3
Hitting the Right Note

3.1 Prefrontal Cortex Versus the Rest of the Brain

Why the prefrontal cortex may be particularly important for subjective perceptual experiences was outlined in Chapter 2. In particular, the concerned areas include the dorsolateral prefrontal areas and the frontal polar cortex (Figure 3.1). These regions are identified as important because they seem to have survived the controls for all three kinds of confounders discussed in this volume: stimulus, report, and performance.

However, there have been relatively few studies controlling for the confounders of task-performance capacity. So we have to keep this caveat in mind. Perhaps future studies using methods that are more sensitive may reveal that some other areas, possibly outside of the prefrontal cortex, may survive the control of this confounder too.

At the theoretical level, it is worth emphasizing that global theories do not fixate on these prefrontal areas alone. Other prefrontal regions, such as the anterior cingulate, have also been included in previous discussion (Dehaene 2014; Mashour et al. 2020). Specific parietal areas are densely connected to their lateral prefrontal "counterparts," so, functionally, they often work together (Katsuki 2012). As such, these areas are likely relevant too, according to global theories. So the divide between local and global, really concerns early sensory regions, versus later areas in the association cortex (including both the prefrontal and parietal areas).

Having said that, the prefrontal cortex is often a key focus in current debates. One of the main reasons is that some authors have made strong statements on its exclusion from the neural correlates of consciousness (NCC) (Koch et al. 2016). For example, it has been suggested that the entire frontal lobes, together with the insular, hippocampus, amygdala, and claustrumc—that is, all the areas in the anterior portion of the entire brain—are all not constitutively involved in *any* kind of subjective experience, and not just visual experiences (Koch 2018). Presumably, this includes emotions, sense of volition, conscious recollection of past memories, hunger, interoception, and bodily arousal, for example. This may seem surprising in the light of current

In Consciousness We Trust. Hakwan Lau, Oxford University Press. © Hakwan Lau 2022.
DOI: 10.1093/oso/9780198856771.003.0004

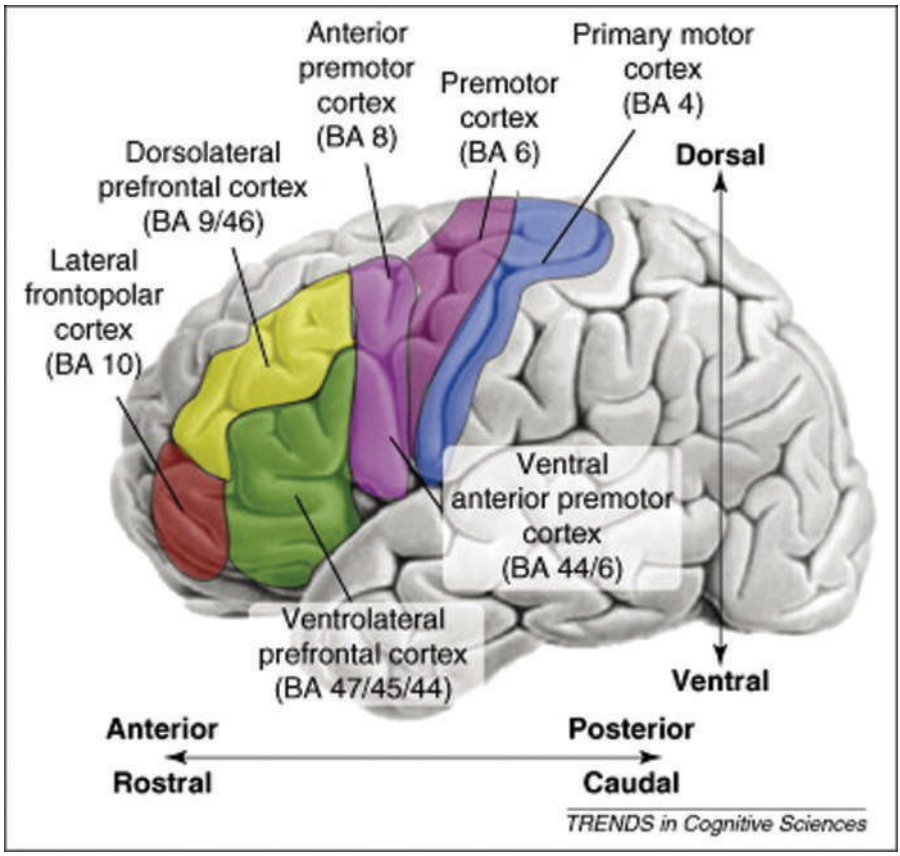

Figure 3.1 Different subregions within the prefrontal cortex

standard textbook knowledge (Gazzaniga 2009; Kandel 2013). These same authors are also not so clear about exactly where the NCC may be. At times they invoke phrases such as *posterior hot zone* or *posterior cortex*. But these are neither standard nor specific anatomical labels. So, in the context of visual awareness, their main agenda seems to be just to rule out other anterior areas. Given the evidence reviewed in Chapter 2, these claims may be puzzling. Here we discuss how such confusion might have come about.

3.2 Conceptual Confusions About Lesions

One main argument against the prefrontal cortex as part of the NCC has to do with lesion effects (Pollen 2008; Koch et al. 2016). Damage to the prefrontal cortex does not lead to functional blindness nor total loss of sense of smell,

taste, touch, and so on. So the prefrontal cortex can't be critically important for subjective experiences, the argument goes.

But how do we account for the activity found in the prefrontal cortex in neuroimaging and invasive recording studies on conscious perception? Interestingly, the authors who make the lesion argument often say that such activity may be due to the reporting confounder (Koch et al. 2016). That is, the activity may be driven by the fact that subjects had to report about what they saw, rather than conscious experience per se. But there seems to be a logical contradiction, as pointed out by Michel and Morales (2019): if the prefrontal activity reflected report rather than subjective experience, shouldn't damages to the area abolish the subjects' ability to make these reports? If so, wouldn't these subjects perform poorly in perceptual tasks, as they fail to make reports? And yet, this is typically not what was found in these patients. As pointed out by these same authors (Koch et al. 2016), the patients can do these tasks very well.

The way to resolve this contradiction is to recognize: lesions do not always abolish the relevant mechanisms causally supported by the region (Jonas and Kording 2017). We have already mentioned this point in Section 2.3 of Chapter 2. This really is standard knowledge not only in cognitive neuroscience, but biology in general. In particular, the notion of degeneracy refers to the fact that in biological systems, multiple parallel mechanisms can support a similar function (Mason 2010). These mechanisms are unlikely to be exactly identical in every aspect, so strictly speaking, this is not to say evolution has given us straight-up redundancy. But all the same, within the context of a specific function, these mechanisms are close enough that they can often serve as "backup" for one another, at least to some extent. So knocking out a substrate for one mechanism may not completely abolish the function. The lack of abolishment of function does not mean that the substrate is causally irrelevant.

Specifically, we expect the level of degeneracy to be high for systems of high complexity, as defined by how different units are connected to each other (Tononi, Sporns, and Edelman 1999). The dorsolateral prefrontal areas are widely connected to many other regions (Petrides and Pandya 2002), and some of their connected areas, such as those in the parietal cortex that, in turn, have rather similar connectivity profiles (i.e., the pattern of connections with respect to all other areas) (Katsuki 2012). So these areas are in a sense central in the entire network, and yet they aren't uniquely irreplaceable. When we look at such a "circuit diagram," we should already expect these areas to show higher levels of degeneracy than others. Recent neuroimaging work has exploited this insight, to predict the severity of lesion effects based on the

connectivity profile of the damaged brain region in the network (Alstott et al. 2009; Lim, Hutchings, and Kaiser 2017).

