

To appear in Challenges of the Technological Mind: Between Philosophy and Technology,
ed. Paulo Alexandre e Castro, Palgrave
Macmillan

CONSCIOUSNESS, THEORY, AND MENTAL APPEARANCE

David Rosenthal

Philosophy and Cognitive Science

Graduate Center, City University of New York

Email: davidrosenthal1@gmail.com

ORCID: <https://orcid.org/0000-0002-9323-0424>

Word count: 7,370

Abstract

I contrast one-factor views of consciousness with two-factor views. On two-factor views a state's being conscious consists in something distinct from the state itself; one-factor views deny that. Examples of one-factor views are the first-order theory of Fred Dretske, the first-order approach of Thomas Nagel, and Ned Block's conception of phenomenal consciousness. Examples of two-factor views are a higher-order theory of consciousness, such as what I have defended elsewhere, and the global-workspace theories of Bernard Baars and of Stanislas Dehaene and Lionel Naccache.

I argue that one-factor views preclude not only a useful explanation of consciousness but also any informative description of what it is for a mental state to be conscious. Because of that and other related factors, one-factor views are strongly anti-theoretical. I consider several indicators of this anti-theoretical attitude, including advocacy of a hard problem or explanatory gap and the appeal to allegedly pretheoretic intuitions.

Some have found one-factor views appealing because they reject any coherent contrast between the mental appearance and the mental reality of conscious states. But denying that distinction derives from a misunderstanding of our commonsense conception of consciousness. And independently of that it is also theoretically indefensible. Examining the rejection of a distinction between the mental appearance and mental reality of conscious states helps us appreciate the shortcomings of a one-factor view.

I. Two Approaches to Consciousness

The current literature about consciousness often taxonomizes views into first order (FO) and higher order (HO). But a more revealing division, I'll argue, is more general than

that, and relies of distinguishing between what I'll call one-factor (1F) and two-factor (2F) views.

1F views hold, with Thomas Nagel (1974), that the property of a mental state's being conscious is intrinsic to that state. By contrast, higher-order (HO) and global-workspace theories (GWT) are 2F views, which hold that a state's being conscious is a matter of some factor distinct from the state itself.

HO theories all explain what it is for a state to be conscious by appeal to one's being aware of that state. FO theories deny that any such HO awareness (HOA) figures in a state's being conscious. So the views of Nagel, Fred Dretske, and many others are FO: Being conscious is intrinsic to the state.

On GWT, by contrast, a state is conscious if its content is available for global downstream effects. GWT is also a FO view, since it does not rely on any HOA (Baars 1997; Dehaene & Naccache 2001). But both HO and GWTs explain consciousness by appeal to a factor that's distinct from the state, a HOA or global availability for downstream processing. Indeed, it is natural to fold the basic idea of HO theories into a GW theory, as Lionel Naccache arguably does (2018). So both HO and GW theories operate in similar ways, despite their being taxonomized differently on a distinction between FO and HO theories.

This distinction between 1F and 2F theories parallels the distinction sometimes drawn between local and nonlocal theories of consciousness (e.g., Michel & Doerig 2022). On local theories so described, the neural correlate of a perceptual state's being conscious is local to the relevant area of perceptual cortex. On a nonlocal theory, the neural correlate is elsewhere in cortex, typically in prefrontal cortex.

But as I'll argue shortly, it's important to cast theories of consciousness in distinctively psychological terms, since consciousness itself is a distinctively psychological phenomenon. The distinction between local and nonlocal theories is cast not in terms psychological terms, but by appeal to underlying neurological correlates. By contrast, both 1F and 2F theories are cast in distinctively psychological terms. So in what follows I'll stick with that distinction.

Most contemporary discussions of consciousness tend to focus on perceptual consciousness. But just as the distinction between FO and HO theories applies across the board to mental states of every type, as Myrto Mylopoulos (2018) shows for the case of agential awareness, so too the distinction between one-factor and two-factor views applies to what we say about the consciousness of mental states of every mental type..

I'll argue here that 1F views face serious difficulties, difficulties that are built into those views, and so cannot be in any way avoided, adjusted for, or overcome. So any progress we might make in explaining consciousness and developing a serious scientific treatment will have to rely on 2F views.

And within the two types of two-factor views already considered, In addition GWT faces compelling counterexamples: There are many peripheral sensations that are conscious but lack any downstream psychological effects. And in the opposite direction, there are many thoughts and desires that have substantial downstream

effects on psychological processing even though though se thoughts and desires are altogether unconscious. Because of that, HO theories should be our preferred type of 2F view. Still, it will be useful to cast the relevant issues in terms of the distinction between 1F and 2F views.

One especially salient difficulty that 1F views face is how to explain the way conscious states differ from mental states that aren't conscious. If a state's being conscious is intrinsic to that state, it's hard to see how a state of that same mental type could occur without being conscious.

