CHAPTER 7

The Search for the Neural Correlates of Consciousness

"The astonishing hypothesis is that 'You,' your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules."
—*Francis Crick,* The Astonishing Hypothesis

"In examining things with care, I seem to have evidently discovered that the part of the body in which the soul immediately exercises its functions is not in any way the heart; nor is it either the whole brain, but only the most internal of its parts, which is a certain very small gland [the pineal gland]"
—*René Descartes,* The Passions of the Soul

"Every remotely reasonable theory of consciousness must … find a role for the brain. That is, no remotely reasonable theory can afford to claim that correlating conscious states with brain states is completely irrelevant to a full understanding of consciousness."
—*Jakob Howhy,* The Search for Neural Correlates of Consciousness

Chapter Overview

7.1. Introduction

Recall the **difference question**: given that mental states can be conscious or unconscious, what's the difference that makes the difference? In Chapter 5 we looked at cognitivist theories of consciousness, which attempt to answer the difference question by identifying a form of cognitive functioning that is present when mental states are conscious and absent when they are not. We looked at several cognitivist proposals, such as mental states are conscious when they are PANIC states, when their content is part of the global workspace (GWT), or when they are the target of higher-order representations (higher-order theory).

Importantly, the difference question is subject to a neural interpretation as well. That is, given that mental states can be conscious or unconscious, what's the *neural* difference that makes the difference? Presumably, there is something happening, somewhere in the brain, when mental states are conscious, that is not happening in the brain when mental states are not conscious. This 'something' is what researchers call a **neural correlate of consciousness**, or NCC for short. A **neural theory of consciousness** attempts to identify and explain the neural correlates of consciousness. Note that we say identify *and* explain. It might be interesting to discover that the brain is always in state *B* when we undergo a conscious experience *C*, but researchers are ultimately interested in explaining *why B* and *C* should be related in this way.

What is the relationship between cognitivist theories of consciousness, of the sort we looked at in Chapter 5, and neural theories of consciousness like the ones we'll introduce in this chapter? Ideally, these two kinds of theories will be complementary and will inform each other. So suppose, for example, that we have compelling reasons to accept some version of the higher-order thought theory, according to which consciousness depends on thought-like representations of one's own mental states. This would direct us to look for the neural basis of consciousness in the prefrontal cortex (PFC), the area of the brain that is thought to be responsible for self-awareness and metacognition. Or suppose that we had compelling reasons to accept the Global Workspace Theory, according to which conscious states belong to a global workspace that various cognitive systems can access. This would direct us to look for the neural correlate of the global workspace—presumably, a brain structure with widespread neural connectivity to other parts of the brain, allowing for access and information-sharing. In this way, cognitive theories of consciousness can influence and guide NCC research.

At the same time, empirical work on the NCC can help us decide between cognitivist theories. So suppose we discovered that consciousness can survive extensive damage to the prefrontal cortex.[1] That would lead us—or should lead us, anyway—to revisit our commitment to the higher-order theory of consciousness. After all, if we were to discover that consciousness can occur in subjects who lack the neural equipment for higher-order representations, then that would undermine higher-order theories of consciousness. Likewise, if we discovered an NCC in a relatively isolated area of the brain, this might cause us to reconsider theories like GWT, which seem to require widely distributed neural connectivity. These examples are just meant to be illustrative. The important point here is that cognitivist and neural theories are not competing theories of consciousness. They are trying to answer the same question—the difference question—at different levels of analysis, and ultimately they should cohere in the sense that the form of cognitive functioning most closely associated with consciousness should be enabled by the neural mechanisms most closely associated with consciousness.

Identifying the NCC could provide a clue as to how consciousness arises in the brain and could eventually lead to a satisfying scientific explanation of consciousness. After all, if we found the NCC to be a particular type of neural activity in a particular area of the brain, we could then think about what it is

1 There is evidence of this, but it is controversial. See Boly et al. (2017) for discussion of the evidence and Lau and Rosenthal (2011) for a response to the evidence on behalf of higher-order theories.

about this type of neural activity, in this part of the brain, that is different from neural activity in other parts of the brain, such that consciousness should light up when it is present.

