



# A socio-historical take on the meta-problem of consciousness

Hakwan Lau, Matthias Michel

## ► To cite this version:

Hakwan Lau, Matthias Michel. A socio-historical take on the meta-problem of consciousness. *Journal of Consciousness Studies*, 2019, 26 (9-10), pp.136-147. halshs-03937595

**HAL Id: halshs-03937595**

**<https://shs.hal.science/halshs-03937595>**

Submitted on 13 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A socio-historical take on the meta-problem of consciousness

Hakwan Lau<sup>1,2,3,4</sup>, Matthias Michel<sup>5</sup>

1. Department of Psychology, UCLA
2. Brain Research Institute, UCLA
3. Department of Psychology, The University of Hong Kong
4. State Key Laboratory for Brain and Cognitive Sciences, The University of Hong Kong
5. Sciences, Normes et Démocratie, Sorbonne Université

Correspondence: [hakwan@gmail.com](mailto:hakwan@gmail.com), [matthias.michel.curtil@gmail.com](mailto:matthias.michel.curtil@gmail.com)

**Abstract:** The intuition that consciousness is hard to explain may fade away as empirically adequate theories of consciousness develop. We review socio-historical factors that account for why, as a field, the neuroscience of consciousness has not been particularly successful at developing empirically adequate theories. Based on this we argue that the meta-problem may be a self-fulfilling prophecy, created in part because we inadvertently focused too much on the so-called ‘hard problem’, limiting scientific progress.<sup>1</sup>

---

<sup>1</sup> This is a commentary on Chalmer’s target paper “The meta-problem of consciousness” (2018).

## Explanatory intuitions and progress in consciousness science

It is often said that consciousness is hard to explain, but why do we *think* that consciousness is hard to explain? As Chalmers remarks, “intuitions holding that consciousness is hard to explain” (p.12) are central to this meta-problem. Here, the central intuition seems to be that *whatever physical mechanisms* we imagine, it remains unintelligible that they could explain consciousness.

However, intuitions about what is intelligible and what’s not depend on the scientific theories available. An example of this is the Newtonian theory of gravitation (Chang, 2009; De Regt and Dieks, 2005). When the theory was first formulated, the fact that it implied *action-at-distance* was regarded as a serious problem, most notably by Huygens, Leibniz, and Newton himself. However, as De Regt and Dieks (2005) remark, between 1700 and 1850, the possibility of action-at-distance became well accepted, precisely because the Newtonian theory was *empirically* successful. De Regt and Dieks (2005) suggest that, at this time, proponents of action-at-distance theories simply didn’t see the problem with supposing the existence of action-at-distance, given the empirical success of the Newtonian theory. As Chang (2009) writes:

There are many other cases from the history of science showing the questionable effect of metaphysical beliefs that are leftovers from earlier scientific theories: Einstein’s insistence on determinism against the Copenhagen interpretation of quantum mechanics, Erwin Schrödinger’s objection to quantum jumps, the Cartesians’ objection to Newtonian action at a distance, and the like. The scientific reactionaries protested that the new theory did not make sense, that it was unintelligible. These cries of unintelligibility largely ceased to resonate with new generations of scientists, who grew up with the new theory and found it just as intelligible or unintelligible as any other theory. (p.67-68)

We ask the reader to consider that the same thing could happen with consciousness science: the intuition that explanations of consciousness are unintelligible might progressively fade away as scientists develop empirically adequate theories of consciousness. But importantly, at the very least, one should see the opposite as obvious: if empirical science doesn’t develop adequately, the intuition that consciousness is unexplainable should remain as strong as ever, which could account for the meta-problem.

By “empirically adequate”, we mean a theory that would allow us, for instance, to: determine which brain regions are important for consciousness, whether attention is necessary for consciousness, how consciousness disappears during dreamless sleep, determine how many items one can be conscious of at once, determine which non-human animals are conscious, determine whether we have conscious experiences of seeing colors

in the visual periphery, explain why visual masking can make people unconscious of visual stimuli, identify with a satisfying degree of accuracy when patients who are unable to communicate are conscious or not, determine the relations between consciousness and self-consciousness, or agency, determine which cognitive activities require consciousness, and so forth. We can make progress on these questions without confronting the so-called “hard problem” head on. And yet, such progress may turn out to be beneficial if not crucial for the problem.

