1
Game Plan and Definitions

1.1 An Overshadowed Literature

In the last chapter, I said that my point was to lower expectations. But the wise reader may see that it could backfire too. After all, it has been famously remarked that nothing worth reading has ever been written on consciousness (Sutherland 1989). It may be easy to be conventional and boring, but can any degree of scientific rigor ever be achieved on the topic, really?

Truth is, nothing would please me more than if I ended up inadvertently attracting a bandit of fierce critics to methodically tear my views apart. As a field, we can benefit from having more critics.

Thankfully, though, I do not have to defend the cognitive neuroscience of consciousness all by myself. Over the past couple of decades, a community of active researchers dedicated to doing solid work on the topic has emerged. This work is sometimes overshadowed by more "exciting," revolutionary proposals, especially in the popular media. So I take the opportunity to review the relevant literature here.

However, even within this group of researchers who identify themselves as cognitive neuroscientists, ideas and theories abound. Having many ideas is often a good thing, but they are only useful to the extent that we have enough decisive experiments and quality data to arbitrate between them. Unfortunately, I cannot say that this is currently the case, in part for reasons explained in the last chapter. As such, any attempt at providing a comprehensive review risks producing nothing but a list of "who said what when." I am tempted to do so for diplomatic reasons. But ultimately that would not be particularly useful for the reader. So let me take you through a shortcut instead.

1.2 Global Theories

According to global theories of consciousness, subjective experiences arise when the relevant information is broadcast to many regions in the brain. The philosopher Dan Dennett once likened the phenomenon to "fame in the brain" (1991).

In Consciousness We Trust. Hakwan Lau, Oxford University Press. © Hakwan Lau 2022.
DOI: 10.1093/oso/9780198856771.003.0002

The idea traces back to Bernard Baars's global workspace theory (1989), according to which the brain has specialized "modules," including those for language, long-term memory, motor control, and perception in specific modalities. These modules are informationally encapsulated (Fodor 1983). They mostly mind their own businesses. But now and then, they need to communicate with each other. They can do so by setting up a direct one-to-one contact, which need not reflect consciousness. For example, when you play your favorite fast-paced ball game, your motor control system is probably very much connected to your visual perceptual system (assuming you are any good at it, and you're in the zone). The relevant reflexes are so fast that they may not be fully conscious.

But most of the time, when we are not in such highly rehearsed situations, how the modules access, store, and coordinate information among themselves is not so clear. When you see a person on the street, you don't automatically engage the motor system to reflexively act. Instead, there is probably some central system, in which the relevant information is stored for all modules to access and edit. This central system is likened to a workspace, a hub for exchange of information. According to the theory, information becomes conscious in the brain if and only if it enters this workspace. So when you consciously see someone on the street, your visual perception module puts that information in the workspace for other modules to access. This allows you to talk about it, act on it, remember it, check if it is coherent with what you hear, and what you smell, for example. That is what consciousness involves: the global broadcast and central executive control of information.

Stanislas Dehaene put these ideas into the context of known neuroanatomy and physiology (2014). According to what he calls the global neuronal workspace theory, the relevant mechanisms for consciousness critically depend on activity in the prefrontal and parietal cortices (Figure 1.1), where neurons have long-range connections with many other regions in the brain. This view is supported by ample empirical evidence, as we will see in the next chapters.

Overall, global theories of consciousness, and their very many variants, are endorsed by numerous active research groups (Cohen et al. 2012; Joglekar et al. 2018; Mashour et al. 2020).

1.3 Local Theories

In contrast, we also have local theories, according to which subjective experiences happen when the right kind of neural activity occurs in the relevant sensory modality. Take vision as an example. According to local theories, we

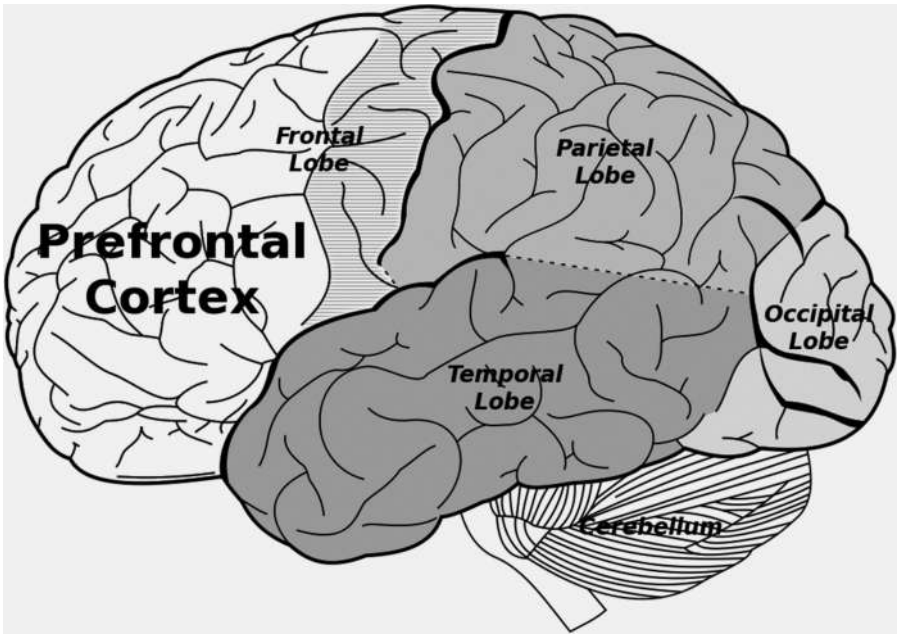

Figure 1.1 Global theories suggest that prefrontal and parietal areas in the brain are causally important for subjective experiences to arise

