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.

DISCARDED

A potential challenge for HOT theory involves the prevailing intuition that visual phenomenology is incredibly rich, and to use Block's (1995) terminology, overflows access in that its richness far exceeds what humans are reasonably capable of attending to. However, there is strong evidence that one's visual phenomenology represents much fewer things than one believes, for instance in the case of change blindness (Dretske, 2007), or the mere fact that despite having an optical blind spot, one typically does not experience a hole in one's vision, suggesting that the generation of higher-order thoughts "fills in" any gaps or inconsistencies in sensory perception. This leads one to question whether higher-order representations modify first-order sensory percepts in every case of phenomenology, and if the process of their formation constructs phenomenal experience without phenomenal experience being there at all.

Introspective illusions are especially hard to explain compared to perceptual illusions. With perceptual illusions, it is possible to provably show that the illusion exists. For instance, recalling the watercolor illusion (Pinna & Reeves, 2006), an easy test to prove that the interior of the shape is totally white is to use a digital RGB detector, hence the effect of the chromatic border bleeding into the interior must be illusory. Similarly with the Müller-Lyer illusion, where one perceptually represents two equal lines as being of different lengths, and the fact that the lines are of equal length can be easily verified via objective measurement using a ruler. Therefore, it is much easier to prove that perceptual illusions are genuinely illusions since there are ways of objectively verifying reality and demonstrating the illusion exists. However, because the supposed illusion of phenomenal experience is inherently private and first-personal, options for verification is limited and limited to the subject themselves.

Final Paper – Illusion and Representational Theories of Consciousness

Broad motivation of the paper: Frankish gives a (somewhat) cursory treatment of the arguments against illusionism, in particular, those who say "what do you mean?! Of course we have qualia!". His arguments for how and why this illusion arises are also generally not very satisfying. I hope to analyze counterarguments to illusionism more carefully, and potential ways of approaching the illusion problem.

<INTRODUCTION>
→ Introduce concept of illusionism (vs. realism, both conservative and radical), replacing the hard problem of consciousness with the illusion problem

- Battle so far has been almost exclusively between the two realist camps, but maybe we must explain why we believe what we believe/judge what we judge about phenomenal properties. If we do that in a way that doesn't posit phenomenal properties at all, what else do we need to explain?

→ Motivation illusionism (1): "Converting" conservative realists to illusionists

- This is an unstable position, threatening to collapse into either dualism or illusionism
- Conservative realists already deny many of our intuitions about phenomenal properties (that they are: ineffable, intrinsic, private, infallibly known)
    - Argue that intuitively you have special states with these special phenomenal properties, yet denies these phenomenal properties, need to explain who they could nonetheless exist and be phenomenal
- If deny all the metaproperties of qualia / metaphysical presuppositions about what makes qualia special, why should this leave us with genuine qualia instead of no qualia at all? What makes these properties genuinely phenomenal rather than quasi-phenomenal?
    - Need to do this in a way that doesn't collapse into either dualism (by making them metaphysically mysterious) or illusionism (by denying that they really exist), which is a hard line to hold
- Should give up explaining qualia because it doesn't exist. Dennett asserts that concept of qualia is confused/inconsistent, nothing could answer to that concept. Should give up on the idea of explaining what it is and instead explain all the capacities/things we think go along with consciousness
    - Dennett: more terminological/semantic than anything
        - Reduction vs. elimination – how committed are you to talking that way before, how well does new theory match up with old theory
    - You have first order sensory states, you're misrepresenting those as endowed with phenomenal qualities

→ Motivation illusionism (2): Connections to HOT Theory

- We need representations of phenomenal properties anyway, so why not just use them for explanation? Any kind of judgement about phenomenal properties goes via representations of them, might as well just use representations