This issue underlies the difficulty that arises with Dretske's (1993) FO claim that a state is conscious if being in it makes one aware of something. But unconscious mental states also make one aware of things, just not consciously aware. The priming effects and forced choice that subliminal perceptions often enable could not be explained without acknowledging that these unconscious perceptions make one unconsciously aware of things.

Dretske (2006) himself came to see this. But Dretske's proposed remedy for the problem illustrates vividly the difficulty in avoiding this shortcoming of 1F views without endorsing some psychological mechanism that is in 2F. And the correction Dretske does propose is plainly 2F; it argues that perceiving is conscious only if one can cite it as a justifying reason for doing something. But one cannot verbally cite something without being aware of it. So this response is not only 2F, but it is in effect HO in requiring a type of HOA.

Nagel explains consciousness by appeal to cognate terms such as 'what it's like', 'subjectivity', 'point of view', and 'perspective'. And on his view we cannot break out of that closed family without losing what is central to consciousness. So we cannot then explain how conscious states differ from unconscious mental states.

And all this points to a deeper difficulty. If being conscious is an intrinsic property of conscious states, confining us to Nagel's small circle of interdefined terms, not only are we unable to explain consciousness in any useful way; we can't even describe in an informative way what it is for a state to be conscious (Rosenthal 1983).

This is reminiscent of the closed curve of terms W. V. Quine (1951) sees as figuring in any attempt to explain what it is for a statement to be analytic. Just as the inability to break out of Quine's closed curve shows that any such explanation of analyticity is uninformative, so too with the inability to break out of the closed curve that Nagel insists must figure in any adequate description of consciousness.

Ned Block's conception of phenomenal consciousness is also a 1F view, denying that it can be explained in terms of anything extrinsic to a phenomenally conscious state itself. And to his credit, Block in effect explicitly acknowledges the difficulty just noted. We can, he concedes, say little if anything about what conscious qualitative character is beyond Louis Armstrong's famous quip about jazz: "If you gotta ask, you ain't never gonna get to know" (1978, p. 281). And more recently: "The best you can do is use words to point to a phenomenon that the reader has to experience from the first person

point of view" (2015, p. 47). Pointing in this context is plainly metaphorical, and so of no relevant help. How does one point to another person's experiences?

This observation generalizes. Those who claim that the mental property of being conscious is intrinsic to one or another type of mental state, or perhaps to mental states of any type, are committed to maintaining that all such mental states are conscious. And this contention precludes any informative way to describe how states of the relevant mental type that are conscious differ from those that are not.

This will encourage many 1F theorists to maintain that if a state is not conscious, it cannot genuinely be a mental state. And if that seems too stark, one might construct special conditions for unconscious mental states, such as perceptions or thoughts, as Block (2017) does for unconscious perception. But the conditions Block offers for an unconscious state to qualify as a perception do not match the conditions that define conscious perceptions. So the states he regards as perceptions in the unconscious case are states of a different mental type from conscious perceptions.

This will likely to so for any attempt to provide conditions for a distinctive type of mental state's occurring unconsciously if one also holds that the property of a state's being conscious is intrinsic to the states in the conscious cases. And if one can't say in any useful or informative way how conscious states differ from states that are not conscious, one also won't be able to say what it is for a state to be conscious in the first place. And then the term, 'conscious', becomes no more than a kind of honorific, a way of saying something nice about mental states.

Given the lack of any informative account of what it is for a state to be conscious that's cast in distinctively psychological terms, Block offers instead a neural implementation, favoring Victor Lamme's (2006) proposal that perceptions are conscious when recurrent processing occurs in sensory cortex. But a neural implementation by itself simply won't do. Consciousness is a psychological phenomenon. So we must explain in psychological terms what it is for a state to be conscious. Neural input may come to influence and adjust our initial explanation, but we must start with something psychological.

Indeed, a psychological account will typically guide our search for a neural implementation, for example, whether to look in sensory or in prefrontal or parietal areas of cortex. Because most 1F views say almost nothing about consciousness in psychological terms, they encourage instead offering accounts that are cast solely in neural terms.

And that can distort things. It suggests looking just for a neural on-off switch for conscious states, bypassing any explanation in psychological terms of what it's like for one, that is, an explanation of how exactly conscious and unconscious mental states differ. 2F views invite a more fine-grained neural implementation, which captures not only the difference between conscious and unconscious states but also, crucially, the very many different ways in which states subjectively occur in consciousness. Whatever the correct neural implementation ultimately turns out to be, it cannot be some simple on-off switch.

The focus on finding a neural correlate of consciousness is due largely to the temptation, amplified by 1F theories, of thinking that consciousness is a mysterious matter that resists any informative description, as with the foregoing remarks quoted from Block. Since consciousness is a distinctively psychological phenomenon, one would expect, and indeed hope for, an informative description cast in psychological terms. But if one deems that impossible, there is perhaps nothing informative left to say except what the neural correlates of consciousness are. (For a thoughtful review of neurobiological correlates and their theoretical status, see Mylopoulos 2022.)

