Models and the test of consciousness

For decades, new theories have been emerging that want to explain what consciousness is. They speak of information integration (Tononi, 2004, 2015), of global availability (Baars, 1988; Dehaene, 2014), recursive predictions (Friston, 2010) or phenomenological structures (Varela, Thompson & Rosch, 1991). Their common conviction is that consciousness can be explained by certain dynamics and forms of organization. But the crucial question remains: How could we tell that such a model not only sounds conclusive, but actually produces consciousness?

In the following, it will be shown that all current models of consciousness are subject to a fundamental test. Four aspects must be considered: (1) its generative power, which has not yet been proven, (2) its metaphysical positions, (3) its circular validation between theory and empiricism, and (4) its predominantly interpretive, non-explanatory function.

1. Empirical Testing and Generative Power

The real touchstone of a model is not its internal consistency, but its generative power. A model endures if it can be shown that it creates a system that actually possesses consciousness. This is not a trivial point. This is because many models are content to provide correlations: they show that certain structures in the brain regularly coincide with consciousness (Mashour, Roelfsema, Changeux & Dehaene, 2020). From this it is then concluded that these structures explain consciousness. But until it is proven that the model itself is capable of producing consciousness, it remains an unproven hypothesis.

Empirical confirmation can only consist in the fact that the respective theory can generate consciousness in a reconstructed or artificially implemented form -- i.e. prove its own generative power. This would only be the case if a system constructed according to the principles of theory produced consciousness. As long as no such system exists that actually produces consciousness on the basis of a theory, any theory of consciousness remains in the realm of metaphysics.

But what could such proof look like? Most likely, only by reconstructing the model in a reconstructed or artificially implemented form. The replica is not an end in itself, but it provides the only way to show that the model is not just a heuristic framework, but a generative structure. If it is possible to construct a system that meets the same conditions and that behaves like a conscious system in tests such as complexity measures, multimodal integration, memory performance or stable self-reference (Casali et al., 2013), then one can assume with good reason that the model actually generates consciousness.

1.1 The Global Workspace Theory as an Example of Correlative Research

Bernard Baars' Global Workspace Theory (GWT) illustrates the problem of correlative findings in an exemplary way. The theory postulates that consciousness arises when information is made available in a global neural workspace and is thus accessible to different cognitive modules. Indeed, empirical studies show that conscious processing is associated with wide-ranging cortical activation, while unconscious processing remains locally limited (Dehaene & Changeux, 2011).

But what exactly is explained here? The theory describes a functional architecture: information is passed on from local processors to a central distributor. This explains the availability of information, but not the phenomenal experience itself. Why should global availability go hand in hand with subjective experience at all? The theory presupposes this connection instead of explaining it. It shows the conditions under which consciousness occurs, not why it occurs.

In addition, it remains unclear what an artificial implementation should look like. If a computer system were constructed according to the principles of GWT -- with a central workspace and competing modules -- would we have created a conscious system? The theory itself does not provide any criteria to answer this question. It operates on the level of functional descriptions without making the leap to phenomenal quality.

1.2 Higher-Order Theories and the Problem of Infinite Regression

Higher-Order Theories (HOT) claim that consciousness arises from higher-order representations: a mental state becomes conscious when it is represented by another mental state (Rosenthal, 2005; Lau & Rosenthal, 2011). This self-reference is supposed to be the characteristic feature of conscious experience.

However, the theory runs into two fundamental difficulties. First, if consciousness arises from a higher-order representation, what makes that higher-order representation conscious? Either it is itself conscious -- in which case we need another representation of even higher order, which leads to an infinite regression. Or it is unconscious -- in which case it remains a mystery how an unconscious process is supposed to produce consciousness.

Second, empirical research shows that higher-order representations correlate with metacognitive judgments (metacognition: the knowledge of one's own cognitive processes), but not necessarily with phenomenal consciousness. There are cases in which people have conscious experiences without reflecting on them and, conversely, make metacognitive judgments without a corresponding phenomenal experience (Berger, 2014). The HOT systematically confuses self-awareness with consciousness in general.