    - Higher order thought picture – when we start thinking about the hard problem / introspecting our sensory states, form representations of those states
    - Why do we need to posit that it comes from some genuine phenomenal property in your brain?
- Example of visual phenomenology
    - Seems extremely rich and stable, even though we can only attend to a small part of it and report even less than that → visual phenomenology overflows access (in Block's terminology) (much richer than our attention)
    - However, there is plenty of evidence that our visual phenomenology is much poorer than we intuitively believe (e.g., change blindness)
        - Represent much fewer things in our visual phenomenology than we believe we represent
    - Our higher-order thoughts about visual states fill in the blanks
        - We have a blind spot, but we don't experience a hole in our vision
    - Why not think that our higher-order thoughts do this in every instance of phenomenology?
        - Higher order states can change your first order states so much, what if they CONSTRUCT your phenomenal experience without any phenomenal experience being there at all?
    - Dennett Change Blindness
    - Byrne Counterargument
        - You could be wrong about what your sensory states are. If you're forming higher-order representations, those could misfire and tell you incorrect things about your sensory state, or even say you have a sensory state when you don't. What would that be like? You presumably wouldn't be able to tell any difference because all your conscious thoughts are mediated through higher-order thoughts, misrepresenting higher-order thought would be as if you were conscious of whatever it says you are conscious of. But now the fact that they're higher order (directed down to first order sensory state) doesn't matter, your conscious experience is just whatever your HOT says it is
        - One direction is to say that's the phenomenality then – not about the FOT, phenomenality arises at HOT level
            - Hard to see how this makes sense – what's the purpose of the FOT then?
        - Another direction is to say that maybe this should push us towards illusionism, thinking that we have HOTs misrepresent out sensory states as phenomenal/having phenomenal character
            - Pushing a higher order theory of consciousness that does away with the supposition that there are any phenomenal properties involved
            - *But why?? Why is it disposed to tell us that we are experiencing them*

        - *Frankish: Creating magic show for you, brain makes you more invested in your future survival – seems unlikely (animals are very invested in survivals without HOTs), surviving the death of your body → Why does this illusion happen? When?*
- TLDR: To use sensory states in cognitive economy, we need to represent them, and that's where we go wrong, when we represent those sensory states as having all these weird properties even though they don't really have those properties, mislead ourselves into thinking our sensory states are phenomenal

<PROBLEMS FOR ILLUSIONISM>

→ Main problem: Need to explain why / how the illusion is created, and **why it is so strong**. Why is it so hard to accept/stomach?

Not that I am objecting illusionism, I think elements of it are quite compelling, but to really have a knockdown objection to critics of illusionism, illusionists need to deal with this head on.

- Kammerer's illusion meta-problem
- (a) *With fiction and perceptual illusions, we can provably tell ourselves that it is an illusion (e.g. using RGB detector on the white patch to tell us it's white even though there has a watercolor bleeding effect), but how does that work with consciousness?*
    - The notion that because you can't verify the illusory nature of phenomenal experience (because it is private)
    - Kammerer has an objection to this – it is not the failure in the ability of to verify the illusion that it is strong (detector example)
        - Talks about good reasons to believe it is faulty – do we have good reasons to believe we are fallible about our own experiences? Wasn't he pushing that phenomenal experience and being aware of phenomenal experience are identical (self-intimating)
        - This doesn't make sense – the detector is yourself. How can we design a way to test if our belief about our undergoing of phenomenal experiences is faulty? Perception doesn't map to experience
        - Look into introspective judgement vs. awareness – razor / icicle example
    - You can only experience the phenomenal from the perspective of introspection (privacy)
        - Dennett: Good to represent your representations – opens up possibility of imagining, latent planning etc. but could also misrepresent. Metacognition is good + something else that gives rise to this illusion
    - Deny the phenomenal properties of qualia while keeping them around
      Conscious states are the states that we mistakenly judge have phenomenal properties. You have conscious states, it's just that they don't have these special phenomenal properties
- (b) Babies, animals: haven't developed those illusory properties, when does this develop? If someone doesn't have higher order thoughts, they don't have this illusion, just go about their day. More "enlightened" than we are, not in the grip of an illusion

    - *What's the difference between us and zombies? Do they lack these metacognitive abilities that give rise to this illusion? Or does this theory make the notion of zombies incoherent?*
- See Kammerer paper

→ Generally hard / struggle to give an evolutionary account of why we (think we) have phenomenal experiences

