

The empirical case against introspection

Rik Peels

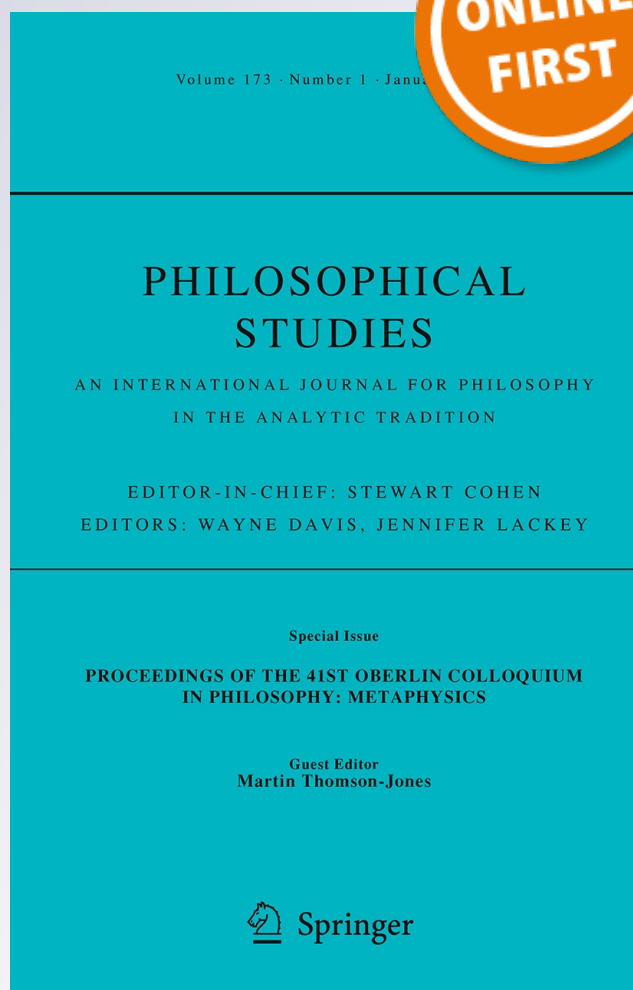
Philosophical Studies

An International Journal for Philosophy
in the Analytic Tradition

ISSN 0031-8116

Philos Stud

DOI 10.1007/s11098-016-0623-5



Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media Dordrecht. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

The empirical case against introspection

Rik Peels¹

© Springer Science+Business Media Dordrecht 2016

Abstract This paper assesses five main empirical scientific arguments against the reliability of belief formation on the basis of introspecting phenomenal states. After defining ‘reliability’ and ‘introspection’, I discuss five arguments to the effect that phenomenal states are more elusive than we usually think: (1) the argument on the basis of differences in introspective reports from differences in introspective measurements; (2) the argument from differences in reports about whether or not dreams come in colours; (3) the argument from the absence of a correlation between visual imagery ability and the performance on certain cognitive tasks; (4) the argument from our unawareness of our capacity of echolocation; (5) the argument from inattentional blindness and change blindness. I argue that the experiments on which these arguments are based do not concern belief formation on the basis of introspection in the first place or fail to show that it is unreliable, even when limited to introspection of phenomenal states.

Keywords Belief formation · Introspection · Phenomenal states · Reliability

The introspection of current conscious experience, far from being secure, nearly infallible, is faulty, untrustworthy, and misleading, not just sometimes a little mistaken, but frequently and massively mistaken, about a great variety of issues.

(Schwitzgebel 2011, 129).

✉ Rik Peels
mail@rikpeels.nl

¹ Philosophy Department, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

1 Introduction

This paper assesses five important empirical scientific arguments against the reliability of introspecting one's phenomenal states. By 'empirical scientific arguments' I mean arguments that are somehow based on empirical scientific research. And I take introspection to be a belief forming mechanism, so that by the 'reliability of introspecting one's phenomenal states', I mean the reliability of *belief formation* on the basis of introspecting one's phenomenal states.

As Anders Ericsson points out (Ericsson 2003, 1–7), the idea that introspection is unreliable can already be found with the early behaviourists, such as Watson (1913). According to many contemporary philosophers and scientists, though, what was merely a suspicion at the time of the early behaviourists has now been backed up by substantial empirical evidence [that this is the view of many contemporary philosophers and scientists is rightly pointed out by Overgaard (2006, 631)]. In response, some philosophers, such as Kriegel (2013), have provided an a priori defence of the thesis that *introspection is reliable* (IR). In this paper, I do not defend IR by giving arguments for it. Rather, I defend it by diffusing five recent empirical scientific arguments against IR. I focus on introspecting *phenomenal states* rather than introspecting such things as *beliefs* or *emotional states*. Below, I explain what phenomenal states amount to.

Whether or not introspection is reliable is important for at least four reasons. First, we usually take it for granted that we can reliably form true beliefs about, say, whether we are hungry, what we believe about a particular foreign policy, or how our environment visually appears to us. If introspection is unreliable, then we should discard what is widely considered to be one of the main common sense sources of knowledge (as Marcel 2003, 167–171, rightly notes).

Second, there is an important philosophical tradition, originating with Descartes (*Meditations* I and II) and Hume (book I of *Treatise on Human Nature*), carrying into the twentieth century (Price 1954, 3) and twenty-first century (Chalmers 2003), according to which introspection is particularly secure or even infallible. If introspection is unreliable, this philosophical tradition ought to be abandoned. As we shall see below, none of the five arguments that I discuss gives us good reason to abandon the infallibility thesis.

Third, an influential view in philosophy and science has it that only the natural sciences are a reliable source of knowledge (e.g. Rosenberg 2011). Elsewhere, I have called this view 'epistemological scientism' (Peels 2016). One important argument in favour of epistemological scientism is that empirical research has shown that IR is false. If this is *not* the case, then the argumentative support for epistemological scientism is weaker than several adherents of scientism think.

Fourth, there is an important debate about whether science should make use of people's first-person reports about their conscious experiences. According to Daniel Dennett, for instance, scientists should not. Rather, they should do what he calls 'heterophenomenology': people's verbal expressions of their beliefs about their conscious experiences should be taken as data to be explained.¹ Introspection itself,

¹ Remarkably, Dennett claims that introspection is unreliable *and* that we have perfect authority about what *seems* to be happening in us (Dennett 1991, 81, 86). This is remarkable, for such seemings seem

though, should not be relied on, for it is unreliable (see Dennett 1991, 2007). However, there is also a group of scientists who claim that science needs introspection for good psychological and neuroscientific research. They adopt different methods: neurophenomenology (Jack and Shallice 2001; Jack and Roepstorff 2003; Lutz and Thompson 2003), front-loaded phenomenology (Gallagher 2003), or descriptive experience sampling (Hurlburt and Heavey 2001). Each of these methods assumes that introspection is reliable. Whether or not introspection is reliable, then, makes an important difference to how we do science, especially psychology and certain branches of neuroscience.

In this article, I discuss five objections to the view that introspection of phenomenal states is generally reliable. This is not to deny that there may be *other* reasons, in addition to these five objections, not to use beliefs based on introspection as data in scientific research. Elizabeth Irvine, for instance, argues that behavioural methods, such as Signal Detection Theory measures, do as good a job as Ramsøy's and Overgaard's Perceptual Awareness Scale, that phenomenal clarity is *not* something that can provide a transferable metric across subjects because it is always evaluated relative to a particular aim or task [this point was already made by Bode (1913)], and that subjective measures, such as introspection, are inevitably biased (see Irvine 2012, 635–645; Irvine 2013, 286–290). Since those objections focus on the use of introspection of phenomenal states *in scientific research*, whereas the other five objections target the reliability of introspection of phenomenal states *in general*, I will focus on the latter here and leave an assessment of the former for another occasion.

The paper is structured as follows. First, I define 'reliability' (Sect. 2) and 'introspection' (Sect. 3). Subsequently, I discuss five arguments to the effect that introspection, of which introspecting phenomenal states is the main example in this paper, is unreliable: the argument on the basis of differences in introspective reports from differences in introspective measurements (Sect. 4), the argument from differences in reports about whether or not dreams come in colours (Sect. 5), the argument from the absence of a correlation between visual imagery ability and the performance on certain cognitive tasks (Sect. 6), the argument from our unawareness of our capacity of echolocation (Sect. 7), and, finally, the argument from inattentional blindness and change blindness (Sect. 8). I argue that the experiments on which these arguments are based do not concern introspection in the first place or fail to show that introspection is unreliable, even when limited to introspection of phenomenal states (Sect. 9).

