Introduction: Reality as One Sees It

I.1 Echo Chambers

Countless eminent scholars have written on the topic of consciousness. Do we really need yet another book about it?

I am writing one because I believe that the science of consciousness is at a crossroads. On the one hand, public interest on the topic remains as high as ever. Our visibility in the media ensures that private donors recognize our challenges. At times they support us generously. Students are excited about the topic. Many are eager to join the field. In many ways, things seem to be going well.

On the other hand, it is unclear where we really stand scientifically. Outside of the field, many of our colleagues don't think much of what we do at all. Some may concede that this is just a matter of the nature of the scientific challenge we face. However, many also believe that the sheer lack of quality and rigor of the work is to blame. Unfortunately, I have to confess, sometimes I think they have a point.

Perhaps the discrepancy between the two different outlooks can be explained by two facts. The first is that scientists often vote with their feet. If they see that some field is hopeless, they may just ignore it, and focus on what they see as more tractable instead. To hear from our critics, we may need to seek them out.

I often like to hear what my 'opponents' have to say. Maybe it is in part my temperament. But I also realize—intellectual benefits aside—there are strategic reasons for engaging in this kind of conversation. The success of a scientific discipline depends not just on sheer empirical and theoretical progress. Often, acceptance by our academic neighbors matters too.

This brings me to the second reason why the negative outlook is often downplayed. Not only do our critics tend to keep their strong opinions to themselves, but those who are within the discipline are also prone to ignoring these negative comments. We often choose to live inside our own echo chambers. In part because, frankly, it doesn't seem to be good for business to emphasize too

In Consciousness We Trust. Hakwan Lau, Oxford University Press. © Hakwan Lau 2022.
DOI: 10.1093/oso/9780198856771.003.0001

much our own shortcomings or to promote them. Criticisms are also generally unpleasant to hear. But even when we aren't so shortsighted and defensive, there is this romantic feeling of going against the grain, that is pretty much shared within the community of consciousness researchers: *Historically, we know that things have been hard. But we shall ignore our critics, and soldier on. Against all the odds, we will eventually get there and prove them wrong . . .*

I.2 Romantic Ambitions

I am no stranger to this romantic feeling. After all, I have spent half of my life in this somewhat controversial field.

In college, I read Dave Chalmers' *Conscious Mind* (1996)*)*. Like the grunge and alternative rock music that was popular around the time of its publication, the book shook my world. Besides the refreshingly clean arguments, I distinctly remember how *cool* it was, that a rising star of a young scholar expressed so beautifully his heartfelt frustration at the many attempts to reduce our subjective experiences down to some physical processes. Those attempts can sometimes lead to "elegant theories," I recall Chalmers wrote, *but the problem does not go away*.

Don't get me wrong—I was, and still am a cognitivist, in the sense that, I believe the best way to understand the brain is to think of it as a biologically instantiated computer. Concepts from electrical engineering and computer science have proven to be great analogies, if not straightforward theoretical constructs, for understanding how the brain functions. That's our bread and butter.

But there is one problem sticking out like a sore thumb. Machines just don't seem to *feel* anything, however sophisticated they are at processing signals. So conscious perception cannot just *be* a matter of processing signals because there are these unexplained *raw feels*. There is *something it is like* seeing the color red. It's more than just picking up some wavelength values of incoming lights. Explaining how these subjective experiences come about is what Chalmers called the Hard Problem, and it does sound like one indeed.

This was, to my impressionable young mind then, on par with Gödel's application of devilishly clever logical analysis to show that mathematics can never be complete (Smullyan 1998). By taking the hard problem of consciousness seriously, we are recognizing that cognitive neuroscience may too be *incomplete*.

Somehow this did not prevent me from going to graduate school to study more cognitive neuroscience. There, many of my fellow students and

professors alike would snicker at the silliness of my lofty philosophical obsessions. We *are* cognitive neuroscientists. Why worry about problems that we can't solve anyway? If a problem is decidedly so 'hard,' why not find something more rewarding and tractable to *do*? Being a good scientist is to be realistic about what we can or cannot do.