The natural and obvious move here is to see whether we can dispel whatever sense there is that consciousness is indeed a mysterious phenomenon that resists informative explanation. And that is indeed that strategy of 2F views, which seek to explain what it is for a state to be conscious by appeal to informative factors extrinsic to the conscious state itself.

Hakwan Lau (2022) has advanced a novel and ambitious 2F view that has deservedly captured widespread attention. A visual state is conscious if there is neural state in prefrontal cortex (PFC) that neurally points to the relevant state in visual cortex. Absent such neural pointing by some state in PFC, the state in visual cortex would occur, but not consciously. Similarly with other types of perceptual state.

There are two difficulties with this view. One is that the pointing state determines only whether the visual state is conscious, not what visual properties it is conscious in respect of. It is the visual state itself, according to Lau, which determines in what way the state is conscious, for example, in respect of what color and shape. But it is arguably theoretically awkward at best to divide things up in this way, so that whether a state is conscious is determined by one neural factor and how it is conscious is determined by another. It would be more natural to explain whether a state is conscious and how it is conscious by appeal to the same considerations.

And there is another worry for this aspect of Lau's account. A perceptual state is not always conscious in respect of exactly the visible input that results from a particular stimulus. Conceptual factors may influence how a perceptual state is conscious, as when a gray banana shape is consciously perceived as yellow. In that case, visual color presumably registers the gray color, but the conscious experience is as of yellow. If the state in visual cortex did determine how the resulting state is conscious, that would not happen. This concern arises for all cases in which subjective awareness does not accurately reflect the mental properties of the FO state, a circumstance that will be discussed independently in section II.

Lau describes this account as a type of HO theory, since the second factor occurs in PFC and operates on the FO visual state. But on a standard HO theory, the HO represents one as being in a FO state of a particular type, and that explains both why the FO is conscious and also why it is conscious in respect of the properties that subjectively appear in consciousness. It's not obvious what theoretical advantage Lau's joint-determination account has, on which those two matters are explained independently.

Lau might urge that his view is simpler and more straightforward. But it's arguable that this is so only from the point of view of the neural underpinnings. And as already argued, a theory of consciousness must stress and rely on those aspects of the situation that are distinctively psychological, since consciousness is itself a distinctively psychological phenomenon.

And there is a second difficulty. On Lau's theory, the HO state in PFC operates simply to point neurally at a FO state in, for example, visual cortex. On a standard HO theory, the HO state would represent the occurrence of that FO state, describing it in psychological terms. We know that states in PFC have rich representational properties, often conceptual properties. But at present we are very much in the dark about how PFC does that. There is no simple neurological map, as there is for the sensory cortical areas, which describes how neural activity at a particular location represents properties of a particular type.

So it is open that the states Lau posits state in PFC as pointing to particular FO states actually represent one as being in the relevant state. It is hard to see how one could settle, empirically or theoretically, whether that is so. And absent some way of doing so, it's not clear that Lau's pointing model provides us with a clear alternative to more standard HO theories.

All told, the foregoing considerations point to 2F views. We must explain consciousness by appeal to phenomena that are psychological, but not themselves conscious, since explaining how one state is conscious by appeal to other states that are themselves conscious would be circular and uninformative.

And that is exactly the strategy of 2F views. GWT appeals to downstream psychological processing, which itself need not be conscious. And HO theories rely on a HOA: A state is conscious if one is aware of it in a suitable way. And again that HOA need not be conscious. The HOA's being conscious would require a yet third-order awareness, which itself is rare.

Indeed, we are seldom subjectively aware of any such HOAs. We know about them, from the reasoning provided by HO theories. We rarely have first-person access to them; instead, they are theoretical posits that do a good explanatory job. On 2F views a state is conscious if that second factor is present, not otherwise. That second factor is what a state's being conscious consists in.

But HO theories and GWT diverge in how credible they are. If one is wholly unaware of a mental state one is in, that state is not conscious. This is basic to our commonsense conception of consciousness, and indeed also central to experimental methodology in consciousness research. And that points to a HO theory, on which a state is conscious only if one is in some suitable way aware of it.

By contrast, it is unclear why one would expect that, as a general matter, a state's being conscious consists in its content being available for downstream processing. Indeed, the counterexamples to GWT noted earlier suggest otherwise. Many conscious states do of course influence downstream psychological processing; they are typically those conscious states that involve some measure of attention. But consciousness and

attention also occur altogether independently of one another (Norman et al 2013). And it is unclear what in the nature of consciousness, independently of attention, would in general result in a conscious state's having an influence on downstream processing.

I'll come back to this issue shortly.

II. Theory and Appearance

A 2F view readily lends itself to scientific treatment, since one can empirically investigate and theorize about each factor independently of the other. For GWT, one can study the potential that various states have for their contents to result in downstream processing, and isolate the considerations, such as signal strength, that affect that potential. For HO theories, one can investigate and theorize about the kind of HOA that figures in a state's being conscious and determine what considerations give rise to that HOA and how.