- If you're a conservative realist trying to explain why we are such that we have conscious experiences, it's a hard problem. Hard to say why we're constituted so as to believe what we believe about the phenomenal. A lot of conservative realists will say it's just a byproduct of some things we're actually selected for. This might seem unsatisfying
- Frankish says if this is an illusion, opens up other kinds of evolutionary explanations: by granting you with this illusion that you're the center of your phenomenal world, maybe evolution gives you doubling down on your drive to protect yourself (not super clear)
    - Derived from Humphrey
- Frankish wants to deny the existence of phenomenal properties while construing conscious states functionally
    - For everyday life, not necessarily. Illusions can be very important, illusions in a sense are real because we experience them. Story you are telling yourself about sensory states (fiction) can be very important, don't necessarily need to give this up but need to recognize that it is fiction *(how?)*
    - **Can phenomenal illusions be useful in any way? We could come up with an evolutionary explanation from there.**
        - Frankish accepts you can introspect and learn things about your representations, mind, etc.
        - Red experience in the fiction of your illusion, get the intention out of that fiction?
    - Perceptual illusions can be useful
        - McGurk effect: ambiguous phoneme, auditory illusion (resolving ambiguity) (double flash is the opposite) – how useful are perceptual illusions for this argument?
        - But this seems to be targeting a different problem: Frankish is saying the "ba-ness" of the ba sound or the flashiness of the flash sound is illusory. This takes illusion in a different sense?
        - Perhaps change blindness is a better example here
    - Introspective illusion – grand illusion of visual perception (think more about this)
        - A lot of literature on the introspective illusion of psychological states (not being acquainted with our own minds) – and I have written about this too!
        - But what about an introspective illusion of phenomenal states? General consensus is that you can't be wrong that you are undergoing a phenomenal experience E (despite being able to be wrong about perceptual objects). It doesn't seem to be very advantageous to be wrong about them (e.g. being wrong about pain percept)

- Epiphenomenal dualism: consistent with physical accounts of dualism, mental properties distinct from physical properties but causally inert, and they exist which is nice
    - Objection: if mental phenomenal are causally inert, no reason / strange coincidence that consciousness evolved. Also intuitively mental phenomena do cause/influence physical events
        - This might collapse into illusionism?
    - Felicitous alignment that bad feelings accompany things that threaten our survival and good things accompany things that are good

Very similar problem as illusionism – need to explain why this coincidence (rather than straight up mystery) arises

<**Some attempts to solve it**>

→ Theory of Introspection / TCE theory (Kammerer 2021)

- Tbh I'm not entirely sure what this theory is talking about
- Claims it can overcome the illusion meta problem

→ Cognitive mechanism that we don't have any choice over that causes this illusion in us. Hardwire into our cognitive psychology.

- **Maybe it's culturally inculcated?**
    - What does this mean? That we are zombies (or pseudo-zombies) gaslighting each other into believing we each / all have qualitative experiences?
- Both Pereboom and Graziano say that our introspective mechanisms, because of some of their hard-wired features, represent us as being in mental states endowed with some special properties – properties that they don't really have. These properties happen to be what we call "phenomenal properties". So, for this reason, it systematically seems to us that we are phenomenally conscious, even though we are not.

→ Frankish / Pereboom – qualitative inaccuracy hypothesis

- According to this hypothesis, our introspective mechanisms systematically misrepresent phenomenal states. They represent them as having phenomenal properties, gifted with a qualitative nature that they lack in reality. Phenomenal properties really are instantiated, but they are misrepresented by introspection as having a qualitative nature that they don't have. However, it is possible to slightly reinterpret this hypothesis and to say that these really instantiated phenomenal properties (devoid of the qualitative nature that introspection presents them as having) really are simply "quasi- phenomenal properties" (Frankish says they are not instantiated at all and are just quasi-phenomenal properties instantiated)
- In the past, Humphrey has proposed an explicitly illusionist theory, according to which conscious experiences reflect internalized expressive responses to stimuli, which interact with incoming sensory signals to generate complex feedback loops. When these loops are internally monitored, Humphrey argued, they appear to possess strange qualitative and temporal properties, creating the illusion of a magical inner world
    - Very global workspace-y – byproduct idea comes in here