But, I thought, the first step toward solving a problem is to recognize that there is one. My fellow students and professors were probably too conventional and conservative. *I shall ignore them, and soldier on.* I was particularly encouraged when I learned that the Nobel laureate Francis Crick suggested we put the sign "Consciousness NOW" in our labs and offices. If such great minds as Crick thought that the time was ripe for attack (1994), we had to be onto something.

Who knows? Perhaps decades later, we would find that we have done all we can within the limits of cognitive neuroscience, and *lo and behold*, indeed, we cannot solve the hard problem. We may need something more. Something like a *revolution*. In that sense, we are working on the edges where things may eventually break down in unpredictable ways. Indeed, they say that we may even have to revise the very foundation of physics to accommodate the occurrence of subjective experiences (Chalmers 1996). There just seems to be no room for such subjective phenomena within the ordinary language of objective science.

Like the opening power chords of Kurt Cobain's song *Smells Like Teen Spirit*, these possibilities seemed so intoxicatingly exciting to my juvenile self.

But later I found out that I was wrong. Not because young people shouldn't dream big. But because I was fundamentally misguided about some historical facts.

I.3 A Convoluted History

Like many currently active researchers in the field, I used to think that the scientific studies of consciousness were somehow "revived" in the 1990s. Or perhaps it really all started around then, with only feeble activity here and there before that was very much suppressed in the heydays of behaviorism. That narrative was so prevalent that, even when I was a graduate student at Oxford, where Larry Weiskrantz was still active in research, I just thought he must be an anomaly.

Weiskrantz coined the term *blindsight*, which refers to the phenomenon that people with specific brain damage can show behavioral signs of successful

visual information processing, such as being able to guess the identity of a visual stimulus, all without having conscious visual experience. Much of the work demonstrating the phenomenon was done in the 1970s and 1980s (Weiskrantz 1986).

And then, of course, in neuroscience textbooks, we all know about the amnesic patient HM, studied by Brenda Milner and others. Patient HM can form new nonconscious memories in the form of motor learning. What seemed to be most problematic was that he was unable to form new conscious memories of events as they occurred to him (what is also called episodic memory). Most of these details were documented as early as in the 1950s (Scoville and Milner 1957). So perhaps, another anomaly?

Yet another classic line of work that is no doubt relevant to consciousness is on split-brain patients. After having the major connections between the two hemispheres surgically severed, information presented to the right hemisphere alone cannot be verbalized. Michael Gazzaniga and colleagues have shown that some of these patients showed behavioral signs of being able to process and act on such information. However, much of that behavior seems opaque to conscious introspection. Again, many of these studies were done well before the 1990s (LeDoux, Wilson, and Gazzaniga 1979).

Far from anomalies—except in the sense of being extraordinarily influential—Weiskrantz, Milner, and Gazzaniga are all household names in neuropsychology. Their groundbreaking discoveries are taught today in undergraduate classrooms around the globe. Gazzaniga himself coined the term *cognitive neuroscience*. He is also often considered to be one of the founding grandmasters of the field. His PhD advisor with whom he did some of the split-brain patient work together was the Nobel laureate Roger Sperry. Sperry too wrote on the topic of consciousness (1965, 1969)—well before the 1990s.

Outside of neuroscience, important work has been done in other areas of psychology too. To give just a few examples, in cognitive psychology, Tim Shallice developed elegant models of attention and conscious control of behavior (1972, 1978). In social psychology, Leon Festinger's work on cognitive dissonance continues to be extremely influential to this date (1957). There, it was proposed that subjects have a need to reduce our internal conflicts. Resolving these conflicts can at times lead to rather unexpected behavior and change of attitudes. All of which seems to happen largely nonconsciously. In psychophysics, that is the quantitative analysis of perceptual behavior, Pierce and Jastrow have written on the relationship between subjective awareness and confidence ratings as early as over a century ago (Peirce and Jastrow 1885).

So, my romanticized vision about the science of consciousness turned out to be based on some serious misunderstanding. I thought studying consciousness was akin to some sort of martyrdom. We were preparing for a revolution, to break away from an unforgiving scientific tradition in which there was no place for consciousness—not before the 1990s anyway. But that was just not true.