1.3 Recursive Processing and the Promise of Prediction

Theories of recursive processing, especially the Predictive Processing Framework, have gained great influence in recent years (Clark, 2013; Hohwy, 2013). They argue that the brain constantly generates predictions about sensory inputs and compares them with the actual inputs. Consciousness arises from the hierarchical integration of these prediction errors.

These theories have the advantage of being able to describe a wide range of cognitive phenomena in a unified framework. They explain perception, learning, attention and motor control as variants of the same basic principle. But it is precisely this explanatory breadth that proves to be problematic: If everything the brain does can be described as minimizing prediction error, then the concept loses its specific explanatory power for consciousness. The theory describes cognitive processing, not phenomenal experience.

In addition, the central question remains unanswered: Why should recursive processing go hand in hand with subjective experience at all? A thermostat also minimizes prediction errors (deviation between the target and actual temperature) without us attributing consciousness to it. The mere complexity or hierarchical depth of the system cannot be the decisive factor, because then it would have to be clarified at what threshold consciousness arises -- a question that theory cannot answer.

2. Metaphysical Positing and Categorical Errors

All theories of consciousness are based on metaphysical positing -- they replace explanation with definition. Some models remain close to empiricism and try to develop a theoretical reconstruction from neurophysiological or information-theoretical findings. Others, on the other hand, understand consciousness as a basic principle of reality, as an emergent property of information or as an ontological category of being. But in their epistemological position, both directions are of equal

importance. They differ only in the nature of their speculation, not in their scope. As long as neither of them provides practical proof that consciousness can actually produce it, they remain variations of the same philosophical project.

2.1 Integrated Information Theory: From Heuristics to Metaphysics

A classic example is Integrated Information Theory (IIT). It claims that consciousness arises from the integration of information and quantifies this with a measure (Φ). What begins as an elegant mathematical model, however, becomes a naïve generalization: any structure that integrates information is declared conscious. In this way, a heuristic hypothesis is transformed into a metaphysical ontology that inflationarily extends consciousness to all systems. The real difficulty -- why integration should produce experience at all -- remains untouched.

IIT goes even further: it claims that Φ is not just a correlate, but identical with consciousness. A system with high Φ is by definition conscious. However, this identification is a metaphysical position, not an empirical discovery. It does not solve the problem of explanation, but defines it away. If we ask why a system with high Φ should be conscious, the theory answers: because consciousness is Φ. This is a tautology, not an explanation.

This becomes particularly problematic with the counterintuitive consequences of the theory. According to IIT, an isolated grid of logical gates that has a high level of information-theoretic integration should have more consciousness than a human brain -- even if this grid never interacts with the outside world and does not perform any function (Aaronson, 2014). The theory must either accept this absurd consequence or admit that Φ alone is not sufficient to determine consciousness.

2.2 The Free Energy Principle and its Categorical Confusions

This is even clearer in the case of the free energy principle (Friston, 2010). Two categorical errors are made here: First, the epistemological principle of uncertainty minimization is derived from the thermodynamic principle of energy balance -- two completely different quantities. Second, according to the Bayesian theorem, the ability to calculate probabilities is shifted from the modeling level into the ontology of the organism, as if the brain were actually calculating probabilities. Both assumptions are metaphysical positions, not empirical findings.

The principle claims that all living systems minimize their free energy, with free energy being interpreted as a measure of surprise or uncertainty. This sounds plausible at first: organisms must keep their internal states stable and avoid surprises in order to survive. But the mathematical formalization of this idea leads to considerable problems.

The term "free energy" comes from thermodynamics and refers to the energy that a system can use for work. In the free energy principle, the same term is used for an information-theoretic quantity -- the difference between a model of the world and the actual sensory data. These two concepts have nothing to do with each other except the metaphorical use of the same term. It is a terminological obfuscation that suggests that there is a deep connection between physical energy and epistemic uncertainty.

In addition, there is the assumption that the brain performs Bayesian inference. This is a model-level description: we can model the behavior of the brain as if it were calculating probabilities. However, this does not mean that the brain actually performs probability calculations. A stone that describes a parabola also does not perform differential equations. The confusion of the description and implementation level is a classic categorical error.

2.3 Quantum theories and the lure of the mysterious

A special category is formed by theories that link consciousness to quantum processes (Penrose & Hameroff, 1995). They argue that classical neural processes are not sufficient to explain consciousness, and instead postulate quantum coherence in cellular structures such as microtubules. These theories have a strong intuitive appeal: two of the greatest mysteries of science -- quantum mechanics and consciousness -- are intertwined as if they could illuminate each other.