consciously see something when and only when there is the right kind of activity in the visual cortex. The rest of the brain isn't really critically involved.

Like global theories, there are many flavors here. To some, what constitutes the right kind of activity within the visual cortex depends on the specific brain regions where the activity happens. For example, according to authors like Rafi Malach (Fisch et al. 2009, 2011) and Stephan Macknik and Susana Martinez-Conde (2008), the key regions for visual consciousness are the extrastriate areas, which are visual cortical areas outside of the primary visual cortex (also known as striate cortex or V1). Ultimately, this may also depend on the special visual feature in question; motion and color may depend on different regions (Zeki 2001).

What may also be critical is the dynamics, or the temporal profile, of the activity. For example, Victor Lamme (2003, 2006) argued that what is critical for conscious experience to arise is recurrent activity, first supported by a feedforward wave, for example, from V1 to an extrastriate area (e.g., middle temporal area, MT), and then followed by feedback to V1 (Figure 1.2).

As in many other subfields of research on the neuroscience of perception, studies of vision tend to dominate somewhat. There may be historical reasons for this (Hubel and Wiesel 2004; LeDoux, Michel, and Lau 2020), as well as

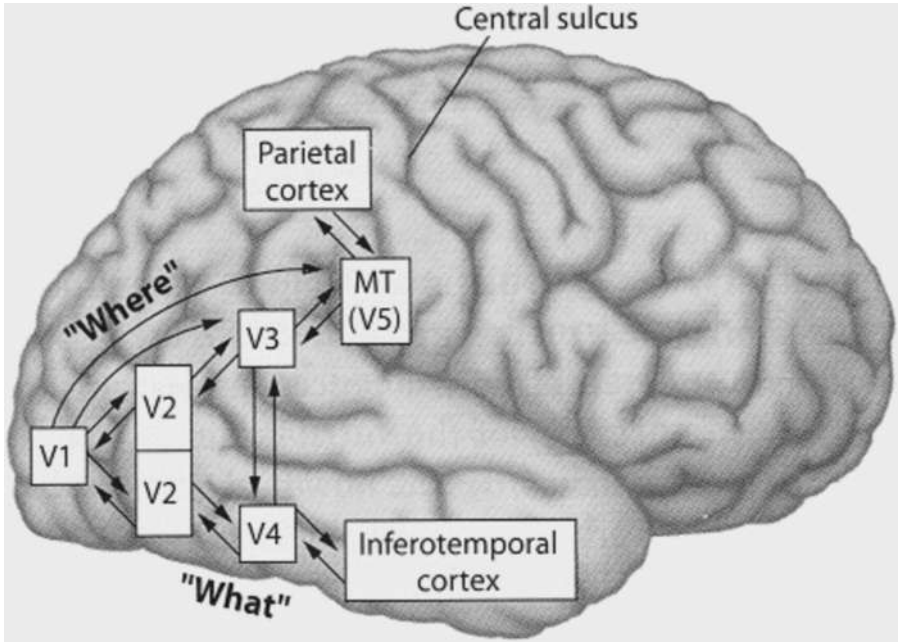

Figure 1.2 Local theories suggest that subjective visual experiences critically depend on specific activity within the visual areas in the brain

considerations of experimental logistics. Some find this unfortunate, and they may be right about that (Smith 2017; Barwich 2020). But regardless, one can think of equivalent ideas in other modalities too. For example, in hearing and touch, there are primary sensory areas in the cortex as well. According to Lamme, feedback to these early cortical areas may also be important. The hope is that once visual consciousness is better understood, the principles derived from this research may generalize more or less to other modalities.

Again, just like global theories, local theories have ample empirical support, as we will see in the next few chapters.

1.4 Theoretical Goal Posts

I mentioned that I would take you through the literature via a shortcut. Here is how: the global and local theories are polar opposites, representing two extreme ends of a theoretical spectrum. By contrasting these two views, we can quickly cover a lot of ground.

As such, the two views are to be treated as somewhat hypothetical guiding points. Like goal posts in a ball game, they work best if they are static. That is

to say, to serve this purpose, I will at times treat each of the family of theories more or less as a singular, stable view. In reality this is often not quite true. Not only are there different versions of global theories, but Dehaene himself has also changed his position on some details over the years (Naccache and Dehaene 2001; Dehaene 2014; King, Pescetelli, and Dehaene 2016), for example. Likewise, Lamme seems not to always insist that feedback to V1 is important; other forms of recurrent activity may also do the job (2016), perhaps.