→ Graziano Attention Schema

- Attention schema is a simplified version of our experiences (distilled/condensed) – fallacious simple (oversimplified) relation
    - You represent things about how your brain works in your brain, but your brain can't represent completely how it works (then it wouldn't have resources to do anything else), just have to have representational bedrock (don't know/care where they come from). Maybe for sensory states, you just don't represent them as coming from any physical process in your brain + they are information rich + fix to belief without cognitive mediation
- Something with all those properties may be sufficient for producing illusion – give you ineffable (information rich) thing that seems to float free without any physical substrate, seems to be the primary thing you're acquainted with
- If you have that, maybe that's all that's required for this impression of phenomenality
- Didn't develop the illusion for the illusion itself, it's not advantageous, it's just a byproduct
    - **Byproduct vs. emergence?**
    - Byproduct – representing the sunset, that's where the felt qualities arise
    - There's something here to be explained even before you get caught up in the philosophy of it. First order representationalist

→ Inspirations from Buddhist philosophy

- The Buddhist philosopher Vasubandhu (4th-5th century CE) put the same point in terms of his analysis of experience in terms of the doctrine of the three natures that constitutes the heart of Yogācāra phenomenology. Vasubandhu, in his Treatise on the Three Natures (Garfield 2009, 2015) argues that every object of experience has three distinct, but interdependent natures (svabhāva): each has an imagined (parikalpita) nature—that we take it to have in virtue of a combination of our prejudices and innate cognitive reflexes; each has a dependent (paratantra) nature, representing its underlying causal structure; and each has a consummate (parinispanna) nature—the nature we see that it has when we empty the dependent of the imagined
- The consummate nature of things, according to Vasubandhu, is, in his phrase, the fact that the dependent nature is empty of the imagined. That is, it is the fact that the set of causal processes in which perception consists is entirely empty of a division into subject and object or into inner and outer. There is just a causal stream constituting the causal interaction of an organism and the world. That stream is experienced as (imagined to be) the apprehension by a subject of a world, delivered exactly as it is into an inner space. But that, Vasubandhu argues, is an illusion, the illusion that the conceptual response to sensory experience represented by the imagined nature is a transparent delivery of experience as it is.
- The idea that patches of blue, or sounds of oboes exist independently as they are experienced by us, and then are reproduced inside us in experience is a fantasy that sounds crazy the moment we make it explicit. So, Vasubandhu, concludes, both the idea that the objects we experience exist external to us, and the idea that we have appearances internal to us, are each products of imagination. Instead, we simply causally interact with a world around us, through sensory systems to whose outputs we

respond conceptually, confusing that conceptual response with immediate awareness. That is what he means when he states that the imagined nature is the imagination of subject-object duality.

Frankish paper – I like the zombie paragraph
https://keithfrankish.github.io/articles/Frankish_Illusionism%20as%20a%20theory%20of%20consciousness_eprint.pdf

https://keithfrankish.github.io/articles/Frankish_Not%20disillusioned_eprint.pdf

Kammerer
https://philarchive.org/archive/KAMCYB
https://philarchive.org/archive/KAMTIO-4

https://jaygarfield.files.wordpress.com/2014/01/illusionism-and-givenness2.pdf
Talks about illusionism and some Buddhist views (could be interesting).

Dan Dennett (2017). ch 14 (Consciousness as an evolved user-illusion)

Dan Dennett (1992) *Consciousness Explained*. Back Bay Books. Obviously can't read all of this! But do have a go at some of it, perhaps especially something of chs 2, 5, 10, 11, 12, 14. Don't let this get in the way of the reading below though.

Roger Fellows & Anthony O'Hear (1993) Consciousness avoided, Inquiry, 36:1-2, 73-91, DOI: 10.1080/00201749308602312

John Dupré (2009). Hard and easy questions about consciousnesss. In Glock & Hyman (eds) Wittgenstein and Analytic Philosophy. pp. 228-249.

Dan Hutto (1995) Consciousness demystified: A Wittgensteinian critique of Dennett's project. The Monist, 78 (4), 464-479. [ignore §2 and §3]

Ludwig Wittgenstein (1958).just §§307-8 of his *Philosophical Investigations.* Blackwell.

https://plato.stanford.edu/entries/consciousness-representational/#Illus

https://plato.stanford.edu/entries/qualia/#Illusional

(I think this is talking about veridical perception though, like hallucinations. Not relevant)
https://www.jstor.org/stable/pdf/24489471.pdf?refreqid=excelsior%3A0901704020b778e210d70a2a5633959c&ab_segments=&origin=&initiator=&acceptTC=1