

2018

# Explaining the Illusion of Phenomenal Consciousness

Daniel S. Shabasson  
*CUNY Graduate Center*

[How does access to this work benefit you? Let us know!](#)

Follow this and additional works at: [https://academicworks.cuny.edu/gc\\_pubs](https://academicworks.cuny.edu/gc_pubs)

 Part of the [Philosophy of Mind Commons](#)

---

## Recommended Citation

Shabasson, Daniel S., "Explaining the Illusion of Phenomenal Consciousness" (2018). *CUNY Academic Works*.  
[https://academicworks.cuny.edu/gc\\_pubs/434](https://academicworks.cuny.edu/gc_pubs/434)

This Article is brought to you by CUNY Academic Works. It has been accepted for inclusion in Publications and Research by an authorized administrator of CUNY Academic Works. For more information, please contact [AcademicWorks@gc.cuny.edu](mailto:AcademicWorks@gc.cuny.edu).

## EXPLAINING THE ILLUSION OF PHENOMENAL CONSCIOUSNESS

Daniel Shabasson

[DanShab@live.com](mailto:DanShab@live.com)

According to illusionism, phenomenal consciousness does not exist. There is nothing “it is like” to see red or feel pain. Most people find illusionism highly counterintuitive and it remains a minority view among philosophers. To increase its intuitive plausibility, we proponents of illusionism must solve what Keith Frankish (2016) has termed *the illusion problem*. We must explain why phenomenal consciousness seems to exist and why the illusion that it exists is so powerful. Focusing on introspective judgments about our color experiences, I propose a theory to solve the illusion problem.

I make several working assumptions in this paper. With Frankish (2012), I adhere to *strong* illusionism and I shall suppose that there are no viable “diet” conceptions of qualia lacking any of the three characteristics of what Frankish calls *classic qualia*—intrinsicity, ineffability, and subjectivity. Thus, I use “qualia” exclusively to refer to classic qualia, and one of my aims will be to explain why we erroneously attribute these three characteristics of classic qualia to our experiences. I also adopt the working assumption that qualia do not exist. I beg no question in doing so, for I do not seek to prove, in the first instance, that qualia do not exist. Instead, I seek to answer the following question: if qualia did not exist, could we nevertheless plausibly explain our powerful disposition to judge that they do? To examine this question, we must adopt the working assumption that qualia do not exist. If we manage to expound a plausible theory to explain the disposition that does not posit the existence of qualia, we can then

decide between that theory and qualia realism according to the relative explanatory power of each.

Some proponents of illusionism maintain that we will need to make significant advances in neurology and cognitive science and learn more about the brain before we can hope to understand the causes of the illusion. But I think this is mistaken. I intend to show here that we can *already* understand the general causes of the illusion—even if important details may need to be filled in—based on three principles. The first principle (section 1) is “Impenetrability,” according to which we have no introspective or conscious access to the processes in our brains, operating sub-doxastically, that engender our introspective judgments about our sensory experiences, nor do we have access to the non-phenomenal (functional, neural, or physical) properties of our experiences that our brains sub-doxastically apprehend to engender our introspective judgments about our sensory experiences. The second principle (section 2), the “Infallibility Intuition,” is our strong disposition to judge our introspective judgments about our sensory experiences to be infallible, or impervious to doubt.<sup>1</sup> The third principle (section 3) is the “Justification Intuition,” an epistemic constraint on judgment/belief. For any agent *a* judging *p*, it must seem to *a* that she is justified in judging that *p* at the time she judges that *p*. Thus, the patent absurdity/incoherence of *a* saying or thinking: *p*, but *I have no reason to judge/hold/believe p*. In section 4, I show that we can explain the illusion of phenomenal consciousness based on these three principles.

---

<sup>1</sup> I use “intuit” and “intuition” throughout this paper as *non-factive*. For example, I argue that we intuit that our introspective judgments about sensory experience are infallible, but I am not committed to such judgments *actually* being infallible.

## 1. Impenetrability

When an agent views a solidly colored red patch, light enters her eyes, strikes the rods and cones of the retina, and a neural signal is transmitted into her brain. The brain processes these neural signals and, under normal conditions, engenders a reddish sensory experience. I shall suppose the reddishness of this experience to be either a functional, neural, or physical property of the brain, but I shall remain agnostic as to which and refer to such properties simply as “non-phenomenal.” Upon introspecting her sensory experience, the agent *classifies* it, forming the introspective judgment *this is a reddish experience*. This introspective judgment is formed by a *sub-doxastic process*, in which the agent’s mind sub-doxastically apprehends or recognizes the signature of the non-phenomenal property of reddishness, forms the judgment classifying the experience as reddish based on this apprehension, and then makes that judgment available to consciousness. Although the agent is consciously aware *that* she forms the introspective judgment that her experience is reddish, she in the dark both with respect to the nature of the sub-doxastic processes of her brain responsible for engendering the judgment, and with respect to the true nature of the non-phenomenal (functional, neural, or physical) property of reddishness sub-doxastically apprehended. That information is *not* available to consciousness. Let us refer to this phenomenon as “Impenetrability.”

### **IMPENETRABILITY:**

An introspective judgment is impenetrable iff the agent is consciously aware that she forms it, but she lacks conscious access both with respect to the nature of the sub-doxastic processes responsible for engendering it, as well as the true nature of the sub-doxastically apprehended non-phenomenal property.

While an agent can use the words “reddish” and “reddishness” when classifying her experience as reddish to refer to the non-phenomenal property of reddishness, she remains fundamentally

ignorant of the true nature of reddishness. She lacks introspective access to the details of the sub-doxastically apprehended property's complex relational and fully effable functional/neural/physical nature.