By contrast, the 1F view that consciousness is an intrinsic property is far less conducive to theory and to empirical investigation. In part this is because there are fewer moving parts. If a property is intrinsic, it's less available for the manipulations that figure in experimentation and theorizing. If one thinks of the mental circumstances in virtue of which the state is conscious as intrinsic to the state, there is only one mental factor to investigate. Indeed, construing a property as intrinsic is akin to seeing it as an essence, and so as something that is simply given, and so resistant to informative explanation.

But there is a deeper reason that 1F views are recalcitrant to theorizing and empirical investigation. If consciousness is intrinsic, it's the last word about the nature of every conscious state. Consciousness then overrides any other information we could possibly have about the mental properties of the state in question. So as already noted, seeing consciousness as intrinsic detaches it from everything we might investigate independently of first-person access, so that there is then nothing objective to say about the nature of consciousness or how it comes to be.

Once we construe being conscious as intrinsic to those mental states that are conscious, what we can know about conscious states in psychological terms is limited, as the foregoing quotations from Block in effect put it, to saying that this is what it's like, where 'this' refers to something accessible only to the subject.

It is this closed-off, consciousness-first picture that by itself gives rise to a sense of mystery about consciousness, which can express itself in the claim that there is an explanatory gap (Levine 2001) or a hard problem (Chalmers 2003). A 2F view, by contrast, opens up rich ties that being conscious has with other mental phenomena, and the availability of such ties dispels those apparent mysteries. More about the hard problem and explanatory gap shortly.

Because on 1F views we know about the nature of conscious states exclusively from what consciousness itself tells us, 1F views are strongly anti-scientific and anti-theoretical. Thus Nagel (1974) insists that if our description of conscious mentality

included anything objective, it would, solely because of including an objective factor, fail to do justice to the nature of consciousness as pure subjectivity.

Other 1F advocates rarely explicitly acknowledge this anti-theoretical aspect of their view. In part that is because they compensate for the lack of any theoretical account in distinctively psychological terms by pursuing direct correlations of conscious occurrences with neurological occurrences, which in effect distracts attention from the lack of a genuinely theoretical account.

But one can see that anti-theoretical attitude inherent in 1F views by the way their advocates simply ignore the theoretical aspect of HO views, and instead measuring all views solely against the subjective appearances. An example of this occurs in connection with the issue about whether a HOA might misrepresent the state it makes one subjectively aware of.

1F advocates note that on a HO theory a HOA might misrepresent a conscious state, and they regard this as decisive against HO theories (e.g. Byrne 1997; Levine 2001). Such misrepresentation does actually occur. For example, in change blindness one is often remains consciously aware of the stimulus or object that changed in the way it was prior to the change (Grimes 1996; Fallon 2020-22). But since the change is often unconsciously perceived (Fernandez-Duque & Thornton 2000; Thornton & Fernandez-Duque 2001; Laloyaux et al 2006), it is plain that visual cortex often registers the changed properties after the change has taken place. So conscious awareness misrepresents the post-change visual state.

Subjective misrepresentation is never evident simply from the subjective appearances themselves, since the subjective appearances are themselves misrepresenting. Though the subjective appearances on which 1F views rest cannot reach beyond themselves, HO theories do reach beyond those appearances. It is that feature of HO theories that 1F views find objectionable. HO theories actually theorize about the subjective appearances themselves.

One can describe such subjective misrepresentation in a way that makes it seem problematic. Suppose a change-blindness subject misses a change of something from green to red, as in the parrot display in Grimes (1996), and continues to see a green object. HO theories could describe that as having a conscious sensation of green. But if the post-change input causes a red state in visual cortex, there is no green sensation to begin with. How can one have a conscious green sensation but no green sensation?

Still, as Daniel Shargel (2016) nicely stresses, describing the situation as being problematic in that way is not only highly misleading, but also altogether optional. The accurate description is that this change-blindness subject has an unconscious red sensation, but it subjectively appears to the subject as though there is a green sensation (see also Berger, 2014). The subjective appearances in such a case simply do not accurately reflect the sensory state of the individual.

And there is nothing obviously problematic about that. Indeed, we can sometimes discriminate in a more fine-grained way than subjective awareness is able to capture,

even though all the relevant perceptual states are conscious. This is evidently what happens in an experiment described by Diana Raffman (2011), in which participants subjectively judge that pairs of color patches are identical in shade even though they match the two samples in a way that accurately reveals the difference between them. In this case subjective awareness simply does not reflect our sensory states with full accuracy, and this doubtless also happens very often in everyday perceptual experience.

1F advocates also note we're rarely aware subjectively of any HOAs, ignoring that HOAs are theoretical posits, and indeed are posited as rarely themselves being conscious. We can again see how 1F advocates reject theoretical thinking about the nature of consciousness. In confining themselves to the subjective appearances, 1F views seriously distort the way we actually think about consciousness and psychological functioning.

