5
What Good Is Consciousness?

5.1 Evolution

Most of us agree that consciousness emerged at some time through our evolutionary history. Many also find probable that it is a relatively recent invention; simple organisms like bacteria are presumably not conscious. But how about insects, fish, birds, rodents, and the like? We will address this question more fully in Chapter 7.

The point here is: even if we accept this evolutionary outlook, it does not mean that consciousness is selected *because* it provides unique functional advantages for our survival—not necessarily.

That's because evolution is a long and complex process. Take a "uniquely human" feature as an example: having chins, that structure sticking out by the edge of our lower jaw. Other mammals do not have it, not even the nonhuman primates. Is it because it allows us to speak? Or is it to protect our relatively fragile throats in melee combats? Or maybe it has something to do with bipedalism—maybe human toddlers fall on their faces so often that they need a bony jaw structure to protect them from damage? It may be amusing to try to come up with stories like these. But there is just no such simple answer (Holton et al. 2015).

Likewise, consciousness could also have evolved mostly as a byproduct (Robinson, Maley, and Piccinini 2015) or for reasons that we do not intuitively expect. Like others, I too find it hard to believe that it really has no function whatsoever. But the point is not that there's decidedly no function. The point is that the answer may not be so obvious from the outset. Beyond "just so" stories, we need direct empirical evidence. Unfortunately, as I will explain, so far most of the evidence we have isn't very strong.

5.2 Libet & Volition

Benjamin Libet is often credited for having challenged the existence of free will from a neurophysiological perspective (Libet et al. 1983). But his famous study

In Consciousness We Trust. Hakwan Lau, Oxford University Press. © Hakwan Lau 2022.
DOI: 10.1093/oso/9780198856771.003.0006

was in part based on an older finding: Kornhuber and Deecke (1965) showed that there was electrophysiological activity preceding simple spontaneous movements. This activity, called the *readiness potential*, was measured by averaging electroencephalograms recorded from the scalp. What was surprising was that this activity can start as early as a whole second before the movement.

Libet reasoned that the early onset may be due to the lack of complete spontaneity on the subjects' parts. But even after discarding blocks of trials in which they reported any recollection of inadvertent "preplanning," the readiness potential was still detectable at up to about half a second prior to movement. That seemed odd: for a truly spontaneous act, it doesn't seem to require that much time to prepare. Confirming this, Libet and colleagues showed that people only reported consciously initiating the movement about a quarter of a second before actual execution. So the brain seems to start nonconsciously initiating the action prior to that.

Perhaps it isn't so bad that our conscious intentions aren't the "first unmoved movers." We should accept that maybe our actions aren't so truly *de novo*. Although they may seem spontaneous to our conscious selves, they may have nonconscious origins too. But if consciousness arises some time before movement execution, perhaps it means we can consciously edit our actions before it's too late. Possibly, we are still a part of the causal chain.

But of course, this would not be possible if the action was already fully determined and predictable at the early stage of nonconscious initiation, before we become consciously aware of the intention. In that case, by the time our conscious intention arises, it would be too late. In the literature, sometimes the words *determined* and *predicted* were used in this context, giving this impression (e.g. Soon et al. 2008). But in fact, the nature of the action at those early nonconscious stages can only be very weakly predicted statistically. So perhaps the conscious intention that arises later can still play a major causal role.

Others wonder if the early nonconscious reflects anything specific to the action at all. Perhaps it is just some general "noise" in the brain. But single cell physiological studies in monkeys have shown that this is unlikely (Romo and Schultz 1987). Prior to spontaneous actions, some neurons within the motor systems fire up to over 2 seconds before the movement onset. These neurons code specific properties of the movements.

So, instead of being driven by nonspecific motor "noise," the readiness potential probably reflects the fluctuation of specific preparatory activity, to be detected consciously when it passes a certain threshold (Nikolov, Rahnev, and Lau 2010; Schurger, Sitt, and Dehaene 2012). This may allow some half-formed, and yet specific, motor plans to reach our awareness, for potential "editing" prior to action execution.

5.3 Free Will as an Illusion

As such, the question seems open regarding whether our conscious intentions play a causal role. When I was a PhD student in Dick Passingham's lab, I set out to test this question. Using functional magnetic resonance imaging (fMRI), we first localized where the representations of conscious intention may be. Like the direct brain stimulation patient study mentioned in Section 3.10 (Fried et al. 1991), we linked intention to a medial prefrontal area called the presupplementary motor area (Lau et al. 2004; Lau, Rogers, and Passingham 2006). This also is roughly the same area from which they recorded single cell activities in the study in monkeys mentioned in Section 5.3 (Romo and Schultz 1987), or just slightly anterior to that. I then used transcranial magnetic stimulation (TMS) to stimulate the area, and found that people's retrospective reports of the onset of intention could actually be modulated even *after* the action was completed (Lau, Rogers, and Passingham 2007). We did control studies to rule out that this finding was just a general memory effect; it was specific to the time around action completion. The effect was also specific to the reported onset of intention, but not to the movement itself, nor the reported onset of other events, like tactile sensations.

After stimulation, subjects reported their intention to have started earlier—as if TMS injected some extra activity into the brain area, which was mistaken as a meaningful intention signal. So, overall they thought they had been intending to make the movement for longer, compared to when their brains were not being stimulated. Importantly, if this interpretation was true, it means that the mechanisms giving rise to consciousness must not be very precise in tracking the intention signal "online." Instead, as the late Daniel Wegner suggested, perhaps it is all loosely reconstructed after the fact. The so-called conscious will may very well be an illusion (2004).