I conclude that these empirical arguments provide no reason to abandon the idea that introspection of phenomenal states is an important source of knowledge, nor to discard the philosophical view that introspecting phenomenal states is infallible, nor to reject the use introspection of phenomenal states in psychology and neuroscience.

Footnote 1 continued

themselves to be phenomenal states. For further discussion on this, see Schwitzgebel (2007); Dennett (2007).

2 Reliability

When I talk about reliability, I mean the reliability of forming and maintaining *beliefs*. I take belief formation on the basis of introspection to be reliable just in case it delivers true belief in most cases, that is, if the process yields true beliefs dependably enough, including across a range of nearby counterfactual cases.² I will assume that we are talking about belief formation on the basis of introspection by an average, adult, properly functioning cognitive subject. This is because the empirical arguments are meant to show that *our* introspection is unreliable, not (merely) that of abnormal subjects, such as people with schizophrenia or subjects with serious brain damage.

Notice that reliability does not entail *indubitability* (the impossibility of doubting that p if p is true), *incorrigibility* (the impossibility for p to be false if one believes that p), or *self-intimation* (the impossibility of not believing that p when p is true). These are interesting issues, but here I focus on reliability.

Now, it seems that there are two important ways to show that introspection is unreliable. First, one could aim to demonstrate that introspection yields inconsistent results. One could, for instance, provide evidence for the thesis that introspection gives rise to conflicting beliefs upon repetition of the experiment in question or to conflicting results depending on factors that seem irrelevant to the truth of whether or not one is in a particular phenomenal state. Second, one could provide an argument to the effect that beliefs based on introspection conflict with what we know about the world or the subjects in the experimental setting in question from sources other than introspection. Below, we will encounter versions of both kinds of argument.

3 Introspection and phenomenal states

The literature on introspection is rife with controversy.³ Instead of delving into that debate, I will spell out three conditions that are widely taken to be necessary for introspection and that will be helpful in assessing the arguments against IR:

1. *First-Person Condition* It is belief formation about oneself rather than about someone else or the world outside oneself.
2. *Uniqueness Condition* It is belief formation that is uniquely available to oneself; others can form beliefs about you by way of empirical investigation, testimony, or visual perception, but they cannot form beliefs about you by way of introspection.
3. *Temporal Proximity Condition* It is belief formation about one's current or immediately past self; beliefs about one's future self or one's past self that is

² Precisely how often a mechanism should deliver true belief in order to count as reliable is a question that I will not attempt to answer here. It is even questionable whether there is an answer to this question.

³ See, for instance, the differences between the self-shaping, expressivist, and transparency accounts as described by Schwitzgebel (2010).

not immediately past are based on memory, testimony, or induction, *not* on introspection.⁴

I will assume that each of these conditions is indeed necessary for introspection. In fact, each of them is embraced by Schwitzgebel (2010) who, as we will see below, is one of the main critics of IR.

A crucial question regards the First-Person Condition: what exactly is belief formation *about oneself*? Belief formation about one's DNA sequence is in a sense belief formation about oneself, but it is surely *not* something we believe on the basis of introspection. Which things about oneself are usually thought to belong to the realm of introspection? It seems there are five main categories:

- A. *Phenomenal states* How things seem to one, say, visually;
- B. *Intentional states* Beliefs, desires, and intentions;
- C. *Sensations* Experiencing pain or pleasure;
- D. *Emotional states* One's being angry or disappointed;
- E. *Motivational states* One's φ -ing because one has reason x or y to φ .

Note that one can also believe that one is in a particular phenomenal state without holding that belief *on the basis of introspection*. For example, if one has been walking around for hours in the Sahara and one has been perceiving only sand, one can rationally believe without any kind of introspection that one is in the phenomenal state of what we could call 'being appeared to sandly', that is, having the kind of phenomenal state that one would have if one perceived sand⁵—a quick induction on the basis of one's memory of one's previous phenomenal states will do the job. One can even hold a belief about, say, what one desires, on the basis of a friend's testimony who seems to know one's desires well. All I am saying is that *if* one holds a belief on the basis of introspection, then that belief will most likely be a belief about things that fall into one of the above five categories.

As I said, in what follows, I focus on *phenomenal* states. This is because introspecting phenomenal states, such as perceiving or smelling an object, is often taken to be perfectly or at least highly reliable. How could I be mistaken about how things, for instance, visually *seem* to me? If the empirical arguments against the introspection of phenomenal states are convincing, then there is little hope for IR in general.

4 Argument #1: introspective measurement and reporting methods

The first argument against IR when it comes to phenomenal states runs as follows. Empirical research shows that in relatively simple experimental settings, one's belief about one's conscious experience, such as one's experiencing a light's

⁴ For more detailed versions of these conditions, see Schwitzgebel (2010). For more on the Uniqueness Condition, see Byrne (2005, 80–81) and Smithies and Stoljar (2012, 4).

⁵ I put it this way, for, according to many philosophers, one can be in that phenomenal state, even if there is no sand around and, thus, one does *not* truly perceive sand. Hallucinations, matrix scenarios, and the like, would be among the possibilities for being in that phenomenal state without actually perceiving sand. *Mutatis mutandis*, the same applies to the introspection of certain other phenomenal states.

coming on or one's perceiving a target letter, *heavily depends on the actual method of reporting*. This suggests that people's beliefs about their own phenomenal states are unreliably formed—that is, in a significantly large portion of cases false.

Assuming that reporting is a way of expressing one's belief, whether or not one believes that one has a particular phenomenal experience depends for a significant part on whether one reports—about whether or not a light came on—by way of verbal account, hand motor response (e.g., button pressing), or eye blink (Zihl and Von Cramon 1980; Marcel 1993, 2003, 174).

Here is how the experiment was set up. In advance, the experimenter asks the subject to respond in a particular way, for instance, by button pressing, and the subject complies. More specifically, the scientist asks the subject to blink his or her right eye, press the button beneath the right forefinger, or say 'yes', in case the subject *feels* that a light has come on. The subject is asked *not* to guess unless she feels she has to. She is then asked to look at a fixation point on a blank field. The stimulus is an increase in illumination of 10° to the right of the fixation point.

It turns out that people's vocal responses are significantly less accurate than their eye blinks and button pressing, at least in the case of conscious reporting, *not* in the case of guessing (which is generally quite accurate). For example, research by Anthony Marcel on normal subjects' report and guessing performance, using the above mentioned three response types, the results are shown in Table 1.

Here, 'percent hits' stands for the percentage of trials in which the light illuminated and in which a 'Yes' response was given, and 'percent false positives' for the percentage of trials in which there was no light and in which a 'Yes' response was given. Each block consisted of 40 trials. Blocks 1 + 2 and 7 + 8 seemed to be randomly selected by Marcel from Blocks 1–8. He does *not* state on which basis these blocks were selected (Marcel 1993).

On the basis of these kinds of experiments, Josef Zihl and Yves von Cramon claim that:

(...) it seems clear that the patients are not able to indicate the presence of a light target by using a simple verbal response (i.e. 'yes'). However, they were able to report whether or not their motor responses were correct after they had achieved a good performance in registration. The patients, therefore, probably have no conscious access to their extrastriate 'vision' but to the effect of this kind of visual capacity on the lid- and hand-motor systems when serving as voluntary effectors. (Zihl and Von Cramon 1980, 297)

Table 1 Percent hits: percent false positives (See Marcel 1993, 174. I have rounded decimals to the nearest whole number)

Type of response	Blocks 1 + 2	Blocks 7 + 8	Guessing (Blocks 9 + 10)
Blink	65:38	78:30	88:20
Finger	63:42	68:33	83:20
Verbal	55:43	58:40	80:25

And Anthony Marcel draws a similar conclusion when he says:

The relevant finding was that in the condition requiring report of experience, but not in the guessing condition, the responses on any one trial frequently dissociated across the different response modes. That is, often subjects were saying ‘Yes, I am aware of a luminance increase’ with their eyes, but were saying *at the same time* ‘No, I am not aware of a luminance increase’ with their manual gesture. (Marcel 2003, 174)

But if it is true that, when one has a particular phenomenal experience, whether or not one *believes* that one has that experience heavily depends on the method of reporting, then, so the argument goes, we have good reason to doubt IR (Marcel 1993, 2003).

