

DETECTING INTROSPECTIVE ERRORS IN CONSCIOUSNESS SCIENCE

ANDY MCKILLIAM

*Center for Traditional and Scientific Metaphysics, Department of Philosophy,
National Taiwan University*

*Monash Centre for Consciousness and Contemplative Studies, Department of
Philosophy, Monash University*

Detecting introspective errors about consciousness presents challenges that are widely supposed to be difficult, if not impossible, to overcome. This is a problem for consciousness science because many central questions turn on when and to what extent we should trust subjects' introspective reports. This has led some authors to suggest that we should abandon introspection as a source of evidence when constructing a science of consciousness. Others have concluded that central questions in consciousness science cannot be answered via empirical investigation. I argue that on closer inspection, the challenges associated with detecting introspective errors can be overcome. I demonstrate how natural kind reasoning—the iterative application of inference to the best explanation to home in on and leverage regularities in nature—can allow us to detect introspective errors even in difficult cases such as judgments about mental imagery, and I conclude that worries about intractable methodological challenges in consciousness science are misguided.

Introduction

Introspection plays a particularly important epistemic role in consciousness science. It is a basic source of evidence. Just as perception provides our basic source of evidence about the world, introspection provides our basic source of evidence about consciousness (Chalmers 2004; Goldman 2000; Jack & Roepstorff 2002; Overgaard 2023). A central task for consciousness science is to determine when and to what extent subject's introspective reports are to be trusted. Sceptics doubt whether this can be achieved. They argue that the privacy of conscious experi-

Contact: Andy Mckilliam <andymckilliam@gmail.com>

ence together with the complexity of introspection as a process makes detecting introspective errors unachievable in practice or even impossible in principle (Irvine 2012: 634; Goldman 2004; Schwitzgebel 2011; Piccinini 2009; Spener 2015; Overgaard & Sandberg 2021).

We cannot dismiss this issue with a casual gesture at the common practice of triangulating introspective reports with behavioural and neural evidence (Bayne & Spener 2010; Jack & Roepstorff 2002; Seth et al. 2008; Timmermans & Cleeremans 2015; Velmans 2009). Without further clarification, triangulation only tells us what to do when these three lines of evidence converge. It tells us that introspective reports can be trusted when they are corroborated by behavioural and neural data. It provides no guidance about how to proceed when introspective reports diverge from behavioural and neural data (Irvine 2021). Nor can we simply quarantine these difficult cases and build a science of consciousness based solely on cases where introspective judgements agree with behavioural and neural data (Chalmers 2010; Bayne & Spener 2010). Many of the central debates in consciousness science today turn on whether introspective reports should be trusted when they conflict with behavioural and neural data: cases like blindsight and perceptual overflow. Should we, for example, take blindsight—a condition caused by damage to primary visual cortex that leaves subjects able to detect, localize and even identify stimuli they deny seeing (Covey 2010)—to be evidence of “powerful forms of unconscious processing” (Brown, Lau, & LeDoux 2019: 765) or as involving “degraded but nonetheless conscious vision [that goes] unacknowledged” by the subject (Phillips 2021: 558)? And when participants in retrocuing paradigms say they clearly saw all the items in the array despite the fact that they can only recall some of them, is this because consciousness overflows cognitive access (Block 2007) or are subjects’ introspective judgments simply mistaken (Kouider, De Gardelle, Sackur, & Dupoux 2010). To answer these questions, we need a framework for detecting introspective errors.

A number of authors have suggested that we can appeal to inference to the best explanation to decide how to proceed in cases where introspective evidence diverges from behavioural and neural evidence (Flanagan 1992; Block 2007; Pauen & Haynes 2021; Michel 2023). However, these discussions have not fully addressed the sceptics’ worries, and as a result, many sceptics remain unconvinced (Schwitzgebel 2011; Irvine 2021). In this paper I will explain how the natural kind method (Shea & Bayne 2010; Bayne & Shea 2020)—the iterative application of inference to the best explanation to home in on and leverage regularities in nature—can address these concerns in a principled way.

In §1, I provide a brief taxonomy of sources of introspective error before explaining why privacy makes detecting introspective error particularly challenging in §2. In §3 I identify the two most pressing reasons for scepticism

toward the possibility of controlling for introspective errors in consciousness science—the primacy worry and the arbitrariness worry—and explain why Matthias Michel’s recent discussion of calibration in consciousness science falls short of addressing them. In §4 I introduce the natural kind method as a framework for detecting measurement errors and explain how it can fill the gaps in Michel’s discussion. And in §5, I apply this framework to a particular case in consciousness science—the task of detecting introspective errors about conscious mental imagery.

1. Sources of Introspective Error

An enduring thought in the philosophy of mind is that some introspective judgements about one’s own states of consciousness are immune to error. David Papineau, for instance, has argued that:

[T]he correctness of standard first-person judgements simply falls out of the special quotational-indexical structure of phenomenal concepts. If I judge phenomenally of some current state of perceptual classification that *it* is like *this*, there is no real room for me to be wrong. (Papineau 2002: 183; see also Chalmers 2003; Gertler 2012; Horgan & Kriegel 2007)

The idea here is this. Suppose you’ve just taken a sip of wine and formed the introspective judgement that the taste experience you are having right now is like *this*, where *this* is a special kind of concept that refers directly to the character of the taste experience in question. Such a direct phenomenal judgment, these authors argue, would be infallible.

As with many ideas in philosophy, this one is contentious. Not everyone accepts that phenomenal concepts have this quotational-indexical structure. What is not contentious however, is that the possibility of error creeps in as soon as we attempt to categorize and communicate our introspective judgments to others for the purposes of conducting scientific research. On this there is more or less a consensus.

Some of these errors may be best thought of as errors in communication. In these cases, subjects’ introspective judgements may be accurate, but error arises in the process of attempting to translate that judgement into a publicly available format.

Some sources of this kind of error are random. Consider slip-of-the-finger errors in studies collecting reports via button press. In these cases, subjects introspect accurately but mistakenly press the wrong button when attempting to communicate their introspective judgment.

Slip-of-the-finger errors and other random sources of communication error are not too difficult to control for. Firstly, because subjects can often tell when they've made one. But more importantly because there is no reason to suppose that subjects are more likely to accidentally report that they saw a stimulus when in fact they did not than the converse. As such, random sources of communication error, can be effectively controlled for by averaging over many trials (Michel 2023).

More problematic are communication errors that stem from variation in how individuals use terms—one person's 'brief glimpse' may be another's 'no experience'—and from the fact that both question-wording (Schwarz 1999) and task complexity (Irvine 2021) can significantly impact how participants report. These are more problematic because subjects are typically unaware of them, and because they are systematic—repeated trials are likely to yield the same error. As a result, scientists cannot control for them by simply running many trials (Michel 2023).

