This is a peer-reviewed manuscript submitted to Melloni, L. & Olcese, U. (Eds.) (Forthcoming), *The Scientific Study of Consciousness– Experimental and Theoretical Approaches.* Springer Nature.

Testing for consciousness beyond consensus cases

Niccolò Negro

School of Psychological Sciences, Tel Aviv University, Israel
niccolonegro@tauex.tau.ac.il
https://orcid.org/0000-0002-1561-799X

Liad Mudrik

School of Psychological Sciences, Tel Aviv University, Israel;
Sagol School of Neuroscience, Tel Aviv University, Israel;
Canadian Institute for Advanced Research (CIFAR), Brain, Mind, and Consciousness Program, Toronto, ON, Canada.
https://orcid.org/0000-0003-3564-6445

This chapter addresses the development of tests for consciousness (C-tests), defined as any protocol or methodology devised to detect specific properties that, if present, would justify higher credence in the belief that the system under test is phenomenally conscious. Though inherently defeasible, C-tests are vital for reducing epistemic uncertainty, balancing ethical and practical considerations regarding the attribution of consciousness to systems like patients with disorders of consciousness, non-human animals, and artificial systems. In this chapter, we first present a taxonomy of current available C-tests, describing how they rely on specific neural and/or psychological properties to reduce uncertainty about the presence of consciousness in various target systems. Second, we clarify the notion of phenomenal consciousness as the target of C-tests, delineating the limits of C-tests in being able to capture it. Third, we address the question of whether a well-established theory of consciousness and/or pre-theoretical intuitions are necessary for validation of C-tests. Fourth, we evaluate several inferential strategies to justify extrapolations of consciousness from consensus to non-consensus cases. Finally, we conclude by describing the iterative natural kind approach as a multidimensional method that integrates multiple tests with weighted evidence. This model would provide probabilistic assessments of consciousness across different populations, offering a more reliable framework for addressing non-consensus cases and providing a valuable aid for practical decision-making.

Introduction

One of the most pressing questions in the study of consciousness is how we can identify consciousness in cases where reports cannot be obtained (Bayne et al., 2024). That is, how can we determine if there is "something it is like" (Nagel, 1974) to be the person, animal, or system of interest: do they have qualitative experiences? This question is theoretically important, as being able to identify consciousness in different cases contributes to our understanding of consciousness and its manifestations. Yet no less crucially, it is also of practical and societal importance, as we face more and more edge cases where it is unclear whether a system is conscious, and if so – how we can tell that it is. These cases range between the clinical domain (e.g., patients with disorders of consciousness), the developmental one (i.e., asking when consciousness emerges in foetuses and infants), cases of non-human animals, and, what has become an increasingly relevant question, artificial intelligence and synthetic systems (e.g., neural organoids or xenobots). This chapter accordingly focuses on tests for consciousness (or "C-tests" for short), defined as any protocol or methodology devised to detect specific properties that, if present, would justify higher credence in the belief that the system under test is phenomenally conscious.

This definition calls for three important clarifications. First, C-tests specifically target phenomenal consciousness, as opposed to other aspects or forms of consciousness (e.g., non-phenomenal cognitive accessibility of information, self-consciousness, etc.), or related phenomena like intelligence. However, given the inherent difficulty to directly probe phenomenal consciousness (Block, 1995, 2019), many tests do so indirectly, by establishing a specific link between phenomenality and other properties, and testing for those properties (more on this in 2.1.). Second, C-tests target consciousness as a global state of a system, as opposed to probing the specific contents of that state. That is, C-tests are supposed to answer the question "is this system conscious now?", rather than the question "what is this system conscious of now?” (but see Fink, 2024). In philosophical terms, C-tests target consciousness as a generic, determinable property (akin to “being coloured”), and not as a specific, determinate mode of (e.g., “being red”, or “being green” – see Bayne, 2007; Bayne & Hohwy, 2016; more on this in 2.2.). Third, C-tests are not devised to provide a definitive answer. They are, rather, probability-raising methods for the attribution of consciousness to various systems. In this sense, attributions of consciousness remain defeasible and uncertain, and the scope of C-tests is to reduce this uncertainty to a degree that can be tolerated in decision-making and practical contexts (more on this in section 4).

This third consideration might accordingly lead one to question the need for C-tests. If we already know that they will not yield a definitive answer, why should we spend time and effort developing them? We argue that C-tests are important for resolving the tension between two opposite forces: the first practical, the second epistemic. The practical force requires us to make decisions that are ethically and practically significant based on whether a certain system is conscious, since it is plausible to think that our obligations towards different systems vary depending on whether those systems are conscious or not[1]. Indeed, many scholars have argued that the moral status of an entity is grounded in that entity being conscious (Lee, 2022; Shepherd, 2018; Singer, 2011; see Tannenbaum & Jaworska, 2018 for a discussion). However, consciousness science is still a nascent field in many ways, and our understanding of consciousness is limited (Francken et al., 2022; Seth & Bayne, 2022; Yaron, Melloni, Pitts, & Mudrik, 2022). This epistemic limitation means that any decision we make based on what we currently know could be mistaken. The importance and urgency of the practical consideration on the one hand, and the epistemic doubt on the other hand, necessitate the development of trustable, albeit defeasible, methods for determining whether a system is conscious or not.

[1] This stance has already led to some practical applications, particularly in the context of animal welfare (Birch et al., 2021; Crump et al., 2022), but there are attempts to extend it to AI ethics (Metzinger, 2021).

C-tests provide such methods. They rely on well-established relations between neural, behavioural, and/or psychological properties and the presence of consciousness in consensus cases, in order to obtain some information about consciousness in non-consensus cases. Thus, good-enough C-tests promise to reduce the epistemic uncertainty of consciousness science, allowing it to better substantiate its practical commitments. This shows how C-tests differ from standard NCC (i.e., neural correlates of consciousness; see Chalmers, 2000) research: in the latter case, researchers manipulate consciousness and look for the underlying neural correlates (so they already know if and when the system is conscious) while for the former case, the researchers conduct the test to find out if the system is conscious.

In this chapter, we limit ourselves to the analysis of the *epistemic* aspects of C-test, focusing on questions related to their target, their validity, and their applicability to different populations. First, we start by presenting some of the most influential C-tests.

1. A taxonomy of C-tests

Many different C-tests have been proposed in the literature, and providing a comprehensive and exhaustive list goes beyond the scope of this chapter (but see Bayne et al., 2024; Dung, 2022, 2023). However, in the table below we provide a list of several influential and widely discussed C-tests, explain how they work, and elucidate what their primary target is. Importantly, the goal of this table is to give an overview of the types of C-tests currently available to consciousness scientists, rather than to provide an evaluation of how compelling these C-tests are.