So there is a risk of misrepresenting these authors. I will try to be as clear as possible in ascribing specific ideas to individual researchers. Beyond that, I have to count on them, along with many other important theorists who are not mentioned in this framework, for understanding. As indicated earlier, my goal here is not to provide a detailed review of all the theories. Rather, it is to summarize the landscape in a gist to orient ourselves. As in any good map, we sacrifice some details. So for each side, we will focus on a representative, prototypical version of the theory. When I say "global theories" or "local theories" I refer to these generic views. They are inspired by the specific authors mentioned in the last sections, but do not necessarily reflect their latest thinking. Once we know the rough orientations, specific versions of their latest views can be better articulated and understood. The rest of the chapter, I hope, will convince the reader of the usefulness of this framework.

Perhaps some theories will fall outside of the spectrum, as they may be considered more extreme than local theories. Not only do they refuse to identify consciousness with some cognitive functions like global theories do, perhaps even the physical substrate proposed may be more abstract than the commonly measured neural activity in a brain region. The substrate may have nothing to do with neurons per se. Perhaps what matters is some general physical properties in the relevant structure. But these are mostly physics-centric theories that are not entirely compatible with the modern language of cognitive neuroscience. As we will see (in Chapters 6, 8, and 9), to the extent that local theories fail, these views will also be in trouble. So we don't need to worry about them too much here.

1.5 The Fine Art of Definitions

Even at this level of convenient abstraction, a tricky conceptual problem arises as we compare the global and local views. Perhaps the two views are different only because they adopt different definitions of consciousness? So they may just be talking past each other?

So far, I have deferred precisely defining the very phenomenon we are after. Some readers may find this odd. Perhaps this should have been done at the very beginning of the book. But issues regarding definitions are sometimes treacherous. They are often better handled after some warming up.

To illustrate the problem, let's change the topic for a moment, to consider the definition of *fish*. In kindergarten, I recall getting upset when my teacher insisted that dolphins are not fish. *But they look like fish, and they swim in the ocean.* Just why was my teacher "correct," and I "wrong"? Turns out, Aristotle actually also classified dolphins as a kind of fish; the kind with lungs (Romero 2012). So one may be tempted to say that my disagreement with my kindergarten teacher was nothing but *a matter of definitions*. She just defined *fish* in a way different from the way I did. We just talked past each other. With Aristotle on my side too, obviously I wasn't so wrong?

While we certainly disagreed on the definition, it doesn't mean that's the end of the argument. Some definitions are better than others. But how to evaluate this is often not so straightforward. In the case of *fish*, modern biologists have decided that it is better to say dolphins are mammals instead. In part, that's because dolphins don't lay eggs, and they don't have scales. But that's hardly the end of the story either. Just why is laying eggs more important as a criterion for being a fish than being able to swim in the ocean? Why doesn't how it looks matter the most?

In the end, biologists decided that a certain taxonomy is better for their purposes. It helps to highlight some facts that are important to them. By adopting their taxonomy, things hang better overall with other pieces of knowledge considered by them to be relevant and established: for example, evolution.

The moral of this story is that definitions are often a matter of ongoing negotiation. At times, they are almost like political debates. They are political in the sense that some definitions serve certain purposes better. But as soon as we talk about purpose, we need to ask: *whose* purpose? Maybe classifying dolphins as mammals fits better with the phylogenetic understanding of the animal kingdom, which in turn allows biologists to make some reliable scientific inductions based on the relevant categorical labels. But my kindergarten self didn't care about that. To my mind then, how it looked was more important. All I needed to know was what belonged to the ocean, rather than the sky or land. I suppose some poets and painters may be on my side too. It may matter little to them what biologists think. But of course, in the end, the biologists had their ways. Collectively, society agreed that they produce more useful knowledge than I did. The poor kid in kindergarten lost the political battle.

1.6 Access Versus Phenomenal

Back to the problem of consciousness. It is a similar situation. To decide what definitions to adopt, we first need to think about what is the relevant purpose: in other words, what is the problem we are trying to solve? This is why, although we didn't talk about definitions, in the introduction, we introduced the "Hard Problem" (Chalmers 1996), that is the challenge of explaining subjective experience in purely mechanistic terms. From there, it should be clear that *if* our goal is to have something meaningful to say about the Hard Problem, what should primarily concern us here would be subjective experience.

By subjective experience, I mean the "raw feels" associated with certain mental processes. Some mental processes are nonconscious, in the sense that they don't feel like anything. We sometimes say there is "*nothing it is like*" to be in those relevant mental states. In fact we mostly don't even realize when such processes are taking place. But some other mental processes are conscious. To consciously see certain things, for example, the color red, involves a certain feel. We sometimes say, there is "*something it is like*" to see red (Nagel 1989). That subjective aspect of the perceptual process is what we are concerned with here. Our overall scientific goal here is to map out the differences between conscious and unconscious mental processes—that is, to figure out why some mental processes are associated with subjective experiences and others aren't.