Perhaps the biggest challenge, however, is the possibility of genuine introspective errors. Communication errors are not the only source of error that scientists need to contend with. Sometimes subjects may simply be mistaken about what it is like to be them (Schwitzgebel 2008). Global scepticism about introspection has little to recommend it. Even Eric Schwitzgebel, who is more sceptical than most about the accuracy of introspection, grants that "some aspects of visual experience are so obvious it would be difficult to go wrong about them" (2008: 235). But we do not have equally secure introspective access to all of our conscious states. Consider the experiences associated with the flow state, or being-in-the-zone. As Uriah Kriegel has pointed out, our introspective access to experiences of flow is less direct than for contents at the focus of attention. The very process of directing introspective attention to the experience of flow requires leaving the state in question (Kriegel 2013: 1171). This makes it hard to get a good introspective 'look' at the target experience. And the absence of strong introspective evidence, we might suppose, increases the likelihood that we confabulate details about our own states of consciousness for the same reasons that the absence of strong perceptual evidence increases the likelihood that we confabulate details about the perceptual environment.

Something similar may be true for introspective judgements about contents outside the focus of attention (Chalmers 2010; Bayne 2015), and immediate retrospection (Hurlburt & Schwitzgebel 2007). In each of these cases, our introspective access to the experiences in question is indirect and the possibility of introspective error needs to be taken seriously.

A slightly different issue arises for introspective judgments about experiences associated with mental imagery and those resulting from direct brain stimulation. In these cases, the absence of a stable stimulus may result in an experience

that is unstable over time—changing from moment to moment (Hohwy 2011). This too may make it hard to get a good introspective ‘look’ at the target experience and increases the likelihood of error. In fact, in these cases subjects often express a good deal of uncertainty about their own introspective judgment.¹

So, while global scepticism about the accuracy of introspective reports is unreasonable, global optimism is credulous. We need to take seriously the possibility that in some cases participants may fail to accurately communicate their introspective judgements to scientists, and that in some cases they may simply be wrong about what it is like to be them.

2. Why Detecting Introspective Errors Is Hard

One reason why detecting introspective errors is hard stems from the fact that introspection, as a method, is private. Typically, science trades in public methods for gathering evidence (Piccinini 2003; 2009). Perception, for example, is a public method. We can all dip our toes into the same pool to gather evidence about its temperature. As a result, we can directly check other’s perceptual judgements. If you judge the pool to be below 20°C while I judge it to be clearly above 20°C, we can conclude that at least one of us is wrong. By contrast, we cannot all introspect into the same mind. I cannot introspect into your mind any more than you can introspect into mine. As a result, introspective judgements can only be compared indirectly.

Alone, this is not a deep problem. Replication in science does not require reproducing findings by using the same methods to investigate the same particular entity. In many cases this is not even possible. If a new antiviral effectively kills a sample of virus, subsequent researchers cannot deploy the same procedure to test the same sample. It is enough to reproduce the findings in different samples of virus, so long as we make sure there is no relevant variation from sample to sample.

In practice however, the fact that we can only compare introspective judgements indirectly presents a serious challenge because we cannot always assume there is no relevant variation from one conscious mind to another. This makes resolving introspective disagreement particularly tricky. Take conflicting introspective judgments about mental imagery. When I imagine my house as seen from the street I can kind of see it. My visual imagery is not crystal clear. It is not like actually looking at my house. But there is some visual phenomenology there. Something like “weak perception” (Pearson et al. 2015). Most people I

1. For an illuminating example of the uncertainty and general confusion that often accompanies introspective judgments about experiences stemming from direct brain stimulation see the transcripts collected by Mégevand and colleagues (2014).

speak to agree. When they engage in mental imagery, they experience something like weak perception too. But not everyone. Some people report that they have no visual experiences associated with mental imagery (Zeman et al. 2015).

What should we make of this disagreement? Is one camp making an introspective error? If so, which one? Or are both camps introspecting accurately and the disagreement stems from variation across the human population?

Answering this question would not be such a problem if we had well-validated non-introspective methods for detecting consciousness. But here we run into another difficulty. Not only is introspection a private method, in consciousness science its targets are private too. States of consciousness are subjective in nature and introspection provides the only direct epistemic access to states of consciousness.

Those who are sceptical about the use of introspection in consciousness science worry that these complications make detecting and correcting for introspective errors in consciousness science either impossible in principle or at least unachievable in practice (Irvine 2021; Goldman 2004; Schwitzgebel 2012; Piccinini 2009; Spener 2015). As Irvine puts it:

...introspection is proposed to be the only method of investigating conscious phenomena... having no other methods with which to compare introspective methods, there is no clear way of establishing when introspective errors are made. (2012: 634)

3. Michel on Calibration in Consciousness Science

In a recent paper Matthias Michel has argued that these sceptical worries about the possibility of detecting introspective errors and achieving “calibration in consciousness science” rest on the mistaken assumption that “one cannot calibrate a procedure by comparing its outcomes with those of a procedure that is known to be inaccurate” (2023: 836).² He argues that scientists have two lines of evidence for consciousness. One line of evidence stems from subjects’ introspective judgements—Type 2 procedures (Michel 2023) or subjective measures as they are often called (Irvine 2021; Spener 2020). The other line of evidence about consciousness stems from perceptual sensitivity to a stimulus as mea-

2. Michel discusses two strategies for calibrating detection procedures in consciousness science, concordance calibration and model calibration (see also Irvine 2021; Spener 2020). In concordance calibration errors are detected by comparing the indications of a measurement procedure with another procedure for measuring the same thing. In model calibration, researchers build a model of the measurement procedure itself, and compare the deliverances of the procedure to those of the model. Here I focus just on concordance calibration.

sured by the signal detection theoretic measure d' —Type 1 procedures (Michel 2023) or objective measures (Irvine 2021; Spener 2020). While neither of these procedures is immune to error, Michel points out that there is no contradiction in using defeasible sources of evidence to detect and correct for errors in other defeasible sources of evidence (Michel 2023: 836). The sceptic's mistake, he suggests, is to assume that detecting and correcting for introspective errors about consciousness would require comparing the outputs of introspection-based procedures with some already well-validated gold-standard, which of course we do not have.

I learnt a lot from Michel's discussion, and at the end of the day I agree with his conclusion: there is nothing so special about consciousness as an object of investigation that prevents us from detecting and correcting introspective errors via comparing them with non-introspective sources of evidence (Michel 2023: 839). But his discussion falls short of addressing the most pressing reasons for scepticism about the possibility of detecting introspective errors, and as a result, it is unlikely to persuade sceptics.

