

# Fact-Introspection, Thing-Introspection, and Inner Awareness

Anna Giustina and Uriah Kriegel

*Review of Philosophy and Psychology* 8 (2017): 143-164

**Abstract.** Phenomenal beliefs are beliefs about the phenomenal properties of one's concurrent conscious states. It is an article of common sense that such beliefs tend to be justified. Philosophers have been less convinced. It is sometimes claimed that phenomenal beliefs are not on the whole justified, on the grounds that (i) they are typically based on introspection and (ii) introspection is often unreliable. Here we argue that such reasoning must guard against a potential conflation between two distinct introspective phenomena, which we call fact-introspection and thing-introspection; arguments for the unreliability of introspection typically target only the former, leaving the reliability of the latter untouched (§1). In addition, we propose a theoretical framework for understanding thing-introspection that may have a surprising consequence: thing-introspection is not only reliable, but outright infallible (§2). This points at a potential line of defense of phenomenal-belief justification, which here we only sketch very roughly.

## Introduction

Let phenomenal beliefs be beliefs about the phenomenal properties of one's own current conscious states. In this sense, beliefs about one's nonconscious states, for example about the dispositional belief that  $13.9 > 11.34$ , are not phenomenal beliefs. Nor are beliefs about one's past conscious states, such as the belief that one's gustatory experience of orange juice yesterday at breakfast was quite pleasant. Likewise, beliefs about another's current phenomenal states, such as one's belief that one's friend feels pain right now, are not phenomenal beliefs in our sense.

Finally, beliefs about the nonphenomenal properties of one's current conscious states, such as the belief that one's experience is occurring exactly a year after the Pope abdicated, are not phenomenal beliefs. For a belief B to be a phenomenal belief, B must concern the phenomenal properties that a concurrent conscious state of the subject herself appears to exhibit.<sup>i</sup>

According to an important philosophical tradition, going back at least to Descartes, phenomenal beliefs have a special epistemic status: they are infallible – having them entails their truth. However, this tradition has been the target of sustained attack, and it has almost become a modern orthodoxy that phenomenal beliefs are *not* infallible.

Some philosophers have called into question not only the infallibility of phenomenal beliefs, but even their tendency to be justified. Indeed, a significant and respectable challenge has been raised against the reliability of phenomenal belief-formation processes, both by epistemologists and philosophers of mind (Chisholm 1957, Schwitzgebel 2008), as well as by cognitive scientists (Nisbett and Wilson 1977). The basic reasoning seems to be that (i) phenomenal beliefs are typically based on introspection and (ii) introspection is often unreliable.

Consider the following pair of compelling arguments from Schwitzgebel (2008). The first is an argument from the ineffability of conscious experience: (1) if introspection were a reliable belief-formation process, then we would be able to formulate detailed correct judgments about the phenomenology of our current conscious experience; but (2) we are not typically able to form such detailed correct judgments; therefore, (3) introspection is unreliable. In support of (2) Schwitzgebel reports several cases (mostly from emotional and visual experience) in which people are uncertain or mistaken about the phenomenology of their current conscious experience. Schwitzgebel's second argument is from introspective disagreement. Consider the philosophical debate about so-called cognitive phenomenology. Some philosophers maintain that conscious thought has a *sui generis* phenomenology, while others maintain that it does not. Arguably, however,

both base their judgments on careful and thorough introspection. If introspection were a reliable belief-formation process, both groups would presumably converge on the same position with respect to such a fundamental phenomenal feature. Therefore, introspection cannot be a reliable belief-formation method.

In the bulk of this paper, we want to argue that there is at least one kind of introspection which is reliable. Our case for this has two parts, one negative and one positive. In §1, we suggest that extant arguments against the reliability of introspection may not apply to one central form of introspection; we draw a distinction between ‘fact-introspection’ and ‘thing-introspection’ that helps see why this is so. In §2, we offer a positive reason to think that at least one kind of introspection – thing-introspection – must be highly reliable; we draw another distinction, between introspection and inner awareness, and develop a theoretical framework that helps see why this is so. The combination of both distinctions yields a coherent package deal with a surprising consequence: thing-introspection is not only reliable, but, if our theoretical framework is accepted, may well turn out to be infallible.

## **1. Fact-Introspection and Thing-Introspection**

Our distinction between fact-introspection and thing-introspection is modeled after Fred Dretske’s (1993) distinction between fact-awareness and thing-awareness (Giustina 2015). Dretske’s distinction targets *perceptual* awareness, but our claim is that a parallel distinction applies to *introspective* awareness. After presenting Dretske’s original distinction (§1.1), we draw our parallel one (§1.2), before showing how it affects the debate over the reliability of introspection (§1.3).

### *1.1. Fact-Awareness and Thing-Awareness*

Suppose you are visually presented with a green tree and everything goes well. It seems that at least two reports are true of you:

- (1) You see the green tree.  
(2) You see that the tree is green.

The surface grammar of (1) and (2) is clearly different, but according to Dretske, the mental states reported in (1) and (2) are also different. The state reported in (1) he calls thing-awareness, because what you are said to be (visually) aware of is a thing – a (green) tree. The state reported in (2) Dretske calls fact-awareness, because what you are said to be (visually) aware of is the fact *that* the tree is green.

A more parsimonious view would be that these are two stylistically different reports, but the underlying mental state they report is one and the same. Three important differences between the reported states, however, rule this parsimonious view out. The first is that fact-awareness is a *propositional* attitude whereas thing-awareness is an *objectual* attitude. The second is that fact-awareness involves the deployment of *concepts* whereas thing-awareness does not. The third is that fact-awareness is *directly expressible* whereas thing-awareness is not. Arguably, the second of these is most fundamental. On the one hand, a content being conceptual appears to guarantee, in a language-capable creature, that it be expressible; for the creature can use terms that *express* the relevant concepts. At the same time, on a common understanding of these notions, concepts are the *constituents* of propositions, so the right combination of suitably connected concepts would *constitute* propositional content. Nonetheless, let us consider each of these in turn.

*Propositional vs. Objectual Attitude.* When *S* is fact-aware that *a* is *F*, the content of *S*'s mental state is the proposition  $\langle a \text{ is } F \rangle$ . Indeed, on Dretske's view, fact-awareness typically takes the form of a *belief*. By contrast, when *S* is thing-aware of *a*, the content of her mental state is simply *a*, and not a proposition involving *a*. Now, granted, it has sometimes been supposed that all ostensibly objectual attitudes are in truth propositional attitudes. But this position is extremely hard to defend (see Forbes 2000, Montague 2007). It is possible to be afraid of a snake without quite being afraid that it will bite you, or indeed that it will harm you at all – for example when the snake appears in a film. So propositional fear is not

*necessary* for objectual fear. Conversely, it seems intuitively plausible that propositional love is not *sufficient* for objectual love: loving one's partner, or one's child, is not typically exhausted by loving that s/he is  $F_1$ , that s/he is  $F_2$ , and so on. Accordingly, there is no a priori reason to suppose that propositional perception (fact-awareness) must be either necessary or sufficient for objectual perception (thing-awareness). In fact, there is a strong reason to suppose the opposite. This is that reports of the former are referentially opaque whereas reports of the latter are referentially transparent. Thus, the following inference appears perfectly valid:  $S$  sees the third tree from the left; the third tree from the left is the tallest tree in the garden; therefore,  $S$  sees the tallest tree in the garden. By contrast, the following appears *invalid*:  $S$  sees that this tree is the third from the left; the third tree from the left is the tallest tree in the garden; therefore,  $S$  sees that this tree is the tallest in the garden. If  $S$  is unaware that the third tree from the left is the garden's tallest, then in seeing that it is the third from the left she would not yet be seeing that it is the garden's tallest.

*Concept Deployment.* For  $S$  to be fact-aware that  $a$  is  $F$ ,  $S$  must have the concept of  $a$  and the concept of  $F$  (Dretske 1993: 265). For example, to be fact-aware that the toast is burning, you must have the concept of toast and the concept of burning. Being thing-aware of  $a$ , however, does not require the mastery of any concept. A cat, for example, can smell the burning toast, and thus be thing-aware of it, even if he does not possess the concept of toast. Likewise, you can taste, and thus be aware of, an under-peaty Ardbeg Uigeadail, without being able to taste, and thus be aware, *that* this is an Ardbeg Uigeadail and an under-peaty one at that. In general, then, the contents of thing-awareness can far outstrip the concepts one possesses – whereas the contents of fact-awareness are constrained by what concepts one has. This suggests that even when we do possess the concept of an  $F$ , in being thing-aware of an  $F$  we do not *deploy* this concept.

