- 61: RIKEN Center for Brain Science, Wakoshi, Saitama, Japan
- 62: Center for Neuroscience Imaging Research, Institute for Basic Science, Suwon, Korea
- 63: Department of Biomedical Engineering, Sungkyunkwan University, Suwon, Korea

- 64: Center for Neural Science and Department of Psychology, New York University, New York, NY, USA
- 65: Department of Psychiatry, New York University Langone Medical School, New York, NY, USA
- 66: Department of Child and Adolescent Psychiatry, New York University Langone Medical School, New York, NY, USA
- 67: Department of Psychology, Lingnan University, Hong Kong
- 68: Department of Psychology, National University of Singapore, Singapore
- 69: Institute of Philosophy of Mind and Cognition, National Yang Ming Chiao Tung University, Taiwan
- 70: Department of Sociology, University of California Los Angeles, Los Angeles, CA, USA
- 71: Center for Brain, Mind and KANSEI Sciences Research, Hiroshima University, Hiroshima, Japan
- 72: Xiberlinc Inc., Tokyo, Japan
- 73: Center for Brain Integration Research, Tokyo Medical and Dental University, Tokyo, Japan
- 74: Department of Psychology, Columbia University, New York, NY, USA
- 75: Department of Linguistics and Philosophy, MIT
- 76: Center for Theoretical Neuroscience, Columbia University, New York, NY, USA
- 77: Department of Neuroscience, Columbia University, New York, NY, USA
- 78: Swartz Program in Theoretical Neuroscience, Columbia University, New York, NY, USA
- 79: Kavli Institute for Brain Science, College of Physicians and Surgeons, Columbia University, New York, NY, USA
- 80: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA
- 81: Department of Humanities and Social Sciences, California Institute of Technology, Pasadena, California 91125, USA
- 82: Computation and Neural Systems Program, California Institute of Technology, Pasadena, CA 91125, USA
- 83: Royal Holloway, University of London
- 84: Department of Psychology, Northeastern University, Boston, Massachusetts, USA
- 85: Department of Philosophy, Northeastern University, Boston, Massachusetts, USA
- 86: Department of Philosophy and Department of Cognitive Science, Carleton University, Ottawa, Canada
- 87: Department of Psychology, University of Florida, Gainesville, Florida, USA
- 88: Psychology Programme, School of Social Sciences, Nanyang Technological University, Singapore
- 89: Department of Physiology and Pharmacology, University of Western Ontario, London, Ontario N6A 5B7, CANADA
- 90: Department of Philosophy, Kent State University, Ohio, USA
- 91: Department of Psychology, Department of Neuroscience, Center for Perceptual Systems, Center for Learning and Memory, The University of Texas, Austin, Texas, USA
- 92: Department of Cognitive Sciences, University of California Irvine, Irvine, CA, USA
- 93: Program in Brain, Mind, & Consciousness, Canadian Institute for Advanced Research, Toronto, Ontario, Canada
- 94: Department of Philosophy and Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, Maryland, USA
- 95: Ernst Strüngmann Institute for Neuroscience in cooperation with the Max Planck Society, Frankfurt, Germany
- 96: School of Psychology, Georgia Institute of Technology, Atlanta, GA, USA