These studies probably set me on a good path of inquiry. I still think the results are more or less correct. Our sense of volitional agency may indeed depend on late-stage cognitive processing, at least in part. That is to say, it is unlikely a direct instant readout of our internal motoric signals, just as the TMS findings suggest. We will revisit this issue in Section 8.5.

But I describe these studies relatively briefly here, because by the time I finished my PhD, I had become frustrated by these experiments. It is a very odd thing to ask people to make these "spontaneous" movements "at any time they like," and to report the onset of "intention" afterwards. If consciousness is to serve some important survival functions, I doubt it concerns primarily situations of this sort.

5.4 Impossible Situation

When I was in grad school, Dehaene's global view had already started to dominate. In particular I remember reading a wonderful piece by Naccache and Dehaene (2001), in which they spelled out some predictions of their theory. It was so clearly written and intellectually honest that it set a standard for all of us to follow.

One prediction was based on the finding that once the subjects understood certain task instructions, they could apply this "task set" to subliminal stimuli. By "task set" we mean the rules of the task, in terms of what sensory stimuli require what motor responses. For example, the task may be to decide if a number was bigger or smaller than five, and to press keys with different hands accordingly. Once they started doing the task, motor activity for the correct hand response could be triggered by visually masked numbers that were invisible. The interpretation is that once the task set was established, information could meaningfully go from a perceptual process to the motor mechanisms, all without consciousness. We call this a "priming" effect: a subliminal stimulus can induce some level of preparedness for the relevant representations. So once a task set was on, a subliminal stimulus can "prime" the relevant motor response.

However, based on the global workspace theory, they predicted that the establishment of the task set itself should require consciousness. That is because the very function of the global workspace is to exert top-down control to coordinate how the perceptual and motor modules are linked. They famously called it an "Impossible Situation" for subliminal stimuli to influence this initial process of setting up the task set (Naccache and Dehaene 2001).

Being a contrarian (and deluded too, as I mentioned in Section 2.11), I set out to challenge that prediction. I reasoned that, in a sense, what they found was that in a single case, a simple function (motor priming by invisible stimuli) could be exercised without consciousness. From there, it is natural to think that consciousness may be required for a more complex function (the establishment of a task set based on instruction). But did they try hard enough to see if the more complex function really couldn't also be exercised nonconsciously? Maybe with a stronger nonconscious signal, it could?

So, the first project I did as a postdoc was to test this (Lau and Passingham 2007). In the experiment we presented a figure to tell the subjects if they needed to do a phonological or semantic task on that trial. In the phonological task they had to judge if a word was disyllabic. In the semantic task they had to judge if the word referred to a concrete object. Because they saw the

instruction figure before the word appeared, during the time between the two they had an opportunity to establish the "task set" quickly. Interestingly, we found that a subliminal instruction figure could also influence task set establishment. Let's say they were presented with a highly visible instruction figure telling them to prepare for a phonological task. When we inserted an invisible figure telling them to prepare for a semantic task before the visible instruction figure providing different instructions, this created a kind of nonconscious conflict. Overall, subjects would still be mostly doing the phonological task as consciously instructed, but they were slower and less accurate—as if they were nonconsciously distracted to do the other task as well. Using fMRI, we showed that when they were "primed" this way, there was less phonological-related activity, and more semantically related activity, in the relevant brain regions. The nonconscious conflicts also led to more activity in areas of the prefrontal cortext, which typically responds to cognitive and response conflicts.

5.5 More "Impossible" Cases

After I moved to Columbia University, my then graduate student Doby Rahnev took this one step further. He presented the subliminal task instruction stimuli outside of attentional focus. And yet, the task set seems to be successfully primed nonconsciously in the same way (Rahnev, Huang, and Lau 2012).

Simon van Gaal has also done similar studies (van Gaal et al. 2008, 2010, 2011; van Gaal, Lamme, and Ridderinkhof 2010). Together with colleagues holding the local view, he mostly focused on response inhibition, another task that is linked to prefrontal higher cognitive functions. So, in these tasks subjects had to make a response quickly. On some trials, there might be a "no-go" or "stop" signal, so they had to withhold the prepared response. They found that a subliminal "no-go" signal could slow down prepared responses too, as if some degree of inhibition took place. It also led to more prefrontal activity, typically reflecting inhibitory control in these settings.

Meanwhile, Dehaene's own lab has also reported various cases in which nonconscious stimuli can influence higher cognitive functions like error detection (Charles et al. 2013; Charles, King and Dehaene 2014) and working memory (King, Pescetelli, and Dehaene 2016; Trübutschek et al. 2017, 2019; Trübutschek, Marti, and Dehaene 2019). As I will explain in Section 5.6, there may be some concerns as to whether these stimuli were truly invisible. Also, in the case of "working" memory, perhaps these are really only cases of brief

sensory memory; one may not be able to actively manipulate the memory content, to protect it against distractors, for example. But all the same, it is remarkable that a strong proponent of the global view himself accepts these cases as showing that nonconscious stimuli can influence such high-level processes.

Together with other colleagues, Ryan Scott and Zoltan Dienes also did some studies showing that nonconscious stimuli likely can go through some central mechanisms (Scott et al. 2018). Again using a priming approach, they showed that nonconscious stimuli from one sensory modality (e.g., hearing) can influence processing in another modality (e.g., vision). For this to work, presumably the nonconscious information has to go through some central mechanisms that link up processes from the different modalities. But if such central mechanisms are what make information conscious, that seems . . . just impossible.