Our hypothesis is that consciousness has been considered unexplainable for so long at least partly because we have done rather poorly in developing empirically adequate accounts of consciousness, lagging behind the rest of the cognitive neurosciences in terms of rate of empirical progress. From this perspective, a response to the meta-problem of consciousness requires one to explain why we have done so poorly relative to reasonable expectation, giving rise to this illusion of inherent scientific inadequacy. In the remainder of this paper, we suggest that this is due in large part to socio-historical factors.

### **Why socio-historical factors matter: a model**

Chalmers expresses skepticism towards socio-historical answers to the meta-problem:

Some might give genealogical accounts of problem intuitions in terms of accidents of cultural history. Perhaps we have all been over-influenced by Descartes, for example. Others might give psychoanalytic explanations, perhaps in terms of fear of death, or our yearning to be special. I am skeptical that explanations of this sort go deep enough... (p. 20-21)

But socio-historical factors do not simply influence our philosophical views about the topic itself. They also crucially influence our ability to produce empirically adequate theories.

We propose a socio-historical model describing a vicious cycle that consciousness science has been in (Figure 1). In a nutshell, when a scientific topic is taboo, it attracts what Sperber (2010) called “gurus”, namely, individuals with high levels of (sometimes unearned) authority who tend to adopt what we call “revolutionary” views. Since they already earned credit from the scientific community in other domains, the gurus can engage in speculation without much repercussion for their careers. Then, these “gurus” are prone to what Kitcher (1995, p. 323) calls “backscratching”. Backscratching happens when individuals express (sometimes unconditional) support for the revolutionary views of gurus. In turn, backscratchers elevate their own status and visibility, and reinforce the attribution of authority to the gurus. Often, backscratchers are not scientific experts on the topic, and together with a lack of incentives, they fail to critically evaluate the gurus’ claims. This can lead to a breakdown of the peer review process, meaning that the empirical rigor as well as

the reputation of the field go down, which also decreases the prospects for funding. This last factor then feeds back into the first factor: because of the lack of funding and the lack of respect within the broader scientific community, the quality of the science suffers and the topic remains a taboo.

Below we will use historical as well as contemporary examples to illustrate how each step of this model applies in the case of consciousness science. If this model is correct, it accounts for our inability to develop an empirically adequate theory of consciousness. In turn, this could at least partly account for the meta-problem.

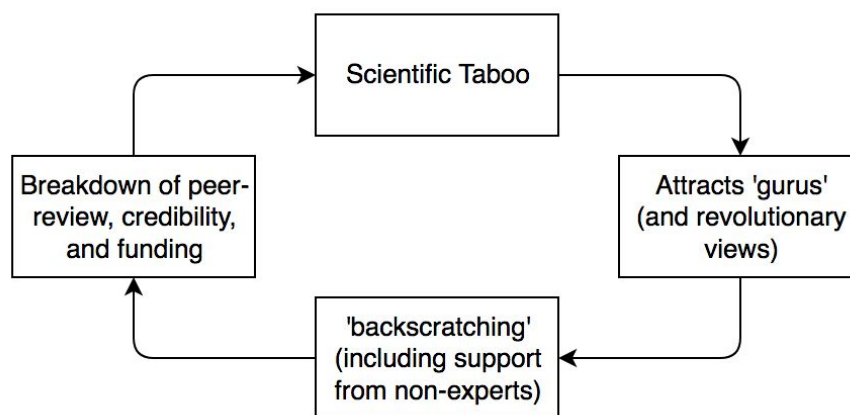


Figure 1. A vicious cycle that reinforces our explanatory intuitions, and accounts for why empirical progress has been limited (which in turns limits the prospects of empirical adequacy). The terms ‘gurus’ and ‘backscratchers’ are taken from Sperber (2010) and Kitcher (1995) respectively.

### Step 1: from taboos to revolutionary ‘gurus’

Baars (2004) wrote: “psychologists avoided consciousness for most of the twentieth century. The central topic in psychological science became taboo. Those with serious interest in it risked professional suicide.” (p.4).