*Direct Expressibility.* Concept possession allows for direct expression in public language. Thus fact-awareness enables one to *express* the contents of one's awareness (Dretske 1993: 266).<sup>ii</sup> When you smell that the toast is burning, your

awareness of the fact that the toast is burning might compel you to run and check the state of your toast; perhaps, while running, you might even cry out ‘The toast is burning!’ Being thing-aware of the burning toast does not have such effects on one’s verbal behavior. For one thing, if you are unfortunate enough to lack the concept of toast, you will be unable to express the content of your thing-awareness of the burning toast. You may still be able to say ‘Something’s burning!’, but that is because you are also fact-aware that *something* is burning. In most cases, of course, we do possess the relevant concepts and can thus give voice to the contents of our thing-awareness. But arguably, even then the only reason we can do so is that our possession of the relevant concepts allows us to form rather immediately the corresponding fact-awareness. For example, because in reality you do possess both the concepts TOAST and BURNING, your thing-awareness of the burning toast enables you to immediately form the fact-awareness *that* the toast is burning, which in turn enables you to assert ‘The toast is burning.’ Here, there is a sense in which the content of your thing-awareness is expressible, but *indirectly*: its expressibility is mediated by the expressibility of your fact-awareness’ content. One can recognize that this is so even without having a complete analysis of the difference between direct and indirect expressibility.<sup>iii</sup>

Dretske (1993: 266) insists that thing-awareness does not entail fact-awareness: *S* can be (thing-)aware of *a* even if *S*’s awareness is not a propositional attitude, does not deploy any concept associated with *a*, and does not impact *S*’s verbal behavior so as to enable direct expression. For example, you can see (and thus be thing-aware of) a *pyura chilensis*, without being aware *that* it is a *pyura chilensis*, without possessing the concept of *pyura chilensis*, and without being able to express what you see.<sup>iv</sup>

For Dretske, what this shows is that thing-awareness is sufficient for conscious awareness. If *S* is thing-aware of *x*, then *S* is in a conscious state of some sort (1993: 270). Accordingly, lack of fact-awareness does not entail lack of consciousness – as long as thing-awareness is present, consciousness is as well. Thus, when you see a *pyura chilensis*, your visual experience has a distinctive

phenomenal character, even though you are not aware *that* what you see is a *pyura chilensis*.

One might wonder whether in such a case there is really no *fact* you are aware of regarding the *thing* you see. For even if you are not aware that what you see is a *pyura chilensis*, you might be aware that it is, say, an animal of some sort, or at least an *entity* of some sort – that you see something rather than nothing. Thus one might object that there is at least one concept, namely the maximally generic concept ENTITY, that is necessarily applied by a subject when she is aware of something (see Dretske 1993: 268).<sup>v</sup> However, there are at least two reasons to resist this requirement. First, whether such a maximally generic concept *exists* is controversial. One fundamental function of concepts is to divide entities: a concept *F* separates the *F*s from the non-*F*s. But a concept which applies to *everything* could not do this. Secondly, it is doubtful that the maximally generic concept must be *possessed* by animals and infants in order for them to be aware of something – arguably, animals and infants do not possess such highly abstract concepts.<sup>vi</sup> It might be replied that infants do possess another kind of concept, namely, the concept of a *thing* (a material object), which can be applied to the objects of perception. The latter is a highly generic concept, though not maximally generic in the absolute sense. It is rather what one might call a *perception-relative maximally generic concept*: it is the maximally generic concept that can be applied to what can be perceived. However, even if infants possess the concept of a thing, it seems unlikely that they *apply* it every time they are perceptually aware of something.

It would seem, then, that thing-awareness does not *require* application of concepts. This is so *even if* perception is theory-laden, as is sometimes claimed (Hanson 1958). If perception is theory-laden, then it is nomically impossible for us to perceive without applying the relevant concepts at our disposal. What this alleged theory-ladenness of perception implies is, at most, that thing-perception and fact-perception are *nomically inseparable*. That is, the laws of nature exclude the occurrence of thing-perception in the absence of fact-perception. Even so, however,

it remains that the two are conceptually or logically distinguishable components or aspects of perceptual awareness.

Importantly, the notion of thing-awareness is not particularly committal on the nature of 'things' in the relevant sense. You can be thing-aware of a tree, but also of a green tree, the tree's greenness, and so on. Apparently, any kind of particular might qualify as a 'thing': an individual substance (the tree), a property-instance or trope (the tree's greenness), a qua-object (the tree qua green), and so on.<sup>vii</sup> In addition, thing-awareness is not supposed to be restricted to *physical* things. As we now turn to discuss, different kinds of mental particular, such as a mental event or conscious episode, might also be the target of thing-awareness.

### *1.2. Fact-Introspection and Thing-Introspection*

As noted, Dretske focuses on perceptual awareness. But there is no reason his distinction could not apply to introspective awareness (Giustina 2015). When you experience hunger, and attend to your experience, you may be (a) introspectively aware *of* your hunger or (b) introspectively aware *that* you are hungry. (Note well: throughout, we are using the term 'hunger' not for the bodily event but for the corresponding *experience*.) The same two kinds of report are available here:

- (3) You introspect your hunger.
- (4) You introspect that you are hungry.

Our claim is that, here too, two different mental states are reported. We label these 'thing-introspection' and 'fact-introspection.'<sup>viii</sup> Our claim is that all three differences between thing- and fact-awareness apply also to thing- and fact-introspection: thing-introspection is an objectual attitude, which does not deploy concepts, and which is not directly expressible; fact-introspection is a propositional attitude that deploys concepts and is directly expressible.

First, fact-introspection is a propositional attitude, whereas thing-introspection is an objectual attitude. When you introspect that you are hungry, or

that your hunger is mild, or that your mild hunger is not as bad as your thirst, the content of your introspective state is always a proposition. Thing-introspection of your hunger, by contrast, does not have a proposition for its content, but an individual item. Relatedly, the following inference is valid: you introspect your hunger; your hunger is a mild hunger; therefore, you introspect your mild hunger. In contrast, the corresponding inference for fact-introspection is clearly invalid: you introspect that you are hungry; your hunger is a mild hunger; therefore, you introspect that you are mildly hungry. If you pay no attention to the intensity of your hunger, you might introspect that you are hungry without introspecting that your hunger is mild.

Secondly, fact-introspection involves the deployment of concepts, whereas thing-introspection does not. It is impossible to introspect that one is hungry without possessing the concept of hunger. Indeed, the concept of hunger seems to be a *constituent* of the propositional content <I am hungry>. But it is possible for us to thing-introspect any number of experiences for which we have no suitable concepts. Clear examples of this are cases in which we have certain types of experience (which certain types of characteristic phenomenology) for the first time. Consider the first time you tasted oyster, or vegemite; the first time you smelled durian; the first time you felt in love, or felt sexual attraction. In all these cases, you were able to thing-introspect your experience, and appreciate its peculiar phenomenology thereby, but without possessing the relevant concept. Accordingly, thing-introspection does not inherently require concept-application. So even when you later do acquire the concepts OYSTER-TASTE, DURIAN-SMELL, etc., you still need not *deploy* them in order to thing-introspect your experiences. To that extent, thing-introspection does not involve *classifying* or *categorizing* one's conscious experience. It is a more direct awareness of the experience itself, in its pure phenomenal appearance so to speak.<sup>ix</sup>

Thirdly, if you introspect that your hunger is mild, then typically you can *describe* your hunger as mild on the basis of this fact-introspection of it – describe it in any public language you master sufficiently well. In doing so, you would be

expressing the contents of your fact-introspection. In contrast, thing-introspecting your mild hunger does not automatically enable you to express what you are introspecting. This may be hard to imagine in the case of mild hunger, since most of us possess the concepts HUNGER and MILD. But consider an episode in which you thing-introspect a subtle olfactory experience of an unrecognized scent. You can *name* that which you are thing-introspecting with a private name, but you cannot describe it in public language without applying concepts to it. As before, when you do possess the relevant concepts, as in the case of HUNGER and MILD, you can probably express that which you thing-introspect, but arguably, this is only because you can immediately form an appropriate fact-introspective state on the basis of your thing-introspection. Someone who thing-introspects a mild hunger and who possesses the concepts HUNGER and MILD is in a position to transition to fact-introspecting that she is having a mild hunger. Arguably, it is only because she is in a position to perform this transition that she can give voice to the contents of her thing-introspection. To that extent, only fact-introspection is directly and independently expressible, whereas thing-introspection is expressible merely indirectly and dependently on the availability of a suitable fact-introspection.