Theoretical expectations aside, studies have also *empirically* demonstrated that after a unilateral lesion to the prefrontal cortex, areas in the opposite hemisphere can dynamically change their activity to compensate for the effect of the lesion (Voytek et al. 2010).

This is of course not to say all prefrontal functions show this kind of resistance to structural damage. One of the most established neuropsychological findings is that damage to specific areas in the left ventral prefrontal cortex (the so-called Broca's area) can completely abolish language production (Musso et al. 2003). But such highly lateralized function is the exception rather than the rule, as far as the prefrontal cortex is concerned. The exceptional nature of such lesion effects is also predicted by the pattern of anatomical connectivity in the concerned language areas, with the same rationale described earlier.

So we should not expect unilateral lesions to prefrontal areas implicated in previous NCC studies to lead to dramatic "knockout" effects, either on report or subjective experience. For these areas (dorsolateral and frontal pole), the anatomical connections and functions are less lateralized (Croxson et al. 2005). Unlike in language, when one side is damaged, the other may serve as "backup," at least to some extent.

3.3 Controversial Case Studies

But how about bilateral lesions? After such lesions, perhaps there is still the possibility that the connected parietal areas can take over as backup. In monkeys, it has been shown that when both the relevant prefrontal and parietal areas were damaged, the animals indeed behaved functionally blind (Nakamura and Mishkin 1986). In fact, such lesions do not need to be bilateral. Unilateral damage to the left hemisphere, together with damage to the optic tract and the corpus callosum, which prevented visual information from going through the other hemisphere, were sufficient to render the animals functionally blind.

In humans, of course, such carefully planned lesions cannot be performed for experimental purposes. But one may expect that if naturally occurring prefrontal lesions (e.g., because of stroke or external trauma) are bilateral and large, the extent to which the parietal cortices can fully compensate for their functions should be somewhat limited. If the global view is right, we should at least expect some noticeable effect of such large bilateral lesions on subjective experience.

Unfortunately, many false claims have been made in the literature concerning these patient cases. Turns out, in most cases, the lesions were

incomplete. The most critical areas concerned here (e.g., dorsolateral prefrontal cortex) were often spared (Odegaard, Knight, and Lau 2017). In a much discussed case (Patient A), it was reported that the patient behaved just as normal, even after complete bilateral removal of the entire frontal lobes (Boly et al. 2017). But the alleged normal behavior is not consistent with other details reported. Rather, "when anything but the most casual attention is directed upon [the patient], peculiarities in his behavior rapidly become manifest" (Brickner 1936).

More importantly, as the figure from the original report clearly shows, large parts of the prefrontal cortex actually remained intact postmortem (Brickner 1952). So the intended surgical lesion, performed back in the days before the guidance of neuroimaging, was likely inadvertently incomplete. The doctor also referred to the lobectomy as "partial" rather than "complete" (subtitle of Brickner's 1936 book about the patient case). And yet, inaccurate versions of these "stories," sometimes involving anatomical claims grossly incompatible with textbook knowledge, continue to be told through secondary sources in the literature.

For a more in-depth review of these controversies, the readers can see Odegaard et al. (2017), including our reply to Boly et al. (2017) in that exchange. In our paper, there is also a link to the video of a patient with truly complete bilateral lesions to the lateral prefrontal areas, as confirmed by modern magnetic resonance imaging. The video showed the patient's behavior as he was tested by our coauthor Bob Knight. It is evident that the patient was largely unresponsive. To the extent that there were some simple reactions to verbal commands, it is difficult to ascertain if they were mere reflexes. This is not to say we can establish that the patient lacked subjective experiences entirely, or even partially. But the point is exactly that: addressing this question empirically in such cases is extremely difficult.

3.4 Choosing the Right Behavioral Measure

So single patient cases can be controversial. Of course, this is true even in some "classic" and well-documented cases, for example, in studies of amnesia. Often, details are revisited, leading to new interpretations (Stanley and Krakauer 2013; De Brigard 2019). Therefore, it is useful to study the phenomena in a larger cohort of subjects, ideally in settings that are more controlled.

Having tried and failed to find enough suitable patients to be tested, in the mid-2000s, I set out to address the question with a different method: transcranial magnetic stimulation (TMS). Using TMS, under the

right setup, we can try to create so-called "virtual lesions" (i.e., temporary suppression of the neural activity in a brain area). This way we can experimentally induce causal interventions in subjects from the general population.

However, as I was planning the study, several senior colleagues warned me that it probably wouldn't work. I appreciated the advice because they were kindly sharing their own experiences with me. The message was that they had tried, and it just didn't work.

I still believe what they said was true. But my question was: *what* didn't work? With TMS to the prefrontal cortex, we don't expect dramatic effects such as complete abolishment of visual functions. Prefrontal functions are not specific to a sensory modality. If the effects are so strong as to completely abolish the functions, even temporarily, perhaps nobody should participate in these studies! As such, at a level of intensity where the stimulation is safe, we just can't expect the subjects to spontaneously tell us something changed drastically. The effects are unlikely to be so big.

More importantly, if we follow the logic of the experiments described in Chapter 2 (Section 2.11), perhaps what one should expect is not that TMS to the prefrontal cortex would impair visual task performance. Rather, the idea is that it may selectively impair self-reports of subjective experience. That is, subjects may say they see the visual targets less clearly, or maybe they are merely guessing on some trials—while their ability to press keys to correctly discriminate the stimulus may not change much, if at all.

Turns out, visual task performance was indeed what my colleagues had focused on in their earlier attempts. This may sound surprising in this context, but in vision science, we tend to focus on task-performance capacity—not as a confounder, but as the measurement of interest. Sometimes we even go out of our way to use analytic tools such as signal detection theory to *remove* the influence of subjective "biases" to focus on "uncontaminated" effects of performance capacity (Peters, Ro, and Lau 2016). The origin of this tradition is complex, in part because of the different purposes of common experiments. There are also historical reasons. But in any case, since my hope was to find a different kind of effect—of subjective reports of experience rather than discrimination task performance—their warning me that it didn't work in their experiments somehow encouraged me to try it out.

3.5 Stumbled Upon Metacognition

It was my collaborator Elisabeth Rounis who ran the study. Together with John Rothwell and others, their lab pioneered a protocol of stimulation known as

"theta-burst." The details aren't important, but after a minute or so of stimulation, they can make a brain area less responsive for up to an hour (Huang et al. 2005). This is handy because usually you don't want to stimulate both sides of the prefrontal cortex simultaneously, for the risk of inducing seizure. But here, one could stimulate one side, wait a bit, and then stimulate the other side. Afterward, we should expect that both sides of the prefrontal cortex become less responsive.

Just as my colleagues expected, TMS did virtually nothing to the visual task performance; the stimulus intensity required for subjects to achieve a near-threshold performance remained the same before and after TMS. However, TMS changed the correlations between the subjective ratings of visibility and performance. In other words, usually after people reported that they saw the stimulus clearly on a trial, they were more likely to be correct in discriminating what the stimulus was. We can say that this across-trials association between subjective ratings and accuracy measures *metacognition*, following the convention in studies of memory and learning (Rhodes and Castel 2009; Metcalfe and Son 2012). Using this measure, metacognition in visual perception was found to be lowered after TMS (Rounis et al. 2010).