Aside from its potential theoretical benefits, identifying the NCC could provide practical benefits as well, such as reducing the rate of clinical misdiagnoses. Patients who retain some level of conscious awareness and feeling are sometimes misdiagnosed as being in a vegetative state (VS), a state devoid of awareness and feeling. Some of these patients have a tragic condition called **locked-in syndrome**, a state of wakefulness and awareness coupled with an inability to move or speak. It is natural to assume that such cases are extraordinarily rare, but the authors of one recent study estimate that one in four patients who have severe brain injuries and cannot respond verbally or physically to commands are actually conscious (Nowogrodzki, 2024). If we could reliably determine the presence or absence of consciousness with a brain scan, errors like these could be avoided. Additionally, identifying the NCC could help us settle 'close calls' when it comes to consciousness in early human development and in non-human animals, at least in those with brains roughly like ours.

Given these potential benefits, it is not surprising that the search for the NCC has been a central preoccupation among neuroscientists over the past thirty years. In this chapter we'll introduce this fascinating work. We'll begin with several important distinctions and then describe some of the main methods researchers use for investigating the NCC. We'll then briefly summarize three influential proposals about the NCC, with a focus on what motivates them and why they remain controversial. We'll conclude with some remarks on the philosophical implications of this scientific research program. Supposing we succeeded in identifying the NCC, what would follow? Would philosophical debates between physicalists and dualists be resolved once and for all? Spoiler alert: no, probably not. Dualists believe that consciousness cannot be fully explained at either the cognitive or neural level. On such views, research into neural correlates may still uncover important associations, but it would not by itself explain what consciousness is or why consciousness exists at all.

Koch's Bet

In 1998, the prominent neuroscientist, Christof Koch, made a bet with the prominent philosopher, David Chalmers. Koch bet Chalmers a case of fine wine that within the next 25 years researchers would identify the neural mechanisms responsible for consciousness. At the annual meeting of the

Association for the Scientific Study of Consciousness, held in June 2023 in New York, Koch admitted defeat and paid up. Despite significant progress in understanding what's happening in the brain when we are conscious, and a slew of theories, Koch conceded that we do not yet understand how the brain produces consciousness.

7.2. What Is a Neural Correlate of Consciousness?

As a first pass, we can say that a neural correlate of consciousness is a brain state that is correlated with a state of consciousness. As we will see, though, this initial characterization is ambiguous in various ways. In this section we'll try to clear up some of these ambiguities.

7.2.1. Access-Consciousness vs. Phenomenal Consciousness

As we know from previous chapters, 'conscious state' can mean many different things. One standard distinction—which we introduced in Chapter 1—is between access-conscious states and phenomenally conscious states, or 'a-conscious' and 'p-conscious' states, as philosophers sometimes refer to them. As the term implies, access-conscious states are defined by their *accessibility* to the subject. A mental state is a-conscious if and only if the subject is *conscious of* being in that state, as evidenced, say, by their ability to report to others that they are in that state ("I believe that Montreal is in Quebec," "I see something up ahead," "I'm feeling a pain in my upper back," etc.). Some of the mental states it makes sense to attribute to a person may not be a-conscious in this sense. One very clear but exotic example is **blindsight**, in which people with lesions to their **primary visual cortex** claim that they cannot see stimuli that behavioral tests indicate they are able to see at a subliminal level (see Chapter 5 for a more detailed discussion of this phenomenon). Phenomenally conscious states, on the other hand, are defined not by their accessibility to the subject but by their phenomenal character; there is something that it's like for the subject who is in a p-conscious state. Of course, many mental states are 'conscious' in both senses; the subject may be conscious of being in a state that he is able to report it, and there may be something that it's like to be in that state (reporting the perceptual experience, 'I see a rainbow in the sky,' for example). The distinction is thought to be useful because of the conventional though somewhat controversial view that mental states may be a-conscious without being p-conscious—perhaps your belief that Montreal is in Quebec

is an example of this—and the even more controversial view that mental states can be p-conscious without being a-conscious. Block (2023) calls this ***phenomenological overflow***, the idea being that we can experience more than we can introspectively attend to or report on (perhaps an experience in your periphery). But we won't address these controversies here. What needs to be noted is that when philosophers and scientists refer to 'conscious states' they can mean either states the subject is conscious of being in (a-consciousness) or states with phenomenal character (p-consciousness).