For subjective experience, other terms I use synonymously include *conscious experience*, *qualitative experience*, *subjective feel*, *raw feel*, *phenomenology*, *phenomenality*, *phenomenal quality*, *phenomenal experience*, *phenomenal consciousness*, and *conscious awareness*. They all mean the same thing. For precision I really should stick to one term only. However, for variety and flow, I sometimes sacrifice absolute precision. Whenever unspecified, *consciousness* refers to this "default" notion of subjective experience, rather than some other notions such as wakefulness or control, which we will discuss in the next two sections.

The philosopher Ned Block famously distinguished phenomenal consciousness from another notion called access consciousness (1995). Access consciousness happens when a relevant piece of information in the brain becomes available for cognition, or for the rational control of action.

Now, this may look like a way to dissolve the debate between global and local theories before it even starts. One could perhaps argue that global theories are really just about access consciousness. And local theories are about phenomenal consciousness. Because these are two different definitions, they are just talking past each other. Of course, the rational control of action may

require the global broadcast of relevant information. But maybe this has nothing to do with how subjective experiences come about. So both theories may be right, without conflicting with each other.

However, this way of thinking assumes that access consciousness may be totally dissociable from phenomenal consciousness. Two definitions can be conceptually different, and yet, in reality, they may just come down to the very same things. For example, *water* can be defined as colorless liquid at room temperature, of certain viscosity, lacking flavor and odor. Or, it can be defined in terms of its precise chemical constituent, H_2O. But they may just end up referring to the very same substance, in this world at least.

Likewise for access versus phenomenal consciousness. Conceptually they sound different enough. But are they really distinct phenomena in the brain? Can we ever have one without the other, entirely? What exactly is subjective experience without the relevant information impacting our reasoning and rational control of action in *any* way? If we truly *feel* pain, how can it not affect our cognition and decision to act at all? If we consciously *see* the color red, how can it not bear any influence on our thinking that there is something red in front of us?

I shall refrain from assuming one way or the other here, regarding the possible dissociation between phenomenal consciousness and access. Chapter 4 will address this as a challenging empirical question. The point here is to say: just because others have defined a notion of consciousness that is allegedly distinct from access does not mean that we have to *accept* the definition. Maybe phenomenal consciousness turns out to always come with at least some degree of access.

Regardless of the empirical outcome, the scientific community may well also come to agree that it is just more *useful* to focus on a notion of subjective experience that isn't entirely distinct from access. There may be some aspects of subjective experience that are distinct from access, but maybe the community would decide that it is really not of our interest. Those who insist on a definition otherwise may end up being like the poor kid in kindergarten who insists that dolphins are fish. How this plays out will depend on our ongoing investigation and negotiation. As we shall see, this will not be trivial at all.

So for now, we will not assume that global and local theories concern distinct phenomena. Although some global theorists sometimes say that their views are about access consciousness (Dehaene et al. 2014), they do not really refer to a kind of consciousness lacking in subjective experience entirely. Subjective experience is what we all really care about, global and local theorists alike. Because the two views are ultimately about the same phenomenon—at least as construed here—they are substantively different theoretical positions.

1.7 Coma Patients and Experimental Confounders

By focusing on subjective experience, we see why some other notions of consciousness are at once highly relevant, and yet not quite useful enough for our scientific purposes.

In everyday life, of course, the common usage of the term *consciousness* mostly has to do with wakefulness: as in, when we have too much (alcohol) to drink, we pass out, and lose consciousness. Patients suffering from traumatic brain injury, such as from car accidents, may also lose consciousness, or even go into a prolonged coma. Likewise, global anesthesia is meant to put people into nonconscious states. Typically we use this notion of consciousness to refer to the individual, or a state that the individual is in. Subjective experiences, on the other hand, are typically associated with specific mental processes occurring in an individual, like the process of visually perceiving something.

But this common notion of consciousness as applied to the general state of the individual is not unrelated to subjective experiences either. When we are unconscious, as in being entirely unawake and unresponsive, we typically cannot enjoy subjective experiences—unless we are in dreams. So *consciousness* in this sense may be defined as having the capacity to have subjective experiences.

Besides having to deal with the exceptional case of dreams, one trouble is that when one is awake with the capacity to have subjective experiences, one is also capable of doing many other things. When one is conscious rather than unconscious, one can remember things, talk about them, think about them, and produce complex behavior. Overall, our brains are presumably processing a lot more information in much more sophisticated ways than when we are unconscious. This is probably true in dreams too, even though we tend not to act out our behavior physically there.

So, if the goal is to scientifically understand the mechanisms for subjective experience, comparing the brain activity of someone conscious against someone who is in a coma would not be so useful. There will be many confounding factors, in the sense that besides having subjective experiences, many other things also differ between the two cases. So let's say if we find that there is more activity in one part of the brain in the conscious over the unconscious individual, we will not know for sure whether this is specifically due to the occurrence of subjective experiences or something else that is also lacking in the unconscious individual (as mentioned previously).