One might reply to this first objection to IR that many of these experiments were conducted with *abnormal subjects*. Some subjects are *hemianopic individuals*—subjects that suffer from a loss of vision in either the whole left or the whole right half of the field of vision—with blindsight (Zihl and Von Cramon 1980). Other subjects suffer from *visual anosognosia*, that is, being blind without realizing that one is blind (Marcel 1993, 176). As I said in Sect. 2 above, we are concerned with *normal, properly functioning subjects’* introspection.

However, *some* of these experiments show that there is dissociation in reporting methods even in *normal subjects*, that is, that even in normal subjects, whether one reports to have had a particular phenomenal experience depends on the method of reporting. Remarkably, there is no dissociation under time pressure, but there *is* dissociation in circumstances in which there is *no* time pressure (Marcel 1993, 172) and it seems not unreasonable to consider circumstances without time pressure as normal. Thus, this first reply to the first objection to IR is unconvincing.

However, there are other problems with the circumstances in the experiment. For, they are still *abnormal* in another respect: the light is shown only very briefly or there is only a *slight* increase in illumination, for instance, an increase in illumination of 10° to the right of the fixation point (Marcel 1993, 182). It is *not* surprising that in such circumstances, we find dissociation in reporting methods. After all, in normal circumstances we have enough time to concentrate and clearly visually or otherwise perceive the object in question. This is not to deny that in the science of consciousness it is sometimes necessary to set up peculiar cases in order to isolate particular variables, such as in empirical research that involves blindsight. However, the point here is that *any* belief formation on the basis of perception—whether it be visual or otherwise—is unlikely to be reliable on some reporting methods if we make perception of the relevant phenomenon difficult enough. Thus, these particular experiments tell us little about whether or not IR is true.

This first reply to the objection merely saves IR, that is, the idea that introspection is *reliable*; sufficiently reliable to generally trust its deliverances. However, I would like to give a second reply to the objection that not only saves IR, but also the stronger thesis that introspection is infallible. The second point, if convincing, does not *show* that introspection is infallible, but it does save IR from this first objection and is compatible with the idea that introspection is infallible.

The second point is this: a plausible way to interpret the results of these experiments, as even Anthony Marcel himself acknowledges (Marcel 1993, 74), is that different ways of reporting—or different dispositions to report—may give one better or worse access to the world, which, in this case, involves an increase in illumination. Thus, being disposed (not) to eye blink may give one better access to changes in light illumination than, say, being disposed (not) to press a button or (not) to give a verbal report. Since perceiving light is a visual experience and since one normally blinks in response to changes in light illumination, this is not at all surprising. But this means that one's reporting method may influence whether one actually *has* the phenomenal experience in question. Thus, even if one's report depends on the method of reporting, the beliefs that those reports express may very well be reliably formed.⁶

According to Marcel, this line of response calls into question the existence of a unitary reflexive consciousness or at least a unitary cognitive subject who has the experiences and reports about them. An account on which there are different kinds of reflexive consciousness that have differential access to a single phenomenal experience might be a better explanation of the data. In attending to the world, there is a single reporter, but there are several of them when one attends to one's own conscious phenomenal experiences (Marcel 1993, 174–175, 183). Here, I will not address whether or not this is indeed a plausible conclusion on the basis of these experiments; I leave that for another occasion. Rather, I would like to stress that whether or not this counts in favour of there being different kinds of reflexive consciousness, we have *not* been given a good reason to think that it counts against IR.

I conclude that the fact that whether and what one reports about one's phenomenal states depends on the reporting method does not count against IR.

5 Argument #2: recolouring the dreamworld

Eric Schwitzgebel gives a detailed overview of scientific studies of the reported incidence of colour in dreams. They consist of questionnaires, dream reports, and REM awakenings. As it turns out, the percentage of the reported incidence of colour in dreams is low in the 1930s, '40s, and '50s, when all or most films were in black and white, whereas they are high from the 1960s onwards, when most or all films were in colour. Here is a selection made from an overview Schwitzgebel (2011, 2–3) gives of English-language studies from the 1930s onwards on the percentage of people claiming to dream in colour or the percentage of dreams experimental subjects described as containing colour, including one study of his own (Table 2).

As Schwitzgebel rightly points out, it is unclear precisely how we should interpret this data. But one thing, he argues, is clear: it is unlikely that people's

⁶ Marcel (2003, 179–181), also points out that there is some evidence that phenomenal states are changed and created as a result of introspecting those states. This may well be correct, but it does not count against IR as long as introspection is reliable (generally true) *given what one's phenomenal state is when one attends to it*. As I pointed out above, we should not confuse the reliability of introspection (the extent to which it produces true belief) with the self-intimatingness of introspection (the impossibility *not* to believe the proposition in question when that proposition is true).

Table 2 Scientific studies of the incidence of colour in dreams

Author	Year	Method	Percentage of people reporting colour dreams	Percentage of people's dreams said to contain colour
Middleton	1933	Questionnaire	26	–
Middleton	1942	Questionnaire	29	–
Hall	1951	Dream reports	–	29
De Martino	1953	Questionnaire	17	–
Kahn et al.	1962	REM awakenings	94	83
Suinn	1966	Dream reports	81	41
Jankowski et al.	1977	REM awakenings	89	62
Frayn	1991	Questionnaire	74	74
Rechtschaffen et al.	1992	REM awakenings	–	80
Schwitzgebel	2003	Questionnaire	81	–
Murzyn	2008	REM awakenings	100	72–89

dreams—whether or not they were in colour—have significantly altered from the 1960s onwards. Here, I assume that Schwitzgebel's arguments for his claim that whether or not the movies one watches are coloured does not significantly influence whether or not one dreams in colour are convincing. Schwitzgebel (2011, 1–15) concludes that these experiments strongly suggest that people's introspective beliefs about their dreams—which he takes to be a particular kind of phenomenal state—are unreliable.

Even if we assume that our dreams have not significantly altered from the 1960s onward, though, there are at least two problems with the argument. First, dreaming is not exactly a normal kind of phenomenal experience, in comparison with, say, visual or auditory perception. Dreams are usually rather vague, incoherent, hard to describe, and even harder to remember. Also, in dreams, we normally do not form occurrent beliefs. That is, in dreams we normally do not consciously consider propositions and consequently form a belief. An exception to this may be so-called *lucid* dreams: dreams in which one is aware that one is dreaming and in which one can, to some extent, manipulate one's imaginary experiences. But none of these experiments were confined to lucid dreams, while only a small percentage of our dreams are lucid. Thus, dreams differ rather drastically from regular perceptual experience. It would not be surprising, then, if reports on whether or not dreams are coloured or contain colour are unreliable due to such social-cultural circumstances as whether movies are in black and white or coloured. This tells us little about introspection of phenomenal states.

Second, the third condition for something to count as introspection that I spelled out above, the Temporal Proximity Condition, rules out from the evidence base the results of questionnaires and dream reports. After all, questionnaires are filled out and dream reports written down at a point in time that is significantly later than the time at which the relevant phenomenal dreaming experiences occur. At that time, the phenomenal dreaming experiences were too distant in the past to still meet the

Temporal Proximity Condition and the subjects' beliefs about them were, therefore, based on memory rather than introspection. Situations like these are comparable to taking a cognitively naïve walk in the forest—one just enjoys the walk and what one sees, rather than explicitly attending to one's phenomenal experiences—and, upon one's return, one is asked what one's phenomenal experiences were. Since one no longer has those experiences, the relevant cognitive mechanisms are visual and auditory perception at the time one takes the walk and memory of the relevant experiences upon one's return, *not* introspection.

Of course, it is *possible* to consider one's phenomenal experiences themselves rather than the birds and trees *while taking a walk in the forest*. In that (abnormal) case, one forms beliefs on the basis of visual and auditory perception and also certain beliefs on the basis of introspection, which is something one can retain in memory. However, dreaming experiences are *not* like that—or if they are, such scenarios are highly unusual. After all, it hardly ever happens (if it ever happens at all) that one consciously considers one's phenomenal experiences *while one is dreaming* and, consequently, forms beliefs about what phenomenal experiences one has, thus forming beliefs on the basis of introspection. Rather, what seems to happen is that one has certain phenomenal dreaming experiences and that these are somehow retained in memory, so that when one awakens and fills out a questionnaire, one forms one's beliefs on the basis of memory rather than (also) on the basis of introspection.