But this link is purely speculative. There is no empirical evidence that quantum effects are persisted in the warm, humid environment of the brain over sufficiently long timescales to play a functional role. Moreover, even the detection of such quantum processes would not explain why they should produce consciousness. The theory replaces one mystery with another without gaining any actual explanatory power. It is an example of how scientific-sounding terms are used to hide metaphysical speculations.

3. The tautological circle

Theories of consciousness operate in a circular relationship between theory and empiricism: they invoke empirical data to support their validity, and then use the same data to confirm the theoretical premises. This creates an epistemological circle that does not allow for independent validation.

As a rule, these theories are composed of empirical markers, which in turn have been obtained independently. But what does it mean to bring these markers together in an overall concept? Empirically, they do not provide anything that is not already shown by the markers themselves. As theories, they therefore have no additional content, but merely the character of a narrative order. Their overall message is exhausted in metaphorically bundling known findings. As long as they do not point beyond these markers and prove their generative power, they remain epistemically empty.

3.1 The problem of post-hoc rationalization

Let's consider a concrete case: The theory of the global neural workspace identifies several neural signatures of conscious processing -- late cortical activation (P3 wave), prefrontal activity, theta oscillations, and long-range synchronization (Mashour et al., 2020). These markers were discovered through experiments in which conscious and unconscious processes were distinguished.

The theory now integrates these findings into an overall model and claims that the global workspace is the mechanism that generates consciousness. But what does this integration achieve? It arranges the markers in a conceptual scheme, but does not add any new empirical information. If we want to test the theory, we look for exactly the same markers that have already gone into its construction. This is not true validation, but post-hoc rationalization.

A real test would have to make novel predictions that would not be possible without theory. The theory would have to tell us under which conditions consciousness should or should not occur under which conditions that have not yet been investigated. Such predictions are rare. Most of the theories are limited to reinterpreting already known phenomena.

3.2 The indistinguishability of competing theories

Another problem of the circle is the fact that different theories of consciousness can explain the same empirical data. The IIT and the GWT can both refer to neuronal correlates of consciousness. Both can explain why patients in certain states are conscious or unconscious. Both are compatible with functional imaging data.

If two theories explain the same data equally well, then the data is not suitable to distinguish between the theories. This is a classic problem of underdetermination: the empirical evidence is not sufficient to justify a clear theoretical interpretation. In this situation, the only thing left to do is to resort to theoretical virtues such as simplicity, elegance, or coherence with other theories -- all subjective criteria that do not allow for an objective decision.

Consciousness research is thus in a precarious position: it constantly produces new data, but this data does not contribute to the clarification of the fundamental theoretical questions. Any theory can assimilate the data without the possibility of falsification. Empirical progress comes to nothing because it does not go hand in hand with theoretical progress.

4. Cognition function and heuristic value

Theories of consciousness serve less scientific knowledge than the interpretation of the world. They tell stories about how the world and experience could be connected, and each give this relationship its own symbolic form. In this sense, all theories of consciousness are of equal rank: they move on the same level of speculative metaphysics. Their differences do not lie in the empirical verifiability, but in the way they are interpreted. In this way, they fulfill a philosophical rather than a scientific function -- they are attempts to conceptualize the inexplicable, not to explain it empirically.

Its greatest heuristic value lies in bringing empirical observations into a consistent structure of interpretation. Those theories that are directly based on empirical data and derive their consistency from established theories of evolutionary biology, psychology and neuroscience make the most valuable contribution here. While they do not provide an explanation of consciousness, they do provide a coherent framework for describing its conditions.

5. Counter-Positions: The Defense of Consciousness Research

The criticism presented here is radical. It claims that all current theories of consciousness have fundamental epistemic deficits and have no scientific explanatory power. It is therefore necessary to examine the strongest counter-arguments and evaluate whether they can refute the criticism.

5.1 The argument of scientific normality

Defenders of consciousness research might initially object that the standards formulated here are unrealistically high. No science can provide complete explanations in its initial stages. Consciousness research is in a phase of theory formation and hypothesis testing, as is typical for young disciplines. It was unfair to demand that she had already completed declarations.