This is why in this book we will not focus too much on coma patients, the state of being in an epileptic seizure, or anesthesia. Understanding these cases has important practical implications. Wonderful experiments have been done on them. But they will not be our empirical starting points, because for our specific purpose of understanding the basic mechanisms for subjective experiences, they suffer from having too many experimental confounders.

This issue of experimental confounders is of central importance. Yet it is often overlooked, even by experts. We will come back to this issue again and again.

1.8 Purposeful Behavior and Experimental Confounders (Again)

Another notion of consciousness, which applies to both the individual, as well as specific mental processes, is purposeful control. When one is fully awake and conscious, one can consciously control one's actions. When one is in a deep coma, one produces no action at all. But in between, there is what Adrian Owen calls the "gray zone" (2019). In such states of semiconsciousness—which can also be achieved by drinking heavily but not *too* heavily to completely pass out—one makes actions that are somewhat routine, as if they aren't under conscious control.

This same notion of consciousness applies to specific mental processes too. Some processes, such as the decision to book a plane ticket through a particular airline to go to a specific destination, tend to come with some sense of volitional control. The individual tends to feel ownership and responsibility for the results of these processes. We say that these decisions are made consciously. Some other processes, on the other hand, may happen relatively quickly and reflexively, such as our attempts to regain balance after almost tripping over a rock on the street. Often, we do not feel that these processes are entirely up to our purposeful control, and we say that the corresponding actions are not fully consciously made.

This notion of consciousness, as in the control of purposeful behavior, may be related to subjective experiences too. Specifically, when applied to mental processes rather than the individual, when we say a process is conscious, it comes with the subjective experience of volitional control, or what is sometimes called a sense of agency. Also, for a process to be conscious in the sense of purposeful control, it is possible that its inputs need to be consciously experienced. That is, when we make consciously controlled actions, we may not actively take into account nonconscious information: that is, information

conveyed by mental processes not associated with any subjective experience. At least, it is likely that in making conscious actions we rely more on conscious rather than nonconscious information.

We will address some of these issues in Chapters 5 and 8. The reasons for deferring them for later is again related to experimental confounders. Conscious and nonconscious actions may well differ in terms of their relationship with subjective experiences, including the very conscious experience of volition. But conscious actions also tend to be more complex, and the corresponding information processes tend to be more powerful and sophisticated. So, by comparing typical conscious and nonconscious actions, we risk having too many experimental confounders, and this would limit what we can learn about the specific underlying processes.

The reader may notice that what is discussed regarding confounders here and in the previous section is somewhat against the spirit of what we discussed in Section 1.6, when we argued that phenomenal and access consciousness may not be empirically dissociated. There, we pointed out that we cannot just define *subjective experience* as having nothing to do with informational access. But if subjective experience turns out to be empirically always linked to such informational access, then how can we consider the latter to be a confounder?

An analogy may help to illustrate this delicate point: in comparing tall people versus short people, we do not say that the length of one's bones is a confounder. That is because being tall *is* to have longer bones. It makes no sense to say, we match the length of all the bones of two individuals, so as to specifically look at their difference in physical height. Likewise, if subjective experience is the very same thing as having global information access—which we suggested in Section 1.6 *may* be the case—then there would be no point in controlling for the latter as a confounder either.

So this issue of confounders is very thorny indeed. Much as we like to think of controlling for confounders as a simple matter of scientific hygiene, the conceptual issues involved are often far from straightforward. Specifically, if we define subjective experience as always having to do with sophisticated information processing, allowing rational access and control, we are automatically loading the dice in favor of global theories. If we define *subjective experience* as having decidedly nothing to do with these sophisticated information processes, we are likewise tilting the table in favor of local theories. This is why we cannot assume one way or the other from the outset. Nor can we end the debate by saying it is just a matter of definitions, followed by a shrug. These issues must be carefully examined on a case-by-case basis, depending on the experiments and phenomena concerned. Each time we set out to control for a confounder, we need to ask: Is this meaningful? Or is it begging the question?

Is it even possible in principle? Ultimately, the experiments need to be convincing, and, unfortunately, plausibility is not always a hard-and-fast objective matter.

1.9 Five Key Issues

Having now cleared the ground about the various conceptual issues and definitions, we can outline the key issues on which we will arbitrate between the global and local theories. As explained earlier, it is best to start with the more straightforward issues and move on from there to the more speculative ones. So the ordering matters here.

The first issue concerns the relevant neural mechanisms, also sometimes called the neural correlates of consciousness (NCC). This may be the most obvious issue because the global and local theories are more or less defined in terms of the NCC. For local theories, the NCC is the activity within the sensory regions of the modality concerned. For global theories, activity outside of the sensory regions is involved. The NCC includes activity in what is sometimes called the "association areas," in the prefrontal and parietal cortices, where neuronal coding doesn't seem to be specific to a single sensory modality.