- 97: Department of Philosophy, University of California, Santa Barbara, California, USA
- 98: Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA
- 99: Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
- 100: Department of Psychology, Crean College of Health and Behavioral Sciences, Chapman University, Orange, CA, USA
- 101: Institute for Interdisciplinary Brain and Behavioral Sciences, Chapman University, Irvine, CA, USA
- 102: School of Optometry & Vision Science, The University of Auckland, 85 Park Road, Grafton, Auckland, New Zealand
- 103: School of Psychology, University of Sussex, UK
- 104: Center for Cognitive and Brain Health, Boston, MA, USA
- 105: University of California Riverside, UCR Brain Game Center for Mental Fitness and Wellbeing, Riverside, CA, USA
- 106: ICREA, Barcelona, Spain
- 107: Universitat Autònoma de Barcelona, Departament de Filosofia, Cerdanyola del Vallès, Spain
- 108: School of Psychology, Faculty of Health and Medical Sciences, University of Surrey, United Kingdom
- 109: Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, the Netherlands
- 110: Institute for Brain and Behavior, Vrije Universiteit Amsterdam, the Netherlands
- 111: Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires - CONICET, Argentina
- 112: Basque Center on Cognition, Brain and Language, San Sebastián, Spain
- 113: Department of Experimental Psychology, Division of Psychology and Language Sciences, University College London, UK
- 114: Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands
- 115: Amsterdam Brain and Cognition, University of Amsterdam, Amsterdam, The Netherlands
- 116: Department of Psychiatry and Addictology, Université de Montréal, Montreal, Quebec, Canada
- 117: Centre de Recherche de l'Institut Universitaire en Santé Mentale de Montréal, Montreal, Quebec, Canada
- 118: Department of Psychology, Vanderbilt University, Nashville, Tennessee, USA
- 119: Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755, USA
- 120: Brain & Behavior Laboratory, University of Hawai'i at Mānoa, HI, USA
- 121: Department of Philosophy, Classics, History of Art and Ideas, University of Oslo, Norway
- 122: Department of Philosophy, University of Houston, 515 Agnes Arnold Hall, Houston, TX 77204, USA
- 123: Santa Fe Institute, Santa Fe, NM 87501, USA
- 124: Consciousness Lab, Institute of Psychology & Centre for Brain Research, Jagiellonian University, Krakow, Poland
- 125: Department of Psychology, University of Girona, Girona, Spain

Standfirst: Theories of consciousness have a long and controversial history. Recently, one well-known proposal – Integrated Information Theory (IIT) – has been labeled as “pseudoscience,” causing a heated open debate. Here we discuss the case and argue that the theory is indeed unscientific because its core claims are untestable even in principle.

Introduction

Theorizing about consciousness is challenging because some of the issues lie at the border between empirical science and philosophical inquiry. One recent example of this challenge is the controversial and yet highly publicized view called the Integrated Information Theory (IIT) of consciousness. IIT seeks to provide an explanation of why certain systems, such as the human brain, generate rich and complex conscious experiences while others do not. It does so by linking consciousness with the ability of a system to integrate information. IIT proposes that consciousness arises when a system's components interact in a way that creates a unified whole, quantified by a custom measure of complexity called 'Phi' (Φ). The higher the Φ value, the more conscious the system is. Central to IIT are the concepts of intrinsic causality, where a system's current state influences its own future and past states, and cause-effect structure, which determines the specific quality of conscious experiences.

This ambitious theory has been promoted as "well-established,"¹ "a gigantic step in the final resolution of the ancient mind-body problem," "the only dominant theory," and "the most promising scientific theory" of consciousness². However, we argue that the core ideas of IIT lack empirical support and are metaphysical and not scientific. Because previous expressions of such concerns about the theory³ have not received sufficient attention, a group of us wrote an open letter that questioned the scientific status of IIT⁴.

Whether IIT is pseudoscience is important because of the practical and ethical importance of understanding consciousness. In response to widespread and polarized reactions to the letter, the limited empirical support and radical claims concerning who and what is conscious, here we expanded on the original letter to explain why it was written and to clarify which claims made by IIT proponents are problematic. We distinguish the core claims of IIT, which we argue are unscientific, from work merely inspired by IIT that uses measures of complexity to investigate brain activity associated with different states of wakefulness. This latter work is uncontroversially scientific but neither supports nor tests the core claims of IIT.

Scientific theories of consciousness should be testable

Scientific theories of consciousness aim to identify specific brain processes and/or psychological functions crucial for subjective experiences. These theories generally make claims that are readily testable with the tools of psychology or cognitive neuroscience (Box 1).

In contrast, IIT postulates that “consciousness is a fundamental feature of the universe,” which it seeks to understand starting from a set of claims (“axioms”) that are said to be “irrefutable” and true of every conscious experience⁵. However, some have criticized these axioms⁶ and argued that they lead to circularity⁷. The axioms have also evolved^{5,6}, casting doubt on their “irrefutability.”