I.4 The "Mindless" Approach

So what happened in the 1990s, exactly? What kind of revolutions were we really talking about? It would be unfair to deny the hugely positive impact of what took place then. In a series of meetings started in the early 1990s in the city of Tucson, Arizona, some of the truly great scientists of our times gathered together and plotted strategies for attacking this age-old problem of consciousness. Besides Crick, another Nobel laureate, Gerald Edelman, was also there. The Wolf Prize winner Sir Roger Penrose (now a Nobel laureate too) also attended. In those meetings a research agenda for a generation was set, and stars were born.

It's not entirely clear how it went, but somehow a misleading narrative also emerged that the modern science of consciousness started more or less right there. There's some truth to the fact that consciousness science as a relatively *organized* activity really flourished from those meetings. Inspired by the Tucson meetings, a couple of journals dedicated to the topic started, another meeting spun off: the Association of the Scientific Studies of Consciousness (ASSC) was created. But it is not true that those were the first-ever academic conferences on consciousness (LeDoux, Michel, and Lau 2020).

Instead, what really happened was that a rich and ongoing history of studies of consciousness in cognitive neuroscience and psychology were somewhat sidelined. Replacing it was a newfound obsession for physics and other natural science disciplines. The creation of ASSC restored that balance to some degree. But back in Tucson, the biannual meetings continue to attract media and public attention. The popular impression is clear: our agenda is to unlock the mystery of consciousness, and to understand our place in nature. It is a challenge for all of the natural sciences. In fact, it is one of the few remaining scientific frontiers that truly matter. Or so the narrative goes.

Perhaps this focus on the "bigger picture" isn't so bad. As yet another Nobel laureate, Rutherford, famously said, in science, there is really only physics (Birks 1962)—*the rest is just stamp collecting*.

But I have never found Rutherford's comment convincing. Good for him that he won the Nobel for chemistry. But as soon as one moves from chemistry to biology, it is well-known that lawlike reductions to physics fail (Fodor 1974; Kitcher 1984). It is all very well that water is H_2O, and hydrogen and oxygen can be defined precisely in terms of atoms, protons, electrons, and things like that. But just how does one give physical, lawlike definitions of biological functions such as digestion or reproduction?

This is not to say that biological functions are not instantiated by purely physical stuff. They are. The problem is there are no laws or equations that you can conveniently write down, in fundamental physical terms, to parsimoniously describe these functions. These functions can be realized by many different forms of physical substrates. A gut can be replaced by a functional equivalent made of rather different materials. Good luck finding fundamental physical laws about digestion. Even if one finds some equations that can fit some current data, more or less, treating them as "laws" of nature is a totally different matter.

So when I heard someone like the physicist Max Tegmark (2015) argue that consciousness may be ultimately about how the physical parts of an organism are arranged together, as if this too can be described in some simple clever equations, I just felt … maybe the 1990s were in part to blame. Perhaps the infamous "decade of the brain" misled some of us to think that understanding the "software" of the brain—that is, that fluffy thing called the "mind"—isn't as cool and impressive as going straight to the hardware. But that would be getting ahead of ourselves. If one ever wants to write down some equations at the level of physics to distinguish between a conscious and an unconscious brain, how about we first try writing down some equations to distinguish between a computer sending emails properly, versus an annoying computer which, upon having the "send" button clicked, just silently saves the drafts in the outbox without ever sending them? Can one really ignore the nitty-gritties of software, algorithms, and the like, and directly derive physical first principles there?

I.5 Before Newton

Some colleagues may feel that I'm just being too pessimistic. After all, in physics, great things have been achieved through theorizing in the abstract. Had Einstein shied away from his bold attempts at deriving the first principles, the world we live in today would be utterly different. Few would describe what

Einstein did as "reverse engineering" of a specific, messy system. It was just pure theory, on the most general and foundational level.

But what if Einstein was born in a different place and time, such as in ancient Greece, where thinkers also wondered about the universe? There, armchair theorizing seems not to have done nearly as much good. One may wonder if that was because Newton and Leibniz had not yet invented the beautiful tools of calculus for them. Perhaps the Greek thinkers did not make more progress because they lacked the precise language of advanced mathematics?