TEST	PRIMARY GOAL	HOW IT WORKS	KEY REFERENCES	PRIMARY TARGET POPULATION
Command following test	Designed primarily for detecting consciousness in clinical settings (in adult humans with disorders of consciousness)	A patient is brought into an fMRI machine and is asked to perform some mental functions that in healthy participant evoke activity in different brain areas (e.g., the Supplementary Motor Area when imagining playing tennis vs. Hippocampus when imagining walking in their own house). If the brain response is similar enough to that of healthy participants performing the same functions, the patient passes the test.	(Bodien et al., 2024; Owen, 2018; Owen et al., 2006)	Humans
Glasgow Coma Scale	Exclusively designed for clinical contexts, typically used as the primary C-test for initial assessment of the patient's level of consciousness	Clinicians' test that probes the patient's responsive behaviour given some criteria organised along three dimensions: eye-opening; motor responsiveness; verbal responsiveness. The score for each criterion is then summed up, and the higher the score, the higher the level of consciousness.	(Teasdale & Jennett, 1974; Teasdale et al., 2014)	Humans
Narrative Comprehension Test	Designed primarily for detecting consciousness in clinical settings (in adult humans with disorders of consciousness)	Patients with disorders of consciousness have their brain activity recorded while watching a movie or hearing a tape. The test is passed if the brain activity and executive functions are similar to those exhibited by healthy individuals.	(Naci, Cusack, Anello, & Owen, 2014; Naci, Sinai, & Owen, 2017)	Humans
Global detection test	Designed primarily for detecting consciousness in clinical settings (in adult humans)	The patient is presented with several sequences of tones. In each sequence, the last tone is an oddball, which creates a local irregularity. In the last sequence there is no oddball, thereby creating a	(Bekinschtein et al., 2009)	Humans

		global irregularity. The test is passed if the patient detects this global irregularity, where this detection is signalled by the P300 component		
Perturbational Complexity Index (PCI)	Designed primarily for detecting consciousness in clinical settings (in adult humans with disorders of consciousness)	A magnetic pulse is transmitted to the brain through TMS. The subsequent brain activity is measured through EEG. The test is passed if the spatiotemporal complexity of the brain activity following the pulse (i.e., its index of "incompressibility") is comparable to that of healthy and conscious (either awake or dreaming) participants.	(Casali et al., 2013; Casarotto et al., 2016; Massimini, Boly, Casali, Rosanova, & Tononi, 2009; Massimini et al., 2005)	Humans
Sniff test	Designed primarily for detecting consciousness in clinical settings (in adult humans with disorders of consciousness)	A pleasant or unpleasant odor is presented to a patient with a disorder of consciousness. The test is passed if the patient performs a sniffing response that is similar to that of conscious subjects.	(Arzi et al., 2020)	Humans
Unlimited Associative Learning (UAL) Test (comprises trace conditioning, cross-modal learning, rapid reversal learning)	Designed primarily for detecting consciousness in non-human animals (although based on properties of human consciousness); at this point, only some aspects have been implemented (e.g., trace conditioning)	The test is passed if an animal shows the ability to learn associations between stimuli, or between stimuli and external events that are well separated in time, in an open and flexible manner.	(Birch, 2022; Birch, Ginsburg, & Jablonka, 2020; Browning & Birch, 2022; Ginsburg & Jablonka, 2019; Grover et al., 2022)	Animals
Metacognition	Designed primarily for detecting consciousness in	The animal performs a task that varies in level of difficulty. Wrong attempts are punished, while correct attempts are rewarded. The	(Perry & Barron, 2013)	Animals

	non-human animals.	animal can decide whether to continue attempting the tasks or opt-out. The test is passed if the animal shows that opting out coincides with task difficulty, suggesting it is capable of metacognition.		
Motivational trade-off	Designed primarily for detecting consciousness in non-human animals	The animal faces a choice between a positive reward (e.g., food) obtained via noxious stimulation (e.g., heated food dispenser) and neutral feedback without noxious stimulation. The test is passed if the animal learns to use contextual information (e.g., the association between the color of the dispenser and its food quantity) to make decisions on how to resolve the trade-off. For example, this is indexed by increased preference towards noxious (i.e., heated) food dispensers over time.	(Gibbons, Versace, Crump, Baran, & Chittka, 2022; Tye, 2017)	Animals
Double dissociation test	Designed primarily for detecting consciousness in non-human animals	The animal is presented with a task that elicits opposite performance signatures when the stimuli are consciously perceived compared to when they are unconscious (specifically, a cue that indexes that the reward will be given in a different location is presented). The test is passed if the animal displays opposite performance signatures in the two conditions. That is, if it is able to learn the contingency so its reactions are faster when a supraliminal cue appears, while no such advantage is found for a subliminal cue.	(Ben-Haim et al., 2021)	Animals

AI Consciousness test (ACT)	Designed exclusively for detecting consciousness in artificial systems (currently, this is only a hypothetical test)	An AI system is trained on a dataset that does not include any referral to the metaphysics of consciousness (and we also add other relevant learning constraints). The test is passed if the AI starts speculating about consciousness.	(Schneider, 2019)	Artificial systems
Chip test	Designed exclusively for detecting consciousness in artificial systems (a hypothetical test that cannot, at this point, be implemented)	A certain brain area (e.g., V4) is substituted with a functionally equivalent silicon chip. Then, the participant introspects to see whether the phenomenological property that area is known to process (e.g., colour), is still there or not. The test is passed if that particular phenomenal aspect is still present, suggesting that now the chip contributes to consciousness and that a system made entirely of silicon chips can be conscious.	(Schneider, 2019)	Artificial systems
Theory-based computational indicator properties	Designed exclusively for detecting consciousness in artificial systems	A list of "indicator properties" of consciousness is derived, given the principles of computationally functionalist theories of consciousness. The more indicators are fulfilled by the system, the more likely it is that the system is conscious	(Butlin et al., 2023)	Artificial systems

As is clearly seen, these C-tests target different properties, at different levels of analysis. For example, some probe purely behavioural and/or linguistic abilities, others target cognitive capacities (e.g., learning, maintenance of information, etc.) or behavioral responses (e.g., sniffing), while others focus on neural properties (e.g., the complexity of brain activity). The tests are also primarily designed to target different populations. Despite these differences, it might be possible to translate some C-tests from one domain of applicability to another, provided that the property of interest is measurable in the new domain (Dung, 2023). For example, the global detection test has been applied also to foetuses (Bayne, Frohlich, Cusack, Moser, & Naci, 2023; Moser et al., 2021), and the PCI might be applied to non-human primates and other mammals. Similarly, the sniff test seems to be applicable to various non-human animal species, but, at present, it cannot target artificial systems, since computers do not typically have noses.

Moreover, some of these tests are related to a different degree to specific theoretical commitments, while others are more theory-neutral. For example, the relation between the PCI and consciousness is underpinned by the idea that conscious states require both signal integration and differentiation, and is accordingly inspired by the integrated information theory (IIT; Albantakis et al., 2023), whereas the global detection test relies on the Global Neuronal Workspace Theory (GNWT; Dehaene & Naccache, 2001; Mashour, Roelfsema, Changeux, & Dehaene, 2020) to draw a connection between consciousness and global integration, arguably indexed by the P300 component. Notably though, the validity of these tests does not directly and exclusively stem from that of the theory that originally motivated it. For example, the PCI shows outstanding performance in predicting the recovery of consciousness in patients with disorders of consciousness (Rosanova et al., 2023), and therefore at this stage it derives its validity more from practical success than from theoretical motivations (such that even if one disagrees with the theoretical background, they could still endorse the test due to its high success rates).

Moreover, the test is also compatible with other theories claiming that consciousness requires widespread connectivity in space and time (Farisco & Changeux, 2023; Sarasso et al., 2021; Storm et al., 2024).

2. What are C-tests targeting?

As a first step towards better understanding the C-tests presented above, we further clarify the target of interest of any C-test, namely phenomenal consciousness as a global state of a system. To that end, we describe the dissociations between phenomenal and access consciousness, between global states and contents of consciousness, and between the presence and capacity for consciousness.