But how big was the effect? Given that TMS did little to discrimination task performance, perhaps we shouldn't expect the effect on visual metacognition to be very large. But quantifying the magnitude of this effect was not trivial. Typical correlation measures do not suffice, as they will be influenced by very many factors not of our interest. Fortunately, I had an extraordinarily brilliant research assistant in my laboratory at Columbia University at that time. Using signal detection theory, Brian Maniscalco developed the measure meta-d', now rather commonly used in studies of this sort (Maniscalco and Lau 2012). With this measure, we can precisely assess how far a subject is away from "ideal" metacognitive efficiency. We found that before TMS, people were nearly perfect. After TMS, this metacognitive efficiency dropped by over 20% (Rounis et al. 2010).

A TMS effect of 20% change in a behavioral measure isn't exactly small. Some have found that the direction of this effect seems to depend subtly on the exact location of stimulation; TMS to the frontal polar area, anterior to where Rounis et al. targeted, boosted rather than impaired metacognition in one study (Rahnev et al. 2016). This is consistent with more recent findings that mechanisms for metacognition likely depend on spatially distributed patterns of activity within the lateral prefrontal and parietal areas (Cortese et al. 2016). We should not think of prefrontal activity as representing the "intensity" of metacognition. The prefrontal cortex is complex, and certainly not a simple signaling device. We will come back to this point in Section 3.11.

Others have challenged that this effect may not be replicable (Bor et al. 2017). But turns out, the alleged nonreplicability was observed only after the authors discarded data that they did not consider "good" enough. If they did not discard the data, a significant positive finding was actually found. I think the intention to focus on "good" data should be applauded, but a computational simulation analysis showed that the criteria chosen for discarding such data do not actually improve the authors' ability to reject false positive results (Ruby, Maniscalco, and Peters 2018). Nor does it improve their statistical power, which was low to begin with, in both the original study and the attempted "replication." At low power, even if the effect is real, the chance of missing it is high. So from a statistical point of view, despite the good intention, it's unclear what was the basis of their claims. And a null result would have been difficult to interpret in any case (because of limited statistical power).

Fortunately, the result that disruption of prefrontal activity can lead to impairment of metacognition has now been observed in studies across many labs and species, including lesions in humans (Fleming et al. 2014), and chemical inactivations (via muscimol injections, which lead to temporary disruptions of neuronal activity in the area) in both monkeys (Miyamoto et al. 2017) and rats (Lak et al. 2014). Martijn Wokke (personal communication) likewise replicated the finding using the same TMS protocol (i.e., theta-burst), targeting the same brain region (i.e., dorsolateral prefrontal area) in humans. Another TMS study has reported that stimulation to this region likely lowered subjective confidence in a visual task, leading to more reported "guesses" (Chiang et al. 2014). Also using TMS, Lapate et al. (2019) found that stimulation to this area impaired metacognition in a perceptual judgment (i.e., whether a face was upright or inverted), but not affective judgment (i.e., concerning the emotions expressed by the face). Targeting slightly different prefrontal areas (i.e., more anterior to dorsolateral prefrontal cortex) with TMS, others have also found selective effects on metacognition in perceptual tasks (Rahnev et al. 2016; Miyamoto et al. 2021).

So the finding seems to hold up. But this is not to say my original study was well-conceived. There were admittedly many flaws. As I mentioned, I did it somewhat out of spite, as a young postdoc trying to show that my senior colleagues were wrong. But thanks to my brilliant collaborators, somehow it worked. This theme is to repeat itself very many times: a rather poorly conceived idea of mine, leading to something empirically replicable and insightful somehow, owing to the sheer luck of having great people working with me (Rounis et al. 2010; Maniscalco and Lau 2012).

3.6 More Windfall: Specific Lesion Effects

To be frank, I was initially not so sure about the TMS result. We set out looking for an effect on subjective experience, based on the neuroimaging studies reviewed in Section 3.5. But we ended up finding an effect on perceptual metacognition. Are the two related at all? In the paper, we said TMS changed "metacognitive awareness." But I have come to admit that it was a bit of a stretch.

Meanwhile, Steve Fleming in London has looked at individual differences in gray matter volume in different brain regions, and found that the people with a larger or denser frontal polar cortex were also better at visual metacognition (Fleming et al. 2010). The frontal polar area in question was just slightly anterior to the dorsolateral prefrontal cortex, the region we targeted in our TMS study (Rounis et al. 2010).

I was intrigued by the finding, but assessing individual differences requires large samples. So I was hoping to test if we could replicate these findings. A very outstanding undergraduate student working in my lab, Liyan McCurdy, took up the task. To make it slightly more interesting, we thought we could compare perceptual metacognition with memory metacognition too. As I mentioned earlier, the idea of treating the correlation between confidence and accuracy as a measurement of metacognition came from studies of memory and learning. So we asked people to study a list of words. Later they were given a pair of words, one new, and one from the list they studied. Subjects had to indicate which was which. After that, they rated confidence. We could use the same analysis method (i.e., meta-d') to assess how close they were to metacognitively ideal performance (i.e., how well their confidence ratings maximally distinguished between their correct and incorrect memory responses).

We confirmed Fleming's finding that individual differences in gray matter volume in the frontal polar area reflected visual metacognition (McCurdy et al. 2013). Memory metacognition was instead reflected by variations in gray matter volume in a medial posterior parietal area known as the precuneus. The two behavioral measures correlated weakly. That is, people who were good at visual metacognition tended to be good at memory metacognition too. But that was explained by the fact that people with larger or denser frontal poles also tended to have a larger or denser precuneus (Figure 3.2). Each type of metacognitive behavior (e.g., memory vs perceptual) seems to have their own structural correlates (e.g., precuneus vs frontal pole).

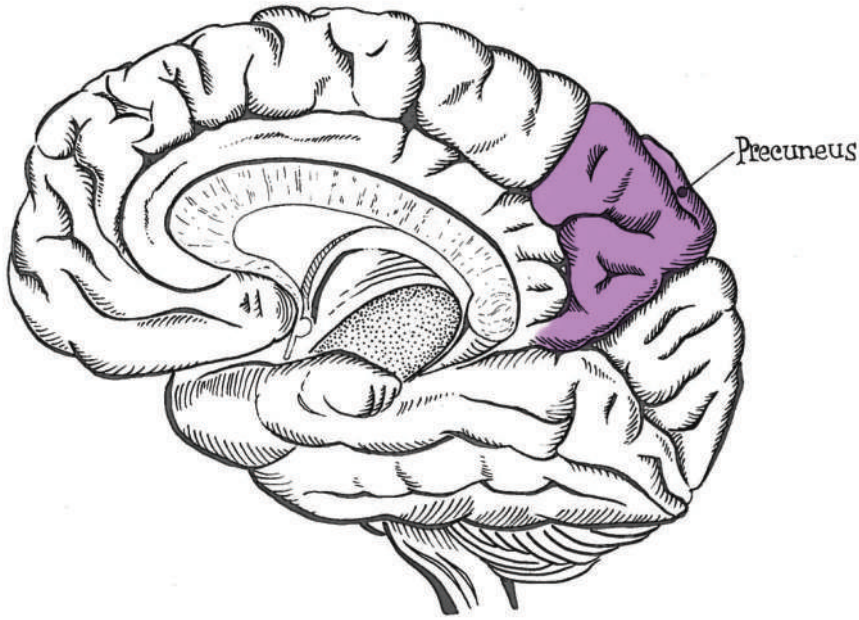

Figure 3.2 A medial parietal region known as the precuenus; besides contributing to memory metacognition, the region seems to also play a role in determining memory vividness (Richter et al. 2016)

Reproduced with permission from Trimble, M. R., and Cavanna, A. E. (2008). The role of the precuneus in episodic memory. *Handbook of Behavioral Neuroscience*, 18, 363-377.

