Olivia Y. Lee
PHIL 186 Final Paper

The Missing Piece: Dispelling the Mystery of Introspective Illusion

Theories of consciousness typically deal with the hard problem of consciousness, the problem of explaining the relationship between physical states and mental states with phenomenal qualities (i.e., "phenomenally conscious" states) (Chalmers, 1996). Such experiences are deemed to have introspective phenomenal qualities, often referred to as the "raw feels" determining what it is like to undergo those experiences. Attempts to tackle this problem have prompted a longstanding debate between two main camps: one seeks to significantly revise or extend current science to accommodate phenomenal properties, and the other seeks to explain their existence with current physical theories. Frankish (2016) refers to these two camps as radical realists (encompassing dualists, neutral monists, mysterians, and "those who appeal to new physics") and conservative realists (encompassing physicalists and representationalists). Notably, both groups accept that phenomenal qualities are manifestly real qualities of mental states, though they differ in their explanations accounting for the existence of said qualities. An orthogonal view to both, however, is the theory of *illusionism* about phenomenal qualities, the view that phenomenal characteristics of mental states are illusory. Rather than trying to account for the existence of phenomenal qualities, illusionism seeks to explain why such qualities seem to exist but in reality, do not, replacing the hard problem with the illusion problem (Frankish, 2016).

Illusionism is not a popular view, despite support from some prominent contemporary philosophers, and is commonly dismissed as "failing to take consciousness seriously" (Chalmers, 1996). Frankish (2016) puts forward the case for illusionism, summarized in Section I of this paper. While Frankish clearly articulates arguments against realism and in favor of illusionism, and certain elements of his argument are indeed compelling, they seem to fall short in an important respect. Rather than targeting specific theoretical weaknesses in arguments for illusionism, as is typically done in debates between realist camps, skeptics simply deem the theory "obviously false" (Goff, 2016), "utterly implausible" (Balog, 2016), and "absurd" (Nida-Rümelin, 2016). The notion that one's introspective grasp of phenomenal states is illusory and there is actually *nothing* it is like, in a qualitative sense, to feel pain, see red, taste wine, or smell lavender, is highly unappealing to many philosophers. The illusionist's response to critics is to deny the existence of the very thing being debated and which critics insist exists, leading to an unsatisfying stalemate. This paper seeks to address the question of why the intuition that phenomenal states exist endures, and why illusionism is so hard to stomach. It contends that beyond explaining how the illusion of phenomenality arises, a robust theory of illusionism must adequately explain the incredible strength of the illusion and the difficulty of freeing oneself from the grip of this enduring intuition. Frankish attempts (with limited success) to address the former but not the latter, and explaining the potency of the illusion is the crucial missing piece for a sound illusionist theory.

Section I of this paper explores motivations for illusionism, drawing from Frankish (2016) with additional appeals to higher-order theories of consciousness. Section II notes that Frankish's arguments for illusionism, while clearly articulated and conceptually compelling at points, does not adequately explain the potency of the illusion that phenomenal qualities exist, which is crucial in addressing the enduring intuition that opponents of illusionism firmly hold. Finally, Section III

explores desiderata for a positive theory of illusionism by analyzing two categories of illusionist theories, drawing connections to related theories of consciousness, such as global workspace theory (Dennett, 2001) and Buddhist philosophy.

I. Motivations for Illusionism

Frankish (2016) outlines motivations for illusionism by sketching arguments against radical and conservative realism while providing positive arguments to demonstrate the appeal of illusionism. It is important to first highlight similarities between illusionism and these views. Illusionism draws from the radical realist's emphasis on the anomalousness of phenomenal consciousness, sharing the intuition that reducing phenomenal qualities to purely physical terms fails to capture their metaphysical richness. Illusionism is also sympathetic to conservative realist's rejection of radical theoretical innovation to accommodate the existence of phenomenal properties. It seeks to reconcile these views by positing that phenomenal properties are illusory: while one can be introspectively aware of one's sensory states, this awareness simultaneously misrepresents these sensory states as having phenomenal properties, despite not actually having said properties. Frankish introduces the notion of *quasi-phenomenal* properties, intermediate representations of sensory states that are physical (or non-phenomenal) but typically misrepresented as phenomenal[1].