2.1. Phenomenal vs. Access consciousness

In a seminal paper, the philosopher Ned Block (1995) introduced a distinction between phenomenal and access consciousness (for criticisms, see Cohen & Dennett, 2011; Dennett, 1995; Naccache & Dehaene, 2007). The former refers to the subjective and qualitative aspect of experience: the warmth of holding a hot cup of tea, the blueness of the sea, or the sweetness of honey. Phenomenal consciousness is not *necessarily* defined in terms of function or of cognitive capacities, and some have gone as far as claiming that it is so detached from cognitive abilities that it might be epiphenomenal (Jackson, 1982). Access consciousness, on the other hand, pertains to information processing and is crucial for cognitive functions. It is that facet of our mental life that makes a certain piece of information available to various output systems, including our ability to report to ourselves that we experience that information.

As mentioned above, this dissociation is widely debated, and even those who accept it, including Block himself, agree that these two types of consciousness most often go hand in hand (Block, 1995). Thus, it is very difficult (and according to some, impossible; see Overgaard, 2018) to find cases of phenomenal consciousness without access consciousness and vice versa (for some attempts, see Amir, Assaf, Yovel, & Mudrik, 2023). If this is indeed the case, it seems plausible to test for the presence of phenomenal consciousness by testing for cognitive markers that are more directly related to access consciousness (e.g., perceptual grouping (Lamme, 2020), integration of information (Mudrik, Faivre, & Koch, 2014), or requiring active maintenance of information for a certain time duration (Baars & Franklin, 2003)). Accordingly, detecting those markers would be indicative of the existence of phenomenal consciousness, in the same way as spotting smoke could inform about the existence of a fire.

Though the logic behind this proposal seems sound, it is not that trivial to implement, for two reasons. First, currently there is still no consensus about which type of cognitive abilities require access consciousness. Second, despite the often co-occurrence of phenomenal and access consciousness, there is ample debate in the literature about whether phenomenal consciousness "overflows" access consciousness (Block, 2011; Knotts, Odegaard, Lau, & Rosenthal, 2019; Kouider, de Gardelle, Sackur, & Dupoux, 2010; Overgaard, 2018). If it does, many tests that target phenomenal consciousness through access consciousness might not be sensitive enough to detect consciousness in those cases in which it is present, but it is not reportable. Thus, until further conceptual and empirical work clarifies what markers correlate with access consciousness, and what are the relations between access consciousness and phenomenal consciousness, this rationale should be taken with a grain of salt.

2.2. C-tests and global states of consciousness

In the literature, examples of global states of consciousness include wakefulness, drowsiness, dreaming, and so on (Chalmers, 2000; McKilliam, 2020). These are usually interpreted as states a system is in, which determine how contents of consciousness can be organized and made available within the mental life of a system, in the same way as the various modes of a car (e.g., sport mode, city mode, etc.) dispose the car to react in different ways to the characteristics of the road and the inputs from the driver (Bayne & Hohwy, 2016).

Such global states of consciousness differ from local states, which are defined by specific contents of consciousness: the former are identified independently of what the system is conscious of, and therefore independently of the specific contents of consciousness. Importantly though, many C-tests rely on the ability to detect specific sensory states (e.g., an auditory stimulus) or sensations (e.g., pain) in order to infer that the tested system is conscious. The reasoning is that in order for the system to experience this sensory stimulus or that sensation, it must have a subjective experience to begin with. That is, to be in a local state of consciousness, the system must be in some global state of consciousness. In this case, testing for specific contents of consciousness can be considered a bridge towards identifying global states of consciousness, which is the final target of any C-test. Notably, C-tests are not intended to be so fine-grained to distinguish between different modes, or global states, of consciousness. That is, they are not aimed at specifying whether the tested system is, for example, awake or dreaming; rather, C-tests are supposed to inform us of whether the system is in a conscious state, whatever that state might be. With the car example, it is as if the test provided an answer to the question "is the car's engine on?", independently of whether it is on city or sport mode.

Because different global states of consciousness might be associated with different functions, C-tests that focus on functions associated with one state might be insensitive to detecting consciousness in other states. For example, if a test focuses on the ability to follow commands, a positive result in such a test justifies raising our credence that the tested system is conscious. But if the ability to follow instructions is instantiated only during one state of consciousness (e.g., wakefulness), and not another (e.g., dreaming), a system could fail the test, despite being conscious. Thus, in many cases, a negative result in a C-test should not necessarily lower the credence in the belief that the tested system is conscious (we further discuss this issue below).

Finally, our discussion so far seemed to imply a dichotomous approach: either a system is conscious (i.e., under a global state of consciousness) or not. However, some consider global states to come *in degrees,* such that they can be ordered based on the level or degree of consciousness (Bachmann & Hudetz, 2014; Koch, Massimini, Boly, & Tononi, 2016; Laureys, 2005; for furhter disucssion, see Overgaard & Overgaard, 2010). For example, it seems intuitive to think that a fully awake person is more conscious than a drowsy person, which is in turn more conscious than a patient in a minimal conscious state.

The notion of levels, or degrees, of consciousness raises many thorny questions: it is unclear how to conceptualize the degrees of consciousness, or which criteria to use to construct such a consciousness scale, or whether the same scale can be valid not only for intraspecies comparisons (e.g., saying that a drowsy adult human is less conscious than a fully awake adult human) but also for interspecies comparisons (e.g., saying that a magpie is less conscious than a monkey; see Bachmann, 2012; Bayne, Hohwy & Owen, 2016 and Lee, 2023 for discussions).

C-tests are not well-poised to answer these issues. Although they are not necessarily binary, as their results often raise/lower the credence we assign for the possibility that the system is consciousness, this depends on our uncertainty, not on the nature of the system's consciousness: if, following a C-test, we confer a 0.6 chance that a system is conscious, that result does not mean that the system is 0.6 conscious (whatever that might mean), but simply that our degree of belief in the system's consciousness is 0.6. The non-binary nature of C-test accordingly depends on our ignorance, not on the nature of the phenomenon we are testing for. This can pose a problem, because if consciousness is graded, C-tests might miss some conscious states because the tests might be designed to detect consciousness states that are above a certain threshold. This again points at an asymmetry with respect to C-tests and their result: in this case, a positive result is more meaningful than a negative one, given the potential concern of insensitivity to states that are below the threshold tested by the test[2].

---

[2] A further problem is that consciousness might not only be graded, but also indeterminate (Godfrey-Smith, 2024; Lee, 2023). If that is the case, then there might be no fact of the matter whether a system is conscious or not. C-tests would be entirely unapplicable to many entities simply because there is no right/wrong answer to the question "is this system conscious?". This means that even our best battery of C-tests could be useless when it comes to entities that are neither determinately conscious nor determinately non-conscious.

2.3. *Presence vs. capacity for consciousness*

C-tests are generally focused on detecting the presence of consciousness, rather than the capacity for consciousness (Bayne et al., 2024; see also Mudrik, Mylopoulos, Negro, & Schurger, 2023). That is, they ask if the system is conscious at the moment of the test, as opposed to asking if it has the necessary conditions for consciousness, or might have the potential to develop consciousness in the future. From a moral perspective, it is not clear whether it is the presence of consciousness or its capacity that matters for a system to have moral worth (Harman, 2003; Singer, 2011), but from the epistemological standpoint, this means that a negative result in a C-test leaves open the question of whether the system *could* be conscious, were certain conditions modified.