The two most pressing sources of scepticism, as I see it, are what I call the primacy worry and the arbitrariness worry. Neither assumes that concordance calibration requires comparison with a gold-standard. In the remainder of this section, I will unpack each of these concerns and explain why Michel's discussion falls short of alleviating them. I will explain how natural kind reasoning can overcome them in section 4.

3.1. *The Primacy Worry*

Michel claims that subjective and objective measures of consciousness are underpinned by independent "basic" epistemic principles. Introspection-based procedures are underpinned by the assumption that "people can usually tell whether they are conscious of something or not" (Michel 2023: 838). Objective procedures are underpinned by the assumption that the better one's perceptual system is able to extract information about a stimulus and make it available for use in guiding thought and action, the more likely it is that one is conscious of it (837).

One source of scepticism stems from the thought—not an unreasonable one in my view—that the assumption underlying objective measures of consciousness is not epistemically basic. It is derived from reflection on our own case, from introspection. We notice that in our own case there is a striking correlation between being conscious of a stimulus and being able to use information about it to guide thought and action. And we infer from this that perceptual sensitivity and availability for planning and action are a defeasible guide to conscious

perception.³ According to this view, introspection provides the only basic source of evidence about consciousness and “all other measures of consciousness are directly or indirectly derived from introspection” (Overgaard 2023; see also Goldman 2000; 2004).

Those who feel the force of the primacy worry do not doubt that defeasible measurement procedures can be used to calibrate other defeasible measurement procedures. What they doubt is that derivative measurement procedures can be used to detect and correct for errors in the procedures from which they were derived. I take it that this is what Morten Overgaard and Kristian Sandberg have in mind in the following passage:

How can we know, for instance, that any [objective] measure ... is actually about the subjective experience of interest—and more so than the subjective report? It seems the only knowledge we could have comes from a prior correlation with introspective observation and report and, accordingly, cannot have any higher precision than the introspective observation/report. (2021: 1–2)⁴

At first glance one might suppose that Michel’s discussion of the history of thermometry demonstrates that the primacy worry is misguided. But on closer examination it does not. Drawing on Hasok Chang’s influential work on the topic (2004), Michel points out that instruments for measuring temperature objectively were initially justified through their conformity with sensations of hot and cold. Despite this, it can be perfectly rational to take these instruments to be more accurate than sensations. To explain how, he points to an often-cited example involving a bucket of water. If you put one hand in a bucket of hot water and the other in cold water, and then put them both in a bucket of lukewarm water, the lukewarm water will feel both hot and cold at the same time. Your thermometer, however, will tell you that the water is actually quite uniform in temperature. In this case, the only reasonable thing to do is to take your thermometer to provide a more accurate guide than your perceptual judgements.

This example is useful. But there are some important disanalogies between it and the cases we care about in consciousness science. Firstly, in the bucket example, the principle of single value is doing a lot of work (Michel 2023). The principle of single value says that, quantum phenomena aside, things typically are not in several incompatible states at the same time. It would be very

3. The fact that Michel encourages readers to consider their own case when motivating the link between consciousness and perceptual sensitivity arguably supports this view (2023: 837, fn 12).

4. A version of this idea can also be found in Chalmers’ influential work (1996: 238–242).

odd, for example, for a bucket of water to be both hot and cold simultaneously. This means that when our left hand is telling us that the bucket is warm while our right is telling us it is cool, we can safely assume that at least one of them is wrong.

In the disputed cases in consciousness science the principle of single value is of little use. To be sure, if a subject were to report that they both were and were not conscious of a stimulus in a particular trial, then we could appeal to the principle of single value to infer that something has gone amiss with their introspective reports. But as Michel rightly points out, it is standard practice in consciousness science to collect many trials across many different subjects. And when comparing introspective judgements across trials and across subjects, we cannot help ourselves to the principle of single value since it is an open question how much subjects' experiences may vary from trial to trial (Hohwy 2011), and how much experiences may vary across subjects (Fink 2018).

Secondly, this example does not actually address the primacy worry. To address the primacy worry we need to explain why there is no contradiction in taking derivative sources of evidence to defeat the sources of evidence from which they were derived. The bucket example does not demonstrate this. In the bucket example our perceptual judgments are defeated by other perceptual judgments, not by the outputs of thermometers. One hand tells us that the bucket is warm while the other tells us that it is cold. But we know that buckets of water are typically uniform in temperature, so our perceptual judgments defeat one another.

Moreover, demonstrating that it is possible for derivative sources of evidence to defeat basic sources of evidence is what is required to resolve the disputed cases in consciousness science. Consider the case of blindsight. Blindsight subjects do not deliver self-defeating introspective reports. They consistently deny seeing the stimuli. The question is whether we should take the fact that information about the stimuli is clearly being used to guide their behaviour to defeat their introspective reports.

Michel's discussion does not engage with this question. Instead, he takes it for granted that the fact that objective measures diverge from introspective judgments in these cases gives us good reason to believe that above chance performance on discrimination tasks is not an accurate guide to consciousness in these cases (Michel 2023: 12).⁵ This leads us to a second source of sceptical worry: the arbitrariness worry.

5. Elsewhere Michel has offered a more detailed discussion of blindsight (Michel & Lau 2021) but as far as I can see the primacy worry remains unaddressed here too.

3.2. *The Arbitrariness Worry*

Another brand of sceptic is happy to grant that there is no in principle barrier to taking behavioural and neural evidence to defeat introspection. Instead, they worry that in the disputed cases in consciousness science—cases like blindsight—deciding which way to go will be arbitrary.

Consider again Michel's claim that blind sight provides researchers with a good reason to believe that objective measures of consciousness are not accurate in some cases. Why should we accept that? If subjective measures and objective measures really are independent measures of consciousness, then one might suppose that in cases like blindsight where they deliver conflicting results the appropriate response is to remain neutral. To see why it will be useful to have the concept of "collective defeat" from Pollock's work on defeasible reasoning on the table (1994). To use Pollock's own example, suppose Smith and Jones, two friends that you consider equally reliable, give you contradictory statements about whether or not it is raining outside. If you have no other evidence to go by and both Smith and Jones strike you as sincere, then the rational position to take is one of neutrality. You should withhold belief. Neither believing that it is raining, nor that it is not. Siding with either Smith or Jones, in this case would be arbitrary. Similarly, one might suppose that in the case of blindsight, the appropriate response is to remain neutral and that siding with either introspective or behaviour evidence would be arbitrary (Irvine 2021).