This resistance to theorizing also affects the way many 1F advocates think about what utility conscious states have that is due exclusively to the property of those states being conscious. Much of a conscious state's utility must of course be due to its content properties, which drive most of its causal connections with behavior and other states, since differences in content properties result in different results (Rosenthal 2008). But confining attention to the subjective appearances encourages running together the content properties with consciousness, which leads to a sense that all the utility is due instead to consciousness.

And though a GWT is not itself 1F, this kind of thinking about the utility of a state's being conscious may well encourage a GWT, on which a state's being conscious is tied to its downstream psychological results. This connection is especially vivid in the work of Daniel C. Dennett on consciousness (e.g., 1991), which itself is arguably 1F in rejecting any HO notion of seeming (Rosenthal 2018), and accordingly ends up explaining consciousness by appeal to its behavioral effects.

Yet another way to see the anti-theoretical aspect of 1F approaches to consciousness pertains to the so-called explanatory gap and hard problem mentioned earlier. It's said that there is a hard problem about consciousness because of the apparent difficulty in explaining how neural processes could give rise to, or actually constitute, conscious experiences. And that difficulty can be described as an explanatory gap that blocks explaining consciousness by appeal to neural processes.

The apparent difficulty in giving such an explanation is usually described by appeal to the intuitive difficulty in seeing how neural processes could result in consciousness. It is said that we can imagine or conceive of those neural processes without consciousness, and that nothing about them seems intuitively to result in consciousness. It doesn't seem to make intuitive sense that they would.

But does it make intuitive, pretheoretic sense that a liquid would result from bonding together two gases, as with water? Or that most of the spatial extension of every solid object should be constituted by nothing solid, but at most wave-mechanical phenomena? We don't assess scientific explanations by whether they make intuitive,

pretheoretic sense, but whether they're embedded in a body of theory that generates solid predictions and fits comfortably with other bodies of theory.

Those who see a hard problem or an explanatory gap will argue that consciousness is a special case, in part because they maintain that we can imagine or conceive of whatever neural processes figure in a proposed explanation without any conscious experiences. But we can also conceive of hydrogen and oxygen being bonded in the right way without there being any water. What we can't conceive is that could happen compatibly with the relevant science that we know. And once we have a good predictive theory of consciousness and apply that standard, we won't be able in that way to conceive of the relevant neural processes without conscious experiences.

The appearance of a hard problem or an explanatory gap is fed in part by the odd view that an explanation of conscious mental phenomena should proceed directly from neural processing. But as already noted, this demand is reasonable only if one has already ruled out anything informative to say about consciousness in distinctively psychological terms. So one shouldn't appeal to the hard problem or explanatory gap in urging that the only explanation of consciousness must be directly in terms of neurological functioning. Indeed, since consciousness is plainly a psychological phenomenon, one would expect any explanation to be cast in the first instance in psychological terms, only moving to the neurological level when the psychological factors are all in place. Skipping over those intervening psychological factors is part of what makes it seem that a neurological explanation of consciousness doesn't make intuitive sense.

And there is more. Explaining consciousness by appeal to neurological occurrences will not make intuitive sense if we start with the subjective appearances of consciousness and attempt to reason to the appropriate neurological factors. It will inevitably seem that we are leaving out what is important to consciousness.

But that is always the within scientific explanations. Consider explaining water by appeal to H_2O . That will seem intuitively credible when we work up from H_2O , together with chemical theory, to explain the macroscopic nature of water. Proceeding in that direction, the explanation makes good intuitive sense. But it would be hopeless to start instead with the macroscopic nature of water and attempt to work back down to the H_2O .

More generally, post-Galilean scientific explanations can engender an unintuitive feel when one starts with specific macroscopic explananda and attempts to reason to the underlying reality posited by the relevant scientific theory. We readily dispel that unintuitive feel if instead we reason from the underlying reality, together with the relevant body of theory, to the macroscopic natures they explain.

The unintuitive feel engendered by taking the macroscopic natures as basic is due in part to leaving out the relevant body of theorizing, but not entirely. It is also because the macroscopic nature of things often does not in itself contain pointers to the relevant underlying reality. Indeed, in a pre-Galilean age, taking macroscopic natures as basic encouraged an Aristotelian approach to science, on which the main goal is to describe and taxonomize those macroscopic appearances.

These considerations readily carry over to explaining consciousness. We must not think of working from the subjective nature of consciousness down to factors that explain that nature, whether psychological or neurological. We must anticipate having psychological and neurological explanations that are good enough to enable us reason from them to specific subjective appearances, as we do with H₂O and water. And the specifically neurological explanations will only be good enough for that when all the relevant psychological factors are also in place. Insofar as claims that there is a hard problem or explanatory gap rest on taking the macroscopic appearances as dictating the possible avenues of theoretical explanation, those mysterian ideas are in effect a throwback to an Aristotelian science.