3. How to validate C-tests?

So far, we have seen examples of C-tests, and analysed their target of interest. We turn now to the epistemological issue of what grants justification to C-tests. Specifically, we ask how they can be validated, whether they need a theoretical underpinning, and whether they are inherently anthropocentric.

3.1. *Consciousness and folk attributions*

One option for validating C-tests is to rely on our intuitions and pre-existing beliefs about the distribution of consciousness. To illustrate, imagine that a test concludes that label-makers or spoons are conscious, while it deems cats and dogs not to be conscious. It seems reasonable to dismiss such a C-test. However, imagine you have applied several C-tests to various entities that we intuitively consider as conscious, and imagine that your indoor plant also passes these C-tests. What do you conclude? Will you put more confidence in the belief that the plant is indeed conscious, or would you just discard the validity of those C-tests?

The relationship between the validity of C-tests and pre-theoretic, folk-psychological, intuitions about the distribution of consciousness is a delicate one. As Schwitzgebel puts it, we seem to be "more confident that there is something it is like to be a dog than we could ever be that a clever philosophical argument to the contrary was in fact sound" (Schwitzgebel, 2020, p. 54). The idea is that intuitions and existing beliefs about which systems enjoy consciousness seem hard to shake, and that this could be instrumental for validating C-tests.

This attitude is not new: for example, Block's criticism of functionalism (Block, 1978) is based on the intuition that a system like the China-brain[3] cannot be conscious. Recently, Liz Irvine (2020) used this strategy to argue against the motivational trade-off C-test. She reviews evidence that even *C. elegans* passes the test, and claims that since *C. elegans* is too simple of a system to be considered as conscious, the test itself is invalid (see Andrews, 2024 for a discussion).

This approach seems compelling, but in a way, it pulls the rug from under the purpose of C-tests to begin with: these tests are supposed to provide some theoretical, knowledge-based guidance on which types of systems are conscious and which are not. But if validating C-tests requires considering pre-theoretical intuitions on which systems are conscious, we face circular reasoning. Basically, we must assume what we want to discover: a belief about the distribution of consciousness.

There are at least two strategies to resolve this problem. On the one hand, one might simply set intuitions aside on the basis that science has often proved our intuitions wrong: the Earth is not flat, and the Sun does not revolve around it, despite what people used to think. Thus, we should

[3] In this argument, Block presents a hypothetical system composed of all the people in China and their connections. Block's idea is that such a system could be made to replicate the functional profile of the brain down to the level of each single neuron (i.e., each neuron's functionality would be replicated by each single Chinese citizen, passing information to other citizens), and yet it seems silly to think that such a system would be conscious.

not rely on intuitions for validating C-tests; instead, the tests would have to draw their validations from other sources (e.g., a well-established theory of consciousness. More on this in 3.2).

On the other hand, it might not be necessary to entirely dispense with pre-theoretical intuitions. Rather, we can limit the application of intuitions on the distribution of consciousness as a source for validation to clear-cut cases (for example, we can safely attribute consciousness to healthy adult humans, and do not attribute it to shoes and socks). Such obvious cases, beyond any reasonable doubt, can indeed serve for calibrating C-tests. Yet, for systems on which there is no clear consensus, folk intuitions would be almost inert. This process might not be necessarily viciously circular, since the intuition-based attribution of consciousness can be used to single out some superficial markers of consciousness in consensus domains; then, those markers can be used to refine our search for consciousness in more contentious domains (see 3.3).

3.2. The theory-heavy approach

According to the theory-heavy approach, pre-theoretical intuitions and folk attributions are irrelevant for the validation processes. Instead, C-tests should be based on existing knowledge generated by the science of consciousness, or – more specifically – on theories of consciousness. If C-tests are derived from such theories, they can draw their validations from them. Accordingly, a C-test is validated inasmuch as the theory that motivates it is valid (Dennett, 1991; Lau, 2022; Tononi & Koch, 2015; see Birch, 2022 for further discussion)

Despite the clear advantages of relying on scientific consideration rather than untested intuitions, the theory-heavy approach has nonetheless attracted many criticisms, based on different grounds (Baetu, 2024; Bayne et al., 2024; Birch, 2022; Shevlin, 2021; but see de Weerd, 2024 for a defence).

Probably the most severe issue is that there is no agreed upon theory of consciousness, and agreement on this front is not in sight (Francken et al., 2022; Seth & Bayne, 2022; Yaron et al., 2022). Thus, C-tests that are validated with the theory-based approach will inevitably be highly controversial, which negatively impacts both on the epistemic and on the pragmatic rationale of C-tests (a possible solution is to construct a test that is based on several theories; see Butlin et al., 2023).

The second problem is that many (if not all) theories of consciousness are constructed upon *human* consciousness[4], and might accordingly not be tuned to detect other forms of consciousness (Block, 2002; Usher, Negro, Jacobson, & Tsuchiya, 2023). Birch (2022) and Shevlin (2021) make this point by focusing on GNWT, which holds consciousness to coincide with a global workspace sharing information between many output systems that can process it and manipulate it (Dehaene, Kerszberg, & Changeux, 1998). Now, the nature of such a global workspace (e.g., the number of output systems, their functional capacities, their informational bandwidth, etc.) seems to be based on the human global workspace, since most of the studies testing GNWT are done on human subjects (for an updated distribution, see https://contrastdb.tau.ac.il). But it seems plausible that different systems might have a different global workspace, with different components and possibly different functions. Thus, C-tests derived from GNWT might only test for human consciousness (Mudrik et al., 2023).

This criticism is relevant to other major theories of consciousness too, as the starting point of most, if not all, current theories is human consciousness. For example, the phenomenological foundation of IIT is based on phenomenology done from the human perspective. Therefore, the theoretical and explanatory apparatus deriving from it is necessarily based on the human case.

[4] An exception might be the Dendritic Integration Theory of consciousness (DIT), which is built on cellular mechanisms of human and rodent consciousness, under the assumption that rodents are conscious (Bachmann, Suzuki & Aru, 2020; Aru, Suzuki & Larkum, 2020).

And so, any C-test that relies on IIT's explanatory apparatus would be anthropocentric too, at least to some extent.

In sum, the strategy of validating C-tests via theories of consciousness faces at least two major problems. The first seems to be a contingent fact about the current state of consciousness science, while the second is inherently conceptual, as it requires justifying extrapolations of non-human consciousness from theories of human consciousness. Although we do not think this problem is impossible to solve, we flag it as a challenge for the theory-heavy approach, and we will return to this issue in section 4.

3.3. The iterative natural kind approach

The third strategy for validating C-tests does not require a theory of consciousness, or at least not a complete one. Instead, this strategy might take advantage of the assumption that consciousness is a natural kind, and that it is related to some observable phenomena.

Natural kinds are ways to "carve nature at its joints", namely ways to categorize reality based on real features, patterns, and/or properties. The periodic table is often cited as an example of how natural kinds (in this case, chemical kinds) work: if I know that gold is that thing with atomic number 79, and that it has certain observable properties (e.g., being yellowish, being malleable, etc.), I can reasonably infer that every instance of an element with atomic number 79 will also share those observable properties, while an element with atomic number 47 will have different properties[5]. From this, I can also infer that if an element has some properties like being yellowish and malleable, then it is probably gold (i.e., it has atomic number 79).