This objection has a kernel of truth: science is a process, not a ready-made collection of truths. But the crucial point is not that consciousness research does not yet have complete answers, but that it does not have clear criteria by which progress can be measured. In other young disciplines -- such as early genetics or immunology -- there were ambiguities and competing theories, but it was clear in principle what kinds of findings would confirm or disprove the theories.

It is precisely this clarity that is missing in consciousness research. We don't know what empirical discovery would solve the problem. We don't even know what a solution would have to look like. This fundamental disorientation distinguishes consciousness research from other scientific disciplines in its early phase. It suggests that we are not only in a preliminary stage of realization, but may be taking a categorically wrong approach.

5.2 The argument of practical successes

A second objection refers to the practical successes of consciousness research. Theories such as GWT and IIT have led to new diagnostic tools that can be used to better assess the state of consciousness of coma patients (Casali et al., 2013). They inform the development of anesthetics and help to understand the neuronal correlates of various mental disorders. This practical usefulness shows that the theories are not mere speculation, but provide real gains in knowledge.

This objection is justified, but does not get to the heart of the criticism. No one denies that consciousness research has identified important correlates. The point is that these correlates do not provide an explanation of consciousness itself. A thermometer is practically useful for diagnosing fever, but it doesn't explain why infections cause fever. Likewise, the neural markers are practically useful for identifying states of consciousness, but they do not explain why these neural states are associated with subjective experience.

The practical successes of consciousness research are real and valuable. They justify the continuation of this research. But they should not be confused with a real explanation of consciousness. There is a difference between clinical usefulness and scientific explanation.

5.3 The Gradual Enlightenment Argument

A third objection is that consciousness is a complex, gradual phenomenon that cannot be understood in a single theoretical breakthrough. The different theories of consciousness each shed light on different aspects of this phenomenon. The GWT explains the availability of information, the IIT explains the unity of experience, the Predictive Processing Framework explains the intentionality and directionality of consciousness. Taken together, these approaches provide an increasingly complete picture.

This position is attractive because it is pluralistic and undogmatic. It avoids the claim that a single theory has to explain everything. But it only postpones the problem. Because now we have to ask: How are these different aspects connected? Is there an overarching theory that explains why information availability, unity, and intentionality occur together? Or are these independent phenomena that are only connected contingently?

In addition, the fundamental difficulty remains: None of these partial explanations shows why the processes described should go hand in hand with subjective experience at all. One can understand all the functional aspects of consciousness -- attention, working memory, self-modeling, behavioral control -- without clarifying why these functions are accompanied by phenomenal experience. This is the essence of the "hard problem" (Chalmers, 1996), and no approach to gradual enlightenment has found a convincing answer to it.

5.4 The argument of methodological impossibility

Perhaps the most important objection is that the "generative force" required here is an impossible criterion. Because even if we were to construct a system that behaves like a conscious system in every way, we could never know for sure whether it is actually conscious. Consciousness, by definition, is only accessible in the first-person perspective. We cannot "look into" the system and check its subjective experience. Therefore, the demand for generative proof is in principle unfulfillable.

This objection touches on the fundamental problem of the "other minds problem" -- the question of how we know that other beings are conscious. It is true that we can never have absolute certainty about the consciousness of others. But it does not follow from this that the demand for generative power is pointless. There is an important distinction between:

(a) a system constructed according to the principles of a theory of consciousness that meets all functional and behavioral criteria for consciousness, and

(b) a system that is merely correlated with certain neuronal correlates of consciousness, without it being clear whether these correlates are causally relevant.

In case (a), we would have at least a strong prima facie reason to believe that the theory is correct, because it would have demonstrated its ability to produce consciousness-like phenomena. In case (b) we have only a correlation, not an explanation. The demand for generative power is therefore not impossible to fulfill, but represents a higher, but in principle achievable level of confirmation.

5.5 The False Dichotomy Argument

A final objection might be that the critique presented here establishes a false dichotomy: either complete explanation or mere speculation, either generative force or epistemic emptiness. In fact, there is a wide spectrum between these extremes. The theories of consciousness may not be complete explanations, but they are not completely useless either. They provide conceptual clarity, structure empirical research and enable more precise questions to be asked.