Accordingly, those who engaged in the early scientific study of consciousness tended to be those with enough job security to be willing to take such a risk. For example, this included neuroscientists John Eccles (Eccles 1965) and Roger Sperry (Sperry, 1969), who were both Nobel Prize winners, as well as other titans in neuroscience, such as Wilder Penfield (1975) or in psychology, such as Leon Festinger (1967). In the early 90s, two more Nobel Prize winners joined in: Francis Crick (Crick & Koch, 1990) and Gerald Edelman

(Tononi and Edelman, 1998). Note that neither of them were originally neuroscientists by training.

Another prominent figure in the 1990s 'revival' was also from outside the field; Sir Roger Penrose is an eminent theoretical physicist. It is notable that Penrose's (1989) contribution to studies of consciousness has been used as an example when Sperber proposed the "guru effect" – an effect by which the authority of the source of a statement circumvents its critical evaluation, when there is little empirical evidence available for the evaluation, or when evaluating this statement is too costly – for instance in the case of Penrose's statements, because evaluating them requires the mastery of too many domains (Sperber, 2010).

We refrain from taking sides regarding Penrose's actual intellectual contribution to the problem. But historically, there is indeed a pattern for eminent scientists entering the field to favour ambitious ideas that are difficult to be scientifically evaluated by the community. Another example is John Eccles, who went from a decades-long careful consideration of the nature of the synapse to writing that:

... mind-brain interaction is analogous to a probability field of quantum mechanics, which has neither mass nor energy yet can cause effective action at microsites. More specifically it is proposed that the mental concentration involved in intentions or planned thinking can cause neural events by a process analogous to the probability fields of quantum mechanics. (1989, p 197)

It is extremely difficult to find any empirical evidence that would specifically support Eccles' claims. Eccles himself mentions "psycho-kinetic experiments" (1951, p. 57). If one rightfully puts this highly controversial 'evidence' aside, it is fair to say that his view was radically underdetermined by the available empirical evidence (Turnbull, 2018).

That one can turn from careful experimental studies to wild metaphysical speculations as soon as one enters the field of consciousness is notable. Another related example may be B.F. Skinner, who is well known for his methodological conservatism (Skinner, 1953). However, later on in his career he speculated rather loosely on consciousness and human nature (Skinner, 1971). Overall, there seems to be a *culture* supporting the indulgent impression that consciousness is such a topic of an entirely unique nature that usual scientific prudence needs not apply.

Let us distinguish between two approaches to scientific problems in consciousness science: the "revolutionary" and the "incremental". On the revolutionary approach, theories are built with the explicit goal of solving the problem of consciousness. On the incremental approach, theories are built with the goal of answering as many basic "why-questions" (van

Fraassen, 1980) as possible, such as: why is this part of the brain active only when one is conscious, or why is it that subjects do not report being conscious of stimuli that are quickly followed by a mask, and so forth. That is, adopting the incremental approach consists in progressively building an empirically adequate theory of consciousness.

Throughout the history of consciousness science, revolutionary approaches have featured prominently, which might have prevented the development of empirically adequate theories, as we hypothesize in the next section. Besides the examples given above, this general pattern is easy to understand: given the enormous stature of some scientists that this taboo topic attracts, typically near their end of careers, many are aiming for a major 'blowout', one last ground-breaking achievement to add to their well-decorated vitae. It makes sense *for them* to take bigger risks than usual.

As such, we emphasize that adopting this revolutionary approach was contingent, and depended on historical and sociological factors rather than on scientific necessity. Instead of trying to "solve" the problem of consciousness head-on, and explain why we are conscious at all, consciousness scientists could have attempted to answer a wide variety of simpler why-questions regarding consciousness, and thereby progressively build empirically adequate theories.

We do not assume that the revolutionary approach has no merits. Instead, our point is that privileging the incremental approach could progressively lead to the disappearance of our intuition that consciousness is unexplainable. Nor do we mean that the revolutionary approach has completely dominated the field; many do work on incremental problems (Block et al., 2014; Michel et al., 2019). However, given the enormous influence that revolutionary views have, they can severely undercut the necessary incremental progress, as we will explain next.