[5] It seems more reasonable to think that consciousness is more similar to a biological kind rather than a purely chemical one. In the life sciences, natural kinds are seen as "homeostatic property clusters", namely clusters of properties that tend to co-occur in virtue of homeostatic mechanisms that cause and sustain those properties (Craver, 2009). In this view, there is no need for a unique and essential property that keeps together all the other properties of the cluster (Bayne & Shea, 2020; Boyd, 2019).

In other words, natural kinds determine a cluster of properties that naturally "hang together", and can accordingly justify the attribution of these properties to any member of the kind.

Assuming that consciousness is a natural kind thus prescribes a specific methodology for consciousness research: first, find the properties that cluster together because of consciousness (in the same way as being malleable and yellowish cluster together in virtue of the element having atomic number 79), then, look for entities that exhibit those properties. Thus, the assumption that consciousness is a natural kind allows us to trace back those property-clusters to the presence of consciousness, even in the absence of an agreed-upon theory of consciousness.

This strategy helps validating C-tests because, as Bayne et al. put it:

> Although the validity of any putative C-test begins with pre-theoretical judgements, it is not simply derived from those judgments but can outstrip them in various ways. Pre-theoretical measures of consciousness […] play a crucial role in the initial stages of our inquiry, but we regard them as open to revision following discoveries about the underlying natural kind (Bayne et al., 2024, p. 8).

The natural kind approach thus grounds its validation procedure on iteration. Although pre-theoretical and superficial markers can help us calibrate C-tests in consensus cases (with the previous example, we can safely discard a C-test that returns as conscious label-makers), the list of properties that cluster together because of consciousness will be refined through a process of iteration that includes, progressively, different domains and different tests. So, we start with running several C-tests on a consensus population (e.g., responsive humans), for which we agree that it is conscious (or not, when being manipulated, for example by anaesthesia), based on folk attributions and pre-theoretical intuitions. We then compute the correlations between the C-tests, to identify a first set of properties that seem to be reliably sustained by consciousness. Then, we progress to a less consensual populations (e.g., patients with disorders of consciousness), and run the same tests again. Based on their agreement in this new population, we can revise the identified cluster of properties, also when applied to the previous, consensual population. This iterative process of running C-tests, looking for their relations, and updating the credences we assign to each test based on its correspondence with the other tests, continues until we are able to identify the relevant properties that cluster together. During this iterative process, C-tests are validated insofar as they test for the putative property-cluster, and the results of those C-tests prompt revision and refinement of the cluster.

Importantly, the natural kind approach can also be criticized for anthropocentrism, since the initial cluster of properties that are supposed to "hang together" with consciousness is also defined first for humans, being the consensus population. This suggests that any C-test will be inevitably anthropocentric, and this could considerably limit our capacity to detect forms of consciousness in systems that differ substantially from us (Bayne & Shea, 2020, p. 80).

The problem can be further described in the following terms: both neuroscientists and philosophers working on consciousness largely agree that when we want to detect consciousness in other humans, behavioural and functional procedures are often not enough. The neural level might be more sensitive, as it can be found even in the absence of outputting an overt behaviour. For this reason, many C-tests rely on neural markers of consciousness, or of functions for which consciousness is held to be necessary (Bayne et al., 2024). However, C-tests based on neural observations can be very specific and effective in the human population, but might not be easily applicable to other populations. This is because such populations are likely to greatly differ in terms of neuroanatomy and neurophysiology even if they exhibit functional and behavioural properties similar to conscious humans. For example, any relationship between a certain neural property and consciousness that was found in humans might not be applicable to non-humans, since it is possible (quite probable, for animals very far from us on the evolutionary scale, as well as for artificial systems) that that neural property is either not instantiated, or very differently instantiated, in the non-human population. This is true for cognitive functions too: it is not clear that a function that can only be performed consciously in humans must be performed consciously in non-humans, since different populations might have different functional profiles (Dung, 2022; Michel, 2019). The question is then how to justify the applicability of C-tests when the iterative cycle prescribed by the natural kind approach reaches distant populations.

4. Justifying inferences in C-tests

A certain degree of anthropocentrism thus seems to be inevitable. If so, how can we pair our preferred methodology for validating C-tests with a valid *extrapolation* process (Baetu, 2024; Sober, 2000)? This is a well-known issue in the social and life sciences, in which a phenomenon studied in a certain population must be projected to a different one (Guala, 2010; Steel, 2007): how can we know that differences between domains are not differences that make the difference, with respect to the phenomenon of interest?

Traditional approaches for tackling this problem include analogical reasoning and Inference to the Best Explanation (IBE), which have also been used as justificatory bases to address the problem of the other minds (Avramides, 2000; Hyslop & Jackson, 1972; Koch, 2019; Melnyk, 1994; Pargetter, 1984). In a nutshell, analogical reasoning justifies the inference that the phenomenon of interest is present in the target domain by exploiting the knowledge that i) some phenomenon-relevant similarities have been detected in two different domains; and ii) the known dissimilarities are not relevant for the tested phenomenon. Instead, IBE exploits the knowledge that the phenomenon best explains certain properties in a known domain in order to infer the presence of that phenomenon in the target domain, when similar properties are observed.

We maintain that extrapolating consciousness to non-consensus cases requires not only a sound method for validating C-tests, but also a clear inferential framework for justifying extrapolations. That is, both the theory-heavy approach and the natural kind approach should be paired with analogical reasoning, IBE, or a combination of both.

Indeed, the strongest form of logical reasoning for extrapolations about consciousness in non-consensus cases might be provided by analogical abduction, which combines aspects of analogical reasoning and IBE (Negro & Mudrik, forthcoming; Schurz, 2008). The idea is that consciousness could be projected from one domain to another only when there is a *structure-preserving mapping* between domains, provided that the mapping targets properties that are known to be explanatorily related to consciousness (either because they *explain* or because they are *explained by* consciousness). Whether these properties are taken to be a cluster of consciousness "symptoms" or properties explanatorily related to consciousness because of theoretical tenets, it depends on whether one endorses the natural kind approach or a theory-heavy approach. The main point is that courtesy of analogical abduction, consciousness in non-consensus cases might be discovered following scientifically sound and conventional practices (e.g., sound waves were discovered by analogical abduction thanks to their similarity with water waves – see Thagard, 1988 and Schurz, 2008).

Although there are important open questions related to how to better understand the functioning of the inferential machinery that can be used to justify extrapolations of consciousness (e.g., how to define the necessary threshold of enough similarity between domains in terms of structure-preserving mappings), pairing such a machinery with an approach for validating C-tests could contribute to the process of discovering consciousness in non-consensus cases.

5. Conclusion: A multidimensional approach to C-tests

In this chapter, we have surveyed some of the main C-tests aimed at testing for consciousness in different populations, and analysed some of the conceptual challenges they face. We showed that different C-tests target different aspects of consciousness and are designed for different target populations, yet are all anthropocentric, to some degree.