This leaves us with the experimental results of REM awakenings. However, all the evidence Schwitzgebel provides from REM awakenings suggests that people dream in colour: this is what they claim upon being awakened during some REM period of sleep.

Now, Schwitzgebel is aware of this threat to his argument. His reply is twofold. First, there were no REM awakenings prior to the 1960s, but there *were* awakenings at random times during the night. Bentley (1915, 202), who used that method, obtained significantly more reports of dreams in grey than of dreams in colours. Second, the scientists doing this research in the 1940s and '50s must themselves, every now and then in their daily lives, have spontaneously awoken from REM sleep and reflected on their dreams. Presumably, says Schwitzgebel, when they did so, most of them judged their just-ended dreams black-and-white. If they had judged them to be coloured, Schwitzgebel suggests, they would have reported that, since their own experiences, consisting of coloured dreams, would then have conflicted with the data they had gathered about their experimental subjects, most of whom claimed to dream in black and white. Since there are no such reports, it is plausible to assume that the scientists themselves thought their dreams to be black and white when spontaneously awakening from REM sleep (Schwitzgebel 2011, 11).

Let us consider these two points in turn. As to the former, only 20–25 % of our sleep is REM sleep. Therefore, waking up people at random times during the night is likely to result in waking people up during their *non-REM* sleep. But then most reports will *not* be given right after the dreaming experience. Thus, for most dreaming reports, the Temporal Proximity Condition will not be satisfied. Hence, even if most reports were grey reports rather than colour reports, not much follows from that with regard to IR. As to the latter point, we simply do not know what the

dreaming experiences of these scientists were. Maybe many of them *did* experience their dreams as coloured when awakening from REM sleep. We should note that even in the 1930s, '40s, and '50s, in dream reports or questionnaires a substantial number of subjects claimed their dreams to be coloured—varying between 9 and 41 %. Thus, experiences of dreaming in colour would not have surprised the researchers and are unlikely to have led to further investigation on their part.

I conclude that the fact that the percentage of the reported incidence of colour in dreams is low in the 1930s, '40s, and '50s, whereas it is high from the 1960s onwards, does not count against IR.

6 Argument #3: visual imagery and cognitive tasks

More than a century ago, in an experimental setting, Francis Galton invited subjects to visualize—to form a mental picture of—an object from a familiar scene, namely their breakfast table. He then asked them a couple of questions about their visualized image, such as whether it was dim or fairly clear, whether the objects in it were pretty well defined or not, and whether the colours were natural and distinct. It turned out that several people, especially his fellow-scientists, claimed to be unable to have such visual imagery, whereas other people claimed to have distinct, detailed and coloured visual imagery, and still other people claimed to have visual imagery that was not particularly vivid or distinct (Galton 1880). More recent research about the properties of people's visual imagery, such as the degree of determinacy, the colour saturation, the picture likeness, the spatial location or flatness, and their vividness, confirms the thesis that people vary in their abilities of visual representation.

Sheehan (1967), for instance, asked 140 males and 140 females to think of seeing (“the sun sinking below the horizon”), hearing (“the sound of escaping steam”) and to consider carefully the image that came to mind. Stimulus items that were suggested in other modalities were “the prick of a pin” (cutaneous), “running upstairs” (kinesthetic), “oranges” (gustatory), “cooking cabbage” (olfactory), and “drowsiness” (organic). The subjects rated the vividness of their imagery on a seven point rating scale that ranged from “no image present at all” (7) to “perfectly clear and vivid” (1). The results were as follows, where ‘M’ stands for the mean vividness ratings and SD for ‘Standard Deviations’ (Sheehan 1967, 387; Table 3).

As this figure suggests and as the other research mentioned confirms, there is good reason to think that people differ significantly in their visual imagery abilities. This is, of course, something a defendant of IR, could readily acknowledge: it may very well be that, for some reason or other, people vary in their ability to have particular mental images.

However, this is only the starting point of the third argument against IR. Angell (1910) and Schwitzgebel (2011, 43–46) have claimed that if individuals widely differ with regard to their imagery abilities, then we should expect there to be a correlation between such imagery ability and the performance on certain cognitive tasks, where the word ‘task’ is used rather broadly, so that, as we shall see, it includes some phenomena in which one is primarily passively involved rather than

Table 3 Mean vividness ratings and standard deviations across items in seven sensory modalities for 140 males and females

	Visual	Auditory	Cutaneous	Kinesthetic	Gustatory	Olfactory	Organic
Male							
M	2.68	3.02	2.89	2.85	3.07	3.52	2.89
SD	1.39	1.54	1.38	1.39	1.44	1.67	1.40
Female							
M	2.47	2.98	2.65	2.87	2.94	3.38	2.75
SD	1.30	1.52	1.36	1.50	1.61	1.73	1.46

actively engaged. Empirical research, however, has found no correlation of this kind. Subjects who in tests such as the Questionnaire upon Mental Imagery (QMI), the Visual Vividness Imagery Questionnaire (VVIQ), and the Differential Attentional Processes Inventory (DAPI) report vivid mental imagery, do not, on average, do better on cognitive tasks the performance of which would seem to depend on one's ability of visual imagery.

For instance, no correlation was found between one's performance on VVIQ and DAPI on the one hand and the degree to which one is easily hypnotisable, that is, easily accesses trance, on the other (Gemignani et al. 2006). And no correlation was found between subjects' scores on QMI and the accuracy of their pictorial memory, that is, the ability to remember visual rather than verbal representations of situations and events in the form of line drawings and pictures, such as coloured photographs (Richardson 1980, 60–63, 67–70, 82–83, 117–142). And, *mutatis mutandis*, the same applies to other tasks, such as visual creativity, mental rotation, mental folding, Gestalt figure completion, and motor control (see, for instance, Paivio 1990, 177–212). Since people with great visual imagery capacities should be expected to do better on these tests, while they do *not* do better, we have good reason to think that their sincere reports about their visual imagery and, thus, their beliefs about their visual imagery are often false. This would cast serious doubt on IR, at least on when it comes to the introspection of phenomenal states.

There are at least three problems with this argument, though. First, as John Richardson rightly points out, subjects cannot make absolute judgements about, say, the vividness of their mental images. After all, they only have their own images to compare them with. Says Richardson:

He [the subject, RP] can judge whether he has a mental image or not (...). He can judge whether one item evokes a more vivid mental image than another (...). However, he is unable to make an absolute judgement of the vividness of a single image (...) it follows that comparisons among experimental subjects in terms of their ratings of evoked mental imagery are neither valid nor meaningful, and it is quite unsurprising that they should fail to predict performance in learning tasks. (Richardson 1980, 125)

The point is: there may well be differences in mental imagery ability. But it is hard, if not impossible, to establish how good people are at mental imagery. For, in order

to establish that, we would need to rely on people's reports about their mental images. Since people are phenomenally acquainted only with their own mental images and *not* with those of others, they are unable to make absolute judgments about their phenomenal states. Thus, they might describe a particular phenomenal state as vivid, while someone else, who is in more or less the same state, would *not* describe the state she is in as vivid. Hence, judgments about the vividness of mental imagery are always and necessarily relative to the subject; we are unable to compare people's mental imagery abilities based on how they rate the evoked mental imagery.⁷

Second, it is not clear why we should expect subjects who report a vivid mental imagery to do better on the kinds of cognitive tasks mentioned above than subjects who report *not* to have vivid mental imagery. Why, for instance, would someone with a vivid visual imagery with high colour saturation of a particular figure do better on mental rotation than someone with a rather dim visual imagery with low colour saturation of that same figure? It seems that mental rotation is an *additional* cognitive capacity and how good one is at that seems to be determined, unsurprisingly, by how good one is at mental rotation. There is insufficient reason to think that having a vivid visual imagery of figures contributes to the capacity of mental rotation. Of course, if one has *no* visual imagery, then mental rotation will be extremely difficult or even impossible and one will perform particularly badly on that task. But this fairly weak thesis was not the point that Galton, Schwitzgebel, and others aimed to establish. Similarly, whether or not one has vivid mental imagery, one's pictorial memory may very well largely rely on one's *ability to remember*.