But one might think, that is a bit too quick. Not all sources of evidence carry the same evidential weight. To return to Pollock's example, it would not be arbitrary to side with Smith if he had actually been outside while Jones had only read the weather report. And as a matter of fact, there are a range of cases where objective measures diverge from subjective measures and yet it is pretty clear how we should proceed. Take Anton syndrome for instance. Patients with Anton syndrome display profound visual deficits accompanied by lesions to the occipital cortex, but when asked about their visual experience they deny having any deficits and confabulate details of the visual scene around them (Maddula et al. 2009). In this case, introspection-based evidence conflicts with non-introspective evidence, and yet, no one thinks that neutrality is the appropriate stance to take. In cases like this it is pretty clear that patients' reports are not to be trusted (Pauen 2023).⁶

But here is where the real force of the arbitrariness worry comes in. Not all cases in consciousness science are this straightforward. There is genuine disagreement, for example, as to whether introspective evidence or behavioural evidence carries more evidential weight in cases like blindsight. What is more,

6. Thanks to an anonymous referee for suggesting this example.

whether or not a group of researchers takes behavioural measures to defeat introspection-based measures is likely to be tightly correlated with which theory of consciousness they pre-theoretically find most attractive. This makes converging toward consensus in consciousness science particularly difficult (Irvine 2013; 2017; 2021; see also Phillips 2018).

So, while there is a lot to like about Michel's discussion, it has not addressed the central worries fuelling scepticism. He has not explained why there need be no contradiction in taking derivative sources of evidence to defeat the sources of evidence from which they are derived. Nor has it been explained why there need be nothing arbitrary in deciding how to proceed in cases where introspective reports diverge from non-introspective sources of evidence. As a result, it would be understandable if sceptics remain unconvinced.

4. Iterative Natural Kind Reasoning: A Method for Detecting Measurement Errors

Natural kind reasoning involves the iterative application of inference to the best explanation to home in on and leverage regularities in nature. It can provide a generalizable epistemic framework for thinking about the development of novel measurement procedures that can address the sceptical worries discussed in the previous section. It is closely related to Chang's notion of epistemic iteration, however Chang himself does not directly appeal to explanatory considerations (2004; 2017).⁷ And it has been discussed as a framework for thinking about the detection of consciousness in disorders of consciousness (Shea & Bayne 2010, Bayne & Shea 2020), non-human animals (Birch 2022), artificial systems (Dung 2023; Mckilliam forthcoming) and in the absence of cognitive access (Shea 2012; see also Mckilliam 2024; Bayne et al. 2024).

Applied to the case of thermometry and painting with a broad brush, natural kind reasoning suggests the following two-step strategy. The first step involves inferring that perceptual judgements of temperature and thermometers are measuring the same phenomenon. We note that there is a striking correlation between how hot or cold something feels to the touch and how much it causes fluids to expand or contract. And we reason that this is best explained by the existence of a natural phenomenon—in this case temperature—that causes both i) sensations of hot and cold, and ii) fluids to expand and contract.

7. Chang discusses explanatory virtues—simplicity, fertility, neatness, explanatory power, etc—in the context of establishing the degree to which successful epistemic iteration qualifies as progress (2004: 226–228). However, it is unclear whether he thinks explanatory considerations play a role in the process of epistemic iteration itself.

In the second step we deploy inference to the best explanation once again to decide between competing explanations for why our perceptual judgement diverges from the thermometer in a particular case. If the best explanation for divergence is that something has gone wrong with our perceptual judgment, then that is what we should conclude.

This framework can allow us to detect measurement errors even in situations where initial sources of evidence are not self-defeating and where the principle of single value cannot be safely assumed. Consider a modified version of the bucket example from earlier. First you dip your hand in one bucket and record your perceptual judgement: it feels cold. At the same time, you also measure it with your thermometer. Then you dip your hand into a second bucket of water and note that this one feels considerably warmer. When you check it with your thermometer however, your thermometer suggests that both buckets are the same temperature. In this case there is nothing internally inconsistent about your perceptual judgements—they do not defeat one another—and we cannot safely assume the principle of single value—the buckets very well may be different temperatures. But we can still leverage inference to the best explanation to decide how to proceed.

One potential explanation for the disagreement between your perceptual judgement and the thermometer is that perceptual judgements of temperature and thermometer readings do not have a common cause. But if that were the case, then the striking correlation between the two would be, as Ian Hacking would say, a preposterous coincidence (1983). Another potential explanation is that something has gone wrong with this particular thermometer reading. But if no confounding factor can be identified that would explain why the usually reliable thermometer got it wrong in this particular case, the best explanation for the divergence in this case is that you have made a perceptual error.

Applied to the study of consciousness, the natural kind method suggests the following strategy. 1) Start with subjects' introspective reports as a defeasible guide to consciousness. 2) Look for behavioural capacities and neural processes that are systematically correlated with introspective reports in a way that is best explained by a common underlying mechanism. 3) Leverage evidence of that underlying mechanism to decide how to proceed in cases where introspective reports diverge from behavioural and neural evidence.

This framework has the resources to address both the arbitrariness worry and the primacy worry discussed in the previous section. There is no contradiction in taking derivative sources of evidence to override the sources of evidence from which they were derived when doing so provides the best explanation of all the data. And we can appeal to explanatory considerations to determine how to proceed in cases where introspective evidence diverges from behavioural and neural evidence.

In the remainder of the paper, I will demonstrate how this framework can be applied to a particular case in consciousness science—introspective judgements about conscious mental imagery. This is an interesting case for a number of reasons. First, a failure to resolve disagreements about mental imagery played a role in the collapse of introspectionist psychology a century ago (Boring 1953). And second, until recently, mental imagery has been seen by some as a paradigm example of the unreliability of introspection as a source of data for consciousness science (Schwitzgebel 2011). Experimental results from the last few years suggest otherwise.

5. Detecting Introspective Errors about Conscious Mental Imagery

The idea that the vividness of mental imagery varies from person to person, and can even be entirely absent is not new. It dates back at least to Francis Galton (1880). However, the past decade has seen a considerable resurgence of interest in the topic and the term ‘aphantasia’ has been coined to refer to those who report an absence of any conscious experiences associated with acts of mental imagery (Zeman et al. 2015).

Today, the vividness of mental imagery is most commonly assessed via the Vividness of Visual Imagery Questionnaire (VVIQ). The VVIQ instructs participants to imagine a range of scenes and then answer a range of questions relating to the vividness of any associated experiences of mental imagery on a 5-point scale ranging from 5: “Perfectly clear and as vivid as normal vision” through 1: “No image at all, you only “know” that you are thinking of an object” (Marks 1973).

As one might expect, self-reported aphantasics tend to score quite low on the VVIQ—typically <20 out of a possible 80. Initially there were serious doubts about the reliability of the VVIQ and self-reports about mental imagery in general (Schwitzgebel 2011). However, in recent years a number of striking correlations are beginning to emerge between introspective judgements about the vividness of mental imagery and a range of physiological effects.