This objection is the strongest of all, for it recognizes the limitations of contemporary theories without completely negating their value. It is true that theories of consciousness structure the field of research and provide heuristic orientation. They make it possible to systematically compare different phenomena and generate new hypotheses.

But the central question remains: Are these theories on the right track to a scientific explanation of consciousness, or do they move within a paradigm that is incapable of capturing the phenomenon in principle? Heuristic usefulness alone does not guarantee that we will approach the truth. Ptolemaic astronomy was heuristically useful for centuries and made precise predictions possible, but was based on a fundamentally wrong view of the world.

The concern is not that consciousness research is progressing slowly, but that it is progressing in a direction that misses the real problem. As long as all theories have the same basic void -- the inability to explain why physical processes are associated with experience in the first place -- this points to a systematic problem, not a temporary immaturity of the discipline.

5.6 Reply: The difference between preliminary incompleteness and fundamental blank space

The objections raised highlight important nuances, but do not change the basic diagnosis. The crucial point is not that consciousness research is incomplete -- that is every science. The point is that it has a principled explanatory gap that cannot be filled by more data or refined theories.

This gap is that all current theories describe functional or structural properties (information integration, global availability, recursive processing), while consciousness is a phenomenal property (how-it-is-to-experience). No description of functional properties, no matter how detailed, can logically imply the existence of phenomenal properties. This is not an empirical problem that could be solved by better measurements, but a conceptual problem.

The demand for generative force is therefore not arbitrary, but follows from this insight: If a theory claims to explain consciousness, it must be able to show that the mechanisms it postulates actually produce consciousness. Until this succeeds, we have no reason to believe that the theory is more than a useful description of correlates.

The practical successes of consciousness research are real and valuable, but they do not justify the theoretical claims that are often associated with them. There is a difference between identifying markers and explaining the phenomenon itself. Consciousness research has made considerable progress in the former, but not in the latter.

6. Consequences and outlook

6.1 Epistemic modesty as a scientific virtue

If the analysis presented here is correct, then consciousness research should find a greater epistemic modesty. This does not mean discontinuing research, but rather specifying its claims. Theories of consciousness should honestly present themselves for what they are: heuristic models that identify correlates and describe functional architectures, but do not provide explanations of phenomenal consciousness in the true sense.

This modesty would have several advantages. First, it would avoid unrealistic expectations and objectify the public debate. Too often it is suggested that science is on the verge of solving the "riddle of consciousness", only to present the same old problems in a new guise. Second, it would promote interdisciplinary openness. Acknowledging that neuroscientific approaches have limits of principle creates space for philosophical, phenomenological, and contemplative perspectives.

Third, it would focus research on productive issues. Rather than claiming to "explain" consciousness, research could focus on mapping the conditions under which consciousness occurs, understanding the neural mechanisms that enable different contents of consciousness, and developing the clinical applications of these findings. All this is valuable work that does not require exaggerated theoretical claims.

6.2 Alternative research strategies

If the current approaches reach their limits in principle, what alternatives are there? One way is to ask the question differently. Instead of asking "What is consciousness?" or "How does consciousness arise?", one could ask: "Under what conditions do systems report conscious experience?" This shift from the ontological to the epistemic does not circumvent the hard problem, but it does make it manageable.

A second strategy is the development of new experimental paradigms. Instead of looking for correlates, one could try to manipulate consciousness in a targeted way -- not only to suppress it (as in anesthesia), but to change it qualitatively. Psychedelic research, neurofeedback, and technologically mediated alteration of consciousness could provide new insights precisely because they push the boundaries of normal states of consciousness (Carhart-Harris & Friston, 2019).

6.3 The role of phenomenology

Current consciousness research is strongly fixated on third-person methods: brain scans, behavioral experiments, mathematical models. But consciousness is primarily a first-person phenomenon. The neglect of systematic first-person research -- phenomenology -- is a serious deficit.

Phenomenological methods can precisely describe the structures of experience: the temporal dynamics of the stream of consciousness, the intentionality of experience, self-givenness, the affective tint of all experience. These descriptions are not subjective in the sense of arbitrary, but intersubjectively comprehensible and systematic (Husserl, 1913; Varela & Shear, 1999).