So, taking together all of the evidence discussed, the "impossible situation" as predicted by Naccache and Dehaene (2001) doesn't seem so impossible after all. I felt smug for a while, as I thought we had falsified the global view. But later I came to think these studies may not really tell us that much.

5.6 The Limits of Subliminal Priming Experiments

We can all agree that the primary function of having legs is for locomotion. It allows us to move around. Lower limbs are probably selected through evolution because of this useful function. But if we remove the legs of a small insect, it may still be able to wiggle around a little. Or maybe it has wings too, so it can fly around.

By analogy, consciousness may also have the primary functions of higher cognitive control, including establishing task sets and response inhibition. Without consciousness, we can exercise these functions *a little bit*, and this is perhaps all that the priming studies showed. Perhaps it is only with consciousness that we can drive these functions fully. So consciousness may be selected through evolution for this reason after all.

This is somewhat related to our tendency to think in terms of necessity. As I have argued in Sections 2.3, 3.2, and 4.3, this is often misguided. In studying the functions of consciousness, we often want to show that it is "necessary" for certain functions, but not others. But it does not work this way. There are several problems.

The first is logical. Consciousness may not be strictly necessary for a function to be exercised to some minimal extent. But it could be necessary for the

function to be exercised *more effectively*. Or maybe there are multiple ways to exercise these functions, and consciousness is one of the more obvious and economical ways, conveniently picked out by evolution. In this sense, testing for necessity just doesn't directly answer our question.

Nor is it very practical. Like attention and many other psychological constructs, consciousness isn't something we can neatly turn on and off. When we apply visual masking, how do we know it is effective? We typically make subjects do forced-choice detection or discrimination tasks to ascertain that they are at chance, meaning no perception occurs. But to statistically prove that something is completely absent is difficult. With very many trials, even very weak stimuli may be detectable or discriminable above chance. Showing a nonsignificant result is easy; one only needs to test it with an insufficient number of trials, or to measure things poorly. But proving that something truly is at chance is a statistical challenge. So, this is the second problem: convincing demonstration of subliminality.

In recent years, some authors have given up on showing that these stimuli are truly invisible in the sense of completely lacking detectability or discriminability. Instead, they took subjective reports of the lack of awareness at face value. Using these methods, various authors have claimed to have found evidence for nonconscious working memory (King, Pescetelli, and Dehaene 2016; Trübutschek et al. 2017, 2019; Trübutschek, Marti, and Dehaene 2019), nonconscious metacognition (Charles et al. 2013; Charles, King, and Dehaene 2014; Jachs et al. 2015), and blindsight in normal observers (Hesselmann, Hebart, and Malach 2011).

I, too, have emphasized the importance of subjective reports. But in most of the studies I reviewed in the previous chapters, the usage was to compare different levels of reported awareness across conditions within the same subjects. The point was to show that the level of awareness differed between the conditions, not that one condition lacked awareness completely. By traditional psychophysics standards, it is not acceptable to infer the lack of awareness based on subjective reports alone (Macmillan and Douglas Creelman 2004). This is because subjects use these reports based on rather arbitrary criteria. If we give people the options "no awareness," "some awareness," and "high awareness," naturally they will say "no awareness" on some of the weakest trials. But this may only reflect that they understand these options in relative rather than absolute terms.

When Megan Peters was a postdoc in my lab at UCLA, she did studies to control for this "criterion" problem (Peters and Lau 2015). Instead of asking people to rate their visibility or awareness on some arbitrary scale, she asked them to compare two consecutive presentations. The logic is: If

a stimulus was truly invisible, one should not be able to distinguish it from another stimulus lacking any useful information. If subjects were asked to bet on their ability to correctly identify one of them, they shouldn't know which one to bet. With this method, she has shown that at least for the case of blindsight in normal observers, it was likely to all due to criterion artifacts—at least for the kind stimuli typically used in previous studies (see also Knotts, Lau, and Peters 2018). That is, there was no convincing evidence for nonconscious perception in normal observers, when the possibility of criterion artifact was taken into account, using her criterion-free forced-choice betting method. So care must be taken before we take subjects' reported lack of awareness at face value.

There is a third problem with subliminal priming studies, which I think is just as troubling. In experiments, when we use, for instance, visual masking to render some stimuli invisible, we are also changing other things. The confounders introduced back in Chapter 2 apply. With the mask on, there is obviously a stimulus difference. A bigger problem, though, is the task-performance capacity confounder. If the invisible stimuli are just weak, then of course many functions will not be driven so well by them. But it may not be the lack of consciousness per se that matters. Perhaps we just haven't found a way to keep a nonconscious signal strong enough.

5.7 Performance Matching & Statistical Power

When I was at Columbia University, a postdoc in my lab, Ai Koizumi, did some work to address this last issue (Koizumi, Maniscalco, and Lau 2015). The idea was similar to the way we deal with task-performance capacity confounders in Chapter 2: she found a pair of stimuli in which task performance was directly matched, and yet they showed some difference in subjective visibility. In the study we actually measured confidence rather than visibility, but I believe that would have worked similarly. Ai-san showed that stimuli perceived to be subjectively stronger didn't really facilitate task set preparation (Koizumi, Maniscalco, and Lau 2015). There were some subtle effects on inhibition strategy, but overall inhibition was no more effective.