Given this context, we expect that well-informed assessments on the probability of consciousness in non-consensus cases could be possible only by adopting a multidimensional model (for similar suggestions, see Bayne et al., 2004; Birch, Schnell & Clayton, 2020; Dung & Newen, 2023). Although there is no space here to fully build the specific features of such a model, we sketch the basic rationale of the model, and leave its development as a direction for future research. This model should include a multitude of C-tests directed at the same target entity, and a weighted system for representing the different degrees of informativeness that different types of evidence carry for that entity (e.g., for a DOC patient, the command following test might be more informative than the sniff test). Moreover, it should include a function that takes the results of various C-tests and their weights as input, and delivers the degree of probability that the target entity is conscious as output.

We expect that such a multidimensional model will have to provide answers to the following questions: what are the test dimensions, or forms, of consciousness, and how are they related to the desiderata - detecting phenomenal consciousness? What are the various global states of consciousness, and how are they typically manifested? What is the best procedure for validating C-tests and for justifying extrapolations? How does the importance of different types of evidence vary with respect to different populations?

Despite these (and possibly many other) open questions, the development of a multidimensional model that combines different C-tests and assigns them with different credences is a research project worth pursuing. The main advantage is the ability to integrate over different tests, even when they yield contradictory results. In such a case, all results will be included, with different weights, resulting in a probabilistic measure of how strong our attributions of consciousness to non-consensus cases are.

This might be an epistemically imperfect strategy for detecting consciousness in non-consensus cases, but its uncertainty-reducing nature, and its empirically-backed methodology would still be useful and productive. This is where the pragmatic aspect of C-tests comes into play. Given that a thorough analysis of this aspect goes beyond the scope of this chapter, we leave the relationship between the epistemic and pragmatic aspect of C-tests for further research and discussion. However, we observe that if consciousness matters for ethical and practical decision-making in various contexts (from AI regulation to clinical settings), then the question of how it is distributed in the world is of utmost urgency. In this context, having an epistemically imperfect methodology is still better than deliberating in the dark.

References

Albantakis, L., Barbosa, L., Findlay, G., Grasso, M., Haun, A. M., Marshall, W., . . . Tononi, G. (2023). Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms. *PLOS Computational Biology, 19*(10), e1011465. doi:10.1371/journal.pcbi.1011465

Amir, Y. Z., Assaf, Y., Yovel, Y., & Mudrik, L. (2023). Experiencing without knowing? Empirical evidence for phenomenal consciousness without access. *Cognition, 238*, 105529. doi:https://doi.org/10.1016/j.cognition.2023.105529

Andrews, K. (2024). "All animals are conscious": Shifting the null hypothesis in consciousness science. *Mind & Language*. doi:https://doi.org/10.1111/mila.12498

Aru, J., Suzuki, M., & Larkum, M. E. (2020). Cellular Mechanisms of Conscious Processing. *Trends Cogn Sci, 24*(10), 814-825. doi:10.1016/j.tics.2020.07.006

Arzi, A., Rozenkrantz, L., Gorodisky, L., Rozenkrantz, D., Holtzman, Y., Ravia, A., . . . Sobel, N. (2020). Olfactory sniffing signals consciousness in unresponsive patients with brain injuries. *Nature, 581*(7809), 428-433. doi:10.1038/s41586-020-2245-5

Avramides, A. (2000). *Other Minds*. London: Routledge.

Baars, B. J., & Franklin, S. (2003). How conscious experience and working memory interact. *Trends in Cognitive Sciences, 7*(4), 166-172. doi:https://doi.org/10.1016/S1364-6613(03)00056-1

Bachmann, T. (2012). How to Begin to Overcome the Ambiguity Present in Differentiation between Contents and Levels of Consciousness? *Frontiers in Psychology, 3*. doi:10.3389/fpsyg.2012.00082

Bachmann, T., & Hudetz, A. G. (2014). It is time to combine the two main traditions in the research on the neural correlates of consciousness: C = L × D. *Front Psychol, 5*, 940. doi:10.3389/fpsyg.2014.00940

Bachmann, T., Suzuki, M., & Aru, J. (2020). Dendritic integration theory: A thalamo-cortical theory of state and content of consciousness. *Philosophy and the Mind Sciences, 1*(II). doi:10.33735/phimisci.2020.II.52

Baetu, T. M. (2024). Extrapolating animal consciousness. *Studies in History and Philosophy of Science, 104*, 150-159. doi:https://doi.org/10.1016/j.shpsa.2024.03.001

Bayne, T. (2007). Conscious states and conscious creatures: explanation in the scientific study of consciousness. *Philosophical Perspectives, 21*(1), 1-22. doi:https://doi.org/10.1111/j.1520-8583.2007.00118.x

Bayne, T., Frohlich, J., Cusack, R., Moser, J., & Naci, L. (2023). Consciousness in the cradle: on the emergence of infant experience. *Trends in Cognitive Sciences, 27*(12), 1135-1149. doi:10.1016/j.tics.2023.08.018

Bayne, T., & Hohwy, J. (2016). Modes of consciousness. In *Finding consciousness: The neuroscience, ethics, and law of severe brain damage.* (pp. 57-80). New York, NY, US: Oxford University Press.

Bayne, T., Hohwy, J., & Owen, A. M. (2016). Are There Levels of Consciousness? *Trends in Cognitive Sciences, 20*(6), 405-413. doi:10.1016/j.tics.2016.03.009

Bayne, T., Seth, A. K., Massimini, M., Shepherd, J., Cleeremans, A., Fleming, S. M., . . . Mudrik, L. (2024). Tests for consciousness in humans and beyond. *Trends in Cognitive Sciences*. doi:10.1016/j.tics.2024.01.010

Bayne, T., & Shea, N. (2020). Consciousness, Concepts and Natural Kinds. *Philosophical Topics, 48*(1), 65-83. doi:https://www.jstor.org/stable/48628586.

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proc Natl Acad Sci U S A, 106*(5), 1672-1677. doi:https://doi.org/10.1073/pnas.0809667106

Ben-Haim, M. S., Dal Monte, O., Fagan, N. A., Dunham, Y., Hassin, R. R., Chang, S. W. C., & Santos, L. R. (2021). Disentangling perceptual awareness from nonconscious processing in rhesus monkeys (Macaca mulatta). *Proc Natl Acad Sci U S A, 118*(15). doi:10.1073/pnas.2017543118

Birch, J. (2022). The search for invertebrate consciousness. *Noûs, 56*(1), 133-153. doi:https://doi.org/10.1111/nous.12351

Birch, J., Burn, C., Schnell, A., Browning, H., & Crump, A. (2021). *Review of the Evidence of Sentience in Cephalopod Molluscs and Decapod Crustaceans*. Retrieved from https://www.lse.ac.uk/business/consulting/reports/review-of-the-evidence-of-sentiences-in-cephalopod-molluscs-and-decapod-crustaceans

Birch, J., Ginsburg, S., & Jablonka, E. (2020). Unlimited Associative Learning and the origins of consciousness: a primer and some predictions. *Biology & Philosophy, 35*(6), 56. doi:10.1007/s10539-020-09772-0

Birch, J., Schnell, A. K., & Clayton, N. S. (2020). Dimensions of Animal Consciousness. *Trends in Cognitive Sciences, 24*(10), 789-801. doi:10.1016/j.tics.2020.07.007

Block, N. (1978). Troubles with functionalism. *Minnesota Studies in the Philosophy of Science, 9*, 261-325.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18(2)*, 227-247.