Indeed, we can readily explain why it seems at all tempting to hold, with 1F views, that consciousness is intrinsic to perceptual states by appeal to a well-entrenched but deeply faulty misconstrual of the implications of post-Galilean science. Many have taken the Galilean dictum that the book of "the universe ... is written in the language of mathematics" (Galilei 1623, passage tr. Popkin 1966, p. 65) to exclude properties such as colors from the natural order. The standard reaction, typically unstated, is to relocate such properties to the mind (Rosenthal 1999).

But colors resist mathematical treatment only as they appear in conscious perception; physical colors as they are on their own, independent of being consciously perceived, pose no difficulty for a mathematical description. And if the color properties we relocate to the mind are construed as they appear in conscious perception, the relocated mental properties will again themselves have consciousness built in. It is this sleight of hand that results in the appeal of the 1F insistence that mental color qualities are intrinsically conscious (Rosenthal, forthcoming, secs. III and VII). And these considerations suggest that rejecting 1F views will be essential to completing the Galileo scientific revolution.

The foregoing can be usefully supplemented by Josh Weisberg's (2024) rich discussion that carefully examines the reasons on offer for thinking that there is a hard problem, in particular, by David Chalmers (2003). Weisberg argues forcefully that those aspects of the appearances of consciousness that might seem problematic result simply from the sensory information accessed in consciousness being compressed and being accessed in an automatic way. But there is nothing problematic about such compression and automatic access. So this explanation dispels the sense that the appearances of consciousness generate a hard problem.

The foregoing considerations call for a careful examination of the popular appeal to intuitions that figures so prominently in arguments for a hard problem or explanatory gap. People don't all report having the same intuitions. More important, if one knows which intuitions somebody avows in this area, one can reliably predict what relevant theoretical positions that person will hold, and conversely.

That strikingly strong correlation suggests that intuitions are simply inviting one-liners that encapsulate a particular theoretical outlook. How could we otherwise explain the overwhelming reliability of the correlation? One could suggest that people start with intuitions and then generate a theory that fits. But where did the intuitions come from if there was no prior tendency to hold the relevant sort of theoretical position?

And if intuitions are encapsulated versions of theoretical positions, we should simply set them aside altogether and concentrate on evaluating the relevant theoretical positions. Since the intuitions themselves actually conceal the relevant theoretical thinking, focusing on them distracts from the theoretical assessment that is all important. In concentrating on what is intuitive, 1F theorists again reveal their anti-theoretical stance.

Scientific theories rarely depart totally from the macroscopic appearances. We do commonly think of things as consisting of much smaller parts; the chemical theory of water simply gives a somewhat surprising account of the relevant parts and how they're combined.

2F theories of consciousness follow suit. On a HO theory, for example, a state is conscious in virtue of one's being aware of being in that state. That fits with, and indeed is equivalent to, the commonsense idea noted earlier that no state is conscious if one is altogether unaware of being in it. What's surprising is that the HOA posited by HO theories is not itself a conscious state. It is an unconscious awareness, on a par with the subliminal awareness we have of stimulus when we perceive them unconsciously. So a HO explanation matches other scientific explanations pretty well, relying partly on commonsense ideas about the nature of what is being explained, but supplementing and adjusting its theoretical posits for explanatory purposes.

This arguably holds also for the specific appeal to higher-order thoughts that I have advanced elsewhere (Rosenthal, 2005, esp. chs 1, 2, 4, and 10-12). The theory posits a HOA in keeping with the commonsense idea that no state of which one is wholly unaware is conscious. But the theory then invokes various theoretical considerations to argue that this HOA is a conceptual state, on a par with the unconscious thoughts we often have.

And the other major type of 2F theory, GWT, also proceeds in just this way. GWT arguably starts from the commonsense idea that many, perhaps most, typical cases of conscious states have some significant utility, and posits that the downstream psychological processing that implements such functionality is what it is for those states to be conscious. Both types of 2F theory operate in the manner of standard scientific theories.

1F views are currently widely dominant in current discussion of consciousness. Given all these downsides, why are they so popular?

1F advocates contend that their picture of consciousness is well-entrenched in common sense, and so does justice to the way we think about consciousness in ordinary, everyday terms. Thus their ubiquitous reliance on allegedly pretheoretic intuition.

1F views do start from a correct observation about our commonsense views, but then wildly overshoot. Consciousness is mental appearance. It is the way our mental lives subjectively appear to us. This is basic to any serious way of thinking about consciousness, whether in common sense or in a scientific theory/

But 1F views extrapolate from that commonsense observation to the extravagant claim that the mental appearances also exhausts the mental reality of consciousness. If this were so, there could be for consciousness no coherent distinction between appearance

and reality when it comes to consciousness. And that is what Nagel famously maintains: "The idea of moving from appearance to reality seems to make no sense" for conscious experiences (p. 444).