In searching for the neural correlates of consciousness, then, are we searching for the neural correlates of a-consciousness or p-consciousness? The answer is that researchers may be interested in identifying either or both, but it's the neural correlates of p-consciousness that tend to generate the most interest and controversy. The reason for this is the nature of the methods scientists use to identify NCC. Suppose we measure brain activity in subjects who are instructed to introspect and report on what they are experiencing. Since any mental states that subjects can introspect and report on are a-conscious, we have a straightforward way of accumulating data on the neural correlates of a-consciousness. But what if we want to know the neural correlates of p-consciousness—or, to put it somewhat differently, the neural activity responsible for the phenomenal character of any conscious state? Again, we can measure brain activity in subjects who are instructed to introspect and report on what they are experiencing, but this raises the worry that we're tracking the neural correlates of a-consciousness and not p-consciousness, as it can be very difficult to tell whether observed brain activity is associated with phenomenal character or with activities like introspection and verbal report. Ideally, we would be able to set up experiments in which we track brain activity in subjects who (i) we know are having phenomenally conscious experiences, but who are (ii) neither introspecting nor reporting on their experiences. But if subjects are neither introspecting nor reporting on their experiences, how can researchers know that their subjects are having phenomenally conscious experiences? This general obstacle to NCC research is worth keeping in mind.[2]

2 In response to these challenges scientists have devised "no report paradigms," which try to dissociate the neural activity that gives rise to consciousness from the neural activity that enables introspection and report. In other words, the no report paradigm tries to dissociate p-consciousness from a-consciousness. But this experimental approach remains controversial. See Tsuchiya et al. (2015), Overgaard and Fazekas (2016), and Block (2019b) for discussion.

The Brain

You don't have to be a neuroscientist to read and appreciate work on the NCC but some basic knowledge of the brain and its parts is bound to help (see Fig. 5). For example, you should know that our brains are made up (in part) of cells called neurons, which send electrochemical signals back and forth. You should know that the brain consists of four main regions: the cerebral cortex, the cerebellum, the diencephalon, and the brainstem. The cerebral cortex is the largest and outermost part of the brain, includes a right and left 'hemisphere,' and is divided into four 'lobes': frontal, parietal, temporal, and occipital. The diencephalon includes important structures like the thalamus, hypothalamus, and limbic system, while the brainstem, which connects the cerebrum and cerebellum to the spinal cord, is divided, from top to bottom, into the midbrain, pons, and medulla oblongata. Note also that the brain is functionally specialized, meaning that these different areas of the brain are dedicated to supporting different forms of bodily functioning and different forms of information processing.

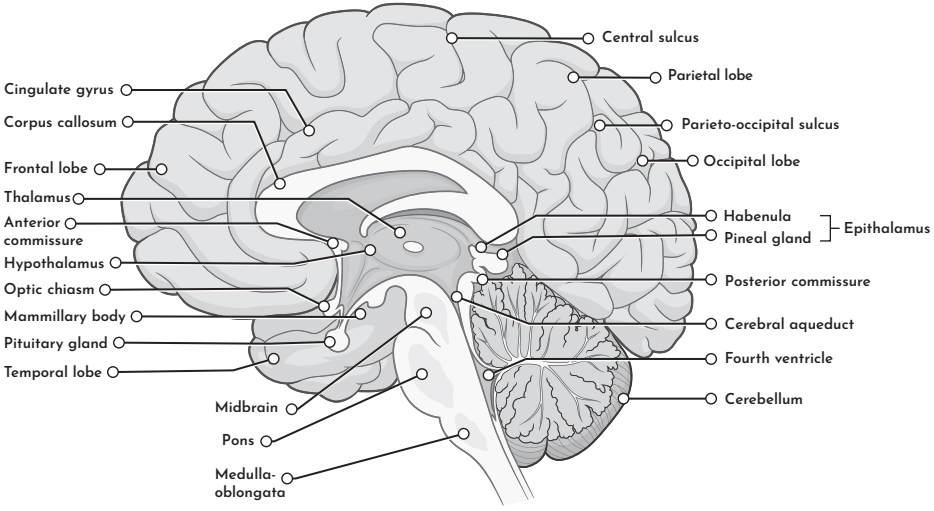

Figure 5
A sagittal view of the human brain.