## **Step 2: from gurus to backscratching**

As was made clear by the examples above, given the ambitious nature of revolutionary claims, views on consciousness are often very difficult to evaluate properly for the scientific community. We suggest that the underdetermination of revolutionary theories by empirical evidence, and the so-called "guru effect" (Sperber, 2010) have led to the persistence of the revolutionary approach.

When theory choice is underdetermined by empirical evidence, it has been suggested that non-epistemic social values take a more significant place in determining which theories scientists as well as non-experts will support (Kuhn 1977). Among those

values is the *authority* of the source of a theory, which may be particularly important given the presence of (even just a few) gurus (Sperber, 2010).

To understand the “guru effect” (Sperber 2010) in this context, it is important to distinguish between two types of authority, following Kitcher’s taxonomy:

there is *unearned* authority that stems from the scientist’s social position (...), the type of authority that arises from being associated with a major institution or from having been trained by a prominent figure; this contrasts with earned authority, that credibility assigned by reflection on the scientist’s performances or through consideration of others’ opinions of those performances. (Kitcher, 1995, p.315)

Using authority as a heuristic might work quite well in many cases: after all, a good explanation for a scientist’s authority is her ability to do good research. However, this heuristic is sometimes dysfunctional: without much empirical data available, our cognitive mechanisms for evaluating the relevance of an idea or theory tend to rely primarily on (unearned) authority as a proxy for relevance (Sperber 2010).

Importantly, this ‘guru effect’ is itself a vicious cycle: the unearned authority of the source of a theory leads to a high number of followers, which themselves reinforce the authority of the source. In accordance with Kitcher’s model of authority in science (1995), we label these followers “backscratchers”. According to Kitcher, backscratching is “a process in which your positive opinion of me raises my respect for you, which, in turn, may further increase you attribution of authority to me” (p.323). The idea is that being part of the restricted community of the guru has an advantage for the backscratchers, namely, they can earn credit by capitalizing on, and reinforcing, the authority of the guru.

For example, Eccles had earned a high level of authority, most of which did not come from a rigorous study of consciousness *per se*. As such, the authority of his claims regarding consciousness was unearned, compared to his authority on, say, synaptic transmission. The well-known relationship between Eccles, a scientist with authority, and Popper, a philosopher *outside* the relevant empirical field, illustrates the backscratching phenomenon. In this case, Eccles benefited from receiving the consideration of a well-known philosopher (Popper & Eccles, 1984). This reinforced his authority: if a respected philosopher supports Eccles’ views, a good explanation seems that there must be something right to Eccles’ claims. On the other hand, Popper, as a philosopher, also benefited from showing that scientists with high levels of authority were taking his views seriously. This mutually beneficial relation forms the basis of the backscratching effect (Kitcher, 1995).

Historically, in the study of consciousness there were other instances where followers from outside of the discipline can support ideas rejected by experts within the field. Such



was the case of Freud, who enjoyed an extremely high level of publicity. When psychiatry was ready to abandon his ideas, support from the humanities ensured that his academic influence remained prominent (Crews, 2017; Eysenck, 1984). This last example is of particular importance, because it illustrates a particularly exaggerated form of the guru and backscratching effects, to which we next turn.

### **Step 3: Breakdown of peer-review, funding, and credibility**

Overall, this constellation of gurus and backscratchers does not constitute a good recipe for empirical success. But as the case of the decline of the so-called 'Freudian empire' illustrates (Eysenck, 1984), sometimes guru and backscratching effects can be further *fermented*, leading not just to mistaken ideas, but rather, complete breakdown of peer review, funding, and credibility. To date, psychoanalysis is hardly considered a respectable scientific discipline, and public funding for its research is virtually non-existent. Several interrelated factors make this kind of decay, or at least its milder variants, particularly likely to happen in the field of consciousness.

The first factor is that because the field is small, just a handful of misguided scientists holding "false paradigms" can support and promote each other, forming a critical mass that would persist for decades to come. In a computational simulation work, Akerlof and Michailat (2018) described exactly such a model, which may be particularly applicable to consciousness science, given the historical culture of lack of scientific prudence discussed above; in their terms, our field has low 'power' to detect and eliminate falsehood.