For example, it has been known for some time that mental imagery can prime which stimuli resolves first in binocular rivalry (Pearson et al. 2008). If you are shown a grid of vertical green lines to your left eye, and a grid of horizontal red lines to your right eye, the resulting perceptual experience is not a mash of the two, but rather oscillates between them. First, you see the red grid, say, then the green grid, then the red grid again. Imagining a green grid prior to rivalry increases the probability that you will see the green grid first while imagining the red grid makes it more likely that you will see the red grid first. In a 2018 study, Keogh and Pearson found that this effect is absent in people who report

no experience of mental imagery. Self-diagnosed aphantasics “showed almost no imagery-based rivalry priming” (Keogh & Pearson 2018: 53).

A second study by the same group of researchers found that aphantasic individuals exhibit a dampened fear response—as measured via skin conductance levels—when reading scary stories (Wicken et al. 2021). While their skin conductance levels were not significantly different from non-aphantasics when viewing scary images, when they read scary stories, their fear response was considerably lower than that of individuals who were capable of forming and experiencing mental images of what they read. This finding fits with the hypothesis that mental imagery serves as an “emotional amplifier” (Holmes et al. 2008), and suggests further correlations might be found between imagery capacities and various psycho-pathologies such as post-traumatic stress disorder. If every time you thought about a past traumatic event the thoughts were accompanied by vivid imagery, we might expect them to be more traumatic than if there was no imagery at all (Ji et al. 2019).

A more recent finding suggests that the brightness of an imagined object is correlated with pupil dilation in much the same way that pupil size is correlated with the brightness of a perceptual object, but not for aphantasics (Kay et al. 2022). While aphantasics display normal pupil responses to perceptual stimuli and cognitive load, they do not display any significant pupil responses when instructed to imagine those same perceptual stimuli.

It is still early days in this research and it is worth noting that not all recent findings are pointing in the same direction here. For example, a recent study by Pounder and colleagues found that Shepard and Metzler’s (1971) classic finding—the time it takes to complete a mental rotation task is correlated with how far one has to mentally rotate the image—was conserved in a group who scored low on the VVIQ (<32) (Pounder et al. 2022). Whether this will replicate with those at the very extreme end of the low score (<20) is unclear. Nor is it entirely clear how to square this with the fact that aphantasics often report deploying non-imagistic strategies when completing mental rotation tasks, and these reports have been corroborated by the fact that their responses are not affected by the oblique effect (Keogh & Pearson 2021).

Despite the fact that many questions remain open, what is emerging is a cluster of correlations linking subjects’ introspective reports of the vividness of mental imagery as measured via the VVIQ and a cluster of physiological effects. What might best explain these correlations? One attractive thought is that there exists a distinct kind of information processing—presumably overlapping with a kind of perceptual processing—that i) facilitates conscious experiences of mental imagery and also ii) primes for subsequent rivalry iii) amplifies emotional responses iv) impacts the size of one’s pupils, and who knows what else.

If there is a distinct form of information processing associated with conscious mental imagery, then we can expect further details to emerge concerning its functional profile and neural basis. But even these early findings give us a battery of tests that we can use to detect introspective errors about experiences of mental imagery. In the remainder of the paper, I sketch two scenarios to outline how this might go.

5.1. Scenario 1—Detecting False Positives

A measurement procedure delivers a false positive when it indicates that a particular condition is present when in fact it is not. When I engage in mental imagery it seems to me that I have something like a weak perceptual experience. This introspective judgment would be a false positive if, in fact, I did not have anything like a weak perceptual experience when engaging in mental imagery. Relatedly, if I score high on the VVIQ when in fact I have no conscious mental imagery, then in this particular case the VVIQ would have delivered a false positive. Can we detect introspective false positives?

Suppose we found that my engaging in mental imagery does not produce any priming effects, and neither does it raise my skin conductance levels above baseline or have any significant effect on the size of my pupils. A number of hypotheses might explain this. One possibility is that we were too quick in drawing conclusions about the functional profile of conscious mental imagery. Perhaps, it is not the case that the kind of information processing that gives rise to experiences of mental imagery also primes for subsequent rivalry, impacts pupil size, and amplifies emotional responses. But if that is true, then the striking correlation between reports of imagery and these physiological effects in others remains unexplained. Another possibility is that something has gone wrong with the experiments. Perhaps, my mental imagery really is eliciting a priming effect and a fear response, and it really is causing my pupils to dilate, we just have not found evidence for these effects due to experimental error. But if we cannot find anything that would explain why these paradigms are failing to find these effects in me, despite finding them in others, then the weight of evidence begins to shift toward a third hypothesis: I am simply mistaken about my own experiences of mental imagery.

True, we cannot be certain of that hypothesis. But that is no different from the situation we find ourselves in elsewhere in science. Unless we insist that introspective judgements about mental imagery are infallible, we should remain open to the possibility that in certain situations, the only sensible conclusion to draw is that subjects are simply wrong about their own conscious experiences.

5.2. Scenario 2—*Detecting False Negatives*

A measurement procedure delivers a false negative when it indicates that a particular condition is absent when in fact it is present. Suppose that when someone engages in mental imagery it seems to them that they had no associated imagery experience. This introspective judgment would be a false negative if, when engaging in mental imagery, they do in fact have something like a weak perceptual experience. Similarly, if someone scores very low on the VVIQ despite having rich sensory experiences associated with acts of imagery, then the VVIQ will have delivered a false negative. Can we detect introspective false negatives?

In this case, things are a little more complicated. Suppose that someone introspectively judges that they have no visual experiences when they attempt to engage in mental imagery and as a result scores low on the VVIQ, but when we run them through our battery of physiological tests, we find that they test positive on all three. Instructing them to engage in mental imagery primes which of two rivalrous stimuli resolves first, reading scary scenes raises their skin conductance levels just as it does with those who experience mental imagery, and their pupils dilate whenever they imagine bright objects. Suppose also that we have ruled out experimental error and the hypothesis that experiences of mental imagery and these physiological effects do not have a common cause. Two hypotheses remain at least *prima facie* plausible. One hypothesis is that they have made an introspective error. They really are consciously experiencing mental imagery, they are just wrong when they say they are not. The other appeals to the possibility of “unconscious mental imagery” (Nanay 2020). According to this hypothesis, the participant is right—they have no conscious experiences of mental imagery—the imagery-based physiological effects are elicited instead by a kind of information processing that can be entirely unconscious.