Not only do we form introspective judgments *classifying* our sensory experiences, but we also form introspective judgments *differentiating* them. For example, an agent viewing a partially red and partially green colored patch simultaneously has different types of sensory experiences, some reddish, some greenish, corresponding to the red and green parts of the patch. She classifies these sensory experiences differently and is consciously aware of the difference in classification, despite knowing nothing about the genuine differences between these types of experiences. She cannot learn introspectively about the specific ways in which the fully effable and relational non-phenomenal properties apprehended sub-doxastically differ. That information is not available to consciousness.

Here's a somewhat metaphorical way of summing up the foregoing claims. Consciousness receives "reports," as it were, from sub-doxastic mental processes, stating how various experiences are to be classified and distinguished by type. Suppose that an agent's reddish and greenish experiences differ with respect to various neural properties. The report that the agent's conscious mind receives omits all the details with respect to the mechanisms by which the different neural signatures or neural patterns of these reddish and greenish experiences have been recognized, classified, and differentiated by her brain, and the specifics of how these reddish and greenish experiences differ neurally. The conscious mind is informed only the *upshot* of the sub-doxastically carried out information processing: how to name and classify various experiences, which experiences are distinguishable from others by type, and which are of the same type. Despite the omission of details and specifics, consciousness accepts the reports it

receives as true—and as I’ll argue in section 2, as *infallibly* true. Here, we have information-processing activities whose precise nature, while highly complex, are to be investigated and described by cognitive science as among the “easy” problems of consciousness.

It should not surprise us that introspective processes about sensory experiences would be impenetrable. Why should we have evolved to consciously apprehend lower-level processing details of brain mechanisms? What survival advantage would that offer us? We negotiate the world by knowing how sensory experiences should be classified and distinguished from others, not by understanding how our brains go about ascertaining this at the neural level. Furthermore, to apprehend the details of what our brains are carrying out sub-doxastically, we would need a *second* introspective system to apprehend the details of these sub-doxastic activities and report them to consciousness. To apprehend the activities of this second introspective system, a *third* introspective system would be required, *ad infinitum*. Unless there were an infinite number of introspective systems in our finite brains, we should fully expect there to be some basic, primitive level or levels of cognitive processing whose activities are not apprehended, monitored, or checked by any further system, and whose activities are therefore impenetrable or inaccessible to consciousness.

## **2. The Infallibility Intuition**

An agent viewing a solidly colored red patch will typically judge that the patch is red. She treats her judgment as fallible. Perhaps, she thinks, lighting conditions are abnormal, and she has misperceived the color of the patch she is viewing. By contrast, she takes her introspective judgment about her sensory experience being reddish to be infallible or impervious to doubt, claiming absolute authority on what her experience is like. She intuits: *if I introspectively judge*

*that my experience is reddish, then it is.* When it comes to how the color *seems* to her, she thinks: *appearance is reality.*

We have the powerful intuition that our introspective judgments about the reddishness (or the greenishness, the yellowishness, the painfulness, etc.) of our sensory experiences are infallible or impervious to doubt. I call this the “Infallibility Intuition.”

### **INFALLIBILITY INTUITION:**

An agent takes her introspective judgments about her sensory experiences to be infallible or impervious to doubt.

The Infallibility Intuition is strongly anchored the folk psychological theory of the mind, and it is also represented in views of some philosophers who have intimated that our introspective judgments about our sensory experiences are *in fact* infallible (e.g., Descartes (1641)). Indeed, a plausible case could be made for infallibility by arguing that introspectively judging a sensory experience to be reddish is precisely what *makes* it reddish, especially if non-phenomenal properties such as reddishness turn out to be functional in nature.<sup>2</sup> However, I shall suppose for the sake of argument that such judgments are *not* infallible. We can nevertheless explain why it strongly seems to us that they are due to their being *highly reliable*. As Daniel Dennett (2002) put it:

“You can’t have infallibility about your own consciousness. Period. But you can get close — close enough to explain why it seems so powerfully as if you do. First of all, the intentional stance (Dennett, 1971; 1987) guarantees that any entity that is voluminously and reliably predictable as an intentional system will have a set of beliefs (including the most intimate beliefs about its doxastic experiences) that are mainly true. So

---

<sup>2</sup> A functionalist about the reddishness of a sensory experience might want to claim that the reddishness of an experience consists (at least in part) in the agent’s disposition, upon introspection, to utter “I am having a reddish experience” and to classify the currently introspected experience together with other experiences the agent had in the past when she viewed red objects (and then having the disposition to behave in a manner consistent with that classification). Hence, an agent’s classifying her sensory experience as reddish is what would *make* it reddish, regardless of how the functional property happened to be neurally realized. Despite the *prima facie* appeal of this argument, I shall remain agnostic with respect to the question whether any introspective judgments are in fact infallible or whether they just strongly seem to be so.

each of us can be confident that in general what we believe about our conscious experiences will have an interpretation according to which we are, in the main, right. How wrong could I be? Not that wrong. Not about most things.”

If Dennett is correct, which I believe he is, in his claim that we are (at least) largely correct in our introspective judgments about our sensory experiences, this would suffice to explain why we have the infallibility intuition. I submit that there would be a powerful evolutionary advantage in taking our introspective judgments as infallible. Harboring doubts about our introspective judgments could cause us to waver in our decisions, hesitating instead of acting decisively to avoid danger or pursue fruitful opportunities. Moreover, a skeptical attitude with respect to introspective judgments would enhance fitness only if our minds possessed both the ability to detect errors and to correct them once detected. Such a system would surely be costly in evolutionary terms and would involve a radical expansion of our introspective capacities, necessitating a further introspective system to monitor our introspective judgments about sensory experiences. In short, the evolution of such an ability, both to detect and correct errors, would be too costly to justify its emergence, given how rarely we are in error with respect to our introspective judgments about our sensory experiences.<sup>3</sup>