Consciousness research that combines third-person data with systematic first-person research would have a richer empirical basis. She would not only ask which brain states correlate with consciousness,

but also what structures the experience itself has and how these are related to neuronal processes. This is a demanding methodological program, but it is necessary if we are to do justice to the phenomenon.

6.4 Consciousness and Artificial Intelligence

The question of whether artificial systems can be conscious becomes particularly explosive due to the analysis presented here. If it is unclear even for biological systems which properties produce consciousness, how are we supposed to decide whether AI systems are conscious?

Current AI systems, including large language models, meet some functional criteria traditionally associated with consciousness: they process information, show flexible behavioral adaptation, can "report" on their own states. But do they also meet the criteria of IIT or GWT? And even if they did, would that prove consciousness or just functional equivalence?

The uncertainty in this question shows the fundamental dilemma: We have no reliable criteria for consciousness that go beyond mere behavior. Every answer to the question "Is this AI conscious?" already presupposes a theory of consciousness. And since all these theories contain metaphysical positing, the answer is ultimately a question of philosophical position, not empirical determination.

This should call for extreme caution, both for claims that AI is definitely not conscious, and for claims that it can or will be. We are in a state of fundamental ignorance that cannot be remedied by technological progress alone.

Result

Theories of consciousness have not yet fulfilled an empirical test. They operate with metaphysical settings, circular references and heuristic metaphors. Its value therefore lies not in its explanatory power, but in its function of symbolically ordering the relationship between the world, life and experience.

The defense of consciousness research rightly points to its practical successes, its heuristic usefulness, and its status as a young science. But this defense does not change the basic diagnosis: All current theories have a fundamental explanatory gap -- they describe functional and structural properties without being able to explain why they go hand in hand with phenomenal experience.

This gap cannot be filled by more data or refined models as long as the basic conceptual architecture remains unchanged. Only when a model can actually generate consciousness -- or when we move on to a radically different question -- does philosophy turn into science.

Until then, consciousness research will remain a discipline with significant empirical successes in the identification of correlates and a persistent theoretical perplexity regarding the explanation of the phenomenon itself. Acknowledging this tension is not a sign of weakness, but of intellectual honesty. It opens up the space for new, possibly more productive approaches and protects against the self-deception of having already solved the problem when we have only reformulated it.

Literature

Aaronson, S. (2014). Why I am not an Integrated Information Theorist (or, The Unconscious Expander). https://www.scottaaronson.com/blog/?p=1799

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge University Press.

Berger, J. (2014). Consciousness is not a property. *Philosophical Studies*, 171(3), 515-537.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18(2), 227-287. https://doi.org/10.1017/S0140525X00038188

Carhart-Harris, R. L., & Friston, K. J. (2019). REBUS and the anarchic brain: toward a unified model of the brain action of psychedelics. *Pharmacological Reviews*, 71(3), 316-344.

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. https://doi.org/10.1126/scitranslmed.3006294

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181-204.

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Viking.

Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200-227.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127-138. https://doi.org/10.1038/nrn2787

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

Husserl, E. (1913). *Ideas on a Pure Phenomenology and Phenomenological Philosophy*. Max Niemeyer Verlag.

Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, 15(8), 365-373.

Mashour, G. A., Roelfsema, P., Changeux, J. P., & Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron*, 105(5), 776-798. https://doi.org/10.1016/j.neuron.2020.01.026

Melloni, L., Mudrik, L., Pitts, M., & Koch, C. (2021). Making the hard problem of consciousness easier. *Science*, 372(6545), 911-912. https://doi.org/10.1126/science.abj3259

Penrose, R., & Hameroff, S. (1995). What gaps? Reply to Grush and Churchland. *Journal of Consciousness Studies*, 2(2), 98-111.

Rosenthal, D. M. (2005). *Consciousness and mind*. Oxford University Press.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5(1), 42. https://doi.org/10.1186/1471-2202-5-42

Tononi, G. (2015). Integrated information theory. *Scholarpedia*, 10(1), 4164. https://doi.org/10.4249/scholarpedia.4164

Varela, F. J., & Shear, J. (1999). First-person methodologies: What, why, how? *Journal of Consciousness Studies*, 6(2-3), 1-14.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. WITH press.