The idea is not without warrant. There are a range of findings that suggest that perception can be both conscious and unconscious (Peters, Kentridge, Phillips, & Block 2017). This is hotly debated of course, but it provides at least conditional support for unconscious imagery. If unconscious perception is possible why not unconscious imagery? There are also some experimental results that can be interpreted as supporting the existence of unconscious mental imagery. It turns out that while imagining a stimulus primes for subsequent rivalry, actively trying not to imagine the stimulus does too. When participants are prompted with ‘red apple’ subsequent rivalry is primed irrespective of whether the participant imagines a red apple or refrains from imagining a red apple (Kwok et al. 2019). Bence Nanay takes these results to “strongly indicate that it is unconscious imagery [rather than conscious mental imagery] that primes binocular rivalry” (2020: 5). This is not the only interpretation of these results of course. But it is not an unreasonable one.

So, we face a puzzle—a puzzle that will be familiar to those working on the methodological challenges associated with measurement in consciousness science (see for example Block 2007; Shea 2012; Philips 2018). Has our subject erroneously introspected that they have no conscious imagery when in fact they do? Or have they accurately introspected that they have no conscious imagery, and this cluster of physiological effects are elicited by unconscious imagery? Can we even answer this question?

We can. The hypothesis that our subject engages in mental imagery that they do not consciously experience predicts that there is further variation within the population. They would not belong with the aphantasics, for they do not display either imagery-based priming or the imagery-based fear response. But nor would they belong with non-aphantasics either since non-aphantasics experience conscious mental imagery and our hypothetical subject does not. Our hypothetical subject would belong to a third category of people, those who can engage in mental imagery, but only unconsciously.

Whether or not this further form of variation within the population exists is, of course, something we can test empirically. If there is this further form of variation within the population then we ought to be able to find corroborating variation at the cognitive and neurobiological level. As it stands there is insufficient evidence to confidently say whether or not this additional variation exists, but either way the evidence turns, we will be able to appeal to inference to the best explanation to determine whether or not our subject's introspective judgment has erred.

Suppose that after a thorough search, no evidence corroborating this further variation is forthcoming. Some people can engage in conscious mental imagery, and some people cannot engage in mental imagery at all, but there does not seem to be anyone who can engage in mental imagery only unconsciously. In this case, the best explanation for all the data will be that our subject's introspective judgements about their own mental imagery are simply mistaken.

Alternatively, suppose we do find corroborating evidence of this additional form of variation at the cognitive and neurobiological level. In that case, we can ask whether or not our subject exhibits the relevant variation. If they do, if our subject exhibits the characteristic features of someone who can engage in mental imagery but only unconsciously, then their introspective judgements about mental imagery would be corroborated. If they do not, if they exhibit all the neural and physiological characteristics of someone who has conscious mental imagery, then we should conclude that their introspective judgement has erred and that they are simply wrong about their own experiences of mental imagery.

6. Conclusion

I have argued that natural kind reasoning—the iterative application of inference to the best explanation to home in on regularities in nature—can be leveraged to detect and correct for introspective errors in consciousness science. If you accept that subjective experience is systematically related to cognition and neural mechanisms in some way or other, then you should also accept that the iterative application of inference to the best explanation should allow us to home in on this systematic relationship and leverage it to detect introspective errors. I have demonstrated how this can work in the case of conscious mental imagery, but the strategy itself is a general one. This is a cause for optimism, not just about the use of introspection in consciousness science but about the prospects of consciousness science more generally. It suggests that the persistent worries that consciousness science faces insurmountable methodological challenges are misguided. We do not need any radical new methods to resolve disputes over the relationship between consciousness and cognition. The iterative application of good old inference to the best explanation will work just fine. Of course, in practice, detecting and controlling for introspective errors will be extremely difficult, but progress in science rarely comes easy and we should expect progress in consciousness science to be no different.

References

- Bayne, Tim (2015). Introspective Insecurity. In Thomas Metzinger & Jennifer M. Windt (Eds.), *Open MIND*. MIT Press. <https://doi.org/10.7551/mitpress/10603.003.0010>
- Bayne, Tim, Anil K. Seth, Marcello Massimini, Joshua Shepherd, Axel Cleeremans, Stephen M. Fleming, Rafael Malach, Jason B. Mattingley, David K. Menon, Adrien Owen, Megan A. K. Peters, Adeel Razi, and Liad Mudrik (2024). Tests for Consciousness in Humans and Beyond. *Trends in Cognitive Sciences* 28(5), 454–466. <https://doi.org/10.1016/j.tics.2024.01.010>
- Bayne, Tim and Nicholas Shea (2020). Consciousness, Concepts, and Natural Kinds. *Philosophical Topics*, 48(1), 65–84. <https://doi.org/10.5840/philtopics20204814>
- Bayne, Tim and Maya Spener (2010). Introspective Humility. *Philosophical Issues*, 20, 1–22. <https://doi.org/10.1111/j.1533-6077.2010.00176.x>
- Birch, Jonathan (2022). The Search for Invertebrate Consciousness. *Noûs*, 56(1), 133–153. <https://doi.org/10.1111/nous.12351>
- Block, Ned (2007). Consciousness, Accessibility, and the Mesh Between Psychology and Neuroscience. *Behavioral and Brain Sciences*, 30(5–6), 481–499. <https://doi.org/10.1017/S0140525X07002786>
- Boring, Edwin G. (1953). A History of Introspection. *Psychological Bulletin*, 50(3), 169–189. <https://doi.org/10.1037/h0090793>

