

To appear in Conscious and Unconscious States and Processes: Examining Their Nature, Similarities and Differences, ed. Juraj Hvorecký, Tomáš Marvan, and Michal Polák, Routledge

METHODOLOGICAL CONSIDERATIONS FOR THE STUDY OF MENTAL QUALITIES

David Rosenthal

Philosophy and Cognitive Science

Graduate Center, City University of New York

425 Riverside Drive, New York, NY 10025-7731, USA

Email: davidrosenthal1@gmail.com

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

Abstract: The currently popular view that mental qualitative character is intrinsically conscious has a significant methodological downside: It precludes any informative account of what conscious qualitative character consists in. And without that, any theoretical or scientific account of its nature is at best severely limited. But that intrinsicist view lacks any independent support. Moreover, it is wholly optional, since a compelling account of mental qualities is available that proceeds independently of consciousness, and is convincingly supported by robust empirical findings. We can also explain why the intrinsicist picture is widely endorsed despite its significant disadvantages and lack of independent support, as well as why many maintain, also without independent support, that the subjective appearances of consciousness exhaust its mental reality.

I. Methodological Issues Facing Intrinsic Consciousness

Thomas Nagel's (1974) "What Is It Like to Be a Bat?" did much to convince a generation and more, in both philosophy and related fields, that mental qualitative character cannot occur without consciousness. A mental state's having qualitative character, Nagel in effect urges, consists in there being something it's like for one to be in that mental state, and he understands there being something it's like for one in terms of consciousness. So a mental state can have qualitative character only if it's conscious. Being conscious is an intrinsic aspect of any mental state that has qualitative character.

And if so, we cannot even understand what mental qualitative character is except in terms of consciousness. Many today see no other way of thinking about mental qualitative character as even conceivable. So influential has Nagel's picture been that consciousness is now often identified with conscious qualitative mentality, sometimes

seeing mental qualities as determinates of consciousness construed as a determinable, so that each mental quality is simply a particular type of consciousness.

Nagel holds that we cannot do justice to the subjective character of consciousness by appealing to anything objective. A grasp of consciousness as genuinely subjective drives out any factor that is not itself subjective. And if we can say nothing objective about conscious qualitative character, it's unlikely that we can say anything that is even at all informative about what it consists in.

The idea that qualitative subjectivity involves nothing objective is a consequence of seeing consciousness as intrinsic to qualitative character. If qualitative character is intrinsically conscious, what consciousness tells us is decisive about the nature of qualitative character. Anything we could come in any other way to know about qualitative character would be overridden if it conflicted with what consciousness says. First-person access is then the last word about qualitative character. Ruling out information about qualitative character derived from any other source amounts to driving out any objective information.

This is reflected in Nagel's discussion, which allows for describing consciousness and qualitative character only by way of a closed family of cognate terms: 'what it's like', 'subjectivity', 'perspectives', and 'points of view'. But these alternative descriptions contain no new information; each is itself to be understood as equivalent to the primary phrase, 'what it's like'. So there is no cognitive progress to be gained from moving from one to another. Being confined to this closed set of cognate notions will recall the closed curve of terms W. V. Quine (1951) highlights as having been traditionally used in an unsuccessful attempt to sustain some notion of analyticity.

The closed curve for analyticity prompted many to abandon any appeal to such a notion. But a parallel reaction to the closed curve pertaining to conscious qualitative character is for most people not a serious option. We take ourselves to have first-person access to consciousness, whatever the nature of that access may consist in. So with rare exceptions, such as illusionists (Frankish 2016; Kammerer 2022), the occurrence of consciousness is not denied.

The upshot for most who endorse Nagel's closed curve is that consciousness comes to be seen as a mysterious, ineffable phenomenon. If one can describe conscious qualitative character only by appeal to cognate subjective terms that have no useful, informative content, consciousness must indeed be strange in nature, unlike anything else we know about. The absence of objective ties to other phenomena undermines informative understanding and makes a mystery of conscious qualitative character.

This sense of mystery is vividly expressed in the idea that an explanatory gap separates consciousness from everything physical (Levine 2001) and that there's a hard problem in explaining how consciousness could have any ties to physical reality (Chalmers 1995). That sense of mystery is also usefully captured by the idea that, since consciousness is our only access to mental qualities, the mental qualities that enable

the perceiving of specific stimulus properties could be undetectably inverted from one individual to another.