From these controversial axioms, proponents of IIT claim to derive an index of the a system's degree of integrated information, Φ , which, they argue, is an objective measure of the degree of consciousness in any system. However, the axioms do not by themselves specify a particular mathematical model. To develop a precise formula for Φ , further assumptions must be made. This challenge becomes evident in the evolving interpretations and understandings of the theory and in the changes in formulas defining Φ over time^{3,8}.

More concerning are other problems that prevent the theory from being empirically validated. An obvious test of IIT would be to measure Φ in humans. To calculate Φ one needs to consider how every component of the system at every spatial and temporal scale—say from the quantum scale to brain regions, and from nanoseconds to days—constrains the past and future states of the system⁵. Given the complexity of these calculations, it is impossible in principle to compute Φ , and hence impossible to identify conscious states—a point proponents of IIT readily admit.

Currently, Φ can only be precisely calculated in idealized systems with a few units³, such as limited grids of logic gates. Simple artificial systems like these are conscious according to IIT, even when the relevant nodes are all inactive⁸.

Hence, the theory readily ascribes consciousness to static systems that do not perform any function. Stating that such systems have ‘consciousness’, when their ‘consciousness’ has no observable consequences, is, in this respect, no different than stating that they have ‘souls’. Based on these implications alone, some consider the theory dismissible^{6,7}. However, proponents of IIT embrace this counterintuitive consequence without calling their theoretical assumptions into question³.

Making theoretical predictions that may not be testable by current methodology does not necessarily render the relevant theories unscientific. However, if the core claims of a theory are not testable even in principle, the scientific status of the theory should be called into question. Nevertheless, some researchers claim to have tested IIT empirically, and have promoted the results as providing partial support for the theory or for its superiority to other theories^{1,2}. Yet, as others have argued, the tests are problematic and not meaningful⁹, because they do not address the core claims of IIT, and the tested predictions do not sufficiently differentiate between IIT and other theories of consciousness.

While IIT may provoke interesting philosophical discussions, we contend that it falls short of the empirical basis required for scientific inquiry. Untestable metaphysical views can of course be discussed, but when they are treated as empirically testable science, this is problematic. In light of the above, we argue that IIT demonstrates characteristics that are hallmarks of pseudoscience (Box 2).

Practical and ethical consequences

Attributing consciousness to systems with minimal complexity has important practical consequences¹⁰. This is because a theory of consciousness may be used to inform practical decisions regarding morally significant legal and medical issues. By promoting to the general public the untestable idea that ' Φ =consciousness', proponents of IIT may ultimately have an unjustified impact on law and policy, including on decisions that involve measures of quality of life, clinical triage, abortion, the rights of non-responsive patients, and welfare considerations for insects, organoids, and artificial intelligences. Given the practical consequences of adopting and promoting any one theory of consciousness over others, we recommend caution and scientific humility—especially for a theory without direct and meaningful empirical support, such as IIT.

IIT-inspired research on brain complexity

It is important to distinguish research that purports to directly test IIT as a theory of consciousness from research that is more loosely inspired by IIT and does not purport to test it¹¹. Research inspired by IIT includes studies on the complexity of neural activity induced by magnetic brain stimulation during different states of wakefulness, sleep, and general anesthesia. These studies show that in some states, such as non-REM sleep and general anesthesia, neural activity has greater uniformity and therefore lower degrees of complexity when compared to wakefulness¹².

While this research is assuredly scientific, interesting, and potentially of practical importance, it does not lend support to IIT. IIT proposes a general theory that identifies which objects are conscious, from humans to static logic gates: it proposes that the 'degree of consciousness' of any object is assayed by a specific measure of complexity, the integrated information Φ . It is not surprising that human states of conscious awareness show greater neural complexity than states of unconsciousness, and this is likely to be true using almost any sensible measure of complexity. Indeed, many other theories of consciousness also predict lower brain complexity during certain non-awake states, such as slow-wave sleep and general anaesthesia¹³. Testing the hypothesis that conscious vs. non-conscious states can be distinguished in humans or some other animals by various measures of neural complexity does not distinguish IIT from these other theories, nor does it test the core hypothesis of IIT that Φ is a general measure of the degree of consciousness of any creature or non-living object. What would be needed to make it a scientific theory is a measure that directly derived from the mathematical apparatus of IIT, akin to how a particular measurement of the motion of a pendulum—like instantaneous velocity—is directly derived from the harmonic oscillation equation. Yet, the measures employed by IIT-inspired studies do not assay Φ specifically, but only some more common notion of complexity¹³. Therefore, these measures are unable to distinguish IIT from other theories.