The second issue concerns the richness of subjective experience. Global theories hold that the content of subjective experience is gated by attentional mechanisms, which is to say, by and large, we only consciously perceive what we are attending to. Therefore, subjective experience is relatively sparse. Outside of attentional focus we do not consciously experience all the details. On the other hand, local theories hold that subjective experience is relatively rich, because capacity limits owing to late-stage processing (e.g., prefrontal broadcast) do not really matter for consciousness. Our subjective experiences are as rich as what early sensory processing can afford.

The third issue is about the functions of consciousness. What are the cognitive advantages of conscious processes, compared to nonconscious processes? Global theories identify consciousness with a powerful cognitive mechanism, the central workspace. Without entering this workspace, the relevant information cannot exercise certain important cognitive functions; for, otherwise, we would not need to have this workspace in the first place. Local theories, on the other hand, make no such commitments. Without having the right kind of activity in the early sensory regions, the information can travel all the way to downstream, late-stage mechanisms without ever becoming conscious. Which is to say, nonconscious processing can be very powerful too. So consciousness may not come with substantive cognitive advantages.

The fourth issue concerns whether other creatures are conscious like we are. That is, we will finally consider the notion of consciousness as applied to an individual. Is consciousness a uniquely human phenomenon? What about very young children? What about primates and smaller animals? On global theories, consciousness ultimately is a higher cognitive mechanism, which some animals may lack. Or at least, like in young children, their global broadcast mechanisms may not be as developed as ours, so even if they were conscious, their capacity for having subjective experiences may be relatively limited. For local theories, once again, these late-stage mechanisms don't matter. Children and some animals may well not have very advanced and developed prefrontal cortices, but this should not limit their conscious experiences.

The fifth issue is similarly controversial, if not more so. It concerns whether machines and robots can ever be conscious. As in the last one, answers to this fifth issue will necessarily be somewhat speculative. If consciousness ultimately is a cognitive mechanism, aligned with what global theories say, one should be able to build the functional equivalent in robots. But this seems to imply that consciousness may already be possible in some current machines, which may seem counterintuitive. Local theories, on the other hand, can hold that the key is having the right kind of biological substrate. This blocks the possibility of consciousness in current robots but offers no principled account as to what makes a biological substrate special.

To summarize, these are the five main questions that we will tackle: 1) Is the NCC global? 2) Is subjective experience sparse rather than rich? 3) Is consciousness important for higher cognitive functions? 4) Is consciousness somewhat limited in young children and primitive animals? 5) Is machine consciousness ever possible? Global theories say *yes* to these five questions. Local theories say *no* to them all. This is why we consider the two views as polar extremes.

1.10 The Need for a Coherent Synthesis

There are of course many other questions one can ask about consciousness. Why focus on these five?

One reason is that I myself actually struggle to come up with other questions of as much contemporary and historical significance as these five, which are at the same time also somewhat tractable at the moment. In part, that's because these questions are logically connected, so there is a factor of synergy.

As we address the first issue regarding the NCC (Chapters 2 and 3), it helps to constrain the answers for the second issue of richness (Chapter 4). That's because if the NCC depends on activity in higher cortical areas (e.g., the prefrontal cortex), then the capacity limits of the relevant late-stage processes may apply. To anticipate, based on the presently available evidence, I will indeed argue that the prefrontal cortex is constitutively involved in the generation of subjective experiences. But the causal role of this involvement may not be global broadcast. As such, it poses some limit to the actual richness of subjective experience. But perhaps it can support an "inflated" sense of richness. That is, the rich details may not be represented as such because the brain may not have the capacity to do so properly. But some mechanisms may exist to fool ourselves, subjectively, that we have these rich details.

So the conscious phenomenology is somewhat rich, but not *really*. This kind of intermediate answer will be a recurring motif. Overall, the empirical evidence is not so kind to either the global or local views.

Likewise, the third issue of functions (Chapter 5) depends somewhat on our take on the first two issues. This is so, especially, because there are tricky methodological issues preventing a clear, direct empirical answer thus far. Given the nature of prefrontal involvement in consciousness, we may expect some functions to be uniquely tied to subjective experiences. But I will argue that these functions are not so general as global theories imply; many high cognitive functions are influenced and controlled by nonconscious information. But consciousness may provide an advantage to some specific functions, such as metacognition and inhibition of some specific process.

As to the fourth and fifth issues, of animals and machine consciousness, they can only be resolved with the help of a theoretical perspective. The earlier "empirical" chapters will be summarized in Chapter 6, which will provide constraints about what a plausible theory should look like. From there, we will outline a view (Chapters 7–9) according to which our brain mechanisms for consciousness may be shared by some mammals. However, some other animals may lack these mechanisms. And yet, in principle, we can build these mechanisms into robots and machines, and make them conscious too. (That is to say, philosophically, I'm a functionalist, as I'll explain in Chapters 6–9.)

Throughout, I will try to be fair to review others' empirical work when they are relevant and decisive. I will no doubt miss many important experiments still, out of sheer ignorance and forgetfulness. The reader will also find that I am evidently biased in favor of reviewing my own work. This is my book after all. So I hope that's okay.