That was all based on statistical analysis, and partly because of that, I confess I did not fully believe the results. One tries to be clean and rigorous about the data collection and analysis, but at the end of the day, I never feel I can trust single studies. The sample size (N = 34) was in line with the standards of the day, but by now we should all recognize its inadequacy. All the same, that's what the data apparently showed. So we wrote up the paper, published it, and hoped to see how it goes (McCurdy et al. 2013).

To my very pleasant surprise, with other colleagues, Steve Fleming later ran a study on patients with unilateral lesions to the prefrontal cortex (2014). Like I mentioned earlier, with unilateral lesions we don't expect very salient effects. But informed by the then recent studies, Steve focused on this same measure of metacognitive efficiency (meta-d') and found that patients with unilateral prefrontal lesions were impaired in visual metacognition by up to about 50%!

To my mind, Fleming et al. (2014) was a landmark study not just because of the magnitude and clarity of the effect. The lesson for me was that perhaps

previous patient studies just did not look for the right behavioral measures. More importantly, Steve found that the subjects were virtually unimpaired in a similar memory metacognition task. This confirmed McCurdy et al.'s prediction (2013), that the prefrontal cortex was primarily important for perceptual, but not memory, metacognition. Also, it means that the patients were not just less able to rate confidence, introspect, or the like. The effect was specific to their ability to monitor the effectiveness of their *perceptual* processes. Whatever is impaired, it seems to have something to do with perceptual experiences specifically.

3.7 A Double Dissociation?

McCurdy et al.'s study (2013) didn't just predict the single dissociation, confirmed by Fleming et al.'s lesion study (2014). If the model supported by the statistical analysis was right, we should also expect that lesions to the precuneus would selectively impair perceptual metacognition more than memory metacognition too.

We never got around to testing this with patients with lesions to the precuneus. But with Sze Chai Kwok's lab, we did a study in which we applied TMS targeting to the area. To my pleasant surprise again, it impaired memory but not perceptual metacognition (Ye et al. 2018), just as McCurdy et al. (2013) predicted.

Despite these findings, it would be dishonest of me to claim that we've solidly proven the double dissociation—that prefrontal interruption only impairs perceptual metacognition, and precuneus interruption only impairs memory metacognition. For example, after researchers injected muscimol to the dorsolateral prefrontal cortex in monkeys, which has the effect of temporarily deactivating the area, the animals were impaired in a memory metacognition task (Miyamoto et al. 2017). Again, the basic primary memory performance was relatively unimpaired. So it was specific to metacognition. But based on McCurdy et al.'s model (2013), one would have expected this to affect perceptual metacognition. And yet a memory metacognition effect was found.

Perhaps one interpretation is that the effect was less selective with chemical inactivations. Because the effect was transient, researchers needed to test the animals immediately after. With lesions, the effect is longer lasting. Testing tends not to be carried out immediately following the brain damage. This may be why lesion studies better reveal what functions can or cannot truly recover in the long-term. For all we know, it is possible that right after the

lesion, Fleming et al.'s prefrontal patients were impaired in both perceptual and memory metacognition too. But over time, memory metacognition recovered while perceptual metacognition didn't. This interpretation is speculative, of course.

3.8 Parallel Versus Hierarchical Architectures

Regardless of the status of the double dissociation, the prefrontal cortex does seem to be causally involved in perceptual metacognition. But is it trivial? *Metacognition* here refers to the monitoring or self-evaluation of an internal process (e.g., perception or memory). Surely, that is done by some "higher-cognitive" areas in the brain (e.g., the prefrontal cortex)?

The findings are not so trivial in several ways. First we have to emphasize that it was somewhat specific to perceptual but not memory metacognition, at least in some studies (McCurdy et al. 2013; Fleming et al. 2014). So it isn't just about self-monitoring or introspection in general. It has something specifically to do with ongoing perceptual experiences.

Also, in the studies reviewed, metacognition often concerns the simple task of giving meaningful confidence ratings. Others have argued that such representations of confidence can be found within the visual cortex (Ma et al. 2006; van Bergen et al. 2015; Walker et al. 2020). So the role of the prefrontal cortex in these simple "metacognitive" tasks is not uncontested.

In fact, the view that perceptual metacognition requires a higher-order monitoring mechanism is contested even by prominent global theorists. Compare the following two different views: in the first view, we can call the perceptual process supporting one's ability to do basic visual tasks (e.g., discriminate or identify the stimulus itself) a first-order process. Let us postulate a second, later-stage process. We can call this a higher-order process, which may monitor the first-order process, in the sense that this later-stage process receives input from the first-order process. Based on this input, the higher-order process evaluates the quality of the first-order process, and thereby generates the metacognitive response (i.e., confidence). This way we can have a selective change in metacognitive efficiency, while holding basic task performance constant; this is achieved by changing the higher-order process alone. We can call this the hierarchical model. See Figure 3.3, right.

An alternative view is: There may be two *parallel* processes instead. They are not in tandem in a hierarchical structure. They are just somewhat independently working side-by-side. One process may contribute to one's ability to do basic visual tasks without informing our metacognition. We can call

this a "nonconscious" process. The other process may contribute to both, which we can call a "conscious" process. This way, we can also have a selective change in metacognitive efficiency, while holding basic task performance constant; this is achieved by changing the balance between the two processes. If the "conscious" process dominates, that is when most of the relevant signals go through it instead of the other process, our metacognition should be well informed. If, on the other hand, most of the signals go through the "nonconscious" process, we can have the same basic task performance overall, but our metacognition may be relatively uninformed. We can call this the parallel model. See Figure 3.3, left.

One of the major proponents of global views, Stan Dehaene, favors the parallel model. This was based on the results from yet another lesion study showing that damages to the prefrontal cortex can impair visual perception (Del Cul et al. 2009). There, unlike in Fleming et al.'s study (2014), basic visual task performance was impaired too. But the impact of prefrontal lesion was more pronounced on the subjective visibility ratings. Dehaene and colleagues wrote out a computational version of the parallel model and fitted it to the patients' behavioral data (see the supplementary materials in Del Cul et al. (2009)). The model accounted for the findings well. The interpretation for these findings is that the "conscious" process is reflected by widespread cortical activities including those in the prefrontal and parietal areas. The "nonconscious" process may be supported by subcortical activities.

In the memory literature, these kinds of parallel models are also popular. In that literature we distinguish between two kinds of memory. When a stimulus remembered from a prior experience is encountered again, we can have conscious episodic recall, where the memory is replayed vividly in our minds.

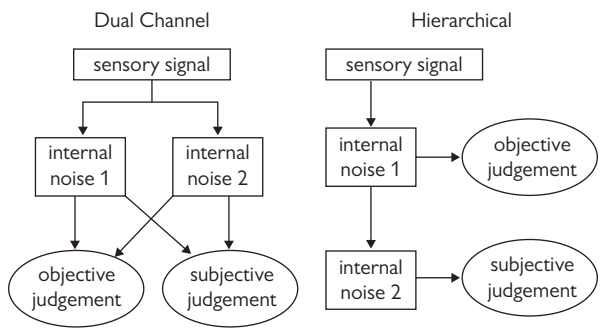

Figure 3.3 Parallel (dual-channel) versus hierarchical models

Alternatively, there may merely be an implicit sense of familiarity. To distinguish between the two kinds of memory responses, one often invokes parallel models akin to the model described herein (Yonelinas et al. 2010). The interpretation is that conscious and nonconscious memories depend on distinct neural pathways.