Third, some of the results of these experiments are actually mixed (e.g. McKelvie 1995 on free recall). Even more importantly, there are also many experiments that *do* claim to establish a correlation between people's imagery abilities and their performance on certain cognitive tasks. For example, subjects who find it easy to imagine their future experience with a particular product that they might purchase are not only more likely to prefer the imagined product than those who have difficulty imagining it (Escalas 2004). It is even the case that people who have difficulty imagining their future experience with a product are less likely to pursue that product than if they had *not* attempted to imagine their future experience with that product (Petrova and Cialdini 2005). And subjects that are good at visual imagery are generally also good at face recognition (Lobmaier and Mast 2008).

Schwitzgebel is aware of these experimental results but provides two considerations that are meant to undermine their evidential weight. On the one hand, the list of experiments that have failed to establish the relevant correlation is somewhat longer:

⁷ This point applies to phenomenal states generally (rather than merely to states of mental imagery): since people are acquainted only with their own phenomenal states, they cannot make absolute judgments about them. This is *not* to deny that people can make absolute judgments about the vividness of external scenes; after all, people *can* perceive external scenes and can come to agree on the vividness of external scenes (in fact we often do). We are never perceptually acquainted with other people's phenomenal states, though, since we have no faculty by means of which we can perceive those states in other people.

And although reports of correlations between the VVIQ and performance on various cognitive tasks presumably involving visual imagery have continued to appear since 1995, so too, perhaps even more often, have studies reporting no significant correlation between the VVIQ and visual or imagery-related tasks. (Schwitzgebel 2011, 170)

I reply that the difference in number is insufficiently large to warrant Schwitzgebel's conclusion. He mentions 21 studies that count in favour of his conclusion, 10 studies that count against his conclusion, and 13 studies with mixed results (Schwitzgebel 2011, 170). Clearly, this does *not* suffice to warrant the conclusion that there is no correlation between visual imagery and performance on the relevant cognitive tasks. The results of these experiments simply fail to clearly point in a particular direction; they are too mixed for that. There might be a good explanation why no correlation was found in the 21 studies, there might be a good explanation for the correlation that was found in the ten studies, and there might be a good explanation for the mixed results in ten studies; at this point, it is hard to tell.

On the other hand, he claims, the positive findings in experiments that *do* claim to establish a certain correlation can be explained in other ways. This is a crucial point, for if it is correct, then the experimental results *do* clearly point in the direction of the absence of a correlation. Schwitzgebel points to three factors that might explain the correlation.

First, psychological variables tend to correlate: someone with a particular social background—such as having had a good education—is likely to be good at *both* visual imagery *and* these various cognitive tasks, even though she will probably not be good at these cognitive tasks *because* she is good at visual imagery. Second, the subject's knowledge of how she did on the first test may affect her performance on the second test: if you know you did well on the visual imagery test, then you may try your utmost at, say, the mental rotation test, whereas if you did *not* do well on the visual imagery test, you may *not* do your utmost best on the mental rotation test. And the same is true, *mutatis mutandis*, when the experiments are done in the reverse order. And, third, the experimenters' expectations might have influenced the performance of the subjects by subtle, non-verbal communications (Schwitzgebel 2011, 46–48). Clearly, that would undermine the evidential weight of these studies.

Let me reply to these alternative explanations of the correlations that were found in the order in which I presented them. First, why should we think that someone with a particular social background—such as a good education—is likely to be good at *both* visual imagery *and* these various cognitive tasks? Surely, people from a particular background (say, rich and highly educated) are more likely to do well on many of these cognitive tasks, given their higher IQ. But why think that people from such a background are better at visual imagery than people from other backgrounds? Schwitzgebel provides no evidence to think that this is the case. Also, if this is what explains the positive results, then why were no positive findings recorded in the other experiments mentioned by Schwitzgebel? If the positive findings are due to the kind of correlation he mentions, then we should expect to find such a correlation in *all* or *many* experiments with a sufficiently large sample, but, apparently, that is *not* the case.

Second, there is no reason to think that the relation described by Schwitzgebel between one's performance on the first test and one's performance on the second test holds. It may just as well be the case that if one does *not* do well on the first test, then one tries even harder on the second test, and that if one *does* well on the first test, one does *not* try as hard on the second test, given that one has already performed well. This may even differ from subject to subject. Again, Schwitzgebel provides no evidence whatsoever to think that the relation he describes holds.

Third, Schwitzgebel rightly notes that the experimenters' expectations sometimes influence the performance of the subjects by subtle, non-verbal communications. However, he provides no evidence that this took place in these particular experiments. More specifically, he does not single out a factor that is *present* in the experiments with positive findings that is *absent* in the experiments with negative findings and without any such evidence, the idea that the experiments influenced the results remains a mere possibility.

I conclude that it is questionable that there is no correlation between visual imagery and various cognitive tasks and that, even if there is no correlation between the two, that does not count against IR.

7 Argument #4: the capacity of echolocation

We are not as good at it as dolphins, bats, and whales, but, contrary to common opinion, we have the capacity of echolocation. Empirical research shows that by echolocation blind people often form true beliefs above chance about the size and distance of physical objects in their surroundings. In fact, similar results have been found with blindfolded *sighted* subjects. Even if they are unpractised, such subjects, whether stationary or moving, can detect features of their physical surroundings, including properties of physical objects that do not themselves produce sound, using the acoustic changes in sounds as they reflect off those objects. Their echolocational capacity is not as well developed as that of blind subjects, but the ratio of true beliefs based on their echolocational capacity is still significantly above chance (Hausfeld et al. 1982; Rosenblum et al. 2000). When one closes one's eyes and approaches a wall, one can hear on the basis of the echo of one's own voice whether or not one is getting close to the wall. Thus, it seems, people are in certain echolocational phenomenal states.

Let me, by way of example, briefly summarize one of the experiments by Lawrence Rosenblum, Michael Gordon, and Luis Jarquin. In this experiment, 20 female and 6 male students, all sighted and reporting good hearing, were presented with a flat piece of drywall. One experimenter issued all the instructions to the subjects, while another manipulated the distance of the wall. The subjects did not have visual access to the test area. There were 20 stationary training trials. Verbal or physical feedback was given after each trial, for both moving and stationary conditions for the wall.

There were 40 critical trials, consisting of 5 moving and 5 stationary trials at each of four distances between the subjects and the piece of drywall:

1. 36 in (0.914 m)
2. 72 in (1.829 m)
3. 108 in (2.743 m)
4. 144 in (3.658 m)

Moving, stationary and distance trials were randomized. Subjects were instructed to echolocate the wall while walking 10 feet up to the starting point of 0 in (0 m)—so the piece of drywall was 36 in, 72 in, 108 in, or 144 in from the starting point. They were given 6 s to echolocate the position of the wall. After echolocating, the wall was removed, and the subjects, still blindfolded, were asked to walk to where the wall had been in their opinion. This is the mean distance traversed in Fig. 1. Subjects were given no performance feedback during the critical trials. They had a short break after 20 trials. The experiment took approximately 100 min for each of the participants. The results were as follows (Fig. 1; Table 4).

Experiments like these suggest that sighted listeners can distinguish the distance of an echolocated wall by using an open-loop walking response, and that listener movement provides a small increase in echolocational accuracy for at least *some* distances (Rosenblum et al. 2000, 197).

Now, the point of the argument that appeals to this kind of research is that most people will initially—that is, before actually trying this out—claim that they cannot detect and do not have any conscious echolocational experience of the shape, distance, size, or texture of silent physical objects in their environment. According to Rosenblum et al. (2000, 202), for instance, “sighted individuals are rarely conscious of their ability to use reflected sound as proprioceptive information.”