Block, N. (2002). The Harder Problem of Consciousness. *The Journal of Philosophy, 99*(8), 391-425. doi:10.2307/3655621

Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences, 15*(12), 567-575. doi:10.1016/j.tics.2011.11.001

Block, N. (2019). What Is Wrong with the No-Report Paradigm and How to Fix It. *Trends Cogn Sci, 23*(12), 1003-1013. doi:10.1016/j.tics.2019.10.001

Bodien, Y. G., Allanson, J., Cardone, P., Bonhomme, A., Carmona, J., Chatelle, C., . . . Schiff, N. D. (2024). Cognitive Motor Dissociation in Disorders of Consciousness. *New England Journal of Medicine, 391*(7), 598-608. doi:doi:10.1056/NEJMoa2400645

Boyd, R. (2019). Rethinking natural kinds, reference and truth: towards more correspondence with reality, not less. *Synthese, 198*(Suppl 12), 2863-2903. doi:https://doi.org/10.1007/s11229-019-02138-4

Browning, H., & Birch, J. (2022). Animal sentience. *Philosophy Compass, 17*(5), e12822. doi:https://doi.org/10.1111/phc3.12822

Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., . . . VanRullen, R. (2023). Consciousness in Artificial Intelligence: Insights from the Science of Consciousness. *arXiv*. doi:https://doi.org/10.48550/arXiv.2308.08708

Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., . . . Massimini, M. (2013). A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior. *Science Translational Medicine, 5*(198), 198ra105-198ra105. doi:10.1126/scitranslmed.3006294

Casarotto, S., Comanducci, A., Rosanova, M., Sarasso, S., Fecchio, M., Napolitani, M., . . . Massimini, M. (2016). Stratification of unresponsive patients by an independently validated index of brain complexity. *Ann Neurol, 80*(5), 718-729. doi:10.1002/ana.24779

Chalmers, D. J. (2000). What is a neural correlate of consciousness? In T. Metzinger (Ed.), *Neural correlates of consciousness: Empirical and conceptual questions.* (pp. 17-39). Cambridge, MA, US: The MIT Press.

Cohen, M. A., & Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences, 15*(8), 358-364. doi:https://doi.org/10.1016/j.tics.2011.06.008

Craver, C. F. (2009). Mechanisms and natural kinds. *Philosophical Psychology, 22*(5), 575-594. doi:10.1080/09515080903238930

Crump, A., Browning, H., Schnell, A., Burn, C., & Birch, J. (2022). Sentience in decapod crustaceans: A general framework and review of the evidence. *Animal Sentience, 32*(1).

Dehaene, S., Kerszberg, M., & Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proc Natl Acad Sci U S A, 95*(24), 14529-14534. doi:10.1073/pnas.95.24.14529

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition, 79*(1), 1-37. doi:https://doi.org/10.1016/S0010-0277(00)00123-2

Dennett, D. C. (1995). The path not taken. *Behavioral and Brain Sciences, 18*(2), 252-253.

Dung, L. (2022). Assessing tests of animal consciousness. *Consciousness and Cognition, 105*, 103410. doi:https://doi.org/10.1016/j.concog.2022.103410

Dung, L. (2023). Tests of Animal Consciousness are Tests of Machine Consciousness. *Erkenntnis*. doi:10.1007/s10670-023-00753-9

Dung, L., & Newen, A. (2023). Profiles of animal consciousness: A species-sensitive, two-tier account to quality and distribution. *Cognition, 235*, 105409. doi:https://doi.org/10.1016/j.cognition.2023.105409

Farisco, M., & Changeux, J.-P. (2023). About the compatibility between the perturbational complexity index and the global neuronal workspace theory of consciousness. *Neuroscience of Consciousness, 2023*(1). doi:10.1093/nc/niad016

Fink, S. B. (2024). How-tests for consciousness and direct neurophenomenal structuralism. *Frontiers in Psychology, 15*. doi:10.3389/fpsyg.2024.1352272

Francken, J. C., Beerendonk, L., Molenaar, D., Fahrenfort, J. J., Kiverstein, J. D., Seth, A. K., & van Gaal, S. (2022). An academic survey on theoretical foundations, common assumptions and the current state of consciousness science. *Neuroscience of Consciousness, 2022*(1). doi:10.1093/nc/niac011

Gibbons, M., Versace, E., Crump, A., Baran, B., & Chittka, L. (2022). Motivational trade-offs and modulation of nociception in bumblebees. *Proceedings of the National Academy of Sciences, 119*(31), e2205821119. doi:doi:10.1073/pnas.2205821119

Ginsburg, S., & Jablonka, E. (2019). *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*: MIT Press.

Godfrey-Smith, P. (2024). Inferring Consciousness in Phylogenetically Distant Organisms. *J Cogn Neurosci, 36*(8), 1660-1666. doi:10.1162/jocn_a_02158

Grover, D., Chen, J. Y., Xie, J., Li, J., Changeux, J. P., & Greenspan, R. J. (2022). Differential mechanisms underlie trace and delay conditioning in Drosophila. *Nature, 603*(7900), 302-308. doi:10.1038/s41586-022-04433-6

Guala, F. (2010). Extrapolation, Analogy, and Comparative Process Tracing. *Philosophy of Science, 77*(5), 1070-1082. doi:10.1086/656541

Harman, E. (2003). The potentiality problem. *Philosophical Studies, 114*(1-2), 173 - 198.

Hyslop, A., & Jackson, F. C. (1972). The Analogical Inference to Other Minds. *American Philosophical Quarterly, 9*(2), 168-176. Retrieved from http://www.jstor.org/stable/20009435

Irvine, E. (2020). Developing Valid Behavioral Indicators of Animal Pain. *Philosophical Topics, 48*(1), 129-154. Retrieved from https://www.jstor.org/stable/48628589

Jackson, F. (1982). Epiphenomenal Qualia. *The Philosophical Quarterly (1950-), 32*(127), 127-136. doi:10.2307/2960077

Knotts, J. D., Odegaard, B., Lau, H., & Rosenthal, D. (2019). Subjective inflation: phenomenology's get-rich-quick scheme. *Curr Opin Psychol, 29*, 49-55. doi:10.1016/j.copsyc.2018.11.006

Koch, C. (2019). *The Feeling of Life Itself: Why Consciousness Is Widespread but Can't Be Computed*. Cambridge, MA: MIT Press.

Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat Rev Neurosci, 17*(5), 307-321. doi:10.1038/nrn.2016.22

Kouider, S., de Gardelle, V., Sackur, J., & Dupoux, E. (2010). How rich is consciousness? The partial awareness hypothesis. *Trends in Cognitive Sciences, 14*(7), 301-307. doi:https://doi.org/10.1016/j.tics.2010.04.006

Lamme, V. A. F. (2020). Visual Functions Generating Conscious Seeing. *Front Psychol, 11*, 83. doi:10.3389/fpsyg.2020.00083

Laureys, S. (2005). The neural correlate of (un)awareness: lessons from the vegetative state. *Trends in Cognitive Sciences, 9*(12), 556-559. doi:10.1016/j.tics.2005.10.010

Lee, A. Y. (2022). Speciesism and Sentientism. *Journal of Consciousness Studies, 29*(3-4), 205-228. doi:10.53765/20512201.29.3.205

Lee, A. Y. (2023). Degrees of consciousness. *Noûs, 57*(3), 553-575. doi:https://doi.org/10.1111/nous.12421

Mashour, G. A., Roelfsema, P., Changeux, J.-P., & Dehaene, S. (2020). Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron, 105*(5), 776-798. doi:10.1016/j.neuron.2020.01.026

Massimini, M., Boly, M., Casali, A., Rosanova, M., & Tononi, G. (2009). A perturbational approach for evaluating the brain's capacity for consciousness. In S. Laureys, N. D. Schiff, & A. M. Owen (Eds.), *Progress in Brain Research* (Vol. 177, pp. 201-214): Elsevier.

Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., & Tononi, G. (2005). Breakdown of cortical effective connectivity during sleep. *Science, 309*(5744), 2228-2232. doi:10.1126/science.1117256

McKilliam, A. K. (2020). What is a global state of consciousness? *Philosophy and the Mind Sciences, 1*(II). doi:10.33735/phimisci.2020.II.58

Melnyk, A. (1994). Inference to the best explanation and other minds. *Australasian Journal of Philosophy, 72*(4), 482-491. doi:10.1080/00048409412346281

Metzinger, T. (2021). Artificial Suffering: An Argument for a Global Moratorium on Synthetic Phenomenology. *Journal of Artificial Intelligence and Consciousness, 08*(01), 43-66. doi:10.1142/s270507852150003x

Michel, M. (2019). Fish and microchips: on fish pain and multiple realization. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 176*(9), 2411-2428. Retrieved from http://www.jstor.org/stable/45211659

Moser, J., Schleger, F., Weiss, M., Sippel, K., Semeia, L., & Preissl, H. (2021). Magnetoencephalographic signatures of conscious processing before birth. *Dev Cogn Neurosci, 49*, 100964. doi:10.1016/j.dcn.2021.100964

Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends Cogn Sci, 18*(9), 488-496. doi:10.1016/j.tics.2014.04.009

Mudrik, L., Mylopoulos, M., Negro, N., & Schurger, A. (2023). Theories of consciousness and a life worth living. *Current Opinion in Behavioral Sciences, 53*, 101299. doi:https://doi.org/10.1016/j.cobeha.2023.101299

Naccache, L., & Dehaene, S. (2007). Reportability and illusions of phenomenality in the light of the global neuronal workspace model. *Behavioral and Brain Sciences, 30*(5-6), 518-520. doi:10.1017/S0140525X07002993

Naci, L., Cusack, R., Anello, M., & Owen, A. M. (2014). A common neural code for similar conscious experiences in different individuals. *Proceedings of the National Academy of Sciences, 111*(39), 14277-14282. doi:doi:10.1073/pnas.1407007111

Naci, L., Sinai, L., & Owen, A. M. (2017). Detecting and interpreting conscious experiences in behaviorally non-responsive patients. *Neuroimage, 145*(Pt B), 304-313. doi:10.1016/j.neuroimage.2015.11.059

Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review, 83*(4), 435-450. doi:10.2307/2183914

Negro, N., & Mudrik, L. (forthcoming). Extrapolating other consciousnesses: the prospects and limits of analogical abduction. *Philosophy of Science*.

Overgaard, M. (2018). Phenomenal consciousness and cognitive access. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1755), 20170353. doi:doi:10.1098/rstb.2017.0353

Overgaard, M., & Overgaard, R. (2010). Neural correlates of contents and levels of consciousness. *Front Psychol, 1*, 164. doi:10.3389/fpsyg.2010.00164

Owen, A. M. (2018). *Into the Grey Zone: Exploring the Border Between Life and Death*: Faber & Faber, Limited.

Owen, A. M., Coleman, M. R., Boly, M., Davis, M. H., Laureys, S., & Pickard, J. D. (2006). Detecting Awareness in the Vegetative State. *Science, 313*(5792), 1402-1402. doi:10.1126/science.1130197

Pargetter, R. (1984). The scientific inference to other minds. *Australasian Journal of Philosophy, 62*(2), 158-163. doi:10.1080/00048408412341341

Perry, C. J., & Barron, A. B. (2013). Honey bees selectively avoid difficult choices. *Proceedings of the National Academy of Sciences, 110*(47), 19155-19159. doi:doi:10.1073/pnas.1314571110

Rosanova, M., Casarotto, S., Derchi, C., Hassan, G., Russo, S., Sarasso, S., . . . Comanducci, A. (2023). The perturbational complexity index detects capacity for consciousness earlier than the recovery of behavioral responsiveness in subacute brain-injured patients. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 16*(1), 371. doi:10.1016/j.brs.2023.01.731

Sarasso, S., Casali, A. G., Casarotto, S., Rosanova, M., Sinigaglia, C., & Massimini, M. (2021). Consciousness and complexity: a consilience of evidence. *Neuroscience of Consciousness*. doi:10.1093/nc/niab023

Schneider, S. (2019). *Artificial You: AI and the Future of Your Mind*: Princeton University Press.

Schurz, G. (2008). Patterns of abduction. *Synthese, 164*(2), 201-234. doi:10.1007/s11229-007-9223-4

Schwitzgebel, E. (2020). Is There Something it's Like to be a Garden Snail. *Philosophical Topics, 48*(1), 39-63.

Seth, A. K., & Bayne, T. (2022). Theories of consciousness. *Nature Reviews Neuroscience*. doi:10.1038/s41583-022-00587-4

Shepherd, J. (2018). *Consciousness and Moral Status*: Taylor & Francis.

Singer, P. (2011). *Practical Ethics* (3 ed.). Cambridge: Cambridge University Press.

Sober, E. (2000). Evolution and the Problem of Other Minds. *Journal of Philosophy, 97*(7), 365. doi:10.2307/2678410

Steel, D. (2007). *Across the Boundaries: Extrapolation in Biology and Social Science*: Oxford University Press.

Storm, J. F., Klink, P. C., Aru, J., Senn, W., Goebel, R., Pigorini, A., . . . Pennartz, C. M. A. (2024). An integrative, multiscale view on neural theories of consciousness. *Neuron, 112*(10), 1531-1552. doi:https://doi.org/10.1016/j.neuron.2024.02.004

Tannenbaum, J., & Jaworska, A. (2018). The Grounds of Moral Status. *Stanford Encyclopedia of Philosophy*, 0-0.

Teasdale, G., & Jennett, B. (1974). Assessment of coma and impaired consciousness. A practical scale. *Lancet, 2*(7872), 81-84. doi:10.1016/s0140-6736(74)91639-0

Teasdale, G., Maas, A., Lecky, F., Manley, G., Stocchetti, N., & Murray, G. (2014). The Glasgow Coma Scale at 40 years: standing the test of time. *Lancet Neurol, 13*(8), 844-854. doi:10.1016/s1474-4422(14)70120-6

Thagard, P. (1988). *Computational philosophy of science*. Cambridge, MA, US: The MIT Press.

Tye, M. (2017). *Tense bees and shell-shocked crabs: Are animals conscious?* New York, NY, US: Oxford University Press.

Usher, M., Negro, N., Jacobson, H., & Tsuchiya, N. (2023). When philosophical nuance matters: safeguarding consciousness research from restrictive assumptions. *Frontiers in Psychology, 14*. doi:10.3389/fpsyg.2023.1306023

Yaron, I., Melloni, L., Pitts, M., & Mudrik, L. (2022). The ConTraSt database for analysing and comparing empirical studies of consciousness theories. *Nat Hum Behav*. doi:10.1038/s41562-021-01284-5