So I was back to thinking that maybe the global view was in some trouble after all. When performance capacity was controlled for, a sheer subjective difference in perception did not lead to functional advantages for these higher cognitive control functions. So consciousness probably wasn't as functionally important as the global view holds.

But I have to acknowledge, there was also a problem in our study (Koizumi, Maniscalco, and Lau 2015): statistical power was rather limited. This is a general issue with our way of addressing the performance-capacity confounders in subjects from the general population. It is very difficult to get a large effect on subjective perception while having task-performance capacity matched. As such, null results need to be interpreted with caution. Just because we did not find functional advantages there for consciousness doesn't mean there aren't such advantages. Maybe we just needed more subjects for a subtle effect to be detected.

Brian Maniscalco and others have recently spelled out a framework for how to design better studies of this sort, with adequate power (Maniscalco et al 2020). But we are yet to see more new studies done this way.

This issue of lack of statistical power is not specific to performance-matching studies. It applies whenever our expected effect sizes are small. Small effects aren't easily detected with a small sample. To verify that they are truly absent, rather than missed because they are so weak, we need a lot of data. So this is a problem for most of the subliminal priming studies too. When the stimuli are masked to become invisible, we can't really expect the effects to be very big. This can be yet another problem with subliminal priming.

With all these caveats in mind, next, I'll highlight a few findings that I think are relatively promising.

5.8 Inhibition & Exclusion

Earlier in Section 5.4, I mentioned that it may be possible to trigger response inhibition nonconsciously. Such findings have been taken by local theorists as challenges to the global view. However, it is possible that consciousness allows inhibition to work better.

Navindra Persaud and Alan Cowey have tested the blindsight patient GY (Persaud and Cowey 2008). GY is the same patient whom we already described in Chapter 2. Blindsight mostly affected GY's right visual field. Persaud and Cowey adopted a version of Jacoby's "exclusion" task (Jacoby, Jones, and Dolan 1998). They asked GY to report the opposite of what was shown. That is, he had to report "up" if the stimulus was actually presented in the lower quadrant, and vice versa. For stimuli presented in the "normal" field, this was easy to do. However, for the "blind" field, not only did GY have some difficulty doing this, curiously, he made more errors as the stimulus was shown at a higher contrast. Under the lack of awareness, the stimulus seemed

to have driven his responses in an automatic fashion, leading to his failure to "exclude" or inhibit the natural and prepotent reaction.

This was only a single patient. But an elegant study from Takeo Watanabe's lab (Tsushima, Sasaki, and Watanabe 2006) also found something similar in an fMRI study with subjects from the general population. In that study the inhibition concerned some irrelevant, distractor stimuli presented in the background. The subjects were required to focus on a central task and to ignore these distractors. It was found that the distractor effect was actually the strongest when the stimuli were at an intermediate strength, around perceptual threshold. The interpretation is, when the distractors were weak, they didn't have much impact. But when they were strong and highly visible, subjects were able to actively inhibit their influence. Congruent with this interpretation, activity in the lateral prefrontal cortex was higher when the distractors were visible. It was as if some inhibitory functions there were only engaged when the distractors were consciously perceived.

These studies are intriguing because they showed that an increase in stimulus strength does not always lead to more effective inhibition. Perhaps inhibitory functions kick in only when consciousness is engaged. But in these studies, the nonconscious stimuli were weak. In Persaud and Cowey (2008), although the stimuli presented to the blindfield were physically as strong as those presented to the normal, sighted field, the associating performance capacity was not matched. To really address the issue of performance capacity confounder, future studies can test whether even very strong nonconscious signals would fail to trigger inhibition as effectively as weak conscious signals.

5.9 Metacognition

With other colleagues (including myself), Navindra Persaud also tested GY's metacognitive abilities between the "normal" and "blind" field (Persaud et al. 2011). Interestingly, GY was willing to bet money on his performance in the "blind" field as much as he was for the "normal" field. Presumably, it means he was "aware" of his good performance there, in a cognitive sense. But over trials, his bets did not track accuracy so well in the "blind" field, as compared to the "normal" field. He did bet higher on his correct trials over his incorrect trials, meaning he probably had some metacognitive ability even in the "blind" field. But this ability was higher in the "normal" field, even for weaker stimuli giving rise to lower task performance than in the "blind" field.

The idea that consciousness may facilitate metacognition is not new (Baars 1988; Shea and Frith 2019). What is intriguing is that, here we have a positive

finding that does not suffer from the task-performance capacity confounders. As in inhibition, the function of metacognition was not completely abolished under the lack of consciousness. But consciousness seems to facilitate it even when task-performance capacity was matched (or when it was lower than in the nonconscious condition).

5.10 Endogenous Attention

Because inhibition and metacognition are functions associated with mechanisms in the prefrontal cortex, one may think that maybe the global view is not so challenged after all. That is, although priming studies showed that nonconscious stimuli can exercise some prefrontal functions (as reviewed in Sections 5.4 and 5.5), perhaps consciousness will always allow us to exercise these functions more effectively. Unfortunately, this is unlikely to be true for all prefrontal functions.

For a counter example, let us reconsider the relationship between attention and consciousness, which is complex, as we have reviewed already in Chapter 4. There, we were mostly concerned with how attention may modulate subjective experience. But one can also ask: For stimuli that are not consciously perceived, can they benefit from attentional cueing too? That is, if we attend to these nonconscious stimuli, do we speed up the processing as much as we do for consciously perceived stimuli? In particular, when an attentional cue is given symbolically , rather than physically around the stimulus in question, we say this is a form of endogenous attention. In this case, the subjects have to process the meaning of the cue, rather than to act reflexively. Endogenous cueing is believed to depend on prefrontal mechanisms.