Those influenced by Nagel sometimes suggest that first-person access, construed as wholly independent of anything objective, does provide useful information about conscious qualitative character. But that is not so. If qualitative character is intrinsically conscious, so that first-person access so construed is our only way to know about conscious qualitative character, such access tells us little if anything informative. First-person access then tells us only that this is what it's like to have a particular qualitative state, where 'this' refers to something accessible only to the subject. "I know it when I see it" won't help if qualities are invertible from one individual to another. So if qualitative character is intrinsically conscious, we can indeed say nothing useful or informative about it. This is an unacceptable price to pay for accepting intrinsicism.

Those who endorse that assumption occasionally are explicit about this consequence. Thus, qualitative character is intrinsically conscious on Ned Block's widely adopted concept of phenomenal consciousness. Block acknowledges that he "cannot define [phenomenal consciousness] in any remotely noncircular way" (1995, 230). As he colorfully puts it elsewhere, we can 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" (Block 1978, 281). "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" (Block 2015, 47). It's not at all clear what such pointing could amount to.

All this raises concerns about whether any serious scientific or theoretical treatment of conscious qualitative character is possible on an intrinsicist view. A scientific account requires a description of the target phenomenon, perhaps informally and in commonsense terms, but still in a way that's informative enough to pick out cases reasonably reliably. But on the intrinsicist picture we can say only that being conscious is an intrinsic aspect of qualitative character. We cannot even say what intrinsic aspect it is. And if the mental qualities responsible for perceiving specific stimulus properties are conceived of as being undetectably invertible, we can say nothing informative about any specific mental qualities.

Some who follow Nagel might be comfortable with these anti-theoretical and anti-scientific consequences. But others, such as Block, are not, and insist that the intrinsicist picture does not rule out a science of consciousness. Block himself proposes a scientific account based on a neural implementation of consciousness, favoring Victor Lamme's (2006) hypothesis that perceptions are conscious just in case recurrent processing occurs in the relevant sensory cortex. Such a localist account fits comfortably with Block's view that consciousness is an intrinsic aspect of qualitative character, since being conscious is then an aspect of the neuronal processing that's also responsible for the specific types of mental qualitative character.

Block may well be unmoved by the demand for an informative description as a condition for establishing a neural implementation, since he holds that no "remotely noncircular" account of conscious qualitative character is possible. If no informative description is possible, it's idle to ask for one.

But we must insist on that demand. Any scientific account of a commonsense phenomenon must show how variation in that phenomenon corresponds to variation in the factors the theory posits, as with degrees of heat and mean molecular kinetic energy. That requires an informative description of the target phenomenon and the ways it varies. For qualitative consciousness, that's variation among the mental qualities and between consciousness and its absence. And undetectable inversion precludes knowing even what quality is at issue. To explain how mental states vary a theory must rely on an informative description of those variations.

Moreover, since the intrinsicist picture is not the only option, we need not forego an informative description. Accounts are also available on which consciousness is not intrinsic to qualitative character. On a higher-order account (e.g., Rosenthal 2005), a state's being conscious consists in one's being aware in a suitable way of being in the state, and on a global-neuronal-workspace account, (Dehaene and Naccache 2001; Naccache 2018), it consists in availability for suitable downstream processing. These theories provide useful, informative descriptions of what it is for a state to be conscious. Both also offer explanations of why first-person reports reflect whether a state is conscious, so that we needn't simply rest with the unexplained idea that such reports express consciousness.

Also, on those accounts some nonlocalist neural implementation, likely involving prefrontal cortex, is likely more credible. Other claims about consciousness may also influence where to look for a neural implementation. If, as some hold, a state's being conscious involves some distinctive utility, or there is a significant tie between consciousness and confidence (on such claims see, e.g., Rosenthal 2008 and 2019, respectively), that might affect which neural implementation makes sense.

So to assess a proposed neural correlate of qualitative consciousness, we need some account of consciousness qualitative character cast in psychological or commonsense terms. That account may just be preliminary or provisional, and neurological findings may well lead us to modify whatever psychological account we start with. But we do need some informative psychological description at the outset.