As with thing- and fact-awareness, thing-introspection does not entail fact-introspection: *S* can thing-introspect her concurrent conscious experience even if *S*'s awareness is not a propositional attitude, does not deploy any concepts associated with her experience, and does not enable direct expression. As before, one might wonder whether one can introspect, say, one's pain without attending it, if not *qua* pain, at least *qua* something, say, *this feeling*. In other words, one might object that there is at least one concept – the concept of a *feeling* (or an *experience*) – that must be applied by a subject when she introspects her current conscious experience. The latter is an *introspection-relative* maximally generic concept in the same sense the concept of a material object is a *perception-relative* maximally generic concept. It is the maximally generic concept that can be applied to what is introspected. The objector's claim is that every act of introspection must apply at least this concept to the item introspected. However, it seems unlikely that infants and animals possess

such an introspection-relative maximally generic concept, though it is also unclear whether animals and infants have the capacity to introspect. More importantly, it does not seem that people necessarily *apply* such a concept when they introspect, even when they *possess* the concept. Moreover, even if application of an introspection-relative maximally generic concept turned out to be necessary in introspection, this would not immediately threaten the thing-introspection/fact-introspection distinction. For we could still distinguish between two kinds of introspection: one that involves *only* the application of such maximally generic concepts, and another that also involves the application of more refined, more discriminating concepts. The underlying difference between introspective acts that *classify* or *categorize* the experiences they take as objects and those that do not would survive. In that case only the *articulation* of the thing-introspection/fact-introspection distinction would have to change: thing-introspection would be any introspective act in which either no concept is applied, or the absolutely maximally generic concept ENTITY is applied, or the introspection-relative maximally generic concept (FEELING OR EXPERIENCE) is. We are inclined to think that thing-introspection does not require the application of *any* concept, but those who resist need not deny the *existence* of thing-introspection – they can simply adopt this modified construal of it.

Also as before, the thing-introspection/fact-introspection distinction is consistent with the notion that introspection is theory-laden, a notion more sparsely embraced than its perceptual counterpart but not without precedent (see Churchland 1979). If introspection is theory-laden, this shows at most that it is nomically impossible for us to undergo thing-introspection without also undergoing fact-introspection; still, the two are conceptually or logically distinct components or aspects of our introspective states. Finally, the term ‘thing-introspection’ is not intended to prejudge the ontological category to which belongs the thing introspected: phenomenal events, phenomenal states, phenomenal property-instances, phenomenal tropes, phenomenal episodes, and phenomenal processes are all eligible referents of thing-introspection.

### 1.3. *Thing-Introspection, Fact-Introspection, and Reliability*

Our main contention is this: typical arguments for the unreliability of introspection target only fact-introspection, leaving the presumed reliability of thing-introspection entirely untouched.

Obviously, we cannot here go through all extant arguments for introspective unreliability.<sup>x</sup> But consider by way of example Schwitzgebel's (2008) compelling observation that we seem unable to describe in any detail our ongoing phenomenal experience. Asking us to introspect our current emotional experience, he writes:

Is it completely obvious to you what the character of that experience is? Does introspection reveal it to you as clearly as visual observation reveals the presence of the text before your eyes? Can you discern its gross and fine features through introspection as easily and confidently as you can, through vision, discern the gross and fine features of nearby external objects? Can you trace its spatiality (or nonspatiality), its viscerality or cognitiveness, its involvement with conscious imagery, thought, proprioception, or whatever, as sharply and infallibly as you can discern the shape, texture, and color of your desk? (Schwitzgebel 2008: 251)

The idea is that we would be able to answer such questions with reasonable confidence if introspection were anywhere near as reliable a belief-formation method as, say, vision is. Since we are unable to answer with any confidence, introspection must in fact be a considerably poorer belief-formation method.

Note, however, that answering the kinds of question Schwitzgebel lists would take the form of uttering indicative sentences that predicate certain phenomenal properties of your emotional experience. For these indicatives to express the content of the mental states bringing them about, the two would have to share the same type of content. Therefore, the mental states in question would have to predicate certain phenomenal properties of your emotional experience. This makes them feature all the hallmarks of *fact*-introspection. First, they have propositional content, since the indicatives that express them have propositional content. Secondly, they have conceptual content, since the predication of phenomenal

properties just is the application of corresponding phenomenal concepts. Thirdly, and most obviously, the mental states expressed by the relevant indicatives are directly expressible. Thus the introspective shortcomings highlighted by Schwitzgebel's barrage of questions are shortcomings of fact-introspection. They do not *immediately* reflect negatively on thing-introspection.

It might be objected that even if the barrage threatens *immediately* only fact-introspection, fact-introspection itself is plausibly formed on the basis of thing-introspection. If so, the relevant introspective judgments depend *ultimately* upon thing-introspection, so our inability to form the requisite judgment ultimately reflects negatively on thing-introspection as well.

In response, we agree that fact-introspection is based on thing-introspection, but point out that it is not based *only* on thing-introspection. At a minimum, fact-introspection involves also the *categorization* or *classification* of that which is thing-introspected, and thereby, the *application of a concept* to what is thing-introspected. In order to introspect that you are hungry, for example, you must not only thing-introspect your hunger, but also apply the concept HUNGER to it. Applying the concept HUNGER to it involves classifying it with other, resembling experiences as belonging to a single category or class of experiences, some (most!) of which you do not currently thing-introspect. It is only such application of a concept, implying categorization or classification of the thing-introspected item, that allows the act of fact-introspection to have a conceptual content. And such conceptual content is a necessary precondition for forming a directly expressible propositional state. One cannot come to entertain the proposition <I am hungry>, let alone fact-introspect that one is hungry, without applying the concept of hunger to the item one is aware of. The application of the concept is thus necessary for forming your fact-introspection.<sup>xi</sup>

(Note well: although the categorization process may be temporally extended, it need not be. On some occasions, you may first thing-introspect your hunger, reflect upon what kind of experience it is, and quite deliberately decide to apply

HUNGER to it. On other occasions, fact-introspection may be formed *instantaneously*, so that concepts are applied to the thing-introspected experience *immediately* and *automatically*. Indeed, this appears to be the paradigmatic case: you introspect your current conscious experience and *directly* fact-introspect that you are hungry. In these paradigmatic cases, the thing-introspection and the classification process may occur simultaneously, as one.<sup>xii</sup> Nonetheless, we can still factorize the fact-introspective state into a thing-introspective component and a classificatory component. For on the one hand, it seems impossible to fact-introspect that one is hungry without thing-introspecting one's hunger, and on the other hand, thing-introspecting one's hunger does not suffice for fact-introspecting that one is hungry.)

The question, then, is whether the fact-introspective shortcomings brought out by Schwitzgebel's line of questioning emanate from shortcomings (i) of thing-introspection, (ii) of its attendant concept-application, or (iii) of both. An initial consideration that suggests (ii) is that it is the concept-application involved in fact-introspection that takes one *beyond* what is immediately given to one in thing-introspection. This involves an implicit commitment to a similarity between what is given in thing-introspection and other experiences not currently thing-introspected. In addition, however, standard cases of uncertain or erroneous fact-introspection seem straightforwardly traceable to misapplication of a concept.

Consider first uncertain fact-introspection. In some cases, you might be dumbfounded by what you thing-introspect, unable for a while to apply *any* relevant concept to it. Suppose you pick up a can of Orangina which, unbeknownst to you, contains high-quality chocolate milk.<sup>xiii</sup> As you take your first sip, you might be so surprised as to be utterly unable to apply any gustatory concept to your taste experience.<sup>xiv</sup> Nonetheless, you are certainly thing-introspectively aware of the phenomenal character of that experience.