But the Newtonian laws of motion were not just written down out of sheer analytical genius. The foundation of mechanics was also built on rigorous empiricism. The laws were accepted, sometimes rather grudgingly by Newton's critics, only because they were *empirically* verified over and over again (McMullin 2001). Although these laws ultimately turn out to be incomplete (as things get extremely small or large), they give theoretical physicists a solid platform on which further derivations and inferences can be made. Had Newton got the basic facts flat wrong, no amount of elegant equations could have saved him.

Today, our mathematics are far more advanced than in the days of Newton's. But, in the science of consciousness, we are still very far from having all the relevant basic facts. It would take some profound misunderstanding of the scientific method for one to think that some such foundational laws can be derived from the sheer comfort of the armchair.

I.6 Responsible Revolutionary Planning

Despite my misgivings, I do not mean to say that ambitious universal theories can never offer any insight for understanding consciousness. The problem is, once we get past all the rigorous-looking abstruse mathematical details, often we find that the underlying assumptions are shaky and controversial (Sloman 1992; Cerullo 2015; Bayne 2018; Pautz 2019). Sometimes, the theorists commit simple logical fallacies and contradict their very own definitions (Lau and Michel 2019a). Or they make neurophysiological and anatomical claims that just seem not quite right by textbook standards (Odegaard, Knight, and Lau 2017).

These are not intrinsic problems of theoretically ambitious approaches; the concerned theorists do not *have* to neglect these details. But the problem is they often do. I worry this reflects some important sociological aspects of the sciences that are too often overlooked (Lau and Michel 2019b).

In consciousness research there has been an unduly heavy focus on personal glory and stardom. Rarely in any area of neuroscience do we think that there are these age-old puzzles, waiting to be solved by some destined genius. That's because science is generally about progress. As we work on a problem, we aim to achieve a better understanding, not to close the book forever. Of course, game-changing discoveries do happen occasionally, but only the most foolish narcissists would *expect* them to happen *in one's own hands*. In the event of such a windfall, one should be grateful for the groundwork already done by those before us.

This is not to say that we must focus on incremental empirical progress alone. The conceptual issues about consciousness are intriguing and are why many of us are here in the first place. But as Thomas Kuhn famously pointed out (1962), scientific revolutions require undeniable evidence—so undeniable as to force us to accept the inadequacy of our current paradigm. This threshold to revolution is ultimately determined by *the scientific community*.

Imagine we have to tell our colleagues in cognitive neuroscience that their approach is decidedly incomplete. Don't we have to first earn their respect? How convincing would that be if they just don't think we are even capable of telling rigorous science from utter nonsense?

And if we are so ambitious as to hope that one day we can tell the physicists to rewrite their textbooks, so as to accommodate our subject matter at their foundational level … how would we look if our colleagues next door point out that we are just flat wrong in the most elementary biological facts, while we are making these grandiose proposals?

Revolutions are exciting. But we don't call for them without the necessary ammunition. I fear though, sometimes we aim too high without being able to actually deliver. Amid all the media glory, we neglect that the *academic* reputation and longevity of the field matters. Meanwhile, career and public funding prospects remain grim for young scientists studying consciousness, especially in the United States (Michel et al. 2018, 2019). Sometimes I wonder: are we so deluded to think that the next generation doesn't matter because we think we can start and finish the said revolution ourselves right here?

I.7 Between the Vanilla and the Metaphysical

So throughout this book I will advocate for a conservative—or perhaps even *boring*—empirical approach for studying consciousness. We shall remain interested in the deeper philosophical issues, but getting the empirical facts right shall ever be our first priority.

This approach is conventional in the sense that it basically is just run-of-the-mill cognitive neuroscience these days. To those already familiar with the literature, one may wonder if this means that current major theories like the global workspace theory already suffice (Dehaene 2014)? The answer is: no. We will introduce some of these views in the next chapter, and expose their inadequacies through Chapters 2–6. That motivates a novel, alternative view.