The argument against radical realism from the illusionist's standpoint aligns with the conservative realist's, since illusionism leans towards a conservative treatment of phenomenal properties[2]. The primary argument hence explores the pressure towards illusionism from conservative realism. Besides invoking canonical arguments against physicalism[3], arguments for illusionism over conservative realism additionally hinge on the instability of the conservative realist's position. Many conservative realists (or weak illusionists) argue that phenomenal properties do not possess the features ascribed to them, i.e., being ineffable, private, and infallible, while maintaining phenomenal properties are real. Strong illusionists believe all introspectable properties of experience are quasi-phenomenal properties, while weak illusionists maintain that phenomenal properties exist. However, the weak illusionist must conceptualize phenomenality in a way that is stronger than quasi-phenomenality, while preserving a conservative treatment of phenomenal properties (i.e., without introducing non-physical characteristics and collapsing into dualism). This is generally considered a difficult position to maintain, and Frankish (2016) argues weak illusionism ultimately collapses into strong illusionism. Similarly, Dennett (1988) posits that there is no consistent definition of qualia, and one's effort is better directed towards explaining the capacities accompanying consciousness, instead of trying to wrangle with this confused concept.

To outline a positive argument for illusionism, consider the Nagelian notion that there is something it is like to be oneself, just as there is something it is like to be a bat (Nagel, 1974). When asked to articulate, for instance, what a red apple is like, one would typically list its perceptible properties, that it is red, sweet, round etc., and similarly with middle C or the texture

[1] The quasi-phenomenal property of pain, for instance, is the physical property that typically triggers introspective representations of phenomenal pain.

[2] in that radical theoretical moves should avoided if possible.

[3] Namely, that phenomenal concepts have an especially intimate connection to their referents and lack *a priori* connections to physical concepts, and conservative realists must face up to the challenge of explaining how these felt qualities can be physical (otherwise known as the explanatory gap).

of leather. When describing what an object is like, one characterizes the objective, not the subjective, experience. For what it is like to be a bat, bats use sonar sensing, a modality humans lack, to perceive the world. If bats could articulate e.g., what a moth is like, they would characterize moths differently from humans because they perceive different objective properties using sonar systems than humans do with visual systems. However, bats still convey, just using observations from a different modality, what a moth is like, not what it is like to perceive a moth via sonar sensing, just as humans articulate what a moth is like, not what it is like to perceive a moth with one's visual system. Nagel converts the question of what the world is like for a bat to the question of what it is like to be a bat, confusing the object with the subject, which Wittgenstein (1953, §308) terms the conjuring trick. It is intuitive to think of experience as constituting an inner domain populated by immediate objects of experience, distinct from an outer domain of objects one is acquainted with through perception. However, this intuition breaks down when trying to describe the 'apple'-iness of an apple (or 'middle-C'-ness of middle C, or 'leather'-iness of leather) purely subjectively, without appealing to any objective, perceptible properties. One's predicates for characterizing experiences are limited to the objects of one's experience, not the subject. Therefore, when asked to describe one's own consciousness, apart from any object of consciousness, this endeavor is impossible; once the object is subtracted, there is nothing left to characterize.

This naturally leads to the zombie problem (Chalmers, 1996): if there is nothing it is like to be oneself, and nothing it is like for one to experience anything, does that simply make one a phenomenological zombie? Realists about phenomenal qualities often invoke the zombie argument to suggest that humans are fundamentally different, as zombies have no inner life and "all is dark inside" (Chalmers, 1996), and this does not seem to be the case of one's own inner life. However, Chalmers defines zombies as being functionally identical to human beings, which implies that zombies have the same beliefs as "phenomenally conscious" humans, including beliefs about their supposed phenomenal consciousness. However, a zombie's belief that it has phenomenal consciousness is false. Since the zombie counterparts are functionally identical to humans, there are two possible cases. First, zombies possess inner lives like phenomenally conscious humans, which is contradicts Chalmers' definition of zombies thus making the concept of zombies incoherent, thereby failing to support realism about phenomenal qualities. Alternatively, humans, like zombies, falsely believe they are phenomenally conscious and misrepresent their experiences as such, in which case humans are zombies, which is the illusionist theory.