Bob Kentridge and colleagues have tested this on patient GY (Kentridge, Heywood, and Weiskrantz 2004). They presented a visual cue in the center of the screen, where GY could see consciously, to indicate to the patient where the subsequent target stimulus was likely to appear. The target was a bar that was either horizontal or vertical, presented to GY's blindfield. Turns out, when the cue was predictive of the location of the target, it helped patient GY respond more quickly—by over 100 milliseconds. Because normal cueing effects for visible stimuli are generally not bigger than this, we can say that the stimulus presented to the blindfield didn't seem to suffer in this specific context, despite the lack of subjective awareness.

These authors have also found that this kind of cueing also works when the attentional cue was presented in the blindfield, such that GY did not see the cue stimulus. However, the effect in this kind of nonconscious cuing, in which

the cue stimulus itself was not consciously perceived, seems to be smaller (Kentridge, Heywood, and Weiskrantz 1999). So it is possible that consciousness may play some role there. But others have also found fairly robust attentional cueing effects, in subjects from the general population, using stimuli rendered invisible (Zhang et al. 2012; Huang et al. 2020).

5.11 Intuitively "Improbable" Situations?

The examples discussed in Sections 5.7–5.10 hopefully help illustrate that assessing the functional advantages associated with consciousness is no trivial matter. We really have to go through the empirical specifics, in a case by case manner.

Why is it that some prefrontal functions seem to benefit from consciousness (i.e., having the relevant stimuli consciously perceived), while others don't? That's probably because the prefrontal cortex is a substantial part of the brain. Many different cognitive functions reside within this brain region. Let us suppose, for a moment, that some specific prefrontal mechanisms are key to consciousness, just like what global theories suggest. So when a stimulus fails to generate a conscious experience, it could be that the entire prefrontal cortex is compromised. Or it could be that just some specific mechanisms within the prefrontal cortex responsible for consciousness are compromised. Alternatively, it could be that all these mechanisms are intact, but the sensory signals somehow fail to reach the prefrontal cortex, because of some early sensory deficits. Because multiple mechanisms are likely involved in the entire chain of processes, without direct evidence it is hard to say what is definitely the problem when a perceptual process fails to lead to a subjective conscious experience.

In other words, it is also difficult to pinpoint what functions will *always* be compromised in nonconscious processing. A perceptual process can fail to be conscious in different ways (Block 2016). To get at this issue, we probably need a theoretical model outlining what really is the mechanism of consciousness. We probably need to understand how the different components interact to support subjective experience. But even then, when that mechanism breaks down, there may be "back-up" systems. Understanding this is complicated business.

Despite that, some of the evidence discussed is relatively informative. It is clear that not all prefrontal functions are enhanced by consciousness. But it seems that even when the performance confound is not an issue, as in blindsight, the lack of consciousness seems to impair metacognition.

What else is unlikely to be performed well by the blindsight patient? Without the relevant direct experimental data, if you allow me to speculate based on my own interaction with GY (Persaud and Lau 2008): I suspect it is generally impossible for any blindsight patient to spontaneously form the belief with certainty that a specific stimulus is presented. When forced, GY makes a guess. But the perceptual process never impinges on his rational decision-making system directly and spontaneously.

Also, although we never did this experiment, I suspect we could have: let's say we set up a forced-choice experiment, where we ask the patient to discriminate between horizontal versus vertical bars. If on some trials we present to him a 45-degree-tilted bar, very clearly, I suspect he would be none the wiser, and just proceeds to make his guess as to whether it is horizontal or vertical. He probably would not say: "Well this was not expected. What is this?" Or if we presented to him something he had never seen before, like a computer generated fractal image, he probably would not say: "I haven't seen anything like this before. Is there an error?"

But these are my guesses. I do not have empirical evidence for these claims. I spell them out here because, in a way, if GY behaved not as I hypothesized, I would actually doubt whether he genuinely lacked subjective experience in his blindfield. That is to say, this would be improbable on a *conceptual* level. There seems to be something intrinsic to what it takes for one to lack subjective experience. But this clearly is a theoretical matter, open to debate. We shall address these more properly in Chapters 7–9 (especially Sections 9.4–9.9).

5.12 DecNef & Threat Reduction

Why do we rely so much on single-patient studies in blindsight? The reason is as explained earlier in Section 5.6: These studies allow us to see what can be achieved with a nonconscious and yet strong internal signal. With masking, conscious visibility can be abolished, but the internal perceptual signal is also mostly gone. With such weak residual signals we are stuck with some small effects, such as those typically reported in subliminal priming studies. These effects are neither easy to interpret nor to replicate.

There is a way around this with subjects from the general population. The trick is to go directly into their brains and look for these nonconscious strong signals. My colleagues in Japan, Kazuhisa Shibata, Yuka Sasaki, Takeo Watanabe, and Mitsuo Kawato, have pioneered a technique called decoded neurofeedback, which we sometimes call DecNef for short (Shibata et al. 2011; Watanabe et al. 2018). The idea is that we can apply

multivoxel pattern analysis (MVPA) to fMRI data online (i.e., in real-time). In Chapter 2 we already described this kind of analysis. Essentially, in MVPA we look for fine-grained patterns in fMRI data within a region, in order to extract meaningful content from them. This differs from the more traditional approaches, which tend to focus on the overall level of activity in an area.