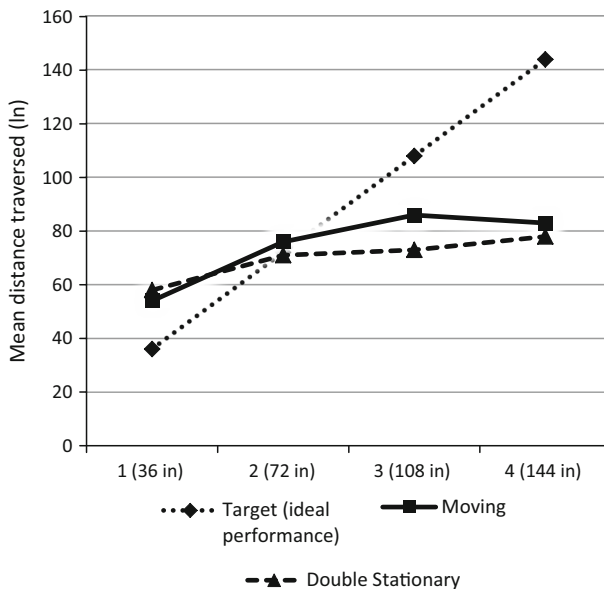


Fig. 1 Mean distance traversed (in) at each of the 4 wall positions as a function of technique

Table 4 Measures of consistency: standard deviations and variable error scores by technique and wall distance

Wall distance	Moving		Double stationary	
	SD	Variable error	SD	Variable error
1	20	54	19	51
2	20	29	18	25
3	20	18	21	20
4	22	16	22	14

Hence, most people falsely believe that they have no echolocational phenomenology (Schwitzgebel 2011, 57–69). Schwitzgebel and Gordon (2000, 235) conclude on the basis of this kind of research that “normal people in normal circumstances can be grossly and systematically mistaken about their own current conscious experience.”

In response, we need to notice at least three things. First, a false belief that one lacks the capacity of echolocation is *not* a false belief based on *introspection*. That we do not have the capacity of echolocation is a general belief about ourselves, presumably based on not remembering any conscious exercise of that ability or based on background beliefs about humans and animals, *not* a belief that is formed after and on the basis of thorough introspection. In order to count against IR, then, the argument should say that people have false beliefs about their particular echolocational phenomenal states, namely the false belief that they are *not* in phenomenal state *X*, the false belief that they are not in phenomenal state *Y*, and so forth. However, if we have the capacity of echolocation and if we have thousands of echolocational phenomenal experiences every day, it seems false that we believe for each of them that we lack that experience. That idea is plausible only if we not only have *occurrent* and *dormant* beliefs, but also tacit or dispositional beliefs (roughly: belief that *p*, while we have never considered *p*), something a large number of philosophers deny—they would say that we *have a disposition* to believe these things rather than that we believe them dispositionally (see, for instance, Audi 1994). But if we lack these beliefs, then we do not hold (a large number of) false beliefs in this regard. Hence, we would merely *fail to truly believe* that we are in phenomenal state *X*, in phenomenal state *Y*, etc., rather than *falsely believe* that we are *not* in phenomenal state *X*, in phenomenal state *Y*, and so forth. Hence, this would count against the *self-intimatingness* of introspection, *not* against the *reliability* of introspection.

Second, I might believe in general that I lack the capacity of echolocation, but, it seems, few rational persons would hold all sorts of beliefs about the absence of echolocational experiences at specific times and places without even trying to find out whether they have the capacity of echolocation (which, presumably, would lead to their giving up those false beliefs). Trying whether one has the capacity of echolocation is, after all, rather easy: one just has to move closer to, say, a wall while making a sound and see whether one can hold true beliefs about how close

one is to the wall merely in virtue of experiencing the echo of that sound. But if people lack these beliefs, they cannot be unreliably formed (nor reliably). Hence, this would *not* count against IR.

Third, we may question whether we are ever in echolocational phenomenal states in the first place. All that follows from the empirical experiments mentioned above is that we are sometimes in certain physical surroundings, that we hear certain sounds in those surroundings, and that those sounds (those auditory experiences) have certain properties such that on the basis of detecting them, we can reliably form true beliefs about, say, the size or distance of certain physical objects in that environment. Now, virtually no person would deny, upon introspection, that she is in those auditory phenomenal states. What she *would* deny, upon being questioned about the issue, is that that tells her anything about the location of silent physical objects in her surroundings, but *that* is not a belief based upon introspection.

Thus, what those who champion this argument need is the claim that there are distinct, echolocational phenomenal experiences that we fail to notice and that we falsely believe to lack. But this is problematic. According to some scientists, our echolocational ability can be explained in terms of our brains unconsciously responding to certain properties or qualities of our auditory experiences. According to others, our echolocational ability ought to be explained in terms of our brains responding to differences in sound waves that reach our skin. But if either of these is correct, it seems false to say that there are echolocational states we can be in: echolocation would then be a brain response to certain physical experiences rather than a conscious state one is in, in the same way as our bodies and brains respond to being exposed to therapeutic ultraviolet light waves, without resulting in our being in any particular conscious state.

I conclude that the fact that we have echolocational capacities and that we are often unaware of the fact that we have them does not count against IR.

8 Argument #5: inattentional blindness and change blindness

The final argument against IR that I will discuss in this paper has to do with so-called inattentional blindness and change blindness. Let us start with *inattentional blindness*. In normal, calm circumstances, people, on the basis of introspection, think that their visual experience consists of a broad, stable field that is detailed, but hazy at the borders. Empirical research shows, though, that the centre of clarity is much tinier than we think it is. It shifts rapidly around a rather indistinct background. It is because our eyes quickly dart around—we turn our eyes many times a second and often constantly reposition our bodies—that we have the illusion that the centre of clarity is fairly large. We manage to move around in particular surroundings either because they are already familiar to us or because we quickly scan the environment, changing the focus of our attention all the time. Inattentive blindness, thus, means that we are blind for any parafoveal stimuli—stimuli beyond the foveal field, that is, the field of our focused attention—and that, contrary to what we think, there is very little in the plenum of our visual field, that is, in the broad, stable visual field that we take ourselves to have.

Closely related to inattentional blindness is *change blindness*. Because we are inattentively blind, we fail to notice changes, even significant changes, in our parafoveal vision area. For example, in some experiments, a computer tracks our saccades. Saccades are the rapid eye movements from point to point, in order to find a new focus. They only take 1/200–1/12 of a second. In the experiment, a computer tracks these saccades and a machine then immediately projects certain dots or letters on that part of the screen that our eyes focus on. Even if the rest of the screen is simply filled with X's rather than other letters (Grimes 1996, 4), our phenomenology is one of a screen covered with an entire text that we can easily read, precisely because our focus changes constantly and our brain makes up for that which is beyond our focus. And the same is true for detailed coloured scenes: we do not see that a prominent building in the city skyline becomes 25 % larger, we fail to notice that two men exchange hats, where their hats have different colours and styles, and so forth. Thus, contrary to what we think, we suffer from change blindness in our visual field. For further reports about such experiments regarding inattentional blindness and change blindness, see Dennett (1969), 139–141; Mack (2003); Rensink et al. (1997) and Noë (2004), 37, 49.

An example of experiments regarding change blindness is Rensink et al. (1997). The authors developed what they call a *flicker paradigm*. In this paradigm, a particular image *A* repeatedly alternates with a modified image *A'*. The sequence was *A*, *A*, *A'*, *A'*, etc., with gray blank fields placed between the successive images. Each of the images was displayed for 140 ms and each blank for 80 ms.

Changes were further divided according to the degree of interest in the part of the scene being changed. Interest was determined via an independent experiment in which five observers provided a brief verbal description of each scene. 'Central Interests' (CIs) were defined as objects or areas that were mentioned by three or more observers, whereas 'Marginal Interests' (MIs) were defined as objects or areas that were mentioned by none of the subjects. The average changes in intensity and color were similar for the MIs and the CIs, but the areas of the MI changes, where the average was 22 deg², were somewhat larger than those of the CI changes, where the average was 18 deg².

The changes were quite large and easy to notice. For example, a prominent object would appear and later disappear, the color would change from blue to red, and so on. In each of the experiments they performed, ten observers participated by freely viewing the flickering display. They were instructed to press a key when they saw a change and to subsequently describe it verbally.

In one of the experiments, it was measured how many alternations were required, on average, for the subject to notice the relevant changes. Here are the results for three sorts of changes (presence/absence, color change, and location change) for scenes of central interest and marginal interest (Fig. 2).