A second factor is media over-promotion, which is generally considered to be more problematic in consciousness science than in other disciplines (Michel et al., 2018, 2019). Since revolutionary views are more impressive, they tend to be more heavily promoted, attracting high visibility. Specifically in consciousness science, since the 1990s the stage has been pretty much set up for an emphasis on and appreciation of stardom (Bartlett, 2018). Media attention for the field loomed so large that at some point Francis Crick wished that journalists would stop writing about the topic (Gorman, 1997). In part, the worry was that empty promises were broadcasted, which would hurt the credibility of the field. But media over-promotion also attracted a high level of attention and participation from non-experts, which further dilutes the expertise of the peer review group, beyond backscratching.

Finally, the third related factor is funding. Historically, and up to this date, public funding for research in consciousness science is weak, especially in the US (Michel et al., 2018, 2019). Lack of funding itself means that less empirically rigorous science can be done.

But there are also secondary effects. Amid lack of public funding, researchers seek to appeal to private donors. In the long term, private funding is generally considered less preferable for the healthy growth of the field, because of its relative reliance on personal connections over rigorous peer review (Michel et al., 2018, 2019). This interacts with the last factor: gurus who loom large in the popular media may earn underserved share of funding as a result, unduly increasing their influences. And this in turn opens up a pathway for funding competition whereby scientific quality is not rewarded, and peer opinions easily neglected.

We note that not all three factors have played an equally large role throughout the history of the field. But the ‘fermentation’ process may not necessarily require all three factors to be present.

#### **Step 4: Completing the vicious cycle, collectively**

We take it that this last step is self-explanatory. Amid a breakdown of healthy peer review, funding, reputation, and job opportunities, many ordinary scientists might be deterred from entrance. The field thus remains a taboo, attracting an unhealthy number of gurus ‘brave’ enough to tackle the problem.

Overall, we do not mean to place the emphasis on the responsibility of particular individuals. In fact, the individuals may not have the explicit intention and strategy to exploit the system as such. Some models show that rational individual scientists might, given some incentive structures, behave in counter-productive ways as groups (Mayo-Wilson, 2011). Our claim has been that the incentive structures are set up amid an historically prevalent culture for this vicious cycle to take place.

Nor do we argue that the historical examples apply generally to the entire field. But because it is a vicious cycle, it only takes a few prominent individuals to engage in the said activity, for this downward spiral to be initiated and re-initiated now and then. From there, it perpetuates itself. This leads to a generally unwarranted emphasis on revolutionary approaches that are decidedly unsupported by the quality of scientific evidence provided.

#### **Concluding remarks**

It is often said that consciousness is an emergent phenomenon. In a sense, all problems of emergence, e.g. life, are hard (O’Connor & Wong, 2015). But others have made peace with such conceptual tensions (Machery, 2012). What we have argued here is that we might have *made* the problem of consciousness particularly hard for ourselves, because we focused so

much on the hardness of the problem, sometimes at the cost of the development of empirically adequate theories – a kind of self-fulfilling prophecy.

Our claim is not that these historical and sociological factors necessarily solve the meta-problem. Whether these factors are sufficient to account for the problem is an empirical matter. Only time will tell if our explanatory intuitions progressively fade away as ever more empirically adequate theories of consciousness develop. Unfortunately, one recent survey (Michel et al., 2018) does seem to suggest that the current state of the field shows properties very much compatible with the model (Figure 1). Some of these observations are also supported by a statement made by as many as 58 current researchers (Michel et al., 2019). As we have argued, the factors described in our model might prevent the development of empirically adequate theories, thereby severely limiting our explanatory intuitions. Further empirical research is of course needed, but we feel that given this context, writing off socio-historical factors from the outset would be a mistake.

## Acknowledgments

We thank Jorge Morales, François Kammerer, and Richard Brown for their helpful comments on earlier versions of this paper.