If the mental appearances of consciousness did exhaust its mental reality, all mental reality would lie behind a wall of subjectivity, separated from anything objective. But there are, in addition to the mental appearances, objective mental realities that underlie and explain the appearances. There are the mental occurrences that subjectively appear to us, states with conceptual content or qualitative character, both of which we can theorize about apart from consciousness (Rosenthal 1986, 2010). Those states are one part of the mental reality that underlies and helps explain the mental appearances.

And there are also the mental occurrences that actually implement the subjective appearances themselves. And these are also an aspect of the objective mental reality of conscious states (Berger 2017; Rosenthal 2022, forthcoming). On GWT, that additional mental reality is the downstream availability of content. On HO theories it is the HOA of the states themselves, which represent the states specifically in respect of those mental properties that figure in our subjective awareness.

1F advocates will object to all this, and claim that distinguishing appearance from reality has no place with consciousness. But all cases of appearing are implemented by something objective. One's appearance in a mirror is not something outside the natural order; it is a matter of how light reflects off particular surfaces. The only reason to think that the appearances of consciousness are not themselves implemented by objective occurrences, which invite straightforward explanation in standard scientific terms, is that it doesn't subjectively seem that the appearances of consciousness are thus implemented.

But it wouldn't seem so subjectively: Objective implementation of all appearances is never evident from the appearing itself. We can't tell from the appearances themselves whether there is any such objective implementation. To determine what the objective implementation is in a case of appearance of any sort whatsoever we must always appeal to factors that lie outside the appearances themselves.

The idea that we could determine how a type of appearance is implemented from considerations internal to the appearances is of a piece with the idea noted above that we can assess the credibility of a scientific explanation of consciousness by seeing whether we can reason from how things seem subjectively to the factors appealed to in the explanation. It is the mistake of thinking that how things appear must govern any account of those appearances.

That is what leads to 1F advocates rejecting any appearance-reality distinction for consciousness, and to the illusion (Mandik 2016; Weisberg 2024) that there is a hard problem or an explanatory gap. Rejecting an appearance-reality distinction for consciousness is groundless, and with it the adoption of a 1F view. We must accept and work with a 2F view.

References

- Baars, Bernard J. (1997), In the Theater of Consciousness: The Workspace of the Mind, Oxford: Oxford University Press.
- Berger, Jacob (2014), "Consciousness is Not a Property of States: A Reply to Wilberg," Philosophical Psychology, 27, 6: 829-842.
- Berger, Jacob (2017), "How Things Seem to Higher-Order Thought Theorists," Dialogue: Canadian Philosophical Review, 56, 3 (September): 503-526.
- Block, Ned (1978), "Troubles with Functionalism," in Minnesota Studies in the Philosophy of Science, IX, ed. C. Wade Savage, Minneapolis: University of Minnesota Press, pp. 261-325.
- Block, Ned (2015), "The Puzzle of Perceptual Precision," in Open MIND, ed. Thomas Metzinger and Jennifer M. Windt, Frankfurt am Main: MIND Group, doi: 10.15502/9783958570726.
- Block, Ned (2017), "Unconscious Perception within Conscious Perception,' 7-9 of 'Does Unconscious Perception Really Exist? Continuing the ASSC20 Debate,' by Megan A. K. Peters, Robert W. Kentridge, Ian Phillips, and Ned Block, Neuroscience of Consciousness, 1, nix015, <https://doi.org/10.1093/nc/nix015>.
- Byrne, Alex (1997), "Some Like It Hot: Consciousness and Higher-Order Thoughts," Philosophical Studies, 86, 2 (May): 103-129.
- Chalmers, David J. (2003), "Consciousness and Its Place in Nature," in The Blackwell Guide to Philosophy of Mind, ed. Stephen P. Stich and Ted A. Warfield, Malden, Massachusetts: Blackwell Publishing Ltd., pp. 102-142.
- Dehaene, Stanislas, and Lionel Naccache (2001), "Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework," Cognition 79, 1-2 (April): 1-37.
- Dennett, Daniel C. (1991), Consciousness Explained, Boston: Little, Brown and Company.
- Dretske, Fred (1993), "Conscious Experience," Mind, 102, 406 (April 1993): 263–283.
- Dretske, Fred (2006), "Perception without Awareness," in Perceptual Experience, ed. Tamar Szabó Gendler and John Hawthorne, Oxford: Clarendon Press, pp. 147-180.
- Fallon, Francis, Alan L. F. Lee, Brian Odegaard, Andrew Haun, and David Rosenthal (2020-2022), "Replication and Extension of Crucial John Grimes Experiment: Change Detection during Saccades," Templeton World Charity Foundation Project 0455.
- Fernandez-Duque, Diego, and Ian M. Thornton (2000), "Change Detection without Awareness: Do Explicit Reports Underestimate the Representation of Change in the Visual System?," Visual Cognition, 7, 1-2-3 (January–March): 324–344.
- Galilei, Galileo (1623), Il Saggiatore, Rome.

Grimes, John (1996), "On the Failure to Detect Changes in Scenes across Saccades," Perception, ed. Kathleen Akins, New York: Oxford University Press, pp. 89–110.