In other cases, you might change your mind about which concept is most appropriate to apply to your experience. As the dentist drills into your

anaesthetized tooth, you might initially classify your experience as a dull pain and only later come to realize that it is really just a sensation of pressure (Rosenthal 2005: 211). Suppose you correctly do so. Then your original classification of your experience was a *misclassification*. That is, your fact-introspection that you were having a dull pain in your tooth was erroneous. Crucially, there need not be any change in the phenomenal character that thing-introspection presents to you when you reclassify your experience as a pressure sensation rather than a dull pain.

In these cases, then, the error and uncertainty in fact-introspection clearly results not from the underlying thing-introspection, but from the categorization process applied to it. For all we have shown here, there might of course be other cases in which it is less transparent whether what went astray was the original thing-introspection or the 'superimposed' categorization. Arguably, however, standard arguments for the unreliability of introspection have failed to provide compelling instances of erroneous fact-introspection in which the error is clearly traceable to wayward thing-introspection, rather than a misclassification or miscategorization thereof.

Obviously, we cannot here go over each and every challenge to the reliability of introspection and show that it fails to challenge specifically the reliability of *thing-introspection*. But we enjoin the reader to think of the standard (or for that matter nonstandard!) challenges and consider for herself whether the problem is not plausibly traced back to the conceptualization, classification, or categorization of the thing-introspected item, rather than to the thing-introspecting itself. We predict that normal adult human thing-introspection is essentially never the main suspect when trying to explain what went wrong in some instance of mistaken fact-introspection. And so we would like to issue the following challenge to critics of introspection: *show us a case in which error or uncertainty in some introspective judgment is more plausibly traceable to thing-introspection than to the attendant concept-application.*

If this challenge cannot be met, then we have here an important asymmetry between introspection and perception. To be sure, a subject may seem to see that

there is a butterfly before her, when in fact there is not, because she does see a flying creature before her and misapplies the concept of butterfly to it. But at least in some cases, a subject may seem to see that there is a butterfly before her, when in fact there is not, because she hallucinates a flying creature where there is none (indeed where there is nothing!). Thus both failures of thing-perception and concept-application may underlie failures of fact-perception. However, as far as we can tell, all failures of fact-introspection result from failures of concept-application. In the next section, we venture an explanation of this asymmetry between thing-perception and thing-introspection.

If all this is right, then thing-introspection is plausibly highly reliable, whatever is the case with fact-introspection. It might be objected that the very notion of reliability has no intelligible application for such a 'thin' phenomenon as thing-introspection. Typically, introspection is understood as a reliable process when it leads to a preponderance of true beliefs. But beliefs are *never* formed by thing-introspection alone, so the notion of a reliable thing-introspection cannot be understood in this way. However, a refined construal of reliability is still applicable here. We can separate the different processes leading to the formation of phenomenal beliefs and ask about the *contribution* of each of these to the likelihood that the ensuing beliefs be true. To claim that thing-introspection is reliable, in this picture, is to say that when beliefs formed partly on the basis of thing-introspection are false, it is preponderantly *not* because of the contribution of thing-introspection to their formation that they are.<sup>xv</sup>

## **2. Thing-Introspection and Inner Awareness**

We have argued that standard challenges to the reliability of introspection do not seriously put in question the reliability of thing-introspection. In other words, there are no good reasons to think that thing-introspection is unreliable. But are there positive reasons to think that thing-introspection is reliable? In this section, we offer such a reason. We do so somewhat dogmatically and speculatively: we *assume*

*without argument* a certain view of the nature of phenomenal consciousness, according to which the occurrence of a conscious experience requires the subject's inbuilt, constitutive awareness of the experience (§2.1); we then offer an account of the relationship between this inbuilt awareness and thing-introspection that suggests the latter is likely infallible (§2.2). Our claim in this section is very much conditional, then: *if* our theoretical framework is accepted, *then* thing-introspection is not only highly reliable, but even infallible. We close this section with further speculation regarding potential capacity of our account of thing-introspection to vindicate the commonsense claim that phenomenal beliefs are reliably formed (§2.3). We suggest that this might be a point in the theory's favor.

### 2.1. Consciousness and Inner Awareness

According to David Rosenthal (1990), conscious states are states we are conscious *of*: in order to have a conscious pain experience, for example, I must be aware of having it. Dretske (1993) opposes this, claiming that conscious states are states we are conscious *with*, not conscious *of*; they are states in virtue of which we are aware of our external environment, though we may be completely unaware of their own occurrence. William Lycan's (1996) view is that *both* are true: conscious states are state we are both conscious of and conscious with. Ned Block's view, meanwhile, is that *neither* is true: sometimes conscious states do not make us aware of anything (Block 1996), and there may well be conscious states we are unaware of (Block 2007). Let us call awareness of the external environment *outer awareness* and awareness of one's own conscious states *inner awareness*. Then the four types of view are as follows:<sup>xvi</sup>



<u>Outer Awareness</u>	Lycan	Dretske
<u>No Outer Awareness</u>	Rosenthal	Block

**Table 1. Modern theories of consciousness: A taxonomy**

What is common to the Rosenthal- and Lycan-style views is that both insist that a conscious state cannot occur without the subject having inner awareness of its occurrence. The claim has an illustrious history: Aristotle writes that ‘knowing, perceiving, believing, and thinking are always of something else, but of themselves on the side (*en parergo*)’ (*Metaphysics* 12.9, 1074b35-6); Locke asserts that ‘Whilst [the soul] thinks and perceives ... it must necessarily be conscious of its own Perceptions’ (ECHU 2.1.12). We are well aware that the view is not uncontroversial, but will assume it henceforth.

The view that conscious experiences involve inner awareness of their own occurrence comes in two main varieties. According to ‘higher-order theory,’ subject *S* is aware of her experience *E* in virtue of being in a numerically distinct mental state *M* (Rosenthal 1990). According to ‘self-representationalism,’ *S* is aware of *E* in virtue of being in *E* itself (Kriegel 2009). Another way to put the difference is the following. When *S* is aware of *E*, *S* must be in mental state *M*, which constitutes the awareness of *E*. According to higher-order theory,  $M \neq E$ ; according to self-representationalism,  $M = E$ . That is, both views require inner awareness of one’s conscious state, but only self-representationalism demands that this inner awareness be *built into* the conscious state itself. Again dogmatically, we are going to adopt here self-representationalism, thus construing the inner awareness of our conscious states as an *inbuilt* awareness.

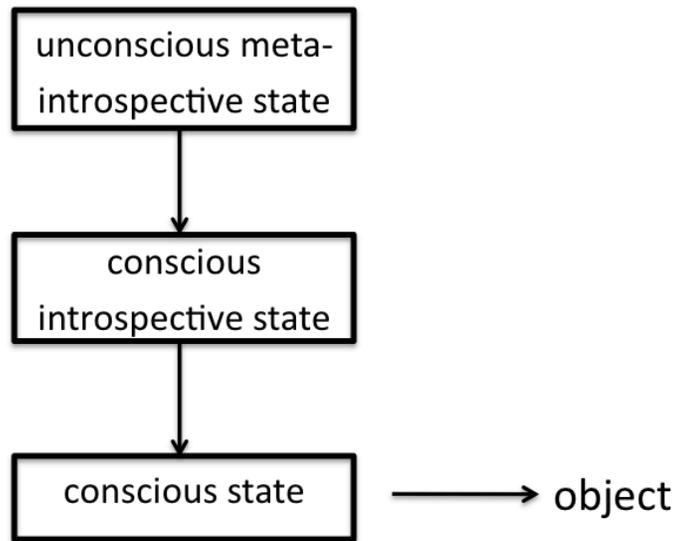
Self-representationalism itself can come in various forms. One crucial difference is between, on the one hand, views according to which a conscious state's self-representation attempts to *track* or *detect* that state's phenomenal properties and, on the other hand, views according to which the state's self-representation *determines* or *constitutes* its phenomenal properties. The two views can be separated by the following Euthyphro-style question: when a normal conscious state is phenomenally reddish and represents itself to be phenomenally reddish, does it represent itself as phenomenally reddish because it is phenomenally reddish or is it phenomenally reddish precisely because it represents itself as phenomenally reddish? The self-detection view accounts for the content of a state's self-representation in terms of that state's phenomenal properties, while the self-determination view accounts for the phenomenal properties in terms of the content of self-representation. Only on the second view is inner awareness of an experience *constitutive* of the experience's phenomenal character.<sup>xvii</sup> Here too, we are going to adopt without argument the second, self-determination view, hence construe inner awareness as constitutive.