The final motivation for illusionism draws upon analyses of higher-order theories of consciousness, which contend that metacognition of representational states is necessary for consciousness. One of Frankish's (2016) positive arguments for illusionism is the apparent anomalousness of phenomenal properties: "if there is even a remote possibility that we are mistaken about the existence of phenomenal consciousness, then there is a strong abductive inference to the conclusion that we are in fact mistaken about it … [f]or our awareness of phenomenal properties would have to be mediated in some way". Assuming the mind is a representational system, phenomenal properties must be represented to be useful in the mental economy[4]. Frankish claims that realists identify phenomenal character with "some functional property of experience such as possession of a certain kind of representational content, or availability to higher-order representation". However, subjects possessing experiences with such functional properties are disposed to judge that their experiences have qualitative character,

[4] i.e., to have any effect on cognitive or affective states like beliefs, memory, and emotional responses.

conflating phenomenality with the representation of phenomenality. Frankish invokes higher-order perception theory to demonstrate the confusion between perceptual awareness of physical vehicles of experience and their illusory intrinsic qualities, which points to quasi-phenomenal rather than phenomenal properties of experience.

Challenges to Rosenthal's (1997) higher-order thought (HOT) theory can also be construed as supporting illusionism. In HOT theory, "[t]he occurrence of a higher-order thought (HOT) makes us conscious of the mental state; so the state we are conscious of is a conscious state" (Rosenthal, 1997). Introspecting on one's sensory (or first-order) states forms HOTs of these states, making the corresponding first-order states conscious. In response to HOT theory, Byrne (1997) poses two main challenges: the inaccurate and targetless higher-order representation problems. He provides examples where HOTs are inaccurate (e.g., the watercolor illusion[5] (Pinna & Reeves, 2006), the Müller-Lyer illusion) or seemingly generated from non-existent sensory input (e.g., blindsight, phantom limb pain). Therefore, one can easily be wrong about one's sensory states, as HOTs could inaccurately report the contents or falsely suggest the presence of sensory states. One would fail to tell the difference since conscious thoughts are mediated by HOTs, so misrepresentation at the higher-order level would result in one being conscious of whatever is dictated by HOTs. Importantly, such misrepresentation is beyond one's control, and is a byproduct of metacognitive processes that make sensory states usable in the cognitive economy.

Byrne's counterarguments may be interpreted to suggest phenomenality arises not at the level of first-order thoughts, but of HOTs, making one's conscious experiences entirely dependent on the contents of HOTs. However, it is hard to make sense of this – what purpose would first-order thoughts serve if their contents are irrelevant to the eventual phenomenal experience? This motivates another interpretation that HOTs misrepresent our sensory states as having phenomenal character. This is exactly the illusionist's argument, which pushes for a metacognitive theory of consciousness eliminating the supposition that phenomenal properties are involved. To summarize: to utilize one's sensory states in the cognitive economy, they need to be represented, and in so doing, higher-order representations misrepresent those sensory states as having phenomenal properties. Therefore, one is systematically misled into thinking one's sensory states are phenomenal when they are not.

II. The Missing Piece in Frankish's Argument

In arguing for illusionism, Frankish clarifies that illusions are not causally inert. Conservative realism endows phenomenal properties with causal roles, which makes it especially attractive over radical realism or illusionism. Frankish claims illusionism does not deny that phenomenal concepts track causally effective properties, but simply denies these properties are qualitative in nature. This argument is crucial; the pain one experiences from accidentally touching a hot stove intuitively causes one to be more careful around hot stoves in the future, and defenders of illusionism must address the fact that supposedly illusory phenomenal elements of experiences have nontrivial effects on behavior. However, this leads Frankish to consider phenomenal properties, with their causal powers, as *intentional* objects, rather than non-physical characteristics lacking causal roles (as radical realists do) or physical objects that deflate their metaphysical

[5] The illusion where one incorrectly perceives the bright chromatic borders of closed shapes as bleeding into the interior of the shapes, which is verifiably proven to be entirely white.

richness (as conservative realists do). Such intentional objects "move us in the same way that ideas, stories, theories, and memes do, by figuring out the objects of our intentional states", serving as a "mental fiction" that is causally powerful but an "unearthly", "magical non-physical inner life" that is ultimately illusory (Frankish, 2016). He uses this to motivate evolutionary functions of consciousness: citing Humphrey (2011), this "internal magic show" endows agents with "a new interest in their existence, inducing them to engage more deeply with their environment (onto which they project phenomenal properties) and creating a sense of self" (Frankish, 2016).