II. Lack of Support

The problems intrinsicism generates are methodological in nature. They involve significant limits on how conscious qualitative character can be investigated. We cannot informatively describe qualitative consciousness, and so have no useful way to

pick out its variations as a target for study. This impedes, and indeed likely precludes altogether, any theoretical account that could help reveal its nature.

Since these methodological constraints result simply from assuming that qualitative character is intrinsically conscious, we must ask whether there is any independent support for that assumption. Those who already see consciousness as ineluctably mysterious may see any alternative as untenable. But one can't sustain the intrinsicist assumption in that way, since the picture of qualitative consciousness as mysterious itself stems from that assumption. And we must not dismiss alternative approaches without convincing support for intrinsicism.

Advocates of intrinsicism typically appeal to what they claim are pretheoretic intuitions. But such an appeal is at best questionable. For one thing, different people report different intuitions; reports of intrinsicist intuitions are by no means universal. But the deeper difficulty is that it's unlikely that the so-called intuitions are independent of tacit theorizing. One can readily predict the intuitions a person will report from the theoretical approach that person favors, and conversely. So it's reasonable to see so-called intuitions as simply inviting one-liners that encapsulate the theoretical approach in question. We should accordingly assess each intuition on the merits of the theoretical approach that intuition channels.

Those who appeal to intuitions claim for them a kind of foundational status, and so as not subject to theoretical evaluation. But it's always irrational to rely on independently unsupported claims of whatever sort in getting at the truth about things. Since advocates of intrinsicism hold that we cannot informatively describe conscious qualitative character, they might also contend that unsupported, foundational intuitions are the best we can hope for. But that claim would be viciously circular, since the apparent ineffability of qualitative consciousness itself stems from the intrinsicist assumption. In no other area of inquiry would we grant special status to claims that wholly lack independent support, and we must not do so here.

And apart from so-called intuitions, it's unclear what support there could be for the intrinsicist assumption. Advocates might appeal to its not subjectively seeming that being conscious consists in anything extrinsic to a conscious state. But that would help only if the subjective appearances were all we had to rely on in assessing the nature of qualitative consciousness. And that constraint in effect simply restates the intrinsicist assumption, providing no independent support for it.

First-person access can be a natural starting point for understanding mental phenomena. It's often useful to start a discussion of mental qualities by describing the subjective awareness we have of them. But even if first-person access is a useful first word about qualitative consciousness, it cannot be the last word. For one thing, first-person access to one's qualitative states can and sometimes does go wrong, as we'll see in section VI. In addition, qualitative consciousness has useful connections with other factors that are accessible independently of first-person access.

First-person access tells us how our mental lives subjectively appear; enabling a phenomenological description of our conscious mental lives. But first-person access cannot tell us about the mental reality that underlies those subjective appearances, or the ties those subjective appearances have with nonmental factors. We can get at these things only if we supplement the subjective appearances with theorizing based both on those appearances and on sound empirical observations and findings.

Seeing consciousness as intrinsic to qualitative character prevents such theorizing, by taking the subjective appearances to exhaust the mental reality of conscious qualitative states. Since intrinsicism sees first-person access as telling us everything there is to know about qualitative consciousness, intrinsicism is, once again, strongly anti-theoretical.

That anti-theoretical stance actually operates to sustain itself. If we can't theorize about conscious qualitative character, we cannot even formulate a challenge to the intrinsicist assumption. So it's crucial to explore theoretical alternatives, and to assess them on their theoretical merits.

As already noted, there are well-developed theoretical alternatives. Higher-order and global-workspace theories both constitute accounts on which the property of a state's being conscious is not intrinsic to the state. And both theories receive considerable support from a range of subjective and theoretical considerations.

III. Explaining the Appeal

The view that mental qualities are intrinsically conscious results in an uninformative conception of both qualitative character and consciousness, and a theoretical dead end about both. But that view is wholly optional. Why then does that view seem inviting to so many? One answer is simply a desire to rely exclusively on first-person access and subjective appearance, and to dismiss any consideration of underlying mental reality. I'll come back to that in section VII.

But there is another source of appeal for holding that qualitative character is intrinsically conscious, which derives from an implication that modern science is thought by many to have about the nature of the physical properties that common sense holds we perceive.