In summary, the view of phenomenal consciousness we are adopting here has three crucial aspects: it construes all conscious states as involving inner awareness of their occurrence, it construes the inner awareness as built into those conscious states, and it construes the phenomenal properties of the conscious state as constituted by the way one is inner-aware of it. On this view, then, for a subject *S* to have a conscious experience *E* with phenomenal property *F* is for *S* to have a constitutive inbuilt inner awareness of *E* as *F*. For want of space, we have defended no part of this view here, but see Kriegel 2009 Ch.4 for a sustained argument for it.

## 2.2. Inner Awareness and Thing-Introspection

Higher-order theories have a straightforward model of thing-introspection. Recall that for the higher-order theorist, *S* has a conscious experience *E* iff (i) *S* has *E* and (ii) *S* has a (suitable) higher-order state *M* that represents *E*. To avoid infinite

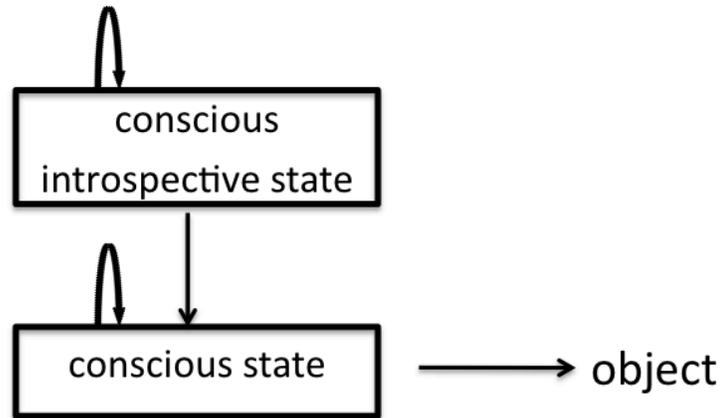
regress, higher-order theories construe  $M$  as a typically *unconscious* state. What happens when  $S$  introspects  $E$ , however, is that  $S$  is *consciously* representing  $E$ , so that  $M$  becomes itself conscious. In order for  $M$  to be conscious,  $S$  must have a third-order state  $R$  that represents  $M$ . Accordingly, for the higher-order theorist, thing-introspection involves the concurrence of *three* distinct mental states: the first-order conscious experience  $E$ , the second-order conscious introspecting  $M$ , and the third-order unconscious state  $R$  (see Rosenthal 1990). Call this the *three-state model* of thing-introspection (see Figure 1).<sup>xviii</sup>



**Figure 1. The three-state model of thing-introspection**

Self-representationalism can offer two different accounts of thing-introspection. One option is to construe thing-introspection as a second-order conscious state  $M$  that, on the one hand, represents the first-order experience  $E$  (this is what makes  $M$  *introspective*), and on the other hand, represents itself (this is what makes  $M$  *conscious*). On this model, thing-introspecting an experience involves the occurrence of *two* numerically distinct states: the first-order conscious experience, which represents both itself and some external object, and the second-order

introspective state, which represents both itself and the first-order state. Call this the *two-state model* of thing-introspection (see Figure 2).



**Figure 2. The two-state model of thing-introspection**

Another option is to construe an introspective state as structurally identical to a regular, non-introspective conscious state, differing only in respect of the distribution of a certain resource, which we may call *attention*. On this view, a regular conscious experience of a green tree represents both the green tree and itself, but does so with a certain characteristic distribution of attention: most of one's attention is dedicated to the tree, and only a small amount is dedicated to the seeing of the tree. What happens in thing-introspection is that this standard attention distribution is reversed: most of one's attention is dedicated to one's seeing of the tree, and a smaller amount to the tree (Kriegel 2009 Ch.5). Consider this toy model: at  $t_1$ ,  $S$  is absorbed with a spectacular sunset, so that  $S$ 's experience  $E$  represents the sunset with 80% of  $S$ 's attentional resources and represents itself with 20% of  $S$ 's attentional resources; at  $t_2$ , the sublime sunset has put  $S$  in an introspective mood, so that  $E$  now represents the sunset with 20% of  $S$ 's attentional resources and itself with 80% of  $S$ 's attentional resources. On this view, the introspective state is the same old first-order experience, but having undergone a

shift in attention distribution. Call this the *one-state model* of thing-introspection (see Figure 3, where relative arrow breadth stands for degree of attentiveness).



**Figure 3. The one-state model of thing-introspection**

In §2.1, we determined to adopt a self-representational account of consciousness. We therefore set aside the three-state model of thing-introspection, which flows rather from the higher-order theory. In addition, however, we find at least two reasons to prefer the one-state over two-state model.

First, if the two-state model were correct, a subject would be aware *twice over* of all her thing-introspected conscious states: once in virtue of those states representing themselves and once in virtue of those states being higher-order represented.<sup>xix</sup> Clearly, however, we do not *experience* ourselves as doubly aware of our thing-introspected states. Now, granted, the two-state proponent might be able to devise some story to explain why we experience a single awareness of our thing-introspected states (perhaps the two awarenesses are somehow ‘fused’?). It is noteworthy, however, that the one-state model does not require any further complications of this sort.

Secondly, the two-state model involves a measure of cognitive overkill. Since we must posit attention to account for independent phenomena anyway (those phenomena studied by attention psychologists), the one-state model requires no new and special posits to account for thing-introspection. The cognitive architecture posited by the two-state model is thus needlessly cumbersome.<sup>xx</sup>

For these reasons, we hereby adopt the one-state model of thing-introspection within the self-representational framework. A crucial feature of this model is that introspective awareness is the same old constitutive, inbuilt awareness that (according to self-representationalism) *every* conscious state involves, though properly augmented with additional attention resources. This means that thing-introspection, too, is constitutive and inbuilt. That is, it is (i) built into the very thing-introspected state and is (ii) constitutive of that state's phenomenal properties.<sup>xxi</sup> Here we want to focus on the second feature. Recall that on the view we have adopted here, an experience is phenomenally reddish *just if* it represents itself as phenomenally reddish; indeed, it is in virtue of representing itself as phenomenally reddish that it *is* phenomenally reddish. In unintrospected experiences, this self-representation is nonattentive. In thing-introspected ones, it is attentive. That is, for one to thing-introspect one's experience as phenomenally reddish just is for one's experience to represent itself *attentively* as phenomenally reddish. Since this self-representation of itself as phenomenally reddish is constitutive of the experience's phenomenal character, thing-introspection, somewhat trivially, cannot get the phenomenal character wrong.

What this means is that thing-introspection is in some sense *infallible*, though no thanks to any special cognitive achievement. To make the point clear, let us distinguish between *de facto* infallibility and *de jure* infallibility. A type of representation is *de facto* infallible when it targets an independent domain of facts and manages always and necessarily to represent those facts correctly. In contrast, a type of representation is *de jure* infallible when the domain of phenomena it represents has no independent existence – the facts are *fixed* by how they are represented by the relevant representations. It is an open question in our minds whether *de facto* infallibility is a real phenomenon; there is certainly something mysterious about the idea that a certain cognitive tool has reached such absolute perfection as to *guarantee* the exclusive production of correct representations. Our present claim, however, is that thing-introspection is *de jure* infallible.

This claim falls out of the combination of two theories. The first is the theory of consciousness according to which all conscious states necessarily involve a constitutive inbuilt awareness; we have offered no argument for this view. The second theory is that a conscious state's thing-introspection is the same old constitutive inbuilt awareness, but properly augmented with attention resources; we have offered some reasons for preferring this to the other model of thing-introspection consistent with our adoptive theory of consciousness. Strictly speaking, then, we have only argued for the following conditional thesis: *if one adopts our favored version of self-representationalism, then there are good reasons to think that thing-introspection is (de jure) infallible.*

### *2.3. Phenomenal-Belief Formation*

The above discussion suggests a certain picture of one central way a normal adult human can form phenomenal beliefs. To a first approximation, such phenomenal-belief formation involves four distinct stages:

- [1] Subject *S* enjoys constitutive inbuilt awareness of the phenomenal properties of her concurrent conscious experience *E*;
- [2] *S* thing-introspects *E*'s phenomenal property-instances;
- [3] *S* fact-introspects that *E* has such-and-such phenomenal properties;
- [4] *S* believes that *E* has such-and-such phenomenal properties.