1.11 Theoretical Upshot

I anticipate that some readers will want to jump straight to Chapter 6 for the summary of the earlier empirical reviews. In a way, the chapter was written exactly for this purpose; I appreciate that some people may not have the time to read books from cover to cover. But I do not recommend skipping the earlier chapters, even for the philosophers. The answers are important. But new findings may come along and change what we know. What will remain useful are the concepts and rationale behind the arguments and interpretations.

For similar reasons, I hesitate to give a soundbite summary of my theoretical views here. If this book has a single take-home message, it is that genuine scientific progress requires us to care about the details. My primary purpose here is to review and synthesize a rather large body of literature, not to profess a narrow viewpoint. But I've also been told that readers tend not to go beyond the first chapters of any book, unless they are sufficiently enticed. So here is my best attempt: Based on the discussion of empirical findings, it should become clear that subjective experience is not entirely disconnected from cognition. There are good motivations for not confounding the two, and the global theorists might have been too quick to assume that consciousness is just a form of strong and stable information processing. All the same, even in experiments in which all reasonable controls are carried out, subjective experiences are somehow linked to at least some degree of impact on the cognitive mechanisms in the prefrontal cortex (Chapters 2 and 3). These mechanisms are also needed to account for the subjective richness of experience (Chapter 4), as well as some empirically observed functional advantages of conscious processing (Chapter 5).

As such, a good theory needs to account for this subtle link between consciousness and cognition, without contradicting empirical data. Introspectively, most authors seem to agree that subjective experiences have this so-called here-and-now quality. They present themselves as reflecting the state of the world, or some ongoings in our bodies, *at the current moment*. This seems to be an indispensable property of conscious experiences. When we are in pain, it is difficult not to worry that something bad is happening to us at the relevant bodily location. Even if we are ultimately convinced that nothing really is wrong physically (it may be a "psychic" or illusory pain), it is difficult to shake off the strong tendency to think about that. This potential tendency seems somewhat intrinsic to the experience.

I will therefore propose in Chapter 7 that a conscious perceptual experience requires not just a representation of the sensory content but also a further

representation to the effect that the sensory representation is reflecting the state of the world right now. That is why a perceptual experience is generally not confused with a memory representation of the same content, which we know does not reflect the world *right now*. The experience of a vivid memory recall is supported by a different kind of further representation. But when this further representation is missing altogether, there should be no subjective experience. This explains why sometimes sensory representations alone do not lead to conscious experiences at all (as in conditions like blindsight or aphantasia).

So this further representation is necessary for conscious experiences to occur. We can call this a higher-order representation. It is generated automatically by a *subpersonal* process. That is, we don't have to think hard to come up with this higher-order representation. It's not a thought in that sense. This higher-order representation serves as a tag or label indicating the suitable epistemic status of the sensory representation, and functions as a gating mechanism to route the relevant sensory information for further cognitive processing. Because such further processing is only a potential consequence, but not a constitutive part of the subjective experience, this sets my view apart from global theories. In other words, consciousness is neither cognition nor metacognition. It is the mechanistic interface right between perception and cognition. Current evidence suggests that such higher-order mechanisms likely reside within the mammalian prefrontal cortex, where the functions of perceptual metacognition are also carried out; I will explain why there is such overlap at the physical implementation level.

The local theorist may reject this notion of consciousness, in favor of a definition concerning purely "raw" experiences, with no constitutive connection to cognition whatsoever. Besides empirical evidence, I will survey some broad theoretical considerations, from the clinical and social sciences (Chapter 8). It is also in this context that we can best understand the nature of emotions, culture, rationality, and free will. Ultimately, the local theorists could insist on using whatever definitions they so prefer. But some definitions will not allow us to speak to these important issues of historical and practical interests. One runs the risk of defining oneself into an obscure corner of isolation—just like that poor kid in the kindergarten mentioned in Section 1.5.

To be fair, likewise, we also need to make sure that our theory does not ignore some local theorists' primary concerns about the subjective character or phenomenal quality of experiences. These issues may have been given less weight within the clinical and social sciences, but philosophers have debated about them for centuries. I will argue in Chapter 9 that our theory can account for the qualitative aspects of experience too.

In philosophy, we often say that there is "something it is like" to be a conscious agent enjoying some specific subjective experiences. I take it that the qualitative aspects of an experience can be understood in terms of its similarity relations with respect to all other possible experiences. The complexity of these exact relations accounts for why it may seem so hard to express the subjective quality of a conscious experience in words. But once these relations are "known," the subjective quality is fully determined. This is all there is to having subjective phenomenology. Red looks the way it does because it is subjectively more similar to pink than to blue, more similar to orange than to silver, and so on (with *all* the relevant similarity relations spelled out in exact terms). It looks *redder* than everything else.