Aristotle (2016) held that colors and other qualitative properties, which he called the proper sensibles, are properties of physical objects that are literally taken on by the soul in perceiving. The qualitative properties in the soul do not merely correspond to or reflect perceived stimulus properties; they are on Aristotle's view literally identical with those stimulus properties.

Today this idea will rightly strike almost everybody as too strange to be taken seriously. But that view was, with variations, widely held for roughly two millennia, until the paradigm shift we associate with modern science resulted in a general acceptance that

everything in nature is mathematically describable. And seeing all physical properties as mathematically describable seemed to raise problems for our commonsense conception of the physical stimulus properties we perceive.

The difficulty stems from a sense that the colors of physical objects as conceived by common sense are not mathematically describable; similarly for the physical properties we perceive by the other sense modalities. If so, those properties, as conceived by common sense, cannot be any part of physical reality.

The standard reaction to this apparent conundrum has been to claim that colors and other properties that common sense regards as properties of physical objects are not physical at all, but only mental. Physical reality contains only mathematically describable correlates of commonsense qualitative properties, such as electromagnetic wavelengths for the colors.

This relocation of commonsense perceptible properties into the mind seems inviting to many, since one can take the claim about mathematical describability as inapplicable to mental phenomena, though the resulting property dualism, howsoever tacit, is for some a lingering source of discomfort. The mental qualities are then conceived on the model of commonsense perceptible properties, as indeed they were with Aristotle; it's just that on this relocation story those commonsense physical qualities are no longer themselves taken to exist.

But property dualism to one side, this widely accepted relocation story has a striking consequence that is rarely recognized. Take commonsense perceptible physical colors. Those properties would resist mathematical treatment only if they were conceived of as we are conscious of them. Apart from the way we're conscious of them, there is no obstacle to a thoroughgoing mathematical description. So conceiving of colors as resisting mathematical description by itself amounts to conceiving of them as we are aware of them in conscious perception. Commonsense physical colors require relocation to the mind only if conceived of as having conscious awareness literally built in.

Since the commonsense physical qualities so conceived have consciousness built in, when they relocated, the resulting mental versions will also have consciousness built in. So thinking of mental qualities as relocated versions of commonsense physical qualities results in thinking of mental qualities as intrinsically conscious. The relocation story, on its own, results in the intrinsicist picture, with all the problems that stem from it (Rosenthal 1999; 2016). The widespread appeal of the relocation story, often merely tacit, underlies the pull so many feel to insist that mental qualitative character is intrinsically conscious.

But the relocation story is itself a mistake. There is no reason to conceive of physical colors and other physical qualitative properties as we are conscious of them. They are simply independent physical properties, and hence mathematically describable. We can then also reject the relocation story, and the resulting conception of mental qualities

as we intrinsically conscious. That opens the way to an account of mental qualitative character altogether independent of consciousness. The following two sections develop and defend such an account.

IV. A Science of Qualitative Character

If the property of a state's being conscious is not intrinsic to that state, we need an account of qualitative character that does not rely in any way on consciousness. And there is independent reason to want such an account. Though in everyday life we tend to focus only on conscious cases of qualitative character, there is compelling evidence, arguably decisive, for qualitative character that's altogether unconscious, as we'll see in section V.

Bracketing whatever connection qualitative character may have with consciousness, qualitative character indisputably has a connection with perceptual discrimination that is highly useful and revealing. A mental quality of red, for example, enables one to discriminate red stimuli from stimuli of any other color; similarly for all other mental qualities of color, and indeed for all mental qualities of every sensory modality.

This tie that qualitative character has to perceptual discrimination provides a way to think about mental qualities that is altogether independent of consciousness, since perceptual discrimination plainly occurs both consciously and unconsciously. Sections IV and V develop and support an account along such lines. We need not conceive of mental qualities as inextricably tied to consciousness; that view is wholly optional.

Moreover, the connection with perceptual role is at least as well-entrenched in common sense and ostensibly pretheoretic intuition as any tie mental qualities might have with consciousness, and likely considerably more so. Somebody untutored in the relevant academic discussion might be puzzled about whether qualitative states are invariably conscious, much less intrinsically so. But if somebody suggested that mental qualities played no role in perceptual discrimination, it would be clear that such a person simply wasn't talking about mental qualitative character at all.