Our overall goal here is to find mechanistic explanations for consciousness, borrowing concepts from electrical engineering and computer science. We try to figure out what may be the relevant computational processes. We infer what these may be, based on a combination of modeling and observing currently measurable neural activity in the brain. We use standard tools like neuroimaging, invasive neuronal recording, and electrical and magnetic stimulations. We ask questions like: What type of activity in what brain region may be important? What kind of cognitive functions are reflected by this activity?

We may not be able to explain everything we need to in the end. But let's see. At least we *first* give it a fair shot before we rush into something more radical. My hope is to convince you that, boring as all this may sound, much light can be shed on consciousness this way.

One may ask though, if the goal is to fully integrate with the modern standards of cognitive neuroscience, why use the term *consciousness* at all? Why not replace it with something less controversial, perhaps already existing in the literature, like, for example, working memory, attention, metacognition, and perception?

The answer is that even if we stay within the language of modern cognitive neuroscience, there is ample room for defining a notion of consciousness (i.e., of subjective experience) that is distinct from these other related concepts. Take for example perception. As we mentioned earlier, blindsight patients have certain visual perceptual capacities. What is lacking is a reported sense of subjective visual experience (Weiskrantz 1986). So, perception-like processes are not always conscious. Understanding perception alone would not be enough to understand consciousness. We also need to understand the mechanisms that render these processes sometimes conscious, and sometimes not.

And likewise, throughout the book we will argue that although consciousness is highly related to mechanisms such as metacognition, working memory, and attention, for example, it is really a distinct phenomenon.

But even if there is a distinct phenomenon, why *call* it consciousness? Why not sidestep the whole historical baggage and create a more precise technical term instead? Trouble is, I'm not sure this kind of eliminativism has ever really worked. No doubt *water* is a somewhat vague term. It is not as precise as H_2O.

But chemists don't tell us to stop saying *water*. Even if they did, I wonder if it would have mattered.

Whether we like it or not, the term *consciousness* is used in many disciplines, from psychiatry to political theory. This is what many people care about. The very notion of consciousness has its roots in the social sciences as well as in mental health research. Often, there are real, meaningful questions to be asked, regarding the role consciousness plays in these contexts. But the current answers often seem murky, not because they have to be. Rather, I fear that we have somehow failed our duty. Between worrying too much about lofty metaphysical problems, or overreacting to the other "vanilla," eliminativist extreme, we simply have not done our job. We have made it sound like no serious scientific claims can be made about the brain mechanisms for consciousness. But the concept isn't going away.

It is time to do our part.

I.8 Chapter Conclusions

Consciousness is the mechanism by which we derive our subjective sense of reality. Ironically, within the consciousness research community, different colleagues don't always see the same reality at all. This division may be particularly salient between the two sides of the Atlantic (Michel et al. 2018, 2019). For various reasons, the cognitive neuroscience of consciousness is doing somewhat better in Europe. But it is unclear how the field can truly flourish if it remains primarily a regional activity. On the Pacific front, we face yet another set of challenges; will the science of consciousness in countries like Japan, China, and Australia become more like what happens in the United States or Europe? Or will it become something totally different altogether?

I started by pointing out that the field is at a crossroads. So what options are we facing? Obviously, one way to go is to do nothing. Judging by how things have gone in the past decade, things may well become increasingly esoteric and theoretically indulgent, especially in the United States. Many may think that's fine. We will probably not run out of "big ideas" any time soon. The popular media, together with a few wealthy private donors, will probably continue to like us all the same.

Alternatively, we can make a case for why it may not be such a bad idea to be a little more aligned with common scientific standards. After all, if we aren't so misguided about the history of the field, we realize that much of the most meaningful work on consciousness has been done this way, rather than in

some unrealistic revolutionary spirit. Perhaps, with some luck, we can eventually break into the scientific mainstream.

As such, one can also say that the point of this chapter is just to lower expectations. To avoid disappointment, perhaps I should warn the reader there will not be any elegant formula allowing you to derive that your teacup is exactly 0.00000247% as conscious as your cat, or anything of equivalent mind-bending proportions. In all likelihood, your metaphysical worldviews will be left unchanged.