Frankish asserts that illusionism permits us to acknowledge both the wonder of phenomenal consciousness and its potency. However, it is unclear whether the above evolutionary argument achieves that. Is the wonder of a "magic show" really necessary to keep agents with biological needs interested in their own survival? It seems unlikely that higher-order thoughts or metacognition, which generates the illusion of phenomenal qualities, is necessary for an agent to be invested in its survival; animals display this interest but do not seem to generate higher-order representations through metacognitive processes. Besides the causal power of this illusory mental fiction not being entirely clear, the need to explain the potency of the illusion is essential to a strong defense of illusionism. In the words of Frankish (2016), why are subjects with experiences possessing functional properties disposed to judge that their experiences have qualitative character, and why does this disposition have the strength that it does?

Beyond explaining how the illusion of phenomenal qualities arises (in a manner seemingly beyond one's control), a robust theory of illusionism must adequately explain the difficulty of freeing oneself from the grip of the enduring intuition that phenomenal qualities are real. This missing piece is crucial to defending illusionism against opponents who assert that phenomenal qualities must exist and cannot shake the intuition that their experiences are non-phenomenal. Note that emphasizing this missing piece is not to highlight a fundamental flaw in the argument for illusionism; rather, it suggests what may be lacking current arguments that is required to advance the case for illusionism, especially against opponents who write it off as impossible or absurd. If found, such an explanation will fortify the argument for illusionism, since it makes a bold claim in denying the existence of experiences deemed intrinsic and fundamental, and if proven theoretically impossible or unsound, would serve to advance alternative views like radical realism.

While it is tempting to explore the appeal of illusionism based on the causal power of intentional objects accounting for evolutionary functions of consciousness, this approach faces challenges in explaining the *value* of such illusions. Kammerer (2019) refers to addressing the link between phenomenality and intrinsic value as the normative challenge for illusionism. Defenders arguing that the illusion of phenomenal consciousness is evolutionarily advantageous must explain why illusions have intrinsic value. Concretely, it must explain why one thinks e.g., being in pain is bad *in virtue of its phenomenal feel*, and why pain sensations must be accompanied by this illusory feeling of awfulness, in addition to all the functional and physical events that pain causes. The illusionist's response to the normative challenge either involves commitments to revisionary normative consequences or explaining why the link between phenomenality and value is false, both of which are difficult. Ultimately, such explanations must argue that introspective illusions are useful and thus selected for, which is a major challenge. Section III hence explores other theories of illusionism to determine desiderata for a robust positive theory of illusionism that both explains how the illusion is created as well as the difficulty of freeing oneself from the grip of it.

III. Desiderata for a Positive Theory of Illusionism

This section explores two categories of approaches to the illusion problem. The first category of theories contends that the metacognitive processes giving rise to the illusion of phenomenal experiences are hard-wired into psychological processes. One has no control over the cognitive introspective mechanisms causing the illusion, which explains its strength. Pereboom's (2011) qualitative inaccuracy hypothesis posits phenomenal properties are indeed instantiated, but that one's introspective mechanisms systematically misrepresent phenomenal states as having a qualitative nature that they lack. This corresponds with Frankish's intuitions, where the phenomenal properties that Pereboom claims are instantiated are equivalent to Frankish's quasi-phenomenal properties. "Genuinely qualitative" phenomenal properties are never instantiated, but persistently seem to be, as one's introspective mechanisms constantly represent oneself as possessing phenomenal states. This supports Frankish's (2016) functionalist view that subjects who represent their mental states are strongly disposed to judge their mental states as having phenomenal properties, thus it systematically seems to the subject that they are phenomenally conscious, even though they are not.