We should not find it surprising that we privilege some judgments, treating them as impervious to doubt. Suppose, after all, we subjected *all* our judgments to potential doubt. A sentient being treating nothing as the basic raw data of experience, the given, could fall into epistemological paralysis, with no safe foundation upon which to construct a stable edifice of belief. To create such a stable edifice, evolution would favor the emergence of minds brute-causally reacting to introspective data, even if this might, on rare occasions, result in misrepresentation. Furthermore, the potential for global doubt in our minds—where *any*

---

<sup>3</sup> Again, we must distinguish being in error about the color of an object we are viewing on the one hand and being in error about the nature of our sensory experiences (i.e., being in error about a neural or functional property of our sensory experience) on the other.

judgment might be subjected to doubt—would present a regress problem: for every class of judgment, we would require a cognitive system to check the reliability of those judgments, and then a further system to check reliability of the judgments of *that* system, *ad infinitum*. To stop the regress, therefore, one should expect to encounter in the mind a most basic, primitive level of judgments not susceptible to doubt. Introspective judgments are plausible candidates for being the most basic sorts of judgments and thus treated by our minds as the infallible, given, raw data of experience, immune from doubt.

Whatever ultimately explains the Infallibility Intuition—the genuine infallibility of such introspective judgments, their mere high level of reliability, or some evolutionary story—the theory of the illusion of phenomenal consciousness I elaborate here does not hinge on *why* we have the intuition or whether the intuition is correct. It requires only that we *have* the intuition, and I think that we do.<sup>4</sup>

### 3. The Justification Intuition

I posit that we have a powerful intuition I shall call the “Justification Intuition,” an epistemic constraint on judgment/belief. For any agent *a* judging that *p*, it must seem to *a* that she is justified in judging that *p* at the time she judges that *p*. Thus, the patent absurdity/incoherence of *a* saying or thinking the following:

*p, but I have no reason to judge/hold/believe p*

---

<sup>4</sup> We lack the Infallibility Intuition with respect to non-introspective judgments. For example, we may initially see an individual from a distance and judge that we see person A, a friend, but on closer inspection we realize it is person B, an enemy. It was crucial to our survival throughout our evolution that we be open to belief revision and treat these judgments as fallible. Likewise, the non-introspective judgment that a solidly colored object is red is treated as fallible because we can sometimes misperceive colors when seen in unusual viewing conditions. We regard these judgments with a degree of skepticism when we have reason to believe our perceptions may be in error, and we can take remedial action by looking at the object from a different angle or under different lighting different conditions, etc. By contrast, we intuit (whether we are right or wrong to do this) that the process of determining how a solid color *seems* to us, the introspective judgment about our own sensory experience, cannot go wrong.

Agent *a* would contradict herself, thinking something incoherent. She would not *really* judge that *p* if it seemed to her that she lacked justification for her judgment at the time she so judges. Judgments and putative or felt supporting justification/evidence/reasons for it are necessary companions.

### **JUSTIFICATION INTUITION<sup>5 6</sup>**

For any agent *a* judging that *p*, *a* intuitively feels that she is justified in judging that *p* at the time she judges that *p*.

This is not to say we can always correctly *identify* the reasons or evidence we take ourselves to possess for any judgment or belief. We are frequently bad at identifying these. Nor do I suggest that reasons, evidence, or other forms of justification must in fact exist for our judgments. We often possess the intuition that there is some reason or other, when there is in fact nothing at all supporting a judgment. And even where we identify some reason *r* for judging that *p*, I do not mean to suggest that *r* must support the judgment *in fact*. Reason *r* could be irrelevant to the truth or falsity of *p*, or pseudo-justificatory. Instead, I am making the far more modest claim that in every case of judging, we *feel or intuit* the existence of some justification to back up our judgment. Having such an intuition of justification is what separates a *bona fide* act

---

<sup>5</sup> Setiya (2008) suggests a similar epistemic constraint on belief (in the context of discussing Bernard Williams' (1976) argument that you cannot intentionally decide to believe a proposition for which it seems to you that you lack sufficient evidence):

*It is impossible to believe that *p* or to be confident that *p* while believing that this degree of confidence or belief is not epistemically justified.*

See also Gibbons (2005): "If you think about Moore-paradoxical statements from the normative perspective, then the same kind of incoherence that is involved in the standard cases also seems to infect the following: *p*, but I have no reason to believe that *p*..."

<sup>6</sup> This constraint does not apply to the sorts of judgments we might better refer to as statements of opinion, which are non-truth-apt.

of judging from an empty mouthing of words.<sup>7</sup> An agent who felt her judgment to be arbitrary and unjustified would *ipso facto* not stand behind it, and thus not judge at all.

Perhaps, then, we can say that it is *conceptually impossible* to judge that *p* and simultaneously think *I have no reason to judge that p*.<sup>8</sup> If a judgment without putative justification is conceptually impossible, it should not be surprising that we have the intuition that every judgment is accompanied by justification. Our intuition would track this conceptual truth. And even if I were wrong about this being a conceptual truth—if it turned out after all to be conceivable that one might judge that *p* without intuiting any justification for one’s judgment—I suspect that having the Justification Intuition would be fitness enhancing and we should therefore expect to find, on evolutionary grounds, that the intuition guides our doxastic dispositions, although I shall not speculate about this matter here.

#### 4. The Illusion of Phenomenal Consciousness

My central claim is that qualia, a.k.a. phenomenal properties, are putative properties whose instantiation in our minds we take to justify our introspective judgments about our sensory experiences.

**QUALIA:** Putative properties whose instantiation in our minds we take to justify our introspective judgments about our sensory experiences.