Below, we offer some preliminary and somewhat speculative remarks on the transitions from [1] to [2], from [2] to [3], and from [3] to [4]. Inevitably, we will leave an enormous amount of detail open. We hope to work out some of the details in future work.

We have already suggested in §2.2 what is involved in passing from [1] to [2], that is, from inner awareness to thing-introspection – namely, a shift in attention. In a typical conscious experience, several objects are represented, but only one is attended to. Your current experience probably involves *attentive visual awareness* of

this paper you are reading, but also various forms of *nonattentive* awareness: nonattentive visual awareness of items in the periphery of your visual field, such as a desk and a book stack; nonattentive auditory awareness of footsteps in the hallway outside your office; nonattentive tactile awareness of the soles of your shoes, and so on. If we are right, then one of these elements is nonattentive *inner* awareness of your conscious experience itself. Thing-introspecting your experience involves just ‘rearranging’ your attentional structure so that you become *attentively* aware of the experience itself and *nonattentively* aware of the paper you are reading.

We have argued in §2.2 that such thing-introspection is infallible, because how it presents the phenomenal properties of the introspected experience is constitutive of what phenomenal properties the experience *has*.

As for the transition from thing-introspection to fact-introspection, as noted in §1.2, the crucial element is the classification or categorization of what one thing-introspects. This is accomplished through the application of a concept to the thing-introspected experience. For example, to form a fact-introspective state on the basis of thing-introspection of one’s hunger, one must apply the concept HUNGER to one’s thing-introspected hunger. On one view, this is *all* that is required to form a fact-introspective state with the content <this is hunger> (or <this is a hunger experience>). On another view, the formation of such a fact-introspective state is more complicated, and requires three independent mental acts: (i) the application of the demonstrative concept THIS to one’s experience, (ii) the application of the descriptive concept HUNGER to the same, and (iii) the performance of a further mental act, ‘predication,’ whereby the concepts THIS and HUNGER are ‘united’ in the state’s propositional content.

It is typically at this stage – in the transition from [2] to [3] – that infallibility is lost. If *S* is classifying her hunger not as a hunger, but as a tickle, then *S* has misclassified her hunger. In doing so, she has implicitly committed to a false proposition: that the experience she is thing-introspecting is a member of kind *K* (when in fact it is not), or that it resembles *S*’s tickle experiences more than it does

*S*'s hunger experiences (when in fact the opposite is true). Typically, then, the concept deployment involved in the transition to fact-introspection introduces the possibility of error. There *may* be quite unusual concepts that do not introduce this possibility. David Chalmers maintains that we have concepts that simply name the phenomenal property-instance we are directly aware of while we are directly aware of it, that go out of existence as soon the property-instance does, and whose content is *constituted* by the relevant property-instance; he calls these 'direct pure phenomenal concepts' (Chalmers 2003: 235). If there are such concepts, then perhaps when *S* thing-introspects a phenomenal property-instance and applies to it such a concept, *S* cannot fall into error. In like fashion, suppose that in addition to the demonstrative concept THIS, which simply demonstrates one's experience, we have a demonstrative concept LIKE THAT, which merely demonstrates the kind of phenomenal quality we are thing-introspectively aware of. Then on the basis of thing-introspecting a phenomenal episode of ours, we can always form a fact-introspective state whose content is <this is like that>. It is an open question in our minds whether such a fact-introspective state is susceptible to error; there are good reasons to think it is not (Giustina 2015). If so, *some* fact-introspection may yet be infallible. Nonetheless, for the great majority of fact-introspective states, infallibility is not in the cards. Contrary to a certain Cartesian tradition, fact-introspecting that you are hungry, that you have a headache, that you are sad, or that you are angry are all susceptible to error. At the same time, it would be quite surprising if fact-introspection were not on the whole quite reliable. After all, it is based on thing-introspection, which is infallible, plus a relatively modest cognitive act of concept-application.

How about the transition from [3] to [4] – say, from fact-introspecting that you feel hungry to *believing* that you feel hungry? Consider, by analogy, perceiving that the table is brown and the belief directly based on it. This kind of transition does not seem to involve any *inference*. Instead, it involves taking the perceptual appearances at face value, that is, committing to things really being the way they perceptually seem to be. We might call this kind of operation *endorsement*. The idea,

then, is that believing that the table is brown can be formed on the basis of endorsing one's perception that the table is brown.<sup>xxii</sup> Crucially, endorsement is *content-preserving*: one can form a belief by endorsement only if the content of one's belief is identical to the content of the endorsed state (in this case, the perception that the table is brown). Consider the following contrast. Suppose you see in northern Mexico many systems of long tunnels with nesting chambers at their end, in which you find pure white eggs. Given sufficient background knowledge, you could infer from this that there are Burrowing Owls living in the area. In this case, you form the belief that there are Burrowing Owls in the area on the basis of perceiving that there are Burrowing-Owl-type nests in the area. But on the same basis you could also form the more cautious belief *that there are Burrowing-Owl-type nests in the area*. More generally, a belief that  $p$  may be based on perception in one of two ways: either it is arrived at by inference from the perception that  $q$  or it may be formed by endorsement of a perception that  $p$ . The belief can be justifiably formed either way, but it is formed through a different operation in each case. Now, just as certain perceptual beliefs can be formed by endorsing perception-that, certain introspective beliefs can be formed by endorsement of introspection-that. One can form the belief that one is hungry by simply endorsing the fact-introspection that one is hungry. In doing so, one would only be committing to things really being the way they introspectively seem to be. That is, one would be taking the introspective appearances at face value.

Endorsing appearances introduces further possibilities of error. Looking at a Müller-Lyer arrow, most of us will (mis)perceive that one arrow is longer than the other, but will refrain from endorsing our perceptual experience. In doing so, we correctly refuse to take the appearances at face value. If  $S$  does endorse her misperception of one arrow being longer than the other, she will have formed a false belief. Thus endorsement may lead to false belief. At the same time, it is noteworthy that on the whole endorsement is a much safer operation than inference. If one correctly perceives that there are Burrowing-Owl-type nests around, one is much likelier to fall into error by coming to believe that there are Burrowing Owls around

(through inference) than by coming to believe that there are Burrowing-Owl-type nests around (through endorsement). Obviously, the added epistemic risk *has a point*, insofar as it expands one's knowledge. Nonetheless, other things being equal an endorsement-formed belief that *p* is more likely to be true, and is thus more epistemically justified, than an inferentially formed belief that *p*. To that extent, beliefs formed on the basis of endorsing fact-introspection are, other things being equal, more epistemically justified than beliefs formed on the basis of inference from fact-introspection. Since phenomenal beliefs are typically formed by endorsement rather than inference, they constitute a particularly secure kind of belief.

Much more remains to be said by way of filling in this model of the formation of (some) phenomenal beliefs. Already this preliminary sketch, however, suggests an important point, namely, that phenomenal beliefs are *noninferential*. After all, their formation relies on the cognitive operations of (i) attention-shifting, (ii) categorization or concept-application, and (iii) endorsement, but not on any *inferential* operation (deductive, inductive, abductive, or other).

This point is in the first instance *psychological*: it concerns how phenomenal beliefs are *actually* formed (at least sometimes, perhaps typically). But it also has *epistemic* significance: it has implications for how *justified* such phenomenal beliefs typically are. If these phenomenal beliefs are ultimately based on inner awareness, and are based on it noninferentially, then they are ultimately based on infallible foundations, and are based on them through operations that are on the whole safer than inference (they involve a lower susceptibility to the introduction of error). If so, then while phenomenal beliefs are not themselves infallible, those of them which are based on the formation process sketched here are likely *the most secure beliefs we have*.<sup>xxiii</sup> (More accurately, they are likely the most secure *empirical* beliefs we have; certain a priori beliefs, such as the belief in the law of contradiction, may be even more secure.)

## Conclusion

We have set to defend the reliability of introspection against the skeptical challenge by pointing at an often neglected phenomenon, thing-introspection, and showing the psychological and epistemic consequences of our account of it. We have argued that *thing*-introspection has not been shown to be unreliable, and may even be infallible. Our defense comprises a negative component and a positive component.