Another theory in this category is Graziano's (2013) attention schema theory, which contends that the brain forms schematic representations of its own attentional processes. Attention schema are simplified representations of attentional processes, abstracting away their complexities and merely representing a simple relation of "awareness" between a subject and a piece of information. Such representations, however, are inaccurate depictions of one's attentional processes and there is no "awareness" relation in the brain, thus phenomenal states are a mistaken construct. This view is related to Dennett's (1991) user illusion example comparing qualitative phenomenal states with a computer's user interface (with icons for files, folders, waste basket, and so on). This fiction is created for the user's benefit to simplify controlling the computer, as it abstracts away the complexities of the computer's programming and hardware. Similarly, representations of phenomenal properties are simplified, schematic representations of underlying brain processes, which simplify introspective processing, but are by no means real. Just as folders and waste baskets do not exist inside computers, phenomenal properties do not exist inside brains.

Related to Graziano (2013), Humphrey's earlier illusionist theories[6] contend that conscious experiences reflect internalized expressive responses to stimuli, which interact with incoming sensory signals to generate complex feedback loops. Internal monitoring of said feedback loops results in the appearance of qualitative properties, creating the illusion of a magical inner world (Humphrey, 2011). Humphrey's proposal is reminiscent of ideas conveyed in Dennett's (2001) global workspace theory that sufficiently rich feedback loops between cognitive modules determine conscious states. While Dennett proposed this theory in the context of psychological and sensory states, it is interesting to consider the prospect that such feedback loops give rise to the impression of phenomenality accompanying said states. This proposal aligns with the intuition that illusory phenomenal properties are not usually apparent unless the subject is actively engaging

[6] Humphrey has since rejected the label "illusionist" and characterizes his view as a realist or "surrealist" one. This is partly because he thinks the claim that consciousness is an illusion invites ridicule, but mainly because his belief that a subject's evaluative responses to stimuli are intentional objects that are very much real for the subject. Frankish would presumably say that Humphrey is fundamentally still an illusionist, as Frankish himself agrees with the view that phenomenal states are intentional objects but questions their supposed qualitative nature.

in metacognitive introspection, or explicitly paying attention to the "feels" of an experience. One does not typically experience the "redness" of red when responding to a stop sign, unless one intentionally focuses on the "redness" of the sign. Additionally, this approach supports the notion that the illusion of phenomenality is not developed for the illusion itself[7], but rather as a byproduct of metacognitive processes.

Another class of theories asserts that subjects engage in mistaken inferential mechanisms of projection, and one's belief in phenomenal properties arises from mistaking properties of external objects for properties of the sensory systems by which they are perceived. These theories invoke similar arguments to Wittgenstein's (1953, §308) conjuring trick from Section I. A particularly interesting theory in this category is Garfield (2016), who outlines the view of Buddhist philosopher Vasubandhu, that "there is a naïve bifurcation of experience into a subjective and objective aspect" by phenomenal realists. Imagined phenomenal properties arise from the projection of subject-object duality[8], and this duality is illusory. In reality, a subject causally interacts with the world through sensory systems whose outputs they respond to conceptually, confusing conceptual responses with immediate awareness. Phenomenal realists argue that there is no appearance-reality gap when it comes to experience: one's inner life is populated by immediate objects of experience with phenomenal properties, distinct from the domain of external objects. Vasubandhu questions the implicit commitment to a subjective-objective duality, arguing it is "an illusory superimposition on a reality that has no such structure" (Garfield, 2016).

Rey (1995) presents another interpretation of projection, noting the strong conviction that other agents possess phenomenal consciousness, which suggests the concept of phenomenal consciousness is sensitive to behavioral factors as well as introspective ones. He offers a Wittgensteinian explanation, which asserts that talk of (phenomenal) "consciousness" (like that of "the sky") has a role within a community's linguistic practice (or "language game"). The concept plays a useful role, reflecting common needs, interests, and moral concerns, but does not pick out a well-defined natural phenomenon explainable by science. If phenomenality is culturally inculcated, the way it is discussed in linguistic practice may further cement the belief that a subject and the agents it interacts with have qualitative experiences. This accords with Frankish's (2016) suggestion that phenomenal concepts are hybrid ones, a product of both individual theorizing and cultural acquisition. That said, and Rey (1995) himself acknowledges, certain aspects of experience (like color experience) are resistant to illusionist explanations of this form.