Interestingly, people are generally unaware of the spontaneous fluctuations of these decoded internal brain signals. For example, we can present red lines and green lines to subjects. These stimuli will activate visual areas to similar overall degrees. But with MVPA, we can distinguish between trials in which red versus green lines are presented. When there is no color stimuli, these same decoded internal signals still fluctuate spontaneously. So, sometimes our brains look as if the line presented was mildly representing green or red. But, of course, we do not consciously "see" this kind of internal fluctuation. In fact, when we asked subjects to directly guess the content of their decoded internal signal, they were generally at chance (Shibata et al. 2011; Watanabe et al. 2018).

But do these decoded internal signals have any cognitive or behavioral impact on the subject? Turns out, they do, in fairly powerful ways. For example, Ai Koizumi did a study in which she paired these red- and green-line visual patterns with unpleasant electric shocks (Koizumi et al. 2016). After a while, subjects showed physiological reactions to these colored patterns alone, even without shock, as if they were physiologically threatened by the line patterns themselves. After that, we used DecNef to see if we could reduce this learned threat response. In one group of subjects, we paired these decoded signals representing red lines with reward, during fMRI. (In another group we targeted green-line patterns as a counter-balanced control.) During this period, no colored patterns were presented. But if the nonconscious decoded brain signal looked as if it was representing red, we gave feedback to the subjects to indicate that they earned some money (on the order of a few cents). This way, we hoped that the brain patterns previously associated with threat will now be "counter-conditioned" with reward. The subjects were blind to the purpose of the study; from their perspective, they just played hundreds of trials of this "mental slot machine," and earned some tens of dollars over a few days.

To our pleasant surprise, the hypothesis worked. After a few hours of DecNef, when presented with the red-line patterns, their physiological reactions were indeed diminished, as if the shock-paired stimuli became less threatening somehow. Importantly, this was specific to the brain signals which

went through the feedback procedure. Subjects showed just as much physiological reactions to the green-line patterns, the decoded signals for which were not rewarded through DecNef.

This served as a promising proof-of-concept. Vincent Taschereau-Dumouchel, a postdoc in my lab at UCLA, thought we could take it one step further. Perhaps we can actually use this to reduce common phobia in people (Taschereau-Dumouchel et al. 2018). To do this we decoded voxel patterns for visual objects and animals, including the commonly feared ones like spiders and snakes. As in Koizumi et al., using DecNef, we paired the decoded internal brain signals with reward. Conceptually it replicated our earlier findings. Physiological-threat responses were reduced when subjects saw the images of the feared animals, specifically for the ones paired with reward only. Because the feedback and randomization procedures were all conducted by the computer, the entire process was double-blind placebo-controlled too.

5.13 Clinical Applications?

Can DecNef one day work in clinical settings for real? We do not know yet. As I'm writing, my lab is currently conducting a clinical trial co-led by my UCLA colleague Michelle Craske, who is an expert on anxiety disorders. One challenge is that to decode these voxel patterns, we typically show many images to the subjects while they are in the fMRI scanner. But for patients who are unable to tolerate seeing these images over and over again, they may as well go through conventional psychotherapy, which is arguably more economical. For phobia and posttraumatic stress disorders, traditional therapy generally works well. The trouble is patients often drop out from these treatments prematurely. This may be understandable, as the treatment usually requires them to encounter the feared objects or to revisit the traumatic events. Many patients find it difficult and unpleasant.

Vincent solved the problem with an ingenious solution, using a technique called hyperalignment to estimate the patient's voxel patterns for the feared animals using data from other "surrogate" subjects (Haxby et al. 2011, 2020). With enough data, he has demonstrated that it works well. This way, the patients could go through the DecNef-based intervention, without ever having to see any images of animals that they are afraid of (except when we tested them for threat reactions, which was only needed for research rather than clinical purposes).

One important caveat is that in the current studies, subjects did *not* report feeling less afraid when they saw images of the targeted animals. The effect was mainly on their physiological reactions, not their subjective reports of fear. Congruent with a prediction made by Joe Ledoux, perhaps it means that nonconscious treatments can ever only change nonconscious physiological responses (2015). To eliminate the conscious experience of fear, we need to focus directly on mechanisms at that level. Or perhaps, as the methods improve, we will also be able to impact conscious experience using nonconscious DecNef. We will need more experiments to find out.

Regardless, these findings show that nonconscious signals can potentially give us much stronger effects than subliminal priming can. Typically in priming studies, a response is sped up or slowed down by some tens of milliseconds. Here, we robustly changed one's physiological-threat reactions in potentially clinically meaningful ways. Even if the impact on the subjective feelings of fear turns out to be limited, reduction of the physiological-threat reactions can perhaps help patients ease into traditional therapy better. In Taschereau-Dumouchel et al., for the animal categories targeted by DecNef, the physiological-threat reactions, in terms of skin-conductance and amygdala fMRI reactivity, were entirely brought down to baseline level. This was true at least right after the intervention. We do not yet know how long these effects last; we are currently in the process of finding out.

We will further discuss the importance and promise of related clinical applications in Chapter 8.

5.14 Chapter Summary: Partially Global?

The literature on nonconscious priming is vast. It has a long and controversial history. And yet, I've been very selective in covering the relevant studies, to the point that this may look unfair and biased. But the reason is that our question isn't about nonconscious priming per se. The question here is whether consciousness may be associated with higher cognitive functions. The global view says *yes*: consciousness is linked to the prefrontal and parietal cortices, where global coordination and exchange of information take place. Higher cognitive functions should be facilitated by these mechanisms.