However, just because a model can fit some data does not mean it is the correct model. There may be other models which can provide a better fit of the data while being just as parsimonious (i.e., simple). So we have to directly compare alternative models quantitatively. This is what Brian Maniscalco did. He computationally implemented many variants of the hierarchical and parallel models. He found that the best-fitting model is hierarchical, not parallel (Maniscalco and Lau 2016). We can interpret this as the late-stage higher-order process being reflected by activity in the prefrontal and possibly also parietal cortices. The first-order process may be reflected better by activity in the visual cortex. This is consistent with many results discussed earlier in this chapter, including the TMS, lesion, and inactivation lesions targeting the prefrontal cortex. A change in the higher-order process mainly changes metacognition but not basic task performance itself.

Therefore, perceptual metacognition seems to be supported by a late-stage mechanism. This is neither trivial nor tautological. It was an empirical finding.

3.9 But Is It Really About Consciousness?

As I hinted at earlier, there is a totally fair criticism for what I have discussed so far. The question we started off with in this chapter is whether the prefrontal cortex plays a causal role in conscious perception. But from the discussion of the findings of Rounis et al. (2010), I have gradually shifted my focus toward metacognition. Based on the results of Fleming et al. (2014), I argue that this is specific to perception but not memory. However, it still is perceptual metacognition, not perception itself. What do these findings have to do with subjective experience?

I acknowledge that we should not treat perceptual metacognition as identical to consciousness at all. I presented the studies here as they happened, in part to highlight that the discovery of the empirical connection is somewhat incidental. Sometimes that's the advantage of doing experiments; the data may tell us unexpected things and stimulate new ideas. Perhaps subjective experience is somewhat linked to metacognition, conceptually. Still, a clear theoretical account is needed to tie it all together. But that has to wait until Chapters 7 and 9, I'm afraid.

For now, perhaps it suffices to say that prefrontal lesions and inactivations do not *only* affect perceptual metacognition. Even if one is not convinced yet that perceptual metacognition has anything to do with consciousness, there are independent reasons to think that the prefrontal cortex is causally relevant somehow.

One finding is what I just described in the Section 3.8: Del Cul et al.'s (2009) lesion study. The experiment was not explicitly about metacognition. Unlike Fleming et al.'s (2014) lesion study, the patients were asked to rate subjective visibility rather than confidence. And yet prefrontal lesions impacted these ratings. Also, as I mentioned, there was a small but positive effect on visual discrimination task performance too.

In the olfactory domain, it has been reported a patient with damage to the orbitofrontal cortex suffered the loss of subjective sense of smell, while having preserved automatic responses to pleasant olfactory stimuli (Li et al. 2010).

In other studies, lesions to the prefrontal cortex impaired the patients' visual detection behavior. These patients showed more conservative bias in detection (Colás et al. 2019). That is, they became less likely to report that they saw the stimulus overall, even if their ability to process the stimuli remained intact. Steve Macknik (personal communication) also observed similar patterns of behavior in a patient with damages to the prefrontal cortex. Another study likewise found that patients with lesions to the dorsolateral prefrontal areas were more likely to miss targets in a visual detection task (Barceló, Suwazono, and Knight 2000).

This is consistent with the finding that TMS also impairs other kinds of detection. For example, change blindness refers to the inability to detect changes that occur in one unexpected part of a large visual scene. In general, subjects are not very good with change detection when the change happens simultaneously with other visual transients such as a flash of a blank screen. But detection of changes was made more challenging when magnetic stimulations were administered to both the prefrontal (Turatto, Sandrini, and Miniussi 2004) and parietal cortices (Beck et al. 2006).

In another study, it was found that TMS to the frontal eye fields *enhanced* visual task performance. Again, the task required detection of a simple stimulus, rather than discrimination (Grosbras and Paus 2003).

Lesions to the prefrontal cortex, especially on the right hemisphere, can also lead to spatial neglect, which refers to patients ignoring stimuli on the left (Karnath and Rorden 2012). Again, this somewhat primarily concerns detection, as patients with neglect seem to have implicit knowledge of stimuli presented on the neglected side (Marshall and Halligan 1988). These symptoms are often associated with lesions to the inferior parietal region, but in fact they

occur after damages to the prefrontal cortex too. Arguably, these effects may tell us more about attention rather than consciousness per se; the relationship between the two will be addressed in the Chapter 4. However, monkeys with lesions to the frontal eye fields seemed to show genuine deficits in detection, even when only a single stimulus was concerned (Latto and Cowey 1971). Lesions to the dorsolateral prefrontal cortex in monkeys also led to more errors in the detection of a single flash of light (Kamback 1973).

What is special about detection tasks? Arguably, they are very much related to metacognition, as suggested by both computational and empirical modeling studies (Ko and Lau 2012; King and Dehaene 2014; Maniscalco, Peters, and Lau 2016; Peters et al. 2017). They are rather unlike discrimination tasks. In a two-choice discrimination task, we only have to compare the evidence in favor of each option. If there is more evidence in favor of one option over the other, we decide that's the correct answer. In detection, it is far less clear where to draw the line to say there is enough evidence for a "yes" answer. That is because a "no" answer is supposed to be signified by a *lack* of evidence. When we lack evidence, we cannot be certain what is going on. Do we lack evidence because the evidence is not there to be found, or just because *we* failed to find the evidence? And just how little evidence is too little? To decide whether a certain level of evidence is enough for detection, we need to have some idea what counts as enough—*typically for oneself under these situations*. That is to say that successful, unbiased detection requires *self-knowledge*.

The last point is not so straightforward. I hope it will become clearer as we move along (especially in Chapters 7–9). But for now, let us at least be clear that disruptions to prefrontal activity do not *only* impair performance in metacognitive tasks. They affect all sorts of other tasks, especially *detection*, which is arguably very much related to awareness: If one truthfully reports "no I do not see something," contra "yes I see something," it would be odd to say there is surely no difference in subjective experience.

3.10 Direct Stimulation

So far we have focused on the effects of disruption of prefrontal activity: lesions, TMS, and chemical inactivations. But how about direct electric stimulation for the purpose of eliciting activity? Wilder Penfield famously applied this technique to different cortical areas, when patients were going through open-head surgery (1958). It was found that stimulations to the sensory areas (e.g., the visual cortex or somatosensory areas) can elicit spontaneous reports of conscious experiences. Such reports were possible because the patients

were kept awake during the surgery; this can help the surgeons map out the functions of the tissues at different locations, via self-reports, to guide the surgery itself.

But what happens when we stimulate the prefrontal cortex? One interesting and well-known finding was about the supplementary motor area, a medial region anterior to the motor cortex. In some cases, when this area was stimulated, patients reported feeling the "urge" to make movements, without actually making them (Fried et al. 1991). This is one of the many reasons why it seems odd to claim that the entire prefrontal cortex isn't constitutively involved in any kind of conscious experience. Certainly, volition, or the experience of motoric intention, is a kind of subjective experience too.

Likewise, stimulation to the orbitofrontal cortex and the anterior cingulate can elicit a wide variety of affective, olfactory, gustatory, and somatosensory experiences (Fox et al. 2018).