Overall, theories contending the illusion is built into introspective mechanisms provide strong support for the difficulty of freeing oneself from the grip of the intuition that phenomenality is real. In addition, theories asserting that the illusion results from mistaken projection of subject-object duality or cultural inculcation address the realist's counterargument that there is no appearance-reality gap in experience. This suggests two desiderata for a positive theory of illusion: the illusion of phenomenality is not only hard-wired into one's representational mechanisms, but also in social perception, effectively addressing the difficulty of ridding oneself of the illusion.

---

[7] i.e., there is no need to suggest that the illusion is evolutionarily advantageous, which was a challenging line of argument as demonstrated in Section II.

[8] the notion of distinguishing between an external world of objects and an inner world of experience

In conclusion, this paper contends that beyond explaining how the illusion of phenomenal qualities arises, a robust theory of illusionism must adequately explain the incredible strength of the illusion and the difficulty of freeing oneself from the grip of the enduring intuition that phenomenal qualities exist. Explaining the potency of the illusion is the crucial missing piece in a sound illusionist theory, and this additional dimension to the illusion problem is an important reason why illusionism is so hard for ardent realists to stomach. Frankish's (2016) argument for the evolutionary function of the illusory "internal magic show" faces fundamental challenges, and analysis of other theories suggests the illusion of phenomenality is not only hard-wired into a subject's introspective representational mechanisms, but also in social perception, which explains the potency of the illusion.

Bibliography

Balog, K. (2016). Illusionism's Discontent. *Journal of Consciousness Studies*, *23*(11-12), 40-51.

Block, N. (1995) On a confusion about a function of consciousness. *Behav. Brain Sci.* 18, 227–247

Byrne, A. (1997). Some like it HOT: Consciousness and higher-order thoughts. *Philosophical Studies* 86 (2):103-29.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory.* Oxford University Press.

Dennett, D. C. (1988). *Quining qualia*. In Anthony J. Marcel & E. Bisiach (eds.), Consciousness in Contemporary Science. Oxford University Press.

Dennett, D. C. (1991). *Consciousness Explained*. Penguin Books.

Dennett D. C. (2001). Are we explaining consciousness yet? *Cognition*, *79*(1-2), 221–237.

Dretske, F. (2007). What Change Blindness Teaches about Consciousness. *Philosophical Perspectives*, *21*, 215–230.

Frankish, K. (2016). Illusionism as a Theory of Consciousness. *Journal of Consciousness Studies* 23 (11-12):11-39.

Frankish, Keith (2016). Not Disllusioned: Reply to Commentators. *Journal of Consciousness Studies* 23 (11-12):256-289.

Garfield, J. L. (2016). Illusionism and Givenness. *Journal of Consciousness Studies* 23 (11-12):73-82.

Rey, G. (1995). Toward a projectivist account of conscious experience. In *Thomas Metzinger (ed.), Conscious Experience. Ferdinand Schoningh*. pp. 123--42.

Goff, P. (2016). Is Realism about Consciousness Compatible with a Scientifically Respectable Worldview? *Journal of Consciousness Studies*, *23*(11-12), 83-97.

Humphrey, N. (2011) *Soul Dust: The Magic of Consciousness*, Princeton, NJ: Princeton University Press.

Kammerer, F. (2019). The Normative Challenge for Illusionist Views of Consciousness. *Ergo: An Open Access Journal of Philosophy* 6.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review* 83 (October):435-50.

Nida-Rümelin, M. (2007). Grasping Phenomenal Properties. In T. Alter & S. Walter (Ed.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.

Pereboom, D. (2011). *Consciousness and the Prospects of Physicalism*, Oxford University Press.

Pinna, B., & Reeves, A. (2006). Lighting, Backlighting and Watercolor Illusions and the Laws of Figurality. *Spatial Vision*, 19, 341-373.

Rosenthal, D. M. (1997). A Theory of Consciousness. In Ned Block, Owen J. Flanagan & Guven Guzeldere (eds.), *The Nature of Consciousness*. MIT Press.

Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford: Blackwell.