---

<sup>7</sup> Such intuited justifications, when we undertake to articulate them, are often expressed with statements as weak as “it strongly seems to be the case,” “it feels to me to be that way though I cannot say why,” or “it is self-evident.” The point is not that we have the ready capacity to identify and articulate some genuinely cogent justification for our judgments, but only that we feel or intuit the present of such justifications upon forming a judgment as an inherent part of the act of judging.

<sup>8</sup> I do not mean to suggest that we are incapable of entertaining *philosophical* doubt with respect to our judgments or beliefs. If I judge that *p*, I intuit that my judgment is justified, but this does not exclude the possibility that I may intellectually, using powers of higher reason, come to doubt that my reasons for judging that *p* are cogent or compelling. In any case, any such cases of higher-level doubt would not typically be simultaneous with the judgment, and moreover, the act of doubt would not be within same module of the mind in which the judgment was formed.

In subsection (I) I explain why we intuit that qualia are instantiated in our minds. In subsection (II), I explain why we take ourselves to *directly apprehend* these instantiated qualia. In subsection (III), I explain why we take our experiences to be “like something.” In subsection (IV), I explain our disposition to attribute ineffability, intrinsicality, and subjectivity to our experiences. In subsection (V), I discuss the evidence in support of my claim that the issue of justification is central to explaining the Illusion of Phenomenal Consciousness. In section (VI), I explain why the folk take illusionism to be counterintuitive. In section (VII), I defend the theory of the illusion presented herein from the objection that the illusion of phenomenal consciousness would itself be a representation with its own phenomenal properties in need of explaining away.

**(I) Why we take qualia—putative properties justifying our introspective judgments about our sensory experiences—to be instantiated in our minds.**

Through rational reconstructions of our intuitive judgment processes with respect to introspective judgment tokens IJR and IJG<sup>9</sup> set out below (“IJR and “IJG” standing for “Introspective Judgment Reddish” and “Introspective Judgment Greenish,” respectively), I shall illustrate why we take qualia to be instantiated in our minds.

**IJR:** *This is a reddish experience*

**IJG:** *This is a greenish experience*

Note that in the rational reconstructions below, the word “I” refers exclusively to the personal-level, doxastic-level, conscious agent, and it excludes the sub-personal, sub-doxastic, infra-conscious, impenetrable processes at work in the agent’s brain. Also note that I am *not* claiming that these rational reconstructions represent the *thought* process of any agent—that would over-

---

<sup>9</sup> It is important to note that IJR and IJG are introspective judgment *tokens*, i.e., particular acts of judging.

intellectualize matters. These rational reconstructions are meant rather to capture what is happening on the level of *intuition* from the agent's first-person perspective.

### **IJR: Rational Reconstruction**

- (i) IJR is true → There is a property instantiated in my mind—call it “P<sup>Reddishness</sup>”—justifying judgment IJR.  
[from the JUSTIFICATION INTUITION]
- (ii) I infallibly know that IJR is true.  
[(from the INFALLIBILITY INTUITION)]
- (iii) Therefore, I infallibly know that P<sup>Reddishness</sup> is instantiated in my mind.

### **IJG: Rational Reconstruction**

- (i) IJG is true → There is a property instantiated in my mind—call it “P<sup>Greenishness</sup>”—justifying judgment IJG.  
[from the JUSTIFICATION INTUITION]
- (ii) I infallibly know that IJG is true.  
[(from the INFALLIBILITY INTUITION)]
- (iii) Therefore, I infallibly know that P<sup>Greenishness</sup> is instantiated in my mind.

I shall now clarify the premises and the conclusions in these rational reconstructions, beginning with premise (i). That premise comes from the Justification Intuition. Again, the Justification Intuition says that an agent will intuit that every judgment she forms is justified at the time she forms it. Hence, if an agent has judged her experience to be reddish, it must seem to her that there is *something about her experience* in virtue of which she is justified in judging it to be reddish. (By *reductio*, if it seemed to her that there were nothing about her experience in virtue of which her judgment that it was reddish was justified, her judgment would seem to her to be

arbitrary and unjustified.) To say that “there is something about her experience” in virtue of which it seems to her that she is justified in judging it to be reddish is precisely to say that it seems to her that *there is some property* of her experience in virtue of which she is justified in her judgment that the experience is reddish. Let us call this property  $P^{\text{Reddishness}}$  for judgments that an experience is reddish, and  $P^{\text{Greenishness}}$  for judgments that an experience is greenish. An agent takes her judgments to be justified if and only if it seems to her that these justificatory properties are instantiated in her mind at the time she forms the judgments. I claim that these putative justificatory properties such as  $P^{\text{Reddishness}}$  and  $P^{\text{Greenishness}}$  are qualia, or phenomenal properties.<sup>10</sup>

Premise (ii) in the rational reconstructions above follows from the Infallibility Intuition, according to which an agent takes all her introspective judgments about sensory experience to be infallibly true.

---

<sup>10</sup> I posit that these phenomenal properties are putatively “*exclusively* justifying,” meaning that, e.g., the instantiation of  $P^{\text{Reddishness}}$  would seem to us to exclusively justify judgments of the type *this experience is reddish*, and foreclose any judgments of different types, such as, e.g., *this experience is greenish*. And the instantiation of  $P^{\text{Greenishness}}$  would seem to us to exclusively justify judgments of the type *this experience is greenish*, and foreclose any judgments with different content, such as, e.g., *this experience is reddish*.

#### **EXCLUSIVELY JUSTIFYING PROPERTY**

A property is exclusively justifying if its instantiation guarantees the truth of one judgment type alone, and categorically excludes the truth of any other judgment types.