But how about other perceptual experiences such as conscious seeing? In particular, the areas concerned here are the lateral and anterior prefrontal regions. Do stimulations to these areas elicit perceptual experiences?

The answer is: rarely, if at all (Raccah, Block, and Fox 2021). There are a few reports that stimulation to the lateral prefrontal areas can elicit spontaneous visual imagery or hallucinations (Bancaud and Talairach 1992; Blanke, Landis, and Seeck 2000; Vignal, Chauvel, and Halgren 2000). But some have argued that these happened only because the stimulation effects spread to early sensory areas. That certainly is possible. Stimulation to an area may have distal effects because brain areas generally do not work in isolation. But once we realize there is such a possibility, how do we know that this did not happen when the early sensory areas were stimulated? How do we know that activity didn't spread to the prefrontal cortex, which ultimately led to the subjective experience?

This issue is not easy to empirically resolve at the moment. The question is not whether stimulating the prefrontal cortex can ever elicit perceptual experiences *at all*. Apparently, it can, at least in some cases. Rather, the right question to ask here may be why is it so much *harder* to elicit such experiences by stimulating the lateral and anterior prefrontal areas compared to the sensory or motor areas? It is admittedly much harder. But the question is why?

One argument could be that some subtle changes in experiences were in fact induced, but the subjects were unable to detect it and to report accordingly. This may sound contrived, but in a way it is exactly what local theorists should accept (Michel and Morales 2019). They claim that the prefrontal cortex is important for attention, access, and report, but not subjective experience per se. So if access is disrupted, there could be unnoticed perceptual changes—a

possibility that localists advocate (as we will see in Chapter 4). Independently, we know that disrupting prefrontal activity can lead to detection failures (as reviewed in the Section 3.9), as well as impairment of metacognitive insight (Section 3.5). So this is not an ad hoc assumption.

I suspect though, limitations of current stimulation protocols are also related. The studies discussed here often involve somewhat arbitrarily chosen stimulation intensity and frequency, which we know works for sensory and motor areas. But we know that the physiology of the prefrontal cortex is different. Maybe in the future we need to design new protocols for the prefrontal cortex tailored for matching the dynamic profiles of its activity, given by concurrent recording from the patient tested. Before we can demonstrate that some suitable prefrontal stimulation protocol can abolish perceptual metacognition as measured in the studies by Rounis et al. (2010) and Fleming et al. (2014), perhaps there just isn't much point in debating whether it could elicit vivid perceptual experiences. In those studies, sensitive psychophysical methods were employed to look for subtle effects at near-threshold. If the stimulation just isn't powerful enough to change the relevant measures under those circumstances, it means that the relevant mechanisms just aren't engaged by the stimulation. In this case, naturally we don't expect changes in subjective phenomenology so salient to be spontaneously reported.

Given that the first argument of undetected change in phenomenology already suffices logically as a reply to the localists, why do I emphasize on this second consideration of potentially limited efficacy of stimulation too? Let me explain with the following analogy.

3.11 Pianos and Trumpets

The piano is a beautifully engineered musical instrument. The keys are arranged spatially, so that the lower notes are elicited by the keys on the left and the higher notes on the right. There is a systematic, one-to-one mapping between keys and notes following a clear logic. Just by watching another person play the piano, a total beginner can already appreciate the spatial layout. If the task is to elicit a single sound of a certain pitch, one may have some success figuring it out with a few trial and error attempts. Even toddlers can make some sounds on the piano.

Contrast this with another wonderful instrument, the trumpet. None of the three "buttons" on the instrument uniquely map to a single note. Some notes can be played with different fingerings. Airflow and the "embouchure" are both critical, and yet neither are easily observable by the audience. One would

have a hard time figuring out how to play a simple scale just by watching someone else play. Beginners sometimes fail to make any sound at all on first attempts.

But both instruments can make great music. Arguably, one generates the sound more directly when playing the trumpet. On the piano, the keys could have been mapped to the notes in a totally different way, had it been engineered differently. The moral is: the sheer ease of triggering a desired effect does not necessarily tell us the full story of the underlying functions and mechanisms.

The sensory cortices are somewhat like the piano. The visual cortex has a spatial layout known as retinotopy. There is a spatial isomorphism, meaning that two points close in space on the retina tend to trigger activity from neurons that are also spatially close to each other in the visual cortex. So there is in effect a map. This spatial map is also found in the somatosensory areas, where there is a logical and isomorphic representation of the body, sometimes known as the homunculus (a little person) (Figure 3.4). Much of that

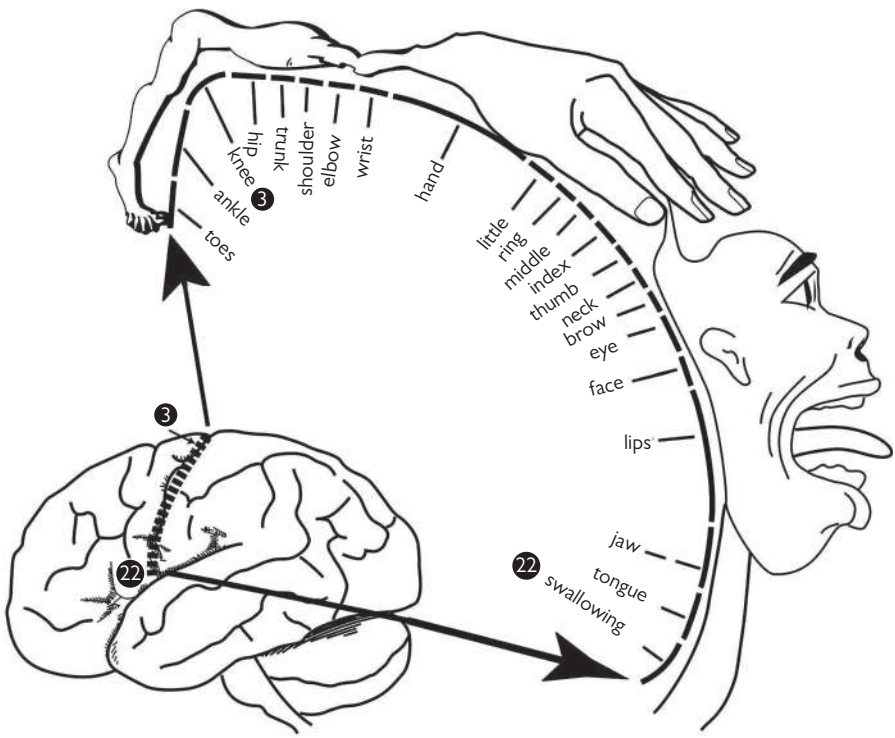

Figure 3.4 Penfield's homunculus: a map of somatosensory representations of different parts of the body in the brain

was detailed in Penfield's work, using the method of direct electric stimulation mentioned in Section 3.10.

Another feature of the sensory cortices is sparse coding, which refers to the fact that at one time, there are relatively few active neurons (Olshausen and Field 2004). To represent something such as an oriented line segment, only a small number of neurons need to fire. This architecture is also sometimes called "labeled lines," as if neurons each carry a label (e.g., "I signal oriented lines at a 42-degree right tilt, for this particular retinal location" or "I signal the face of Barack Obama in a left-facing profile"). This is of course an over-simplification. Things are slightly more complex, where multiple neurons often code similar things, and one neuron is typically involved in multiple representations. But all the same, this describes the basic logic of how things work in the sensory cortices to some extent.