I will further argue that because of the way the mammalian sensory cortices are organized, perceptual signals in the brain are spatially "analog" in a specific sense. I will outline the computational advantages for having these representations organized this way. These explain how we likely evolved to have this functional feature of our brains. Given this analog nature, when the higher-order mechanisms discussed herein correctly address a sensory representation, the relevant similarity relations are all implicitly "known." So when a sensory representation becomes conscious, not only do we have the tendency to think that its content reflects the state of the world right now, also determined is *what it is like* to have the relevant experience—in terms of how subjectively similar it is with respect to all other possible experiences. I submit that this addresses the Hard Problem, better than prominent alternative views.

1.12 Chapter Summary

Here we introduced the local and global views, as useful theoretical goalposts. Between the two extremes, there lies a spectrum on which an empirically plausible middle ground can hopefully be found. A plan is set out for reviewing the literature in the coming chapters, going through the five issues of 1) the NCC; 2) richness of experience; 3) functions of consciousness; 4) consciousness in young children and animals; and 5) the possibility of machine consciousness. These will allow us to arbitrate between the global and local views. To anticipate, we will find that neither position works. But we will learn much from the process of understanding their limitations, respectively. Striking a balance is the key.

We also went through four different notions of consciousness: 1) subjective experience, 2) access consciousness, 3) consciousness as the state an individual is in, and 4) consciousness as purposeful control. Subjective experience

is what we will focus on, but it does not mean that the other notions are entirely distinct. They may empirically turn out to be very much related. We will find out.

Above all, we warned ourselves of some treacherous conceptual issues, regarding definitions and confounders. They are anything but straightforward. If not careful, we may inadvertently tilt the table rather unfairly in favor of one side of the spectrum before the competition even begins.

As in the introduction, here we spent a fair bit of time explaining why certain things will not be discussed much further. The next chapter is where the positive scientific journey really begins.

References

Baars BJ. *A Cognitive Theory of Consciousness*. Cambridge University Press, 1989.

Barwich AS. *Smellosophy: What the Nose Tells the Mind*. Harvard University Press, 2020.

Block N. On a confusion about a function of consciousness. *Behav Brain Sci* 1995;**18**:227–247.

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

Cohen MA, Cavanagh P, Chun MM et al. The attentional requirements of consciousness. *Trends Cogn Sci* 2012;**16**:411–417.

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

Dehaene S, Charles L, King J-R *et al.* Toward a computational theory of conscious processing. *Curr Opin Neurobiol* 2014;**25**:76–84.

Dennett DC. *Consciousness Explained*. Little Brown, 1991.

Fisch L, Privman E, Ramot M et al. Neural "ignition": Enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron* 2009;**64**:562–574.

Fodor JA. *The Modularity of Mind*. MIT Press, 1983. https://doi.org/10.7551/mitpress/4737.001.0001.

Hubel DH, Wiesel TN. *Brain and Visual Perception: The Story of a 25-Year Collaboration*. Oxford University Press, 2004.

Joglekar MR, Mejias JF, Yang GR et al. Inter-areal balanced amplification enhances signal propagation in a large-scale circuit model of the primate cortex. *Neuron* 2018;**98**:222–234.e8.

King J-R, Pescetelli N, Dehaene S. Brain mechanisms underlying the brief maintenance of seen and unseen sensory information. *Neuron* 2016;**92**:1122–1134.

Lamme V. *The Crack of Dawn: Perceptual Functions and Neural Mechanisms That Mark the Transition from Unconscious Processing to Conscious Vision*. In: Metzinger T, Windt JM, eds. Open MIND, 22(T). MIND Group, 2016.

Lamme VAF. Why visual attention and awareness are different. *Trends Cogn Sci* 2003;**7**:12–18.

Lamme VAF. Towards a true neural stance on consciousness. *Trends Cogn Sci* 2006;**10**:494–501.

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 U S A* 2020;**117**:6976–6984.

Macknik SL, Martinez-Conde S. The role of feedback in visual masking and visual processing. *Adv Cogn Psychol* 2008;**3**:125–152.

Malach R. Conscious perception and the frontal lobes: Comment on Lau and Rosenthal. *Trends Cogn Sci* 2011;**15**:507; author reply 508–509.

Mashour GA, Roelfsema P, Changeux J-P et al. Conscious processing and the global neuronal workspace hypothesis. *Neuron* 2020;**105**:776–798.

Naccache L, Dehaene S. Unconscious semantic priming extends to novel unseen stimuli. *Cognition* 2001;**80**:215–229.

Nagel T. *The View From Nowhere*. Oxford University Press, 1989.

Owen A. Into the grey zone: Detecting covert conscious awareness in behaviourally non-responsive individuals. *J Neurol Sci* 2019;**405**:2.

Romero A. When whales became mammals: The scientific journey of Cetaceans from fish to mammals in the history of science. In A Romero and EO Keith (eds), *New Approaches to the Study of Marine Mammals.* InTechOpen, 2012. https://www.intechopen.com/chapters/40763.

Smith BC. Human olfaction, crossmodal perception, and consciousness. *Chem Senses* 2017;**42**:793–795.

Sutherland NS. *The International Dictionary of Psychology*. Crossroad Publishing Company, 1989.

Zeki S. Localization and globalization in conscious vision. *Annu Rev Neurosci* 2001;**24**:57–86.