Furthermore, the comparison of different global states, such as wakefulness and coma, is confounded by many different cognitive and behavioral factors^{6,13}. In particular, much of brain activity during wakefulness is not conscious, so a neural correlate of a waking state is not necessarily a neural correlate

of consciousness as a mental state. Without other controls, it is impossible to draw strong conclusions about consciousness just by comparing different global states, such as wakefulness and coma.

Conclusion

Our main concerns with IIT are twofold. First, it lacks well-defined empirically testable consequences. Second, it explicitly asserts or entails bold non-empirical claims, such as that "the entire cosmos is suffused with sentience"¹², that neurons do not truly exist¹⁴, or that consciousness can be attributed to a static array of logic gates. These bold claims threaten to delegitimize the scientific study of consciousness. We need to be careful what we label as scientific, especially given the practical and ethical significance of consciousness. We appreciate that the word 'pseudoscience' should not be used lightly, but there are good reasons to think that it applies in this case. We sincerely hope that discussing these reasons will move the scientific study of consciousness forward.

BOX 1: Common theories of consciousness and their empirical predictions

An influential method for studying consciousness is to compare what happens at the neuronal and psychological levels in conscious versus unconscious experimental conditions. This method, called 'contrastive analysis', provides the main source of data for current theories, such as global neuronal workspace theory (GNWT), higher-order theories, or the local recurrence theory (LRT) of consciousness. As such, although some of these views have theoretical origins in computer science or philosophy, the core empirical predictions are usually made at levels of observation testable with current methods in cognitive neuroscience. These predictions are often not quantitatively precise, but they are nonetheless constrained by the empirical literature.

For instance, GNWT has long held that a putative frontoparietal network in the mammalian brain should support the broadcast of stimulus-related information during conscious experience. Therefore, if someone reports being conscious of a stimulus without relevant information being represented in this brain network, the theory is put under pressure. One may doubt this particular theoretical proposal, but it is a reasonable theory to test given the importance of frontoparietal areas in functions related to consciousness, such as working memory, attention, and cognitive control.

Other theories also make clear predictions. For instance, LRT predicts that the first feedforward sweep of neuronal processing of visual stimuli occurs unconsciously, and conscious perception only arises when information feeds back from higher visual areas to early visual areas. Several empirical tests have

validated this prediction. Importantly, these empirical studies not only tested LRT but also challenged other theories.

Theories not only make predictions at the neurobiological level, but also at the psychological and cognitive levels. For instance, higher-order theories predict that it should be possible to manipulate subjective experience while holding fixed lower-level sensory responses and perceptual capacities that enable discrimination between similar stimuli.

If proponents of IIT were to give up the core claims (axioms), some weaker version of the theory could perhaps be cast in terms of cognitive neuroscience¹¹ (see also the section in the main text on IIT-inspired research on brain complexity). However, to the extent that IIT's metaphysical commitments remain promoted as key features, this stronger version of IIT, which is our target here, is incompatible with standard approaches in cognitive neuroscience for testing ideas empirically^{6,7}.

BOX 2: Evaluating the scientific status of a theory: applying criteria to IIT

Besides the issue of untestability, the following points also contribute to our doubt about IIT as science.

Overreach: Recent developments have seen the theory expanding into areas that traditionally belong to metaphysics, asserting bold claims about the nature of reality, causation, and the *non*-existence of neurons¹⁴. These assertions are interesting, but they venture beyond the boundaries of empirical science. IIT also ambitiously addresses fundamental existential questions, such as the essence of being and the presence of free will, elevating the theory to an “all-encompassing explanatory framework”¹⁴. According to some, this kind of ambition is a common hallmark of pseudoscientific theories¹⁵, which often claim to provide comprehensive explanations for a wide variety of complex phenomena without sufficient empirical support.