The local theorists say *no*: consciousness takes place within the sensory circuitries. Higher cognitive functions are downstream to consciousness. A percept may be nonconscious because it isn't accompanied by the right kind of local dynamics. But the signal may still be able to get out of the local sensory areas to impact the higher cognitive functions downstream.

To answer this question, priming studies are not always the most relevant. In the social cognition literature, many of these studies concern the lack of attention and explicit reflection, rather than the lack of subjective experience per se (Hassin, Uleman, and Bargh 2005). Even when we restrict ourselves to subliminal priming studies, the aim tends to be to show that a function can be influenced by primes that are not consciously perceived. There have also been strong critiques on whether these primes were truly subliminal. Ignoring these critiques is probably not going to help with our progress. Meanwhile, even when we succeed in establishing some subliminal priming effects, they tend to be small. Small effects are easily confused with null effects and are often hard to replicate.

Above all, logically it is not clear how we should interpret these subtle effects. The global theorists can say that although certain functions can operate somewhat nonconsciously, they work much better when we are conscious of the relevant stimuli. Maybe consciousness is always needed for these functions to fully exercise. This may explain why even proponents of the global view like Dehaene seem to have no trouble acknowledging the possibility of a nonconscious working memory and metacognition. Perhaps working-memory representations are more stable, flexible, and effective only when they are conscious.

All the same, it does raise some questions about how these nonconscious higher cognitive effects are even possible. The content of working memory is supposed to be globally accessible. If that's true, how can it be maintained *at all* outside of the workspace? Presumably it means that there must be some other shortcuts for achieving these global functions. But if there are such shortcuts, why do we need the central workspace in the first place? Or perhaps it means that the global workspace isn't so fully global after all. Is it only *part of* a global mechanism, wherein some other parts can operate nonconsciously?

Furthermore, using methods like DecNef, we are beginning to find more powerful forms of nonconscious effects. This may challenge the global view further. Conceptually, nonconscious signals need not be weak by definition. With stronger signals, eventually we may find that higher cognitive functions can be exercised nonconsciously, perhaps even to relatively high degrees of effectiveness.

Despite these threats to the global view, I reviewed some studies showing that consciousness may allow us to more effectively exercise inhibitory and metacognitive functions. The relative advantages provided by consciousness in these cases are not so trivial, especially in the case of metacognition, in the sense that performance capacity was matched. What may be the mechanistic explanation? We will come back to this question in Chapters 7–9.

References

Baars BJ. *A Cognitive Theory of Consciousness*. books.google.com, 1988.

Block N. The Anna Karenina principle and skepticism about unconscious perception. *Philos Phenomenol Res* 2016;**93**:452–459.

Charles L, King J-R, Dehaene S. Decoding the dynamics of action, intention, and error detection for conscious and subliminal stimuli. *J. Neurosci*. 2014;**34**:1158–1170.

Charles L, Van Opstal F, Marti S et al. Distinct brain mechanisms for conscious versus subliminal error detection. *Neuroimage* 2013;**73**:80–94.

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.

van Gaal S, Lamme VAF, Fahrenfort JJ et al. Dissociable brain mechanisms underlying the conscious and unconscious control of behavior. *J Cogn Neurosci* 2011;**23**:91–105.

van Gaal S, Lamme VAF, Ridderinkhof KR. Unconsciously triggered conflict adaptation. *PLOS One* 2010;**5**:e11508.

van Gaal S, Ridderinkhof KR, Fahrenfort JJ et al. Frontal cortex mediates unconsciously triggered inhibitory control. *J Neurosci* 2008;**28**:8053–8062.

van Gaal S, Ridderinkhof KR, Scholte HS et al. Unconscious activation of the prefrontal no-go network. *J Neurosci* 2010;**30**:4143–4150.

Hassin RR, Uleman JS, Bargh JA. *The New Unconscious*. Oxford University Press, 2005.

Haxby JV, Guntupalli JS, Connolly AC et al. A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 2011;**72**:404–416.

Haxby JV, Guntupalli JS, Nastase SA et al. Hyperalignment: Modeling shared information encoded in idiosyncratic cortical topographies. *Elife* 2020;**9**. https://doi.org/10.7554/eLife.56601.

Hesselmann G, Hebart M, Malach R. Differential BOLD activity associated with subjective and objective reports during "blindsight" in normal observers. *Journal of Neuroscience* 2011;**31**:12936–12944.

Holton NE, Bonner LL, Scott JE et al. The ontogeny of the chin: An analysis of allometric and biomechanical scaling. *Journal of Anatomy* 2015;**226**:549–559.

Huang L, Wang L, Shen W et al. A source for awareness-dependent figure-ground segregation in human prefrontal cortex. *Proc Natl Acad Sci U S A* 2020;**117**:30836–30847.

Jachs B, Blanco MJ, Grantham-Hill S et al. On the independence of visual awareness and metacognition: A signal detection theoretic analysis. *J Exp Psychol Hum Percept Perform* 2015;**41**:269–276.

Jacoby LL, Jones TC, Dolan PO. Two effects of repetition: Support for a dual-process model of know judgments and exclusion errors. *Psychonomic Bulletin & Review* 1998;5:705–709.

Kentridge RW, Heywood CA, Weiskrantz L. Attention without awareness in blindsight. *Proceedings of the Royal Society of London Series B: Biological Sciences* 1999;**266**:1805–1811.