The putative infallibility of introspective judgments entails that phenomenal properties are exclusively justifying properties. We can see this by *reductio*. Suppose hypothetically that a putatively instantiated justificatory phenomenal property,  $P^{\text{Reddishness}}$ , justified the introspective judgment *this experience is reddish*, but that there were a slight chance that instantiation of  $P^{\text{Reddishness}}$  could *also* justify the judgment *this experience is greenish*. If this were the case, then we could not see our judgment that the experience is reddish as infallible, for we could not exclude the possibility, based on verifying that  $P^{\text{Reddishness}}$  is instantiated in our mind, that our judgment had the content *this experience is greenish*. If we deem our introspective judgment as infallible, the justification we have for that judgment cannot be seemingly equivocal, such that it would appear to us an open question as to whether some other judgment type could be true based on the instantiation of the justifying property. That would be inconsistent with infallibility. Infallibility requires rather that the apprehension of a particular sort of evidence *guarantees* that a particular type of introspective judgment is true, and categorically excludes the truth of other types of judgments. Thus, for each putatively justificatory property backing any judgment, we take the justificatory property as justifying that judgment alone and not potentially justifying any other sort of judgment. We take  $P^{\text{Reddishness}}$  as exclusively justifying introspective judgments of the type *this is a reddish experience*, and  $P^{\text{Greenishness}}$  exclusively justifying introspective judgments of the type *this is a greenish experience*.

The Justification and Infallibility intuitions, embodied in premises (i) and (ii) of the rational reconstructions, together entail the conclusion, (iii), according to which any agent forming an introspective judgment about her sensory experience takes herself to know infallibly that a justificatory phenomenal property such as  $P^{\text{Reddishness}}$  or  $P^{\text{Greenishness}}$  is instantiated in her mind.

## (II) Why do qualia seem *directly apprehensible*?

In the above subsection, I argued that when we introspect our sensory experiences we intuit that we infallibly know that justificatory phenomenal properties, putatively justifying our introspective judgments about these sensory experiences, are instantiated. That was the conclusion of the rational reconstructions. What is missing thus far is an explanation for why it seems to us that we *directly apprehend* these phenomenal properties. Phenomenal properties seem to present themselves to us immediately and directly without any initiative or mental effort by our minds. We do not experience ourselves as having to actively reflect and effortfully introspect to see that these properties are instantiated.

Let us say that a property is *directly apprehensible* iff no effortful reflection or chain of inferences, except for a simple logical inference from premises deemed to be infallibly known or self-evident, is required to (seemingly) infallibly verify the instantiation of the property whenever it is instantiated.

### **DIRECT APPREHENSIBILITY**

A property seems directly apprehensible iff no effortful reflection or chain of inferences, except for a simple logical inference from premises deemed to be infallibly known or self-evident, is required to (seemingly) infallibly verify the instantiation of the property whenever it is instantiated.

We judge qualia to be directly apprehensible according to this definition, and the instant theory explains why. Consider again the rational reconstruction of our intuitive reasoning processes with respect to introspective judgment token IJR, reproduced below:

**IJR: Rational Reconstruction**

- (i) IJR is true → There is a property instantiated in my mind—call it “ $P^{\text{Reddishness}}$ ”—justifying judgment IJR.  
[from the JUSTIFICATION INTUITION]
- (ii) I infallibly know that IJR is true.  
[(from the INFALLIBILITY INTUITION)]
- (iii) Therefore, I infallibly know that  $P^{\text{Reddishness}}$  is instantiated in my mind.

We treat premises (i) and (ii) as infallibly known, for they embody principles—the Infallibility and Justification Intuitions—which are integral parts of the way we conceptualize the world. They have the status of conceptual truths for us. From those two premises, conclusion (iii), according to which we infallibly know that a justificatory property,  $P^{\text{Reddishness}}$ , is instantiated in the mind, follows by a simple logical inference. Hence, we directly apprehend  $P^{\text{Reddishness}}$  according to the above definition of direct apprehensibility.

Consider the absurdity of thinking the following:

*I am certain that IJR is true, and I am certain that if IJR is true, necessarily property  $P^{\text{Reddishness}}$  is instantiated in my mind, but I must effortfully reflect on whether  $P^{\text{Reddishness}}$  is instantiated in my mind.*

To think this would be absurd. From the fact that I intuit that my introspective judgment IJR is doubtlessly true, it follows logically and immediately, per the Justification Intuition, that

$P^{\text{Reddishness}}$  is instantiated in my mind. No reflection is required to reach this conclusion.<sup>11</sup> The justifying phenomenal property's instantiation appears to me equally as doubtless and certain as the truth of the classifying introspective judgment I take its instantiation to justify.

### (III) Why do we judge that our sensory experiences to be “like something”?

Thus far, I have argued in subsection (I) that we intuit that there are justificatory properties supporting our judgments instantiated in us, and these are qualia, a.k.a. phenomenal properties. Moreover, in subsection (II) I have explained why it seems that we directly apprehend these instantiated properties. But why do we experience our sensory experiences as *being like something*? I'll now turn to this crucial question.

As I have claimed above, I take properties  $P^{\text{Reddishness}}$  and  $P^{\text{Greenishness}}$  to justify different judgment types, the former (exclusively<sup>12</sup>) justifying judgments of the type *this experience is reddish*, and the latter judgments of the type *this experience is greenish*.  $P^{\text{Reddishness}}$  and  $P^{\text{Greenishness}}$  must seem different to me, and this can be shown by the following *reductio*. Suppose that  $P^{\text{Reddishness}}$  and  $P^{\text{Greenishness}}$  did *not* seem different to me. I would not be able to tell them apart. Thus, it would seem to me that the very same evidence could be the basis for judging either that an experience was reddish or greenish. Now, suppose I judged a particular experience to be reddish. How could I feel that this judgment was justified if the evidence seemed to me to equally well support the judgment that my experience was greenish? My judgment either way—either that the experience was reddish or greenish—would be experienced by me as an arbitrary choice between two equally good alternatives, essentially a stab in the dark. However, the

---

<sup>11</sup> And consider, hypothetically, that I *did* need to pause to reflect if my judgment about my sensory experience was justified; that in and of itself would be a form of doubt, a wavering of sorts, which would be inconsistent the Strong Infallibility Intuition.