- Brown, Richard, Hakwan Lau, and Joseph E. LeDoux (2019). Understanding the Higher-Order Approach to Consciousness. *Trends in Cognitive Sciences*, 23(9), 754–768. <https://doi.org/10.1016/j.tics.2019.06.009>
- Chalmers, David (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.
- Chalmers, David (2003). The Content and Epistemology of Phenomenal Belief. In Quentin Smith and Aleksandar Jokic (Eds.), *Consciousness: New Philosophical Perspectives* (220–271). Oxford University Press.
- Chalmers, David (2004). How Can We Construct a Science of Consciousness? In Michael Gazzaniga (Ed.), *The Cognitive Neurosciences iii* (1111–1119). MIT Press.
- Chalmers, David (2010). *The Character of Consciousness*. Oxford University Press.
- Chang, Hasok (2004). *Inventing Temperature: Measurement and Scientific Progress*. Oxford University Press. <https://doi.org/10.1093/0195171276.001.0001>
- Chang, Hasok (2017). Epistemic Iteration and Natural Kinds: Realism and Pluralism in Taxonomy. In Keneth S. Kendler and Josef Parnas (Eds.), *Philosophical Issues in Psychiatry IV: Classification of Psychiatric Illness* (229–245). Oxford University Press. <https://doi.org/10.1093/med/9780198796022.003.0029>
- Cowey, Alan (2010). The Blindsight Saga. *Experimental Brain Research*, 200(1), 3–24. <https://doi.org/10.1007/s00221-009-1914-2>
- Dung, Leonard (2023). Tests of Animal Consciousness are Tests of Machine Consciousness. *Erkenntnis*. <https://doi.org/10.1007/s10670-023-00753-9>
- Galton, Francis (1880). Statistics of Mental Imagery. *Mind*, 5(19), 301–318.
- Gertler, Brie (2012). Renewed Acquaintance. In Declan Smithies and Daniel Stoljar (Eds.), *Introspection and Consciousness* (89–123). <https://doi.org/10.1093/acprof:oso/9780199744794.003.0004>
- Goldman, Alvin (2000). Can Science Know When You’re Conscious? Epistemological Foundations of Consciousness Research. *Journal of Consciousness Studies*, 7(5), 3–22.
- Goldman, Alvin (2004). Epistemology and the Evidential Status of Introspective Reports. *Journal of Consciousness Studies*, 11(7–8), 1–16.
- Hacking, Ian (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511814563>
- Hohwy, Jakob (2011). Phenomenal Variability and Introspective Reliability. *Mind & Language*, 26(3), 261–286. <https://doi.org/10.1111/j.1468-0017.2011.01418.x>
- Holmes, Emily A., John R. Geddes, Francesc Colom, and Guy M. Goodwin (2008). Mental Imagery as an Emotional Amplifier: Application to Bipolar Disorder. *Behaviour Research and Therapy*, 46(12), 1251–1258. <https://doi.org/10.1016/j.brat.2008.09.005>
- Horgan, Terry and Uriah Kriegel (2007). Phenomenal Epistemology: What Is Consciousness That We May Know It So Well? *Philosophical Issues*, 17, 123–144. <https://doi.org/10.1111/j.1533-6077.2007.00126.x>
- Hurlburt, Russell and Eric Schwitzgebel (2007). *Describing Inner Experience?* MIT Press. <https://doi.org/10.7551/mitpress/7517.001.0001>
- Irvine, Elizabeth (2012). Old Problems With New Measures in the Science of Consciousness. *The British Journal for the Philosophy of Science*, 63(3), 627–648 <https://doi.org/10.1093/bjps/axso19>
- Irvine, Elizabeth (2013). *Consciousness as a Scientific Concept: A Philosophy of Science Perspective*. Springer. <https://doi.org/10.1007/978-94-007-5173-6>

- Irvine, Elizabeth (2021). Developing Dark Pessimism Towards the Justificatory Role of Introspective Reports. *Erkenntnis*, 86(6), 1319–1344. <https://doi.org/10.1007/s10670-019-00156-9>
- Jack, Anthony and Andreas Roepstorff (2002). Introspection and Cognitive Brain Mapping: From Stimulus–Response to Script–Report. *Trends in Cognitive Sciences*, 6(8), 333–339. [https://doi.org/10.1016/s1364-6613\(02\)01941-1](https://doi.org/10.1016/s1364-6613(02)01941-1)
- Ji, Julie L., David J. Kavanagh, Emily A. Holmes, Colin MacLeod, and Martina Di Simplicio (2019). Mental Imagery in Psychiatry: Conceptual & Clinical Implications. *CNS Spectrums*, 24(1), 114–126. <https://doi.org/10.1017/S1092852918001487>
- Kay, Lachlan, Rebecca Keogh, Thomas Andrillon, and Joel Pearson (2022). The Pupillary Light Response as a Physiological Index of Aphantasia, Sensory and Phenomenological Imagery Strength. *Elife*, 11, e72484. <https://doi.org/10.7554/eLife.72484>
- Keogh, Rebecca and Joel Pearson (2018). The Blind Mind: No Sensory Visual Imagery in Aphantasia. *Cortex*, 105, 53–60. <https://doi.org/10.1016/j.cortex.2017.10.012>
- Keogh, Rebecca, Marcus Wicken, and Joel Pearson (2021). Visual Working Memory in Aphantasia: Retained Accuracy and Capacity With a Different Strategy. *Cortex*, 143, 237–253. <https://doi.org/10.1016/j.cortex.2021.07.012>
- Kouider, Sid, Vincent De Gardelle, Jérôme Sackur, and Emmanuel Dupoux (2010). How Rich is Consciousness? The Partial Awareness Hypothesis. *Trends in Cognitive Sciences*, 14(7), 301–307. <https://doi.org/10.1016/j.tics.2010.04.006>
- Kriegel, Uriah (2013). A Hesitant Defense of Introspection. *Philosophical Studies*, 165(3), 1165–1176. <https://doi.org/10.1007/s11098-013-0148-0>
- Kwok, Eugene, Gaele Leys, Roger Koenig-Robert, and Joel Pearson (2019). Measuring Thought-Control Failure: Sensory Mechanisms and Individual Differences. *Psychological Science*, 30(6), 811–821. <https://doi.org/10.1177/0956797619837204>
- Maddula, Mohana, Stuart Lutton, and Breffni Keegan (2009). Anton’s Syndrome Due to Cerebrovascular Disease: A Case Report. *Journal of Medical Case Reports*, 3, 1–3.
- Marks, David (1973). Visual Imagery Differences in the Recall of Pictures. *British Journal of Psychology*, 64(1), 17–24. <https://doi.org/10.1111/j.2044-8295.1973.tb01322.x>
- Mckilliam, Andy (2024). Natural Kind Reasoning in Consciousness Science: An Alternative to Theory Testing. *Noûs*, 1–18. <https://doi.org/10.1111/nous.12526>
- Mckilliam, Andy (forthcoming). Do Mechanisms Matter for Inferences About Consciousness. *Australasian Journal of Philosophy*.
- Mégevand, Pierre, David M. Groppe, Matthew S. Goldfinger, Sean T. Hwang, Peter B. Kingsley, Ido Davidesco, and Ashesh D. Mehta (2014). Seeing Scenes: Topographic Visual Hallucinations Evoked by Direct Electrical Stimulation of the Parahippocampal Place Area. *Journal of Neuroscience*, 34(16), 5399–5405. <https://doi.org/10.1523/jneurosci.5202-13.2014>
- Michel, Matthias (2023). Calibration in Consciousness Science. *Erkenntnis*, 88(2), 829–850. <https://doi.org/10.1007/s10670-021-00383-z>
- Michel, Matthias and Hakwan Lau (2021). Is Blindsight Possible Under Signal Detection Theory? Comment on Phillips (2021). *Psychological Review*, 128(3), 585–591. <https://doi.org/10.1037/rev0000266>
- Nanay, Bence (2020). Unconscious Mental Imagery. *Philosophical Transactions of the Royal Society B*, 376(20190689). <https://doi.org/10.1098/rstb.2019.0689>
- Overgaard, Morten (2023). Methodological Reductionism or Methodological Dualism? In Search of a Middle Ground. *Phenomenology and the Cognitive Sciences*. <https://doi.org/10.1007/s11097-023-09939-6>