Kentridge RW, Heywood CA, Weiskrantz L. Spatial attention speeds discrimination without awareness in blindsight. *Neuropsychologia* 2004;**42**:831–835.

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

Knotts JD, Lau H, Peters MAK. Continuous flash suppression and monocular pattern masking impact subjective awareness similarly. *Atten Percept Psychophys* 2018;**80**:1974–1987.

Koizumi A, Amano K, Cortese A et al. Fear reduction without fear through reinforcement of neural activity that bypasses conscious exposure. *Nat Hum Behav* 2016;**1–6**. https://doi.org/10.1038/s41562-016-0006.

Koizumi A, Maniscalco B, Lau H. Does perceptual confidence facilitate cognitive control? *Atten Percept Psychophys* 2015;77:1295–1306.

Kornhuber HH, Deecke L. Changes in the brain potential in voluntary movements and passive movements in man: Readiness potential and reafferent potentials. *Pflugers Arch Gesamte Physiol Menschen Tiere* 1965;**284**:1–17.

Lau HC, Passingham RE. Unconscious activation of the cognitive control system in the human prefrontal cortex. *J Neurosci* 2007;**27**:5805–5811.

Lau HC, Rogers RD, Haggard P et al. Attention to intention. *Science* 2004;**303**:1208–1210.

Lau HC, Rogers RD, Passingham RE. On measuring the perceived onsets of spontaneous actions. *J Neurosci* 2006;**26**:7265–7271.

Lau HC, Rogers RD, Passingham RE. Manipulating the experienced onset of intention after action execution. *J Cogn Neurosci* 2007;**19**:81–90.

LeDoux J. *Anxious: Using the Brain to Understand and Treat Fear and Anxiety*. Penguin, 2015.

Libet B, Gleason CA, Wright EW *et al.* Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain* 1983;**106**(Part 3):623–642.

Macmillan NA, Douglas Creelman C. *Detection Theory: A User's Guide*. Psychology Press, 2004.

Maniscalco B, Castaneda OG, Odegaard B et al. The metaperceptual function: Exploring dissociations between confidence and task performance with type 2 psychometric curves. PsyArxiv 2020. https://doi.org/10.31234/osf.io/5qrjn.

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

Nikolov S, Rahnev DA, Lau HC. Probabilistic model of onset detection explains paradoxes in human time perception. *Front Psychol* 2010;**1**:37.

Persaud N, Cowey A. Blindsight is unlike normal conscious vision: Evidence from an exclusion task. *Conscious Cogn* 2008;**17**:1050–1055.

Persaud N, Davidson M, Maniscalco B et al. Awareness-related activity in prefrontal and parietal cortices in blindsight reflects more than superior visual performance. *Neuroimage* 2011;**58**:605–611.

Persaud N, Lau H. Direct assessment of qualia in a blindsight participant. *Consciousness and Cognition* 2008;**17**:1046–1049.

Peters MAK, Lau H. Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. *Elife* 2015;**4**:e09651.

Rahnev DA, Huang E, Lau H. Subliminal stimuli in the near absence of attention influence top-down cognitive control. *Atten Percept Psychophys* 2012;**74**:521–532.

Robinson Z, Maley CJ, Piccinini G. Is consciousness a spandrel? *Journal of the American Philosophical Association* 2015;**1**:365–383.

Romo R, Schultz W. Neuronal activity preceding self-initiated or externally timed arm movements in area 6 of monkey cortex. *Exp Brain Res* 1987;**67**:656–662.

Schurger A, Sitt JD, Dehaene S. An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proc Natl Acad Sci U S A* 2012;**109**:e2904–e2913.

Scott RB, Samaha J, Chrisley R et al. Prevailing theories of consciousness are challenged by novel cross-modal associations acquired between subliminal stimuli. *Cognition* 2018;**175**:169–185.

Shea N, Frith CD. The global workspace needs metacognition. *Trends Cogn Sci* 2019;**23**:560–571.

Shibata K, Watanabe T, Sasaki Y et al. Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation. *Science* 2011;**334**:1413–1415.

Soon CS, Brass M, Heinze H-J et al. Unconscious determinants of free decisions in the human brain. *Nat Neurosci* 2008;**11**:543–545.

Taschereau-Dumouchel V, Cortese A, Chiba T et al. Towards an unconscious neural reinforcement intervention for common fears. *Proceedings of the National Academy of Sciences* 2018;**115**:3470–3475.

Trübutschek D, Marti S, Dehaene S. Temporal-order information can be maintained in non-conscious working memory. *Sci Rep* 2019;**9**:6484.

Trübutschek D, Marti S, Ojeda A et al. A theory of working memory without consciousness or sustained activity. *Elife* 2017;**6**. https://doi.org/10.7554/eLife.23871.

Trübutschek D, Marti S, Ueberschär H et al. Probing the limits of activity-silent non-conscious working memory. *Proc Natl Acad Sci U S A* 2019;**116**:14358–14367.

Tsushima Y, Sasaki Y, Watanabe T. Greater disruption due to failure of inhibitory control on an ambiguous distractor. *Science* 2006;**314**:1786–1788.

Watanabe T, Sasaki Y, Shibata K et al. Advances in fMRI real-time neurofeedback. *Trends Cogn Sci* 2017;**21**:12;997–1010.

Wegner DM. Précis of the illusion of conscious will. *Behav Brain Sci* 2004;**27**:649–659; discussion 659–692.

Zhang X, Zhaoping L, Zhou T et al. Neural activities in V1 create a bottom-up saliency map. *Neuron* 2012;**73**:183–192.