<sup>12</sup> See footnote 10, *supra*, for why I claim these phenomenal properties would be taken to *exclusively* justify their respective judgment types.

Justification Intuition bars me from seeing my judgments as arbitrary or unjustified. Therefore, the justificatory properties for my judgments of different types, such as  $P^{\text{Reddishness}}$  and  $P^{\text{Greenishness}}$ , must seem different to me.

The Infallibility Intuition also explains why  $P^{\text{Reddishness}}$  and  $P^{\text{Greenishness}}$  must seem different to me. If the same evidence could result in judgments of various types, such judgments would not strike me as infallible. Again, suppose that the same evidence could be the basis for judging either that an experience was reddish or greenish, and suppose I ended up judging a particular experience of mine to be reddish. How could I feel that this judgment was correct if the evidence seemed to me to equally well support the judgment that the experience was greenish? My judgment either way—either that the experience was reddish or greenish—would be experienced by me as an arbitrary choice between two equally good alternatives, a stab in the dark. Therefore, the judgment would not seem to me to be infallibly true. But I *do* experience my introspective judgments as infallibly true. So  $P^{\text{Reddishness}}$  and  $P^{\text{Greenishness}}$  must seem different to me.

In short, the combined force of the Justification and Infallibility intuitions entails that  $P^{\text{Reddishness}}$  and  $P^{\text{Greenishness}}$  must seem different to me. Since they seem different to me, each putative phenomenal property, considered individually, must seem like something.  $P^{\text{Reddishness}}$  must seem like one thing, and  $P^{\text{Greenishness}}$  like another. For consider the absurdity of saying or thinking:

*$P^{\text{Reddishness}}$  and  $P^{\text{Greenishness}}$  seem different to me, but neither  $P^{\text{Reddishness}}$  nor  $P^{\text{Greenishness}}$ , considered individually, seems like anything to me.*

The above is absurd, incoherent. Two entities that seem different to me cannot be such that I see them as not being like anything considered individually. *It is a conceptual truth that if I see two things as distinct, I must see each individually as being like something, such that, as it were, the*

*something the one seems like is not identical to the something the other seems like.* Thus,  $P^{\text{Reddishness}}$  seems like something, and  $P^{\text{Greenishness}}$  seems like something else. Each of those putative phenomenal justificatory properties is tied to my reddish and greenish sensory experiences and to the introspective judgments I form about them. This explains why my reddish and greenish experiences appear to have distinct what-it's-likenesses.

Now, phenomenal properties such as  $P^{\text{Reddishness}}$  and  $P^{\text{Greenishness}}$  do not exist and are never in fact instantiated. Hence, neither of these justificatory properties is *really* distinct from the other, nor is either individually like anything. I consciously apprehend only the bare fact that reddish and greenish experiences are distinct (even if, sub-doxastically, my brain discerns a genuine difference in the non-phenomenal properties of reddish and greenish experiences). How I think about these nonexistent phenomenal properties is governed by the same logic governing how I would think about existent properties. I would not distinguish between two existent properties or entities unless it seems to me that I discerned a difference, unless each seemed to me to be “like” something different, meaning that each seemed like something considered on its own. Therefore, it also seems to me that the non-existent phenomenal properties  $P^{\text{Reddishness}}$  and  $P^{\text{Greenishness}}$  each seems like something considered on its own.

#### **(IV) Why do we attribute Ineffability, Intrinsicity, and Subjectivity to our experiences?**

Why do we attribute the three characteristics of classic qualia Frankish (2012) identifies—intrinsicity, ineffability, and subjectivity—to our experiences?

##### ***a. Ineffability***

According to the theory I have elaborated above, we cannot put what phenomenal reddishness is like into words because, in fact, the property does not exist, even though, for

reasons elaborated above, the property seems to us to exist, to be directly apprehensible, and seems to be “like” something. Of course, there *are* indeed justificatory properties backing our introspective judgments about sensory experience, but these are *non-phenomenal* properties. However, due to Impenetrability, we can say nothing about them via introspection. They are not available to consciousness.

***b. Intrinsicity***

By “intrinsic,” I mean *non-relational*. There are two ways for a property to be non-relational.

The first way of being non-relational I shall call *internal non-relationality*. An internally non-relational property is atomic and simple, meaning that it has no describable or discernable parts. Without parts, there can be no description of the property in terms of the relation of the parts to one another. An example of an internally relational property would be circularity. A circle has parts and various discernable elements. We have the radius, the circumference, and the various points making up the circle, which are related to one another in space according to mathematical description. Now, unlike circularity, which is a genuine property, phenomenal reddishness does not exist. Hence, phenomenal reddishness, being non-existent, cannot have parts. Of course, we might *fictionally* imagine of a non-existent entity that it had parts. But we do not, as a matter of fact, do this with reddishness. If you reflect on reddishness and try to discern any parts, you will fail. Our minds do not *invent* an intricate story about reddishness having parts, and in this sense the illusion of phenomenal consciousness is a *minimalist illusion*: no illusory details of what phenomenal properties are like are represented, even fictionally.