- Overgaard, Morten and Kristian Sandberg (2021). The Perceptual Awareness Scale—Recent Controversies and Debates. *Neuroscience of Consciousness*, 2021(1), niabo44. <https://doi.org/10.1093/nc/niabo44>
- Papineau, David (2002). *Thinking About Consciousness*. Clarendon Press. <https://doi.org/10.1093/0199243824.001.0001>
- Pauen, Michael (2023). Mental Measurement and the Introspective Privilege. *Phenomenology and the Cognitive Sciences*. <https://doi.org/10.1007/s11097-023-09931-0>
- Pauen, Michael and John-Dylan Haynes (2021). Measuring the Mental. *Consciousness and Cognition*, 90, 103106. <https://doi.org/10.1016/j.concog.2021.103106>
- Pearson, Joel, Colin W. G. Clifford, and Frank Tong (2008). The Functional Impact of Mental Imagery on Conscious Perception. *Current Biology*, 18(13), 982–986. <https://doi.org/10.1016/j.cub.2008.05.048>
- Pearson, Joel, Thomas Naselaris, Emily A. Holmes, and Stephen M. Kosslyn (2015). Mental Imagery: Functional Mechanisms and Clinical Applications. *Trends in Cognitive Sciences*, 19(10), 590–602. <https://doi.org/10.1016/j.tics.2015.08.003>
- Peters, Megan A. K., Robert W. Kentridge, Ian Phillips, and Ned Block (2017). Does Unconscious Perception Really Exist? Continuing the ASSC20 debate. *Neuroscience of Consciousness*, 3(1), nix015. <https://doi.org/10.1093/nc/nix015>
- Phillips, Ian (2018). The Methodological Puzzle of Phenomenal Consciousness. *Philosophical Transactions of the Royal Society B*, 373, 20170347. <https://doi.org/10.1098/rstb.2017.0347>
- Phillips, Ian (2021). Blindsight is Qualitatively Degraded Conscious Vision. *Psychological Review*, 128(3), 558–584. <https://doi.org/10.31234/osf.io/gdk6m>
- Piccinini, Gualtiero (2003). Epistemic Divergence and the Publicity of Scientific Methods. *Studies in History and Philosophy of Science Part A*, 34(3), 597–612. [https://doi.org/10.1016/s0039-3681\(03\)00049-9](https://doi.org/10.1016/s0039-3681(03)00049-9)
- Piccinini, Gualtiero (2009). First Person Data, Publicity and Self-Measurement. *Philosophers' Imprint* 9, 1–16. <http://hdl.handle.net/2027/spo.3521354.0009.009>
- Pollock, John L. (1994). Justification and Defeat. *Artificial Intelligence*, 67(2), 377–407. [https://doi.org/10.1016/0004-3702\(94\)90057-4](https://doi.org/10.1016/0004-3702(94)90057-4)
- Pounder, Zoë, Jane Jacob, Samuel Evans, Catherine Loveday, Alison F. Eardley, and Juha Silvanto (2022). Only Minimal Differences Between Individuals With Congenital Aphantasia and Those With Typical Imagery on Neuropsychological Tasks That Involve Imagery. *Cortex*, 148, 180–192. <https://doi.org/10.1016/j.cortex.2021.12.010>
- Schwarz, Norbert (1999). Self-Reports: How the Questions Shape the Answers. *American Psychologist*, 54(2), 93–105. <https://doi.org/10.1037/0003-066x.54.2.93>
- Schwitzgebel, Eric (2008). The Unreliability of Naive Introspection. *Philosophical Review*, 117(2), 245–273. <https://doi.org/10.1215/00318108-2007-037>
- Schwitzgebel, Eric (2011). *Perplexities of Consciousness*. MIT press. <https://doi.org/10.7551/mitpress/8243.001.0001>
- Schwitzgebel, Eric (2012). Introspection, What? In Declan Smithies and Daniel Stoljar (Eds.), *Introspection and Consciousness* (29–48). Oxford Academic. <https://doi.org/10.1093/acprof:oso/9780199744794.003.0001>
- Seth, Anil, Zoltán Dienes, Axel Cleeremans, Morten Overgaard, and Luiz Pessoa (2008). Measuring Consciousness: Relating Behavioural and Neurophysiological Approaches. *Trends in Cognitive Sciences*, 12(8), 314–321. <https://doi.org/10.1016/j.tics.2008.04.008>

- Shea, Nicholas (2012). Methodological Encounters with the Phenomenal Kind. *Philosophy and Phenomenological Research*, 84(2), 307–344. <https://doi.org/10.1111/j.1933-1592.2010.00483.x>
- Shea, Nicholas and Tim Bayne (2010). The Vegetative State and the Science of Consciousness. *British Journal of Philosophy of Science*, 61(3), 459–484. <https://doi.org/10.1093/bjps/axp046>
- Shepard, Roger N. and Jacqueline Metzler (1971). Mental Rotation of Three-Dimensional Objects. *Science*, 171(3972), 701–703. <https://doi.org/10.1126/science.171.3972.701>
- Spener, Maja (2015). Calibrating Introspection. *Philosophical Issues*, 25, 300–321. <https://doi.org/10.1111/phis.12062>
- Spener, Maja (2020). Consciousness, Introspection, and Subjective Measures. In Uriah Kriegel (Ed.), *The Oxford Handbook of the Philosophy of Consciousness* (610–635). Oxford University Press.
- Timmermans, Bert and Axel Cleeremans (2015). How Can We Measure Awareness? An Overview of Current Methods. In Morten Overgaard (Ed.), *Behavioral Methods in Consciousness Research* (21–46). Oxford Academic. <https://doi.org/10.1093/acprof:oso/9780199688890.003.0003>
- Velmans, Max. (2009). *Understanding Consciousness*. Routledge. <https://doi.org/10.4324/9780203882726>
- Wicken, Marcus, Rebecca Keogh, and Joel Pearson (2021). The critical role of mental imagery in human emotion: Insights from fear-based imagery and aphantasia. *Proceedings of the royal society B*, 288(1946), 20210267. <https://doi.org/10.1098/rspb.2021.0267>
- Zeman, Adam, Michaela Dewar, and Sergio Della Sala (2015). Lives Without Imagery – Congenital Aphantasia. *Cortex*, 73, 378–380. <https://doi.org/10.1016/j.cortex.2015.05.019>