Perceiving requires discriminating among perceptible properties of stimuli, such as colors, shapes, sizes, sounds, odors, tactile pressures and textures, and so forth. By manipulating stimuli, we can test for just noticeable differences (JNDs), differences between stimuli so small that if the stimuli were physically any closer one would be unable to distinguish them at all.

We can then use those JNDs to construct a quality space (QS) that captures all the properties in a range of stimulus properties that an individual can discriminate. Such a space will have as many dimensions as there are independent ways in which those stimulus properties are discriminable. For color these will be hue, saturation, and brightness; similarly for the spatial stimulus properties discriminable by vision, and for the properties discriminable by every other perceptual modality.

Conscious perceptual discrimination plainly rests on such differences in mental quality. We consciously discriminate stimuli by being in conscious perceptual states whose qualitative character differs in ways that correspond to discriminable differences among those stimuli. So in the conscious case, the QS that maps discriminable stimulus properties will also map the conscious qualities that enable one to discriminate those stimulus properties. We can fix the mental qualities by the stimulus properties they enable us to discriminate. And since we can describe those stimulus properties by relative location in a QS, we can also fix the operative mental qualities that way.

This gives us a theory of what the mental qualities are, so far just for the case in which the mental qualities are conscious. Mental qualities are properties of conscious perceptual states, mapped by a QS of consciously discriminable stimulus properties. On this QS theory, mental qualities in the conscious case are fixed by relative location in a QS built solely from discriminative ability.

JNDs are sometimes construed as minimally noticeable differences between the mental qualities themselves. That construal is appealing if one sees mental qualities as intrinsically conscious, since subjective appearance is then our only access to mental qualities. But that won't do for an account of mental qualities that's independent of consciousness.

Independently of that, there are significant methodological advantages to construing JNDs in terms of discriminable stimulus properties, rather than the mental qualities themselves. Manipulating the stimulus properties one is presented with provides some independent control over relative discriminability relations, which is unavailable if we construe JNDs simply in terms of the mental qualities themselves. Multidimensional scaling, which relies solely on subjective similarity judgments, also lacks any independent control, and introduces additional unreliability by relying on distant similarity comparisons. And there is no downside, methodological or otherwise, to construing JNDs in terms of discriminable stimulus properties.

We have a theory for mental qualities in conscious perceiving. But perceptual discrimination also occurs unconsciously, as in subliminal perceiving. Mental qualities conceived as intrinsically conscious could not figure in unconscious discrimination. But that provides a compelling methodological reason to reject that conception. Perceptual discrimination plainly does occur without being conscious, and mental qualities are the properties in virtue of which we consciously discriminate. So rejecting the intrinsicalist claim gives us a unified approach.

There are reasons apart from unity to invoke unconscious mental qualities. There must be mental properties of some sort that enable unconscious discrimination, properties that vary in ways that correspond to unconsciously discriminable differences among stimuli. In the conscious case, those properties are the mental qualities. What could lead one not to see them as such in the unconscious case? Indeed, what other kind of mental property could they be? Mental qualities seem to be the only serious candidate.

An advocate of intrinsic consciousness might contend that those properties aren't mental at all, but simply subpersonal. But that has a striking methodological downside. If unconscious discrimination doesn't rely on mental qualities, but only on some type of subpersonal property, those subpersonal properties would likely also be responsible, on their own, for conscious perceptual discrimination. Conscious qualities would then be altogether irrelevant for perceptual discrimination, again an unreasonable price to pay for accepting intrinsicism.

And as already mentioned, there is compelling empirical evidence that mental qualities do occur without being conscious, as we'll see in section V. But for now, note that QS theory dispels the sense of mystery that some see in mental qualitative character. On QS theory we can trace explanatory ties between neural states and mental qualities, understood independently of consciousness.

Moreover, on QS theory undetectable inversion of mental qualities is literally inconceivable. Inversion could be undetectable only if there were an axis around which the QS is symmetrical; inversion around any asymmetric axis would be readily detectable. But an axis of symmetry would result in the two sides defining identical relative locations. So on QS theory the stimulus properties on the two sides would then be identical. A QS cannot have an axis of symmetry. QS theory implies that we cannot even conceive of undetectable inversion of mental qualities.