The second sort of non-relationality we may call *external non-relationality*. A property that is externally non-relational has no discernable relational properties with respect to any

other property, object, or individual outside of it. Phenomenal properties such as phenomenal reddishness seem to lack these sorts of external relational properties. This is reflected in the folk's<sup>13</sup> intuition in the inverted spectrum thought experiment, where the folk intuit that a spectrum inversion would be conceivable yet undetectable. It would be undetectable, per the folk, because no relational property could ever be articulated to ascertain whether two individuals had their phenomenal spectra inverted. Thus, relational properties, such as facts about what sorts of memories or feeling two individuals might associate with phenomenal color sensations, have no bearing on whether the phenomenal colors are inverted. If for me reddishness is associated with hot and my bluishness with cold, and you have the exact opposite associations, this is, per folk intuitions, insufficient to ascertain whether there is an inversion. After all, you might experience the raw feel of reddishness, just as I do, but you simply associate different things with that raw feeling. Conversely, we can conceive of the possibility that your qualia are inverted but your associations with those qualia are inverted as well. Either way, what one associates with one's qualia, per folk intuitions, do not settle the matter. What is intuitively significant is only what your experience subjectively and inwardly *seems like* to you, not how you react, what you say, what other properties you associate with individual qualia, or any other sort of external reaction to the experience. Phenomenal properties seem intrinsic: we cannot get at them, identify them, ascertain their essences, in terms of relational properties.

What explains why we take phenomenal properties to be intrinsic in this second sense, as externally non-relational? Impenetrability explains it. The complex relational properties of the *non-phenomenal* properties, which is apprehended sub-doxastically, cannot be seen by the

---

<sup>13</sup> I include proponents of illusionism among *the folk*. Even those philosophers who think illusionism is true, such as I, continue to have powerful non-illusionist intuitions. It is just that, for philosophical reasons, proponents of illusionism do not trust these intuitions. These intuitions never go away no matter what position one may take on the reality (or unreality) of phenomenal consciousness.

conscious mind. At the conscious level, we receive a report from the sub-doxastic mind, saying that the experience is reddish, *but no further details are received*. Reddishness is experienced at the conscious level as a mere *that*, with the label “reddish” attached. Of course, given that this label is supposed to distinguish this property from greenishness and all other experience types, it must seem to us that reddishness is like something *specific*, unlike that specific thing that greenishness or yellowishness or pain is like. As elaborated above, if we did not see the properties putatively evidencing our distinct judgments as distinct, and each being like something, all our judgments differentiating various sensory experiences would strike us as arbitrary and unjustified. Thus, the mind experiences an intuition/feeling of reddishness having a specific nature, even though there are no details filled in. The result is the intuition that we experience an intrinsic property: one that has its own very specific nature, but whose specific nature is not constituted by its relation to anything else (either outside of it or inside of it). The specific nature of the property seems unspecifiable *simpliciter*, or ineffable.

**c.     *Subjectivity***

Whereas an objective property/fact is one that can be recognized to obtain by any agent, regardless of the agent’s perspective, a subjective property/fact is one that can be recognized only from a particular perspective. The folk’s intuition in the inverted spectrum case is that phenomenal properties are subjective entities. The folk’s intuition is that no inter-subjectual comparison between my reddish and greenish experiences, and yours, is possible. Hence, it is *per se* impossible ever to know whether the spectra are inverted. Despite this, there is supposed to be some fact of the matter—a subjective fact—about whether the spectra are inverted.

Now, there is simply no such thing as a subjective fact. Why do we intuit that facts about phenomenal properties, facts about our experiences, are subjective? Part of the explanation is

that since those putative phenomenal properties are seemingly both intrinsic and ineffable, we have no way to individuate them or differentiate between them *except by experiencing them on introspection*. And introspection is, as a conceptual matter, *essentially* bound to a single introspecting subject. Only I can introspect *my* experiences. You cannot do experience my experiences. If you could experience them, introspect on them, at that moment of joint introspection we would be one subject, rather than two distinct subjects. At least, this is my intuition, and I suspect it is the intuition of the folk as well.

Since we take phenomenal properties to be revealed and knowable only through acts of introspection, which distinct subjects cannot carry out jointly or share, distinct subject cannot exchange information about what these phenomenal properties are like. Thus, we judge that phenomenal properties are knowable only to a lone, single subject.

**(V) What is the evidence that the issue of Justification is central to explaining the Illusion of Phenomenal Consciousness?**

I claim that the issue of *justification* is the key to the mystery as to why we take our experiences to be like something. After all, the Justification and Infallibility intuitions, which are central elements of my theory, are epistemic notions relating to the notion of justification of judgment/belief.

Here is a consideration I take to militate in favor of my claim that the issue of justification is the key to the explanation of the illusion of phenomenal consciousness: *If we are asked to say how we know our experience is reddish, or how we know we are in pain, we say “because it seems reddish to me” or “it seems painful to me.”* We allude to the instantiation of a putative phenomenal property as the evidence/justification for our introspective judgments. I see this as powerful evidence that the putative phenomenal property is there to fill a “justificatory

gap,” as it were, arising out of the Impenetrability of the introspective judgments, a gap between the genuine justificatory process our minds apprehend only sub-doxastically on the one hand, and a putative phenomenal property we take to be the justificatory property on the other.<sup>14</sup> The gap is filled because the Justification Intuition disposes us to expect conscious apprehension of the evidence for our introspective judgments. The putative phenomenal property allows our minds to make sense of their judging activities. Without imagining that there exist phenomenal properties, our introspective judgments would appear to us to be senseless and arbitrary.<sup>15</sup>

#### **(VI) Why does illusionism strike the folk as counterintuitive?**

If we take these non-existent phenomenal properties as justifying our introspective judgments, as I claim, we should not find it surprising that the folk find illusionism absurd. For illusionism posits that what we take as the justification for our introspective judgments is non-existent. Therefore, Illusionism seems to call into question the reliability of our introspective judgments, judgments which we take to be infallible. Thus, illusionism clashes directly with the Infallibility Intuition, and therefore seems counterintuitive.

Of course, the unreality of the phenomenal states we posit as justifications for our introspective judgments does not *really* call the reliability of these introspective judgments into question, since with every introspective judgment our brains reliably (although not infallibly) apprehend a non-phenomenal property sub-doxastically. It is just that because of Impenetrability,

---

<sup>14</sup> Indirectly, in referring to the non-existent phenomenal property, we are in fact referring to the underlying non-phenomenal property forming the genuine justificatory basis of the introspective judgment.