This general pattern very much breaks down in the prefrontal cortex. There, neurons have much larger receptive fields, meaning that they are not so spatially specific. Sometimes they do not concern spatial locations at all. Also, many neurons fire to the same stimuli, only to varying degrees; and the same neuron also seems to respond to many different things (Fusi, Miller, and Rigotti 2016). To read out from the neuronal activity what the subject is perceiving, one often has to aggregate information from many neurons. The lines aren't so clearly "labeled." One needs to use advanced computational methods to "decode" the content (Mante et al. 2013; Rigotti et al. 2013). Instead of providing simple signals for the presence of specific external stimuli, much more complex computations seem to be carried out in the prefrontal cortex.

Now, who wants to complain that learning to play the trumpet is hard? Of course it is, and so is the prefrontal cortex. But we shouldn't write off something just because it is complicated, especially when we are studying something as complicated as consciousness. This is not to say I encourage the reader to accept things which we do not understand either. In Chapters 7 and 9, we will try to understand better the role of the prefrontal cortex in consciousness at a theoretical level.

3.12 Chapter Summary

A famous fable has it that one night a drunk man looked for his keys on the street (Kaplan 1964). Turns out, he lost them in a park far away. When asked why he didn't look for the keys where we lost them, he replied: "But the streetlights are here!"

Vision neuroscientists love the visual cortex. In large part that's because the clear anatomical organization allows for precise physiological measurements to obtain strong experimental effects. Though I too work with vision scientists, I was initially trained in a "prefrontal lab" (Lau et al. 2004). The differences in scientific culture and expectations are often underappreciated. But just because some measurements are more challenging to obtain robustly shouldn't mean we write off the subject altogether. Somebody needs to study the more difficult, messier things.

Here I have given concrete arguments about why criticisms against the role of the prefrontal cortex in consciousness are problematic. They are conceptually unsound, when we reflect on what we really mean by the NCC, and on biological principles such as degeneracy. Given these, we should not expect to find strong effects of disruption of prefrontal activity. Despite these caveats, there is actually considerable positive evidence. These empirical findings are just falsely written off sometimes.

This concludes our first issue, regarding the NCC: the local theorists don't seem to be quite right. But the view that the NCC reflects a global broadcast may also not be entirely convincing either. In Chapter 2 we summarized that the neural signature for subjective experience may be somewhat subtle in the prefrontal cortex. Here we should also concede that the effects of lesions and stimulation are often modest. When they were found, they mostly concerned detection and metacognition. It is true that more salient effects affecting perception in general can be found in the early sensory areas. But those may be due to performance-capacity confounders (see Chapter 2). This is why we continue to pay so much attention to the prefrontal cortex, despite these relatively subtle findings—they are more specific for our purpose of understanding consciousness.

In Chapter 4, we will move on to the next issue, the relationship between attention and consciousness, where we may likewise find an intermediate answer.

References

Alstott J, Breakspear M, Hagmann P et al. Modeling the impact of lesions in the human brain. *PLOS Comput Biol* 2009;**5**:e1000408.

Bancaud J, Talairach J. Clinical semiology of frontal lobe seizures. *Adv Neurol* 1992;**57**:3–58.

Barceló F, Suwazono S, Knight RT. Prefrontal modulation of visual processing in humans. *Nat Neurosci* 2000;**3**:399–403.

Beck DM, Muggleton N, Walsh V et al. Right parietal cortex plays a critical role in change blindness. *Cereb Cortex* 2006;**16**:712–717.

van Bergen RS, Ma WJ, Pratte MS et al. Sensory uncertainty decoded from visual cortex predicts behavior. *Nat Neurosci* 2015;**18**:1728–1730.

Blanke O, Landis T, Seeck M. Electrical cortical stimulation of the human prefrontal cortex evokes complex visual hallucinations. *Epilepsy Behav* 2000;**1**:356–361.

Boly M, Massimini M, Tsuchiya N et al. Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *J Neurosci* 2017;**37**:9603–9613.

Bor D, Schwartzman DJ, Barrett AB et al. Theta-burst transcranial magnetic stimulation to the prefrontal or parietal cortex does not impair metacognitive visual awareness. *PLOS One* 2017;**12**:e0171793.

Brickner RM. *The Intellectual Functions of the Frontal Lobes: A Study Based Upon Observation of a Man After Partial Bilateral Frontal Lobectomy.* Macmillan, 1936.

Brickner RM. Brain of patient A. after bilateral frontal lobectomy: Status of frontal-lobe problem. *AMA Arch Neurol Psychiatry* 1952;**68**:293–313.

Chiang T-C, Lu R-B, Hsieh S et al. Stimulation in the dorsolateral prefrontal cortex changes subjective evaluation of percepts. *PLOS One* 2014;**9**:e106943.

Colás I, Chica AB, Ródenas E et al. Conscious perception in patients with prefrontal damage. *Neuropsychologia* 2019;**129**:284–293.

Cortese A, Amano K, Koizumi A et al. Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. *Nat Commun* 2016;**7**:13669.

Croxson PL, Johansen-Berg H, Behrens TEJ et al. Quantitative investigation of connections of the prefrontal cortex in the human and macaque using probabilistic diffusion tractography. *J Neurosci* 2005;**25**:8854–8866.

De Brigard F. Know-how, intellectualism, and memory systems. *Philos Psychol* 2019;**32**:719–758.

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

Del Cul A, Dehaene S, Reyes P et al. Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain* 2009;**132**:2531–2540.

Fleming SM, Ryu J, Golfinos JG et al. Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain* 2014;**137**:2811–2822.

Fleming SM, Weil RS, Nagy Z et al. Relating introspective accuracy to individual differences in brain structure. *Science* 2010;**329**:1541–1543.

Fox KCR, Yih J, Raccah O et al. Changes in subjective experience elicited by direct stimulation of the human orbitofrontal cortex. *Neurology* 2018;**91**:e1519–e1527.

Fried I, Katz A, McCarthy G et al. Functional organization of human supplementary motor cortex studied by electrical stimulation. *J Neurosci* 1991;**11**:3656–3666.

Fusi S, Miller EK, Rigotti M. Why neurons mix: High dimensionality for higher cognition. *Curr Opin Neurobiol* 2016;**37**:66–74.

Gazzaniga MS. *The Cognitive Neurosciences*. MIT Press, 2009.

Grosbras M-H, Paus T. Transcranial magnetic stimulation of the human frontal eye field facilitates visual awareness. *Eur J Neurosci* 2003;**18**:3121–3126.

Huang Y-Z, Edwards MJ, Rounis E et al. Theta burst stimulation of the human motor cortex. *Neuron* 2005;**45**:201–206.

Jonas E, Kording KP. Could a neuroscientist understand a microprocessor? *PLOS Comput Biol* 2017;**13**:e1005268.

Kamback MC. Detection of brief light flashes by monkeys (Macaca nemestrina) with dorsolateral frontal ablations. *Neuropsychologia* 1973;**11**:325–329.

Kandel E. *Principles of Neural Science*, fifth edition. McGraw Hill Professional, 2013.

Kaplan A. *The Conduct of Inquiry: Methodology for Behavioral Science*. Routledge, 1964.

Karnath H-O, Rorden C. The anatomy of spatial neglect. *Neuropsychologia* 2012;**50**:1010–1017.

Katsuki F. Unique and shared roles of the posterior parietal and dorsolateral prefrontal cortex in cognitive functions. *Front Integr Neurosci* 2012;**6**. https://doi.org/10.3389/fnint.2012.00017.

King J-R, Dehaene S. A model of subjective report and objective discrimination as categorical decisions in a vast representational space. *Philos Trans R Soc Lond B Biol Sci* 2014;**369**:20130204.

Koch C. What is consciousness? *Nature* 2018;**557**:S8–S12.

Koch C, Massimini M, Boly M et al. Neural correlates of consciousness: progress and problems. *Nat Rev Neurosci* 2016;**17**:307–321.

Ko Y, Lau H. A detection theoretic explanation of blindsight suggests a link between conscious perception and metacognition. *Philos Trans R Soc Lond B Biol Sci* 2012;**367**:1401–1411.

Lak A, Costa GM, Romberg E et al. Orbitofrontal cortex is required for optimal waiting based on decision confidence. *Neuron* 2014;**84**:190–201.

Lapate RC, Samaha J, Rokers B et al. Perceptual metacognition of human faces is causally supported by function of the lateral prefrontal cortex. *Commun Biol* 2019;**3**:360. https://doi.org/10.1038/s42003-020-1049-3.

Latto R, Cowey A. Visual field defects after frontal eye-field lesions in monkeys. *Brain Res* 1971;**30**:1–24.

Lau HC, Rogers RD, Ramnani N et al. Willed action and attention to the selection of action. *Neuroimage* 2004;**21**:1407–1415.

Lim S, Hutchings F, Kaiser M. Modeling the impact of lesions in the brain. In: Ooyen AV, Butz-Ostendorf M eds. *The Rewiring Brain* Academic Press, 2017:465–484.

Li W, Lopez L, Osher J et al. Right orbitofrontal cortex mediates conscious olfactory perception. *Psychol Sci* 2010;**21**:1454–1463.

Maniscalco B, Lau H. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious Cogn* 2012;**21**:422–430.

Maniscalco B, Lau H. The signal processing architecture underlying subjective reports of sensory awareness. *Neurosci Conscious* 2016;**2016**. https://doi.org/10.1093/nc/niw002.

Maniscalco B, Peters MAK, Lau H. Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Atten Percept Psychophys* 2016;**78**:923–937.

Mante V, Sussillo D, Shenoy KV et al. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 2013;**503**:78–84.

Marshall JC, Halligan PW. Blindsight and insight in visuo-spatial neglect. *Nature* 1988;**336**:766–767.

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

Mason PH. Degeneracy at multiple levels of complexity. *Biol Theory* 2010;**5**:277–288.

Ma WJ, Beck JM, Latham PE et al. Bayesian inference with probabilistic population codes. *Nat Neurosci* 2006;**9**:1432–1438.

McCurdy LY, Maniscalco B, Metcalfe J et al. Anatomical coupling between distinct metacognitive systems for memory and visual perception. *J Neurosci* 2013;**33**:1897–1906.

Metcalfe J, Son LK. Anoetic, noetic, and autonoetic metacognition. In: Beran MJ, Brandl J, Perner J, Proust J eds. *Foundations of Metacognition*. OUP Oxford, 2012:289–301.

Michel M, Morales J. Minority reports: Consciousness and the prefrontal cortex. *Mind Lang* 2019;**35(4)**:493-513. https://doi.org/10.1111/mila.12264.

Miyamoto K, Osada T, Setsuie R et al. Causal neural network of metamemory for retrospection in primates. *Science* 2017;**355**:188–193.

Miyamoto K, Trudel N, Kamermans K et al. Identification and disruption of a neural mechanism for accumulating prospective metacognitive information prior to decision-making. *Neuron* 2021;**0**. https://doi.org/10.1016/j.neuron.2021.02.024.

Musso M, Moro A, Glauche V et al. Broca's area and the language instinct. *Nat Neurosci* 2003;**6**:774–781.

Nakamura RK, Mishkin M. Chronic "blindness" following lesions of nonvisual cortex in the monkey. *Exp Brain Res* 1986;**63**:173–184.

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

Olshausen B, Field D. Sparse coding of sensory inputs. *Curr Opin Neurobiol* 2004;**14**:481–487.

Penfield W. Some mechanisms of consciousness discovered during electrical stimulation of the brain. *Proc Natl Acad Sci U S A* 1958;**44**:51–66.

Peters MAK, Ro T, Lau H. Who's afraid of response bias? *Neurosci Conscious* 2016;**2016**:niw001.

Peters MAK, Thesen T, Ko YD et al. Perceptual confidence neglects decision-incongruent evidence in the brain. *Nat Hum Behav* 2017;**1**. https://doi.org/10.1038/s41562-017-0139.

Petrides M, Pandya DN. Association pathways of the prefrontal cortex and functional observations. In: Stuss DJ, Knight RT eds. *Principles of Frontal Lobe Function*; OUP Oxford, 2002:31–50.

Pollen DA. Fundamental requirements for primary visual perception. *Cereb Cortex* 2008;**18**:1991–1998.

Raccah O, Block N, Fox KCR. Does the prefrontal cortex play an essential role in consciousness? Insights from intracranial electrical stimulation of the human brain. *J Neurosci* 2021;**41**:2076–2087.

Rahnev D, Nee DE, Riddle J et al. Causal evidence for frontal cortex organization for perceptual decision making. *Proc Natl Acad Sci U S A* 2016;**113**:6059–6064.

Rhodes MG, Castel AD. Metacognitive illusions for auditory information: effects on monitoring and control. *Psychon Bull Rev* 2009;**16**:550–554.

Richter FR, Cooper RA, Bays PM et al. Distinct neural mechanisms underlie the success, precision, and vividness of episodic memory. *Elife* 2016;**5**. https://doi.org/10.7554/eLife.18260.

Rigotti M, Barak O, Warden MR et al. The importance of mixed selectivity in complex cognitive tasks. *Nature* 2013;**497**:585–590.

Rounis E, Maniscalco B, Rothwell JC et al. Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cogn Neurosci* 2010;**1**:165–175.

Ruby E, Maniscalco B, Peters MAK. On a "failed" attempt to manipulate visual metacognition with transcranial magnetic stimulation to prefrontal cortex. *Conscious Cogn* 2018;**62**:34–41.

Stanley J, Krakauer JW. Motor skill depends on knowledge of facts. *Front Hum Neurosci* 2013;**7**:503.

Tononi G, Sporns O, Edelman GM. Measures of degeneracy and redundancy in biological networks. *Proc Natl Acad Sci U S A* 1999;**96**:3257–3262.

Turatto M, Sandrini M, Miniussi C. The role of the right dorsolateral prefrontal cortex in visual change awareness. *Neuroreport* 2004;**15**:2549–2552.

Vignal JP, Chauvel P, Halgren E. Localised face processing by the human prefrontal cortex: Stimulation-evoked hallucinations of faces. *Cogn Neuropsychol* 2000;**17**:281–291.

Voytek B, Davis M, Yago E et al. Dynamic neuroplasticity after human prefrontal cortex damage. *Neuron* 2010;**68**:401–408.

Walker EY, Cotton RJ, Ma WJ et al. A neural basis of probabilistic computation in visual cortex. *Nat Neurosci* 2020;**23**:122–129.

Ye Q, Zou F, Lau H et al. Causal evidence for mnemonic metacognition in human precuneus. *J Neurosci* 2018;**38**:6379–6387.

Yonelinas AP, Aly M, Wang W-C et al. Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus* 2010;**20**:1178–1194.