Incompatibility with physics: Doubts have been raised regarding IIT's compatibility with current physics⁵, a possible concern for any neurobiological theory. The theory's proponents acknowledge these compatibility issues⁵, yet IIT continues to be promoted as a promising or well-established scientific theory.

Misrepresentation and promotion: Despite its lack of scientific support, IIT is often presented to the public as a direct competitor to other scientific theories of consciousness². This misrepresentation is amplified by media coverage and public endorsements, which may mislead the general public about the theory's empirical validity. While misrepresentation itself does not make the theory unscientific, it suggests that the label ‘pseudoscience’ may be appropriate.

The pseudoscience criterion: Standard dictionary definitions and philosophical discussions¹⁵ describe pseudoscience as featuring unscientific claims (e.g., untestable in principle) that are falsely presented as scientific. This criterion applies to IIT since it portrays untestable assertions as empirically grounded science.

Competing interests

The authors declare no competing interests.

Acknowledgments

Co-author Daniel C. Dennett sadly passed away on 19th April 2024 in the final stages of preparing this paper for submission.

References

1. Cogitate Consortium *et al.* An adversarial collaboration to critically evaluate theories of consciousness. *bioRxiv* 2023.06.23.546249 (2023) doi:10.1101/2023.06.23.546249.
2. Lee, A. L. A reply to Dr. Erik Hoel's post regarding the overpromotion of IIT. <https://osf.io/preprints/psyarxiv/g35ak/> (2023).
3. Open letter to NIH on Neuroethics Roadmap (BRAIN initiative). <http://inconsciousnesswetrust.blogspot.com/2020/05/open-letter-to-nih-on-neuroethics.html> (2019).
4. IIT Concerned *et al.* The Integrated Information Theory of Consciousness as Pseudoscience. (2023) doi:<https://doi.org/10.31234/osf.io/zsr78>.
5. Albantakis, L. *et al.* Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms. *PLOS Computational Biology* **19**, e1011465 (2023).
6. Merker, B., Williford, K. & Rudrauf, D. The integrated information theory of consciousness: A case of mistaken identity. *Behavioral and Brain Sciences* **45**, e41 (2022).
7. Herzog, M. H., Schurger, A. & Doerig, A. First-person experience cannot rescue causal structure theories from the unfolding argument. *Conscious Cogn* **98**, 103261 (2022).
8. Hanson, J. R. & Walker, S. I. On the non-uniqueness problem in integrated information theory. *Neuroscience of Consciousness* **2023**, niad014 (2023).
9. Lau, H. Where is the 'posterior hot zone'? Open Review of Ferrante et al (2023): "An Adversarial Collaboration to Critically Evaluate Theories of Consciousness" (by the ARC-Cogitate Consortium). (2023) doi:<https://doi.org/10.31234/osf.io/93ufe>.

10. Jeziorski, J. *et al.* Brain organoids, consciousness, ethics and moral status. *Semin Cell Dev Biol* **144**, 97–102 (2023).
11. Michel, M. & Lau, H. On the dangers of conflating strong and weak versions of a theory of consciousness. *Philosophy and the Mind Sciences* **1**, (2020).
12. Casali, A. G. *et al.* A theoretically based index of consciousness independent of sensory processing and behavior. *Science translational medicine* **5**, 198ra105-198ra105 (2013).
13. Sitt, J. D., King, J.-R., Naccache, L. & Dehaene, S. Ripples of consciousness. *Trends in Cognitive Sciences* **17**, 552–554 (2013).
14. Tononi, G., Albantakis, L., Boly, M., Cirelli, C. & Koch, C. Only what exists can cause: An intrinsic view of free will. Preprint at <https://doi.org/10.48550/arXiv.2206.02069> (2023).
15. Hansson, S. O. Science and Pseudo-Science. in *The Stanford Encyclopedia of Philosophy* (ed. Zalta, E. N.) (Metaphysics Research Lab, Stanford University, 2021).