<sup>15</sup> Importantly, again, the property we allude to is not the *real* justification—it is not what really justifies us in our judgments, but only what we fancy justifies them because while the Justification Intuition demands we apprehend a justification, at the same time Impenetrability prevents us from providing a genuine justification because we cannot see the true causes, the brain activities, and the non-phenomenal properties they sub-doxastically apprehend, engendering the judgment.

we fail to know anything about that non-phenomenal property. We seem to see instead an *Ersatz* phenomenal property (and by “Ersatz” I mean something like the original German meaning: *stand-in, substitute, surrogate, or replacement* for the real thing). If Impenetrability were not the case, we would be able to directly see the genuine justifications for our introspective judgments about our sensory experiences. But given impenetrability, we are stuck with apprehending fictional *Ersatz* justifications for our judgments. The conscious mind must make do with a stand-in property, the ghostly shadow of the non-phenomenal property we cannot consciously apprehend.

### **(VII) The Illusion of Phenomenal Consciousness Involves no Misrepresentation**

Derk Pereboom, himself sympathetic to illusionism, points out a worry for Illusionism in his 2016 (185).

“... Illusionists agree that the what-it’s-like features of sensory states, what I call the qualitative natures of phenomenal properties, are illusory in that they don’t exist. **But what of the illusion itself? Experiencing an illusion requires that the illusions themselves, which are kinds of introspective representations, will exist. ... However, these illusions themselves will have phenomenal properties, or so it would seem. ... Accordingly, the objector may claim that there will be something it’s like to have an illusion of phenomenal greenness, and it’s the same as what it’s like have a sensation of green.** If this is right, then the illusionist’s strategy won’t get us very far (see Pereboom, 1994, pp. 582–4; Alter, 1995; and Chalmers, 1996, p. 142, for versions of this objection). [Bolds mine]

To sum up Pereboom’s worry: even if my phenomenal reddish state is not like something, my misrepresentation of it as being like something must itself be like something.

The theory I propose is immune from Pereboom’s worry, because on my theory, the illusion of phenomenal consciousness is not a “kind of introspective representation.” *The illusion involves no representation or misrepresentation at all.* The illusion is constituted rather by a set of dispositions to draw false inferences, to posit non-existent properties that are ineffable, intrinsic, subjective, directly apprehensible, and “like something.” But there is no

representation of how one phenomenal property is different from any other, not even fictionally. If there were such a representation, we would be able to say something about how this difference is represented to us. We have the *mere conviction* that there are differences, without representing any.

Consider an optical illusion, such as the Müller-Lyer illusion. This sort of illusion is not constituted merely by a conviction. With the Müller-Lyer illusion, we can describe the misrepresentation: the mind misrepresents lines of equal lengths as being of different lengths. The lines being of different lengths is a genuinely conceivable states-of-affairs. But how does the mind represent reddishness as being different from greenishness? It doesn't. That's why we have ineffability—if a difference were represented, we could say what the difference consisted in. The mind is simply convinced there is a difference it apprehends, but it does not represent that difference.

To construe the illusions posited by illusionism as consisting in (mis)representations is to fail to recognize the strength of strong illusionism's commitment to the non-existence of phenomenal properties. According to illusionism, phenomenal properties are far more robustly unreal than Sherlock Holmes, who even though non-existent, is fictionally depicted in his fictional world as having identifiable and describable features. A reddish experience is like nothing whatsoever, not even fictionally in the mind of an agent subject to the illusion. That is why I draw a total blank when asked to describe what reddishness is. I do not represent of the experience as being like anything. I just have the raw but extremely powerful conviction the experience is like something, and different from what a greenish experience is like. An illusion without misrepresentation can be just as strong as an illusion that misrepresents (and perhaps even stronger). If the theory I present herein is true, the illusion of phenomenal consciousness is

quite different in nature from optical illusions, which always involve misrepresentations. Perhaps the illusion of phenomenal consciousness is *sui generis*.<sup>16</sup>

## REFERENCES

- A. (1998). *Descartes Meditations*. Cambridge University Press.
- Frankish, Keith. (2016). Illusionism: special issue. Exeter: Imprint Academic.
- Frankish, Keith. (2012). "Quining diet qualia." *Consciousness and Cognition*, 21(2), 667-676.
- Gibbons, John (2005). "Knowledge, The Priority of." Encyclopedia of Philosophy. Retrieved April 20, 2018 from Encyclopedia.com: <http://www.encyclopedia.com/humanities/encyclopedias-almanacs-transcripts-and-maps/knowledge-priority>
- Pereboom, Derk, (2017). *Illusionism and Anti-Functionalism about Phenomenal Consciousness*, pp. 184-198, in Frankish, K. (2016). Illusionism: special issue. Exeter: Imprint Academic.
- Setiya, Keiran, (2008). "Believing at Will," *Midwest Studies in Philosophy*, 32 (1):36-52.
- Williams, Bernard. (1976). "Deciding to believe," in *Problems of the Self*, Cambridge University Press, pp.136-151.

---

<sup>16</sup> I suspect that the illusion of the flow of time is likewise an illusion without misrepresentation. Time is not genuinely represented as flowing or moving, since the notion that time flows is not merely false, but somehow *incoherent*. The flow of time is inconceivable, and so cannot be represented. Likewise, in the case of qualia, I argue that qualia are impossible objects that we do not and cannot represent. The illusion consists merely in a disposition to argue for a certain set of impossible properties possessed by our experiences, and to be convinced that this disposition is coherent and that when we speak a certain way we get at some truth, that we apprehend something about the underlying nature of our experience.