## References

- Akerlof, G. A., & Michailat, P. (2018). Persistence of false paradigms in low-power sciences. *Proceedings of the National Academy of Sciences*, 115(52).
- Baars, B. J. (2003). The double life of B.F. Skinner: Inner conflict, dissociation and the scientific taboo against consciousness. *Journal of Consciousness Studies*, 10(1), 5-25.
- Bartlett, T. (2018) Out of their minds: wild ideas at the ‘Coachella of consciousness’. *The Guardian*, 19 Jun 2018.  
(<https://www.theguardian.com/news/2018/jun/19/out-of-their-minds-wild-ideas-at-the-coachella-of-consciousness>)
- Block, N., Carmel, D., Fleming, S. M., Kentridge, R. W., Koch, C., Lamme, V. A. F., Lau, H., Rosenthal, D. (2014). Consciousness science: Real progress and lingering misconceptions. *Trends in Cognitive Sciences*, 18(11), 556–557.
- Chang, H. (2009). Ontological Principles and the Intelligibility of Epistemic Activities. In H. W. De Regt, S. Leonelli, & K. Eigner (Eds.), *Scientific Understanding* (pp. 64–82). University of Pittsburgh Press.
- Chalmers, D. (2018). The Meta-Problem of Consciousness. *Journal of Consciousness Studies*, 25(9–10).
- Crews, F. (2017). *Freud: The Making of an Illusion*. Metropolitan Books.
- Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars*

- in the Neurosciences, 2(263–275), 203.
- De Regt, H. W., & Dieks, D. (2005). A contextual approach to scientific understanding. *Synthese*, 144(1), 137–170.
- Eccles, J. (1951). Hypotheses Relating to the Brain–Mind Problem. *Nature*, 168(4263), 53–57.
- Eccles, J. (1965). *The Brain and the Unity of Conscious Experience*. CUP Archive.
- Eccles, J. (1989). *Evolution of the brain: Creation of the self*. London: Routledge.
- Eysenck, H. (1986). *Decline and Fall of the Freudian Empire*. Harmondsworth: Penguin Books.
- Festinger, L., Ono, H., & Burnham, C. A. (1967). Efference and the conscious experience of perception. *Journal of Experimental Psychology*, 74(4, Pt.2), 1-36
- Gorman, J. (1997). Consciousness Studies: From Stream to Flood. *The New York Times*. APRIL 29, 1997.  
(<https://www.nytimes.com/1997/04/29/science/consciousness-studies-from-stream-to-flood.html>)
- Kitcher, P. (1995). *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*. Oxford University Press.
- Kuhn, T. S. (1977). Objectivity, value judgment, and theory choice. In *The Essential Tension: Selected studies in scientific tradition and change* (pp. 320–339). University of Chicago Press.
- Machery, E. (2012). Why I stopped worrying about the definition of life... and why you should as well. *Synthese* 185 (1):145-164.
- Mayo-Wilson, C., Zollman, K., & Danks, D. (2011). The Independence Thesis: When Individual and Social Epistemology Diverge. *Philosophy of Science*, 78(4), 653-677.
- Michel, M., Fleming, S. M., Lau, H., Lee, A. L. F., Martinez-Conde, S., Passingham, R. E., ... Liu, K. (2018). An Informal Internet Survey on the Current State of Consciousness Science. *Frontiers in Psychology*.
- Michel, M., Beck, D., Block, N., Blumenfeld, H., Brown, R., Carmel, D., ... Yoshida, M. (2019). Opportunities and challenges for a maturing science of consciousness. *Nature Human Behaviour*, 3(2), 104–107.
- O'Connor, T., & Wong, H. Y. (2015). Emergent Properties. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.
- Penfield, W. (1975). *The Mystery of the Mind: a Critical Study of Consciousness and the Human Brain*. Princeton University Press.
- Penrose, R. (1989). *The Emperor's New Mind*. Oxford University Press.
- Popper, K., & Eccles, J. (1977). *The Self and its Brain*. Berlin: Springer.
- Sperry, R.W. (1969). A Modified Concept of Consciousness. *Psychological Review* 76 (6): 532–36.
- Skinner, B.F. (1953) *Science and human behavior*. New York: Macmillan.
- Skinner, B.F. (1971) *Beyond Freedom and Dignity*. Penguin Books.
- Sperber, D. (2010). The Guru Effect. *Review of Philosophy and Psychology* 1 (4):583-592.

- Tononi, G., & Edelman, G. M. (1998). Consciousness and Complexity. *Science*, 282(April).
- Turnbull, M. G. (2018). Underdetermination in science: What it is and why we should care. *Philosophy Compass*, 13(2), 1–11.
- van Fraassen, B. (1980). *The Scientific Image*. Oxford University Press.