Lamme, Victor A. (2006), "Towards a True Neural Stance on Consciousness," Trends in Cognitive Sciences, 10, 11 (November): 494-501.

Laloux, Cédric, Arnaud Destrebecqz, and Axel Cleeremans (2006), "Implicit Change Identification: A Replication of Fernandez-Duque and Thornton (2003)," Journal of Experimental Psychology: Human Perception and Performance, 32, 6 (December): 1366-1379.

Lau, Hakwan (2022), In Consciousness We Trust: The Cognitive Neuroscience of Subjective Experience, Oxford: Oxford University Press (open access: <chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/http://fdslive.oup.com/www.oup.com/academic/pdf/openaccess/9780198856771.pdf>).

Levine, Joseph (2001), Purple Haze: The Puzzle of Consciousness, New York: Oxford University Press.

Mandik, Pete (2016), "Meta-Illusionism and Qualia Quietism," Journal of Consciousness Studies, 23, 11-12: 140-148.

Michel, Matthias, and Adrien Doerig (2022), "A New Empirical Challenge for Local Theories of Consciousness," Mind & Language, 37, 5 (November): 840-855.

Mylopoulos, Myrto (2015), "Consciousness, Action, and Pathologies of Agency," in Disturbed Consciousness: New Essays on Psychopathology and Theories of Consciousness, ed. Rocco J. Gennaro, Cambridge, Massachusetts, and London: MIT Press, pp. 75-102.

Mylopoulos, Myrto (2022), "Neurobiological Theories of Consciousness," in Mind, Cognition, and Neuroscience: A Philosophical Introduction, ed. Benjamin D. Young and Carolyn Dicey Jennings, New York and London: Routledge, pp. 280-293.

Naccache, Lionel (2018), "Why and How Access Consciousness Can Account for Phenomenal Consciousness," Philosophical Transactions of the Royal Society B: Biological Sciences, 373, 1755 (September 9): Article 20170357.

Nagel, Thomas (1974), "What Is It Like to Be a Bat?," The Philosophical Review, 83, 4 (October): 435-450.

Neander, Karen (1998), "The Division of Phenomenal Labor: A Problem for Representational Theories of Consciousness," Philosophical Perspectives, 12: 411-434.

Norman, Liam J., Charles A. Heywood, and Robert W. Kentridge (2013), "Object-Based Attention without Awareness," Psychological Science, 24, 6 (June): 836-843.

Popkin, Richard Henry (1966), The Philosophy of the Sixteenth and Seventeenth Centuries, New York: Free Press.

Quine, W. V. (1951), "Two Dogmas of Empiricism," The Philosophical Review, 60, 1 (January): 20-43.

Raffman, Diana (2011), "Vagueness and Observationality," in Vagueness: A Guide, ed. Giuseppina Ronzitti, Dordrecht: Springer, pp. 107-122.

Rosenthal, David (1983), "Reductionism and Knowledge," in How Many Questions? Essays in Honor of Sidney Morgenbesser, ed. Leigh S. Cauman, Isaac Levi, Charles Parsons, and Robert Schwartz, Indianapolis: Hackett Publishing Co., pp. 276-300.

Rosenthal, David (1986), "Intentionality," Midwest Studies in Philosophy X (September): 151-184.

Rosenthal, David (1999), "Sensory Quality and the Relocation Story," Philosophical Topics 26, 1/2 (Spring and Fall): 321-350.

Rosenthal, David (2005), Consciousness and Mind, Oxford: Clarendon Press.

Rosenthal, David (2008), "Consciousness and Its Function," Neuropsychologia, 46, 3: 829-840.

Rosenthal, David (2010), "How to Think about Mental Qualities," Philosophical Issues, 20 (October): 368-393.

Rosenthal, David (2018), "Seeming to Seem," in The Philosophy of Daniel Dennett, ed. Bryce Huebner, New York: Oxford University Press, 2018, pp. 133-164.

Rosenthal, David (2022), "Mental Appearance and Mental Reality," in Qualitative Consciousness: Themes from the Philosophy of David Rosenthal, ed. Josh Weisberg, Cambridge University Press, pp. 243-271.

Rosenthal, David (forthcoming), "Methodological Considerations for the Study of Mental Qualities," in Conscious and Unconscious States and Processes: Examining Their Nature, Similarities and Differences, ed. Juraj Hvorecký, Tomáš Marvan, and Michal Polák, London and New York: Routledge.

Shargel, Daniel (2016), "The Insignificance of Empty Higher-Order Thoughts," Journal of Cognition and Neuroethics, 4, 1 (April): 113-127.

Thornton, Ian M., and Diego Fernandez-Duque (2001), "An Implicit Measure of Undetected Change," Spatial Vision, 14, 1 (January): 21-44.

Weisberg, Josh, Explanatory Optimism about the Hard Problem of Consciousness, New York and London: Routledge, 2024.