The negative component addresses skeptical arguments against the reliability of phenomenal beliefs. The general form of those arguments tends to be this: phenomenal beliefs are ultimately based on introspection; introspection is unreliable; therefore, phenomenal beliefs are unreliable. We have argued that such arguments only target fact-introspection, and leave the question of the reliability thing-introspection untouched. If error occurs in fact-introspection, then it is more likely not due to a flaw in introspection proper, but rather to miscategorization.

The positive component of our defense is grounded in a certain view of phenomenal consciousness, according to which the occurrence of a conscious experience requires the subject's inbuilt, constitutive awareness of the experience. By consequence, our positive claim is a conditional one: if our theoretical framework is accepted, then thing-introspection is not only reliable, but quite likely infallible. We have suggested that our account of thing-introspection allows for a model of phenomenal-belief formation process which could pave the way for a defense of phenomenal beliefs. Arguably, that our phenomenal beliefs are typically justified is an article of common sense. To that extent, it is a strength of any theory of belief formation that it vindicates the notion that phenomenal beliefs are typically justified.

It must be pointed out, however, that our discussion does not quite defeat skeptical arguments about phenomenal beliefs. For we have only argued for the claim that *thing*-introspection is highly reliable (and may turn out to be infallible). We have *not* shown that the attendant categorizing process leading from thing-

introspection to fact-introspection and the phenomenal belief is reliable. Therefore, we have not demonstrated that, contrary to what skeptics argue, phenomenal beliefs are on the whole justified. Nevertheless, the proposed model of phenomenal-belief formation process might lay the basis for a more developed defense of the reliability of phenomenal beliefs. We hope to pursue this wider project in future work.

Our discussion has important implications for both psychology and epistemology. Regarding the former, since the demise of introspectionist psychology about a century ago, introspection has been considered illegitimate in psychological inquiry. And appeal to introspection does raise a number of genuine issues, in particular as concerns replicability and peer disagreement (Watson 1913, Bayne and Spener 2010). If we are right, however, then modulo these issues, at the core of introspection is a phenomenon which is fundamentally trustworthy, one it might be foolish to leave entirely unexploited.

As for epistemology, our discussion may offer new support for a certain traditional version of foundationalism. On the view we have in mind, all inferentially justified beliefs are justified in virtue of being correctly inferred from *noninferentially* justified beliefs, and noninferentially justified beliefs come in only two varieties – beliefs about one’s ongoing conscious experience and certain a priori beliefs. Traditionally, these foundational beliefs were construed as infallible (BonJour 2001). Such foundationalism has faced two paramount challenges: to show how noninferential justification is possible, and to fend off challenges to the infallibility of beliefs about one’s ongoing conscious experience. Our discussion suggests a fairly developed model of the noninferential justification involved in phenomenal beliefs, thus potentially addressing the first challenge.<sup>xxiv</sup> As regards the second challenge, our discussion can be seen as recommending moving away from the claim that phenomenal beliefs are infallible to the more modest claim that phenomenal beliefs are the most secure we have; this latter claim seems sufficient to justify treating them as foundational, especially given that they are noninferentially justified.<sup>xxv</sup>

## References

- Aristotle. *Physics*. Trans. R. Waterfield. Oxford: Oxford University Press, 1996.
- Aristotle. *Metaphysics*. Trans. W.D. Ross. Oxford: Clarendon Press, 1908
- Bayne, T. and M. Spener 2010. 'Introspective Humility.' *Philosophical Issues* 20: 1-22.
- Block, N.J. 1996. 'Mental Paint and Mental Latex.' *Philosophical Issues* 7: 19-50.
- Block, N.J. 2007. 'Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience.' *Behavioural and Brain Sciences* 30: 481-499.
- Bonjour, L. 2001. 'Toward a Defense of Empirical Foundationalism.' In M. DePaul (ed.), *Resurrecting Old-Fashioned Foundationalism*. Lanham, Ma: Rowman and Littlefield.
- Brentano, F.C. 1874. *Psychology from Empirical Standpoint*. Trans. A.C. Rancurello, D.B. Terrell, and L.L. McAlister. London: Routledge, 1973.
- Chalmers, D.J. 2003. 'The Content and Epistemology of Phenomenal Belief.' In Q. Smith and A. Jokic (eds.), *Consciousness: New Philosophical Perspectives*. Oxford and New York: Oxford University Press.
- Chisholm, R. 1957. *Perceiving: A Philosophical Study*. Ithaca, NY: Cornell University Press.
- Churchland, P.M. 1979. *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press.
- Dretske, F.I. 1993. 'Conscious Experience.' *Mind* 102: 263-283.
- Fine, K. 1982. 'Acts, Events and Things.' In W. Leinfellner, E. Kraemer, and J. Schank (eds.), *Sprache Und Ontologie*. Vienna: Holder-Pichler-Tempsky.
- Forbes, G. 2000. 'Objectual Attitudes.' *Linguistics and Philosophy* 23: 141-183.
- Giustina, A. 2015. *Toward a Defence of Phenomenal Knowledge*. MA Thesis, EHES.
- Hanson, N.R. 1958. *Patterns of Discovery*. Cambridge: Cambridge University Press.
- Kriegel, U. 2009. *Subjective Consciousness: A Self-representational Theory*. Oxford: Oxford University Press.
- Locke, J. *An Essay Concerning Human Understanding*. Ed. P.H. Nidditch. Oxford: Oxford University Press, 1975.
- Lycan, W.G. 1996. *Consciousness and Experience*. Cambridge MA: MIT Press.
- Montague, M. 2007. 'Against Propositionalism.' *Noûs* 41: 503-518.
- Nisbett, R.E. and T.D. Wilson 1977. 'Telling More Than We Can Know: Verbal Reports on Mental Processes.' *Psychological Review* 84: 231-259.

- Rosenthal, D.M. 1990. 'A Theory of Consciousness.' ZiF Technical Report 40, Bielfeld, Germany. Reprinted in N.J. Block, O. Flanagan, and G. Guzeldere (eds.), *The Nature of Consciousness: Philosophical Debates*. Cambridge MA: MIT Press, 1997.
- Rosenthal, D.M. 2005. *Consciousness and Mind*. Oxford: Clarendon Press.
- Schwitzgebel, E. 2008. 'The Unreliability of Naïve Introspection.' *Philosophical Review* 245-273.
- Sellars, W. 1956. 'Empiricism and the Philosophy of Mind.' In H. Feigl and M. Scriven (eds.), *Minnesota Studies in the Philosophy of Science, I: Foundations of Science and the Concepts of Psychology and Psychoanalysis*. Minneapolis: University of Minnesota Press.
- Shoemaker, S. 1996. *The First-Person Perspective, and Other Essays*. Cambridge: Cambridge University Press.
- Watson, J.B. 1913. 'Psychology as the Behaviorist Views It.' *Psychological Review* 20: 158-177.

---

<sup>i</sup> One more specification is needed to home in on the beliefs we are interested in against the background of reductive physicalism. According to reductive physicalism, every phenomenal property is identical to some physical property. Wearing a top-shelf cerebroscope connected to an information processing system, one might form beliefs about physical properties of one's concurrent brain states that are *in fact* phenomenal properties. This kind of belief is not, however, the kind of belief we are interested in. So for a reductive physicalist, we propose that a phenomenal belief be construed as a belief about the phenomenal properties of one's concurrent conscious states *under a phenomenal description* (or presented under *a phenomenal mode of presentation*). For the sake of simplicity, we will ignore this further specification in what follows.

<sup>ii</sup> As noted, concept possession is necessary for language expression. Thus, what is fundamental to fact-awareness is application of concepts – enabling language expression is a derivative feature. However, stressing direct expressibility helps to understand what kind of concepts are deployed in fact-awareness. Dretske himself seems to suggest this: 'Generally speaking, the concepts necessary for awareness of facts are those corresponding to terms occurring obliquely in the clause (the that-clause) describing the fact one is aware of.' (1993: 265, fn 8)

<sup>iii</sup> It might be argued that, in fact, thing-awareness *is* directly expressible in public language: a subject who has the concept TOAST, and is aware of a toast, can directly express her thing-awareness simply by saying 'toast,' even if she is not aware of any fact concerning the toast. However, in this case what is directly expressed is not the subject's thing-awareness of the toast, but rather her fact-awareness that what she is aware of is a toast (i.e., the concept TOAST applies to it).