As the experiment shows, it takes subjects surprisingly long time to notice significant changes in images of real-world scenes. When structures are of central rather than marginal interest, the identification of changes occurs quicker, but it usually still takes many alternations before subjects notice. On the basis of this experiment and similar experiments, Rensink, O'Regan, and Clark conclude:

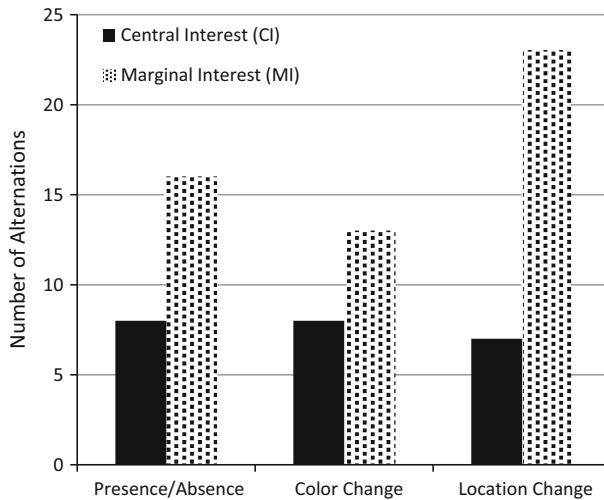


Fig. 2 Number of alternations required for identification of change under flicker conditions

these results indicate that—even when sufficient viewing time has been given—an observer does not build up a representation of a scene that allows him or her to perceive change automatically. Rather, perception of change is mediated through a narrow attentional bottleneck, with attention attracted to various parts of a scene based on high-level interest. (Rensink et al. 1997, 368)

Now, several philosophers and scientists have claimed that experiments along these lines show or, at least, give us good reason to think that introspection is unreliable when it comes to our phenomenal states. Here is their argument: upon being questioned, people sincerely claim and, thus, given the results of these experiments, *falsely believe* that their foveal field is rather large. Also, they falsely believe that during the time period in question no changes take place in what is in fact their parafoveal visual area. This means that people's belief formation on the basis of introspection is unreliable, both when it comes to how large the foveal field is and when it comes to whether or not certain changes took place in the parafoveal visual area. Says Schwitzgebel:

Most of the people I have spoken to who attempt these exercises eventually conclude to their surprise that their experience of clarity decreases substantially even a few degrees from the center. Through more careful and thoughtful introspection, they seem to discover—I think they really do discover—that visual experience does not consist of a broad, stable field, flush with precise detail, hazy only at the borders (...) Most of my interlocutors confess to error in having originally thought otherwise. If I am right about this, then most naive introspectors are badly mistaken about their visual phenomenology when they first reflect on it. Even though they may be patiently considering their experience as it occurs, they will tend to go wrong unless they are warned and coached against a certain sort of error. And the

error they make is not a subtle one; the two conceptions of visual experience differ vastly (Schwitzgebel 2011, 126).

A similar argument can be found in Dennett (1991, 356, 364, 467–468), Irvine (2012, 627–631), and O'Regan (1992, 484).

One could respond to an argument like this by simply denying that we believe that our foveal field is fairly large. The idea would then be that we *do* believe that our perception of our environment is fairly detailed, but that we do *not* believe that all that visual detail is available to us *at once*, without changing our visual focus. According to Alva Noë, for instance, “[i]t is no part of our phenomenological commitments that we take ourselves to have all that detail at hand in a single fixation” (Noë 2004, 57). The problem, though, is that empirical research shows that, upon introspection, most people *do* seem to believe this, that is, most people believe that they are in all these phenomenal states at once (or that they are in one phenomenal state that comprises all this detail). After all, most subjects display surprise upon being presented with the results of the experiments (thus also Dennett 2001) and most of Schwitzgebel’s subjects, as is clear from the quote from Schwitzgebel given above, confess to error on this point.

My own reply to this argument is, therefore, different. It is twofold.

First, in cases of inattentive blindness (I return to cases of change blindness in the second point below), subjects do *not* have a false belief or false beliefs about what their phenomenology is: they are in fact in the phenomenal states that they believe themselves to be in. Rather, they are (at most) mistaken in thinking that they are in all those phenomenal states *at once*—below, I return to the question whether they actually *are* mistaken about that. Similarly, in cases of change blindness, they *are* in all the phenomenal states they take themselves to be in, but, at most, they are not in all of them at once. Thus, they (at most) falsely believe that they see (or at least have the visual experience of) a whole text at once, whereas what they really see (what their visual experiences really are experiences of) are small bits of texts that follow each other rapidly, while their short-term memory keeps the previous visual experiences in mind.

Those who claim that introspection is reliable or that introspection of phenomenal states is reliable do *not* claim that just *any* kind of belief about our phenomenal states is reliably formed. I do not know of anyone who claims that beliefs about the exact temporal boundaries of visual experiences that follow each other rapidly (usually, several within a second) are among them. But even if they are, that would not constitute a problem. For, if such beliefs are among the beliefs about one’s phenomenal states that one forms on the basis of introspection, in order for the subjects to form such beliefs, we would need to ask them or they would need to ask themselves fairly specific questions (or consider very specific propositions), such as: do my phenomenal experiences of what seems to me to be my foveal vision occur all at once or do I quickly change my focus, thereby receiving new experiential input in a rapid sequence? As Schwitzgebel admits, such introspection or the same kind of introspection with a more specific focus hardly ever takes place among subjects that are not trained to perform such introspective action and when subjects *do* actually perform such introspective action, that usually leads to true

beliefs or at least not to false ones, because subjects who consider the matter in detail realize that they have insufficient evidence to think that they are receiving all their phenomenal experiences of what seems to them to be their foveal vision all at once.

Second, these experiments show that, in cases of inattentive blindness, the subjects, to the extent that they hold beliefs about this, are mistaken in thinking that they are in all these distinct phenomenal states at once. Again, it is questionable whether they believe this, but let us assume that they *do* believe this. It seems that in that case, they are *not* mistaken in thinking that *it phenomenally seems to them* that they are in all these phenomenal states at once. That is how their visual field appears to them: phenomenally, they seem to experience all these things at once. Apparently, phenomenology is misleading here in that, as a matter of fact, people's eyes saccade quickly, but subjects are *not* mistaken about the phenomenology itself, because that *is* how things visually appear to them. Their belief, if they hold such a belief, that they *are* in all these phenomenal states at once is false, but their belief that *it phenomenally seems to them* that they are in all these phenomenal states at once is actually true.

Similarly, subjects may be mistaken, that is, hold false beliefs about certain effects or their absence in their physical surroundings, such as whether or not a prominent building in the city skyline becomes 25 % larger or that two men exchange hats with different colours and styles, but it seems they are *not* mistaken in believing that *it phenomenally seems to them* that there is no such change. But that means that, to the extent that they hold beliefs about this, they actually believe *truly* that they are in a phenomenal state in which there is no change in their surroundings.

I conclude that experiments that show that we suffer from inattentive blindness and change blindness do not count against IR.

9 Conclusion

Let us recapitulate. I have *not* aimed to show that introspection, in general, is a reliable doxastic mechanism or a set of reliable doxastic mechanisms. I have not even tried to demonstrate that belief formation on the basis of introspecting *phenomenal states* is a reliable mechanism. Rather, my aim has been to show that five main empirical scientific arguments against IR are unconvincing. Some of the arguments, such as the argument from introspective measurement, claimed that introspection is unreliable, because it gives conflicting verdicts, depending on which measurement method is used. Other arguments, such as the argument from (un)coloured dreams, were arguments to the effect that beliefs based on introspection conflict with our background knowledge. For each of the five arguments that I discussed, I argued that the experiments in question do not justify any claims about introspection at all or, to the extent that they do, they tell us important things about introspection, but do not count against IR.

There are several ways one could go from here. Let me mention three of them. First, one could argue that there are convincing a priori arguments against IR. Second, one could argue that there are no convincing a priori arguments against IR

either and that, in the absence of good arguments against IR, we should believe that introspection is reliable. Third, one could argue that there are no convincing a priori arguments against IR and that there are even good arguments—a priori, empirical, or both—for IR. I have taken this third route elsewhere.⁸ Clearly, the second and third options can be combined.

Of course, the conclusion that the empirical scientific arguments against IR are unconvincing is interesting all by itself. But, coming back to what I said earlier in this paper, we should notice that it is also important for other reasons. It means that from an empirical scientific point of view, there may be no good reason to mistrust introspection. And that means that there would be no good empirical scientific reason to think that the philosophical tradition that says that introspection is particularly secure or even infallible is mistaken—at least, if my second reply to the first objection is convincing. Also, if the empirical arguments against IR do not work, then that does serious damage to epistemological scientism. Epistemological scientism, after all, says that only the natural sciences are a reliable source of knowledge. And adherents of that variety of scientism usually appeal to the scientific, empirical arguments in support of that claim. If those arguments are convincing, then the most important pillar for their claim that introspection, at least that of phenomenal states, is unreliable has come crumbling down. This means that they will have to withdraw to a significantly weaker claim, such as that the sciences *more reliably* deliver knowledge than other sources of knowledge, such as introspection of phenomenal states, or that the sciences deliver knowledge, whereas *some* common sense sources of belief, such as memory or moral deliberation, do not. What I have argued also means that what are often taken to be empirical-scientific reasons *not* to use beliefs formed on the basis of introspection as data for, say, psychological or neuroscientific research are no good reasons at all *not* to do so—assuming, of course, that they are used discriminately.