But does this mean that we will have nothing meaningful to say about the *hard problem* after all? I hope not. I think we will. Let's find out. But to do so, I'm afraid you have to read to the end.

References

Bayne T. On the axiomatic foundations of the integrated information theory of consciousness. *Neurosci Conscious* 2018;niy007.

Birks JB. *Rutherford at Manchester*. Heywood, 1962.

Cerullo MA. The problem with phi: A critique of integrated information theory. *PLoS Comput Biol* 2015;**11(9)**:e1004286.

Chalmers DJ. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford Paperbacks, 1996.

Crick F. *The Astonishing Hypothesis: The Scientific Search for the Soul*. Pocket Books, 1994.

Dehaene S. *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Penguin, 2014.

Festinger L. *A Theory of Cognitive Dissonance*. Stanford University Press, 1957.

Fodor J. Special sciences (or: The disunity of science as a working hypothesis). *Synthese* 1974;**28(2)**:395–410.

Kitcher P. 1953 and all that: A tale of two sciences. *Philos Rev* 1984;**93**:335–373.

Kuhn TS. *The Structure of Scientific Revolutions*. University of Chicago Press, 1962, DOI: 10.7208/chicago/9780226458106.001.0001.

Lau H, Michel M. *On the Dangers of Conflating Strong and Weak Versions of a Theory of Consciousness*. Philosophy and the Mind Sciences, 2019a, DOI: 10.31234/osf.io/hjp3s.

Lau H, Michel M. *A Socio-Historical Take on the Meta-Problem of Consciousness*. Imprint Academic, 2019b, DOI: 10.31234/osf.io/ut8zq.

LeDoux JE, Michel M, Lau H. A little history goes a long way toward understanding why we study consciousness the way we do today. *Proc Natl Acad Sci USA* 2020;**117**:6976–6984.

LeDoux JE, Wilson DH, Gazzaniga MS. Beyond commissurotomy: Clues to consciousness. *Neuropsychology* 1979;**2**:543–554.

McMullin E. The impact of Newton's Principia on the philosophy of science. *Philos Sci* 2001;**68**:279–310.

Michel M, Beck D, Block N et al. Opportunities and challenges for a maturing science of consciousness. *Nat Hum Behav* 2019;**3**:104–107.

Michel M, Fleming SM, Lau H et al. An informal internet survey on the current state of consciousness science. *Front Psychol* 2018;**9**:2134.

Odegaard B, Knight RT, Lau H. Should a few null findings falsify prefrontal theories of conscious perception? *J Neurosci* 2017;**37**:9593–9602.

Pautz A. What is the integrated information theory of consciousness? *J Conscious Stud* 2019;**26**:188–215.

Peirce CD, Jastrow J. On small differences in sensation. *Biogr Mem Natl Acad Sci* 1885;**3**:73–83.

Scoville WB, Milner B. Loss of recent memory after bilateral hippocampal lesions. *J Neurol Neurosurg Psychiatry* 1957;**20**:11–21.

Shallice T. Dual functions of consciousness. *Psychol Rev* 1972;**79**:383–393.

Shallice T. The dominant action system: An information-processing approach to consciousness. *The Stream of Consciousness* 1978:117–157.

Sloman A. The emperor's real mind: Review of Roger Penrose's the emperor's new mind: Concerning computers, minds and the laws of physics. *Artif Intell* 1992;**56**:355–396.

Smullyan R. Gödel's incompleteness theorems. In: Goble L, ed. *The Blackwell Guide to Philosophical Logic*. John Wiley and Sons Ltd; 2001:72–89.

Sperry RW. Brain bisection and mechanisms of consciousness. In: John C. Eccles, ed. Brain and Conscious Experience. Springer Verlag; 1965:298–313.

Sperry RW. A modified concept of consciousness. *Psychol Rev* 1969;**76**:532–536.

Tegmark M. Consciousness as a state of matter. *Chaos Solitons Fractals* 2015;**76**:238–270.

Weiskrantz L. *Blindsight: A Case Study and Implications*. Oxford University Press, 1986.