<sup>iv</sup> Here is how Dretske (1993: 266) puts it: 'Ignorance of what armadillos are or how they look may prevent someone from being conscious of certain facts (that the object crossing the road is an armadillo) without impairing in the slightest one's awareness of the things – the armadillos crossing roads – that (so to speak) constitute these facts.'

<sup>v</sup> Dretske himself is aware of this possible objection: 'One can, to be sure, see armadillos without seeing that they are armadillos, but perhaps one must, in order to see them, see that they are (say) animals of some sort. [...] If this sounds implausible (one can surely mistake an animal for a rock or a bush) maybe one must, in seeing an object, at least see that it is an object of some sort. To be aware of

---

a thing is at least be aware that it is... how shall we say it? ... a thing. *Something or other.*' (1993: 268; italics original)

<sup>vi</sup> Indeed, Dretske makes quite a similar point: 'It seems most implausible to suppose infants and animals (presumably, conscious of things) have concepts of this sort. If the concept one must have to be aware of something is a concept that applies to *everything* one can be aware of, what is the point of insisting that one must have it to be aware?' (1993: 268-269)

<sup>vii</sup> A qua-object is distinguished from a property-instance in being a bona fide concrete particular. Thus one difference between the tree-qua-green and the tree's greenness is that you can bump into the former but not into the latter. For the idea of a qua-object, see Fine 1982 (though the notion itself goes back to Aristotle's discussion of the relationship between the man and the musical man in *Physics* I.7).

<sup>viii</sup> We do recognize the measure of infelicity in the term 'thing-introspection.' What we are introspectively aware of are clearly not 'things' in the ordinary sense of physical concrete particulars. As we have seen above, however, a 'thing' in the relevant sense can belong to any ontological category. This includes mental states, events, and processes, which *are* the proper objects of what we call thing-introspection. If the reader finds the label too grating, we enjoin her to substitute 'item-introspection' for our 'thing-introspection.' We use the latter only to underline the analogy with Dretske's distinction.

<sup>ix</sup> It is worth noting the existence of an intermediary report between 'S introspects her hunger' and 'S introspects that she is hungry'; namely, 'S introspects her hunger *as a hunger.*' This latter report is ostensibly a report of an *objectual attitude* but also one of a *concept-deploying* state. Accordingly, we would not admit it under the rubric of thing-introspection (as we use the term). We are open to two possibilities. The first is that in addition to fact-introspection and thing-introspection there is a third kind of introspection. The second is that 'S introspects her hunger as hunger' is a misleading report of fact-introspection: what it reports is that *S* introspects *that* her experience is a hunger experience.

<sup>x</sup> Among them are the argument from ineffability of conscious experience (Schwitzgebel 2008), the argument from introspective disagreement (e.g., Bayne and Spener 2010), the fraternity initiation case (Shoemaker 1996), and the dental drilling case (Rosenthal 2005). Some of them are analyzed in Giustina (2015), and replied to in terms of the thing-introspection/fact-introspection distinction. As noted, we cannot go through all of them here; we plan to address them, using the thing-introspection/fact-introspection distinction, in future work.

<sup>xi</sup> On a natural model, a propositional content such as <I am hungry> involves as constituents (i) the concept of hunger, (ii) the indexical I-concept, and (iii) a predicative link between these two (whatever that amounts to). Since (i) is a constituent of the propositional content, it is impossible to entertain the content without possessing and applying the concept of hunger.

<sup>xii</sup> We do not mean to commit here to the substantive possibility of instantaneous concept application. Perhaps the process running from an experiential state to the application of a concept to that state must take some time, even if it is automatic. Our present point is just that our account can accommodate both options.

<sup>xiii</sup> Thanks to Josh Weisberg for this example.

<sup>xiv</sup> You might apply the concept(s) DISGUSTING VISCOUS EXPERIENCE, but this is not a particularly gustatory set of concepts: DISGUSTING is rather an aesthetic concept, VISCOUS is rather a tactile concept, and EXPERIENCE is more generic than gustatory concepts.

---

<sup>xv</sup> From this perspective, when one mistakenly fact-introspects that one's experience is such-and-such, there is a sense in which the mistake is actually not one of *introspection proper*, but of some downstream processes. Imagine a patient whose eyesight is irreproachable, but who, due to brain lesion, routinely misapplies shape concepts to what she sees. As a result, she seems to see *that* the building is square when in fact it is rectangular, seems to see *that* the stop sign is round when in fact it is octagonal, and so on. This patient's perceptual fact-awareness is defective, leading to a preponderance of nonveridical visual judgments. And yet there is a sense in which *vision proper* is perfectly reliable in her – there is nothing wrong *with her eyesight*. It is not clear to us that fact-introspection is as unreliable as it is often claimed to be (Giustina 2015), but even *where* it proves untrustworthy, we maintain that the situation resembles that of the above patient: *introspection proper* is perfectly reliable, and it is only the downstream processing of introspective information that is defective.

<sup>xvi</sup> We suppress here a modal operator: clearly, Dretske can allow that *some* conscious states are such that their subject is aware of them, Rosenthal can allow that some are such that in virtue of being in them we are conscious of something else, and so on. But what they will deny is the *necessity* of the presence of such consciousness-of (in Dretske's case) or consciousness-with (in Rosenthal's).

<sup>xvii</sup> Might the two come apart? There is no reason why not. Consider the following variation on the inverted spectrum thought-experiment. Imagine two subjects whose spectra are not inverted but whose inner awareness is 'inverted': looking at a red strawberry, both enter a visual state that represents the strawberry as red, but one's inner awareness represents her visual state as phenomenally red while the other's represents hers as phenomenally green. According to the first version of self-representationalism, the conscious state is phenomenally red but happens to misrepresent itself. According to the second version, the conscious state is phenomenally green, because all there is to a conscious state being phenomenally green is that it represents itself as so.

<sup>xviii</sup> Rosenthal himself does not seem to recognize thing-introspection in his account of introspection. But a higher-order theorist could, in principle, make room for it. This is especially feasible for a higher-order *perception* rather than higher-order *thought* theorist (Lycan 1996).

<sup>xix</sup> For a similar consideration see, e.g., Brentano 1874: 121-6.

<sup>xx</sup> One might object that second-order states must be posited anyway: for example, we can have beliefs about beliefs. However, there is a considerable asymmetry between the two-state model and the case of a belief about a belief. In the former case a new kind of state is introduced (namely, the introspective second-order state); in the latter the second-order state is not of a new kind (it is just another belief).

<sup>xxi</sup> One might wonder how thing-introspection is built into the introspected state, and constitutive of its phenomenal properties, given that it requires additional attention resources. It is worth pointing out that thing-introspection is inbuilt and constitutive of the introspected state *as thing-introspected* – as opposed to the relevant conscious state as non-thing-introspected, namely, simply *as conscious*. The additional attention involved in thing-introspection contributes to determine the phenomenal character of the thing-introspected state. Therefore, for any conscious state C, the phenomenal character of C as thing-introspected might be different from that of C as non-thing-introspected.

<sup>xxii</sup> A completely different but coherent view in this area is that fact-awareness and belief are in reality two different descriptions of one and the same state – perceiving that the table is brown *just is* believing that the table is brown. (Dretske himself often sounds like that.) We set aside this view for purely strategic reasons: we indulge our opponent's supposition that an extra step is needed here in order to form phenomenal beliefs, and show that *even so*, phenomenal beliefs are on the whole justified.

---

<sup>xxiii</sup> In addition, there may be a small subset of phenomenal beliefs that *are* infallible, namely, phenomenal beliefs with the content <this is like that>. Whether this is so will depend on whether it is possible to endorse a fact-introspective state with that content without introducing the possibility of error – something we have not discussed here.

<sup>xxiv</sup> Obviously, many challenges would have to be faced up to before this account can be taken to be complete. For example, we have said nothing here about how we might address the problem of ‘myth of the given’ (Sellars 1956), that is, the problem surrounding how a non-conceptual state could justify a conceptual one, or even enter the so-called space of reasons.

<sup>xxv</sup> For comments on a previous draft, we are grateful to Jonathan Farrell, Brie Gertler, and Tom McClelland. We have also benefited from a presentation at Taiwan Medical University; we are grateful to the audience there, in particular Alex Byrne, Austen Clark, Sam Coleman, and Kevin Kimble.