The fact that the arguments discussed in this paper fail to show that introspection is unreliable, then, is important for both philosophy and science.

Acknowledgments For their helpful comments on earlier versions of this paper, I would like to thank Lieke Asma, Leon de Bruin, Jeroen de Ridder, Geertjan Holthrop, Naomi Kloosterboer, Kelvin McQueen, Emanuel Rutten, Eric Schwitzgebel, Robin Scott, Hans van Eyghen, and René van Woudenberg. Publication of this article was made possible through the support of a grant from Templeton World Charity Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of Templeton World Charity Foundation.

References

- Angell, J. R. (1910). Methods for the determination of mental imagery. *Psychological Monographs*, 13, 61–108.
- Audi, R. (1994). Dispositional beliefs and dispositions to believe. *Noûs*, 28(4), 419–434.
- Bentley, M. (1915). The study of dreams: A method adapted to the seminary. *The American Journal of Psychology*, 26(2), 196–210.
- Bode, B. H. (1913). The method of introspection. *Journal of Philosophy*, 10(4), 85–91.
- Byrne, A. (2005). Introspection. *Philosophical Topics*, 33(1), 79–104.

⁸ See Rik Peels, “An A Priori Defence of the Reliability of Introspection”, *unpublished manuscript*.

- Chalmers, D. J. (2003). The content and epistemology of phenomenal belief. In Q. Smith & A. Jokic (Eds.), *Consciousness: New philosophical perspectives* (pp. 220–272). Oxford: Oxford University Press.
- Dennett, D. C. (1969). *Content and consciousness*. London: Routledge.
- Dennett, D. C. (1991). *Consciousness explained*. London: The Penguin Press.
- Dennett, D. C. (2001). Surprise, surprise. *Behavioral and Brain Sciences*, 24(5), 982.
- Dennett, D. C. (2007). Heterophenomenology reconsidered. *Phenomenology and the Cognitive Sciences*, 6(1-2), 247–270.
- Ericsson, K. A. (2003). Valid and non-reactive verbalization of thoughts during performance of tasks: Towards a solution to the central problems of introspection as a source of scientific data. *Journal of Consciousness Studies*, 10(9–10), 1–18.
- Escalas, J. E. (2004). Imagine yourself in the product: Mental simulation, narrative transportation, and persuasion. *Journal of Advertising*, 33(2), 37–48.
- Gallagher, S. (2003). Phenomenology and experimental design: Toward a phenomenologically enlightened experimental science. *Journal of Consciousness Studies*, 10(9–10), 85–99.
- Galton, F. (1880). Statistics of mental imagery. *Mind*, 5(19), 301–318.
- Gemignani, A., Sebastiani, L., Simoni, A., Santarcangelo, E. L., & Ghelarducci, B. (2006). Hypnotic trait and specific phobia: EEG and autonomic output during phobic stimulation. *Brain Research Bulletin*, 69(2), 197–203.
- Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In E. Akins (Ed.), *Perception* (pp. 89–110). Oxford: Oxford University Press.
- Hausfeld, S., Power, R. P., Gorta, A., & Harris, P. (1982). Echo perception of shape and texture by sighted subjects. *Perceptual and Motor Skills*, 55(2), 623–632.
- Hurlburt, R. T., & Heavey, C. L. (2001). Telling what we know: Describing inner experience. *Trends in Cognitive Sciences*, 5(9), 400–403.
- Irvine, E. (2012). Old problems with new measures in the science of consciousness. *British Journal for the Philosophy of Science*, 63(3), 627–648.
- Irvine, E. (2013). Measures of consciousness. *Philosophy Compass*, 8(3), 285–297.
- Jack, A. I., & Roepstorff, A. (2003). “Editorial introduction: Why trust the subject?”, Trusting the subject: The use of introspective evidence in cognitive science, Vol. 1, *Journal of Consciousness Studies* 10.9-10, v-xx.
- Jack, A. I., & Shallice, T. (2001). Introspective physicalism as an approach to the science of consciousness. *Cognition*, 79(1-2), 161–196.
- Kriegel, U. (2013). A hesitant defense of introspection. *Philosophical Studies*, 165(3), 1165–1176.
- Lobmaier, J. S., & Mast, F. W. (2008). Face imagery is based on featural representations. *Experimental Psychology*, 55(1), 47–53.
- Lutz, A., & Thompson, E. (2003). Neurophenomenology: Integrating subjective experience and brain dynamics in the neuroscience of consciousness. *Journal of Consciousness Studies*, 10(9-10), 31–52.
- Mack, A. (2003). Inattention blindness: Looking without seeing. *Current Directions in Psychological Science*, 12(5), 180–184.
- Marcel, A. J. (1993). Slippage in the unity of consciousness. In G. R. Block & J. Marsh (Eds.), *Experimental and theoretical studies of consciousness* (pp. 168–186). Chichester: Wiley.
- Marcel, A. J. (2003). Introspective report: Trust, self-knowledge and science. *Journal of Consciousness Studies*, 11(7-8), 167–186.
- McKelvie, S. J. (1995). The VVIQ and beyond: Vividness and its measurement. *Journal of Mental Imagery*, 19(3), 197–252.
- Noë, A. (2004). *Action in perception*. Cambridge, MA: The MIT Press.
- O’Regan, J. K. (1992). Solving the “Real” mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, 46(3), 461–488.
- Overgaard, M. (2006). Introspection in science. *Consciousness and Cognition*, 15(4), 629–633.
- Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.
- Peels, R. (2016). A conceptual map of scientism. In J. Rik Peels & R. van Woudenberg (Eds.), *Scientism: A philosophical exposition and evaluation*. New York: Oxford University Press (Forthcoming).
- Petrova, P. K., & Cialdini, R. B. (2005). Fluency of consumption imagery and the backfire effects of imagery appeals. *Journal of Consumer Research*, 32(3), 442–452.
- Price, H. H. (1954). *Perception*. London: Methuen.
- Rensink, R. A., O’Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8(5), 368–373.

- Richardson, J. T. E. (1980). *Mental imagery and human memory*. London: MacMillan Press.
- Rosenberg, A. (2011). *The atheist's guide to reality: Enjoying life without illusions*. New York: W.W. Norton.
- Rosenblum, L. D., Gordon, M. S., & Jarquin, L. (2000). Echolocating distance by moving and stationary listeners. *Ecological Psychology*, *12*(3), 181–206.
- Schwitzgebel, E. (2007). No unchallengeable epistemic authority, of any sort, regarding our own conscious experience—Contra Dennett? *Phenomenology and the Cognitive Sciences*, *6*(1-2), 107–113.
- Schwitzgebel, E. (2010). “Introspection”, Stanford Encyclopedia of Philosophy. <http://plato.stanford.edu/entries/introspection/>, first published February 2nd 2010; substantive revision September 4th 2010, last visited November 24th 2015.
- Schwitzgebel, E. (2011). *Perplexities of consciousness*. Cambridge, MA: MIT Press.
- Schwitzgebel, E., & Gordon, M. S. (2000). How well do we know our own conscious experience? The case of human echolocation. *Philosophical Topics*, *28*(2), 235–246.
- Sheehan, P. W. (1967). A shortened form of Betts' Questionnaire upon mental imagery. *Journal of Clinical Psychology*, *23*(3), 386–389.
- Smithies, D., & Stoljar, D. (2012). Introspection and consciousness: An overview. In D. Smithies & D. Stoljar (Eds.), *Introspection and consciousness* (pp. 3–25). Oxford: Oxford University Press.
- Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*(2), 158–177.
- Zihl, J., & von Cramon, D. Y. (1980). Registration of light stimuli in the cortically blind hemifield and its effect on localization. *Behavioural Brain Research*, *1*, 287–298.