QS theory dispels the sense of mystery in mental qualities construed independently of consciousness. But how about when consciousness is added? If being conscious were intrinsic to mental qualities, there could be no informative explanation of how it is that they're conscious. But if mental qualities are not intrinsically conscious, a state's being conscious is due to some factor that's independent of the state. And describing that factor then enables an informative explanation of how it is that some qualitative states are conscious and others not.

Consciousness is a psychological phenomenon. So an explanation of what it consists in must appeal to factors that are themselves psychological. But those factors cannot also be conscious. Otherwise, the explanation would be uninformatively circular, as with Nagel's closed family of terms. Also, the lack of an intermediate psychological level between consciousness and the neurological would ineluctably produce a sense of mystery. Appealing to some psychological factor that is not itself conscious is the strategy of higher-order (HO) theories of consciousness.

Suppose one thinks or perceives something, but one is wholly unaware of doing so. That thought or perception plainly is not conscious. This folk-theoretic platitude governs our commonsense understanding of what it is for a mental state to be conscious, as well as most experimental investigation. And that platitude is equivalent to a necessary condition for a state to be conscious: For a state to be conscious, one must in some way be aware of being in that state. I've elsewhere called this the transitivity principle (TP), and it's endorsed in some form by all HO theories.

TP gives us a necessary condition for a state to be conscious, but not a sufficient condition. Still, we can close in on a sufficient condition if we determine in what way one is aware of a state when that state is conscious. There are several important constraints. One is that the HO awareness (HOA) must not rely on any conscious mediation, for example, on any inference that one is aware of. It cannot be that I simply believe what you tell me about myself.

Also, HOAs are themselves rarely conscious. We know about HOAs not by first-person access, but as theoretical posits that do a good explanatory job in distinguishing conscious from unconscious mental states. Indeed, for a HOA to be conscious would require a third-order awareness, which itself is likely very rare. This constraint avoids an uninformative circularity.

Elsewhere I have argued that the HOA is very likely a higher-order thought (HOT) to the effect that one is in the target state. Positing HOTs explains why, for creatures with the relevant linguistic ability, a state is conscious just in case it's verbally reportable, and also why, for such creatures, a verbally expressed thought is always conscious (Rosenthal, 2005, chs. 2, 10, respectively). But for present purposes we can stick with positing a generic HOA, and remain noncommittal about the HOT hypothesis.

A HO theory of consciousness fits well with QS theory. Taken together, they explain what it is for a state to be qualitatively conscious without construing mental qualities as intrinsically conscious. And on this combined theory, HOAs represent the mental qualities that figure in a state's being conscious in respect of their relative location in the relevant QS. HOAs describe conscious qualitative states in QS terms.

On our commonsense ways of thinking, subjective awareness does represent differences among conscious mental qualities in respect of their relative location in a QS. When we describe what it's like to have some qualitative experience, we typically compare that experience to other experiences of currently available objects or to characteristic experiences of familiar objects. We describe what it's like for us in terms of the objects our qualitative states enable us to discriminate. We are subjectively aware of mental qualitative character in the relational, comparative terms of QS theory.

Those who see mental qualities as intrinsically conscious typically also claim that they're atomic in nature, and so reject any comparative account, such as that of QS theory. But seeing mental qualities as atomic is tempting only if consciousness is our only access to them. Once we take into consideration their role in perception, their comparative nature becomes manifest and undeniable.

We describe what it's like to have an experience by appeal to other experiences of currently available or otherwise familiar objects. This fits well with the way the phrase, 'what it's like', actually operates in commonsense, nonacademic contexts. Typically, when we ask somebody what something is like for them, we expect and receive an answer that describes the objective situation being experienced. One might say, for

example, that the sound is like the roar of ocean waves. We expect a description, cast in nonmental terms, of something in the person's perceptible environment.

Still, the phrase, 'what it's like', does describe the mental qualitative character of one's experience. And this is what we would expect on QS theory. Since mental qualities are the properties in virtue of which one discriminates stimulus properties, we describe the qualitative character of our experiences by appeal to the objects one experiences. This not to deny the occurrence of mental qualities or our awareness of them. Contrary to Gilbert Harman's (1990) representationalism, we are typically aware of mental qualities that occur in conscious experience. It is simply to highlight the way we describe those mental qualities in commonsense contexts.

V. Empirical Grounding

A science of mental qualities cast in distinctively psychological terms is possible only if they are not solely accessible by consciousness. Moreover, we can avoid exclusive access by consciousness if we see mental qualities as the mental properties that enable perceptual discrimination for both conscious and unconscious discrimination.

We have compelling theoretical reasons for such an account. But it's important to have empirical findings that provide independent support, especially for the occurrence of mental qualitative character that isn't conscious. And there is ample empirical support. To begin with, even when a qualitative state is conscious, there can be some distinctively qualitative aspect of that state that fails to be conscious.

One example occurs in the classic work of George Sperling (1960). Sperling briefly presented subjects with a matrix of letters, typically three rows of four letters each. After they vanish, subjects could identify only three or four letters. But when an auditory tone cued a particular row after the display had disappeared, subjects could identify three or four in that row. Since subjects never know which row will be cued, they must somehow have retained identities of most of the twelve letters. But how?

Block (2007) thinks subjects retain those identities consciously, presumably because the information is qualitative and Block maintains that qualitative character can't occur except consciously. Because the relevant qualitative character isn't subjectively accessible, Block concludes that phenomenal consciousness overflows cognitive access.

But there is no independent evidence for such overflow. And there is compelling evidence that the type of memory operative here isn't conscious (Irvine 2014). Block can't appeal here to his famous distinction between phenomenal and access consciousness, since their being independent is exactly what's at issue (Naccache 2018). Without independent support for qualitative consciousness that isn't subjectively accessible, we must continue to judge what's conscious by what subjects report.

When asked informally how they retain the information, subjects often say it's by way of conscious images, and Block takes those anecdotal reports "at face value" (2007, p. 488). But that's a mistake. Subjects likely have no clear or reliable idea about how they retain the information, and in particular likely haven't considered the possibility of doing so unconsciously. And subjects' desire to say what they think experimenters expect would likely lead them to appeal to conscious images.

Subjects' perceptions of the letters are all conscious; it's just that most of the shapes relevant to determining letter identity aren't conscious. Since visible shapes are qualitative, the more natural interpretation is that most shapes are retained unconsciously. What overflows cognitive access is simply unconscious qualitative information (e.g., Cova et al 2021).

In other experimental work, Diana Raffman (2011) found that subjects accurately adjust one patch to match the hue of others that differed in hue, though subjects consciously perceived those others as the same. All the patches were consciously seen, but matching revealed differences among them that were seen only unconsciously. Indeed, matching revealed unconscious discrimination more fine-grained than conscious discrimination, casting doubt on whether consciousness enhances discriminative ability, and hence perceptual utility.

Also, Arnaud Beauny and colleagues (2020), using very brief presentations in the μsec range, found a threshold at which subjects could consciously detect stimuli, though they couldn't also consciously identify them. But using forced choice, subjects could identify those stimuli well above chance. Conscious perception enabled detection, but unconscious qualitative aspects were needed for identification.

In all these cases the qualitative states are conscious, but they have qualitative aspects that register only unconsciously. So mental qualitative character is not intrinsically conscious, and consciousness doesn't fully reveal the qualitative nature even of those qualitative states that are conscious.

These examples are all incontrovertibly qualitative: shapes of letters, fine-grained shades of color, and qualitative properties needed to identify stimuli. We must accept that qualitative states can be conscious in respect of some of their qualitative mental properties but not others. This doubtless also occurs outside of experimental settings, as when we consciously see something, though there are qualitative aspects we don't consciously see but which nonetheless affect our behavior.

All this is readily accommodated by the transitivity principle (TP) introduced in section IV. A state is conscious only if one is aware of being in that state. But a state can be conscious in respect of some of its mental properties and not others, and the state will be conscious in respect of just those properties one is aware of the state as having. TP applies to states in respect of specific mental aspects.

There is also compelling evidence of qualitative character in states that are altogether unconscious. Liam Norman and colleagues (2014) found that a masked stimulus primed the mask if they match in surface color, but not in spectral reflectance (on this see also Kentridge 2017). Though the mask was consciously perceived, the masked stimulus was perceived only unconsciously. So a conscious qualitative state was matched with another state that was altogether unconscious. Since the matching occurred in respect of qualitative character, it required qualitative character that occurs unconsciously.

In so-called postdictive effects (Herzog et al 2020; Michel and Doerig 2022) perceptions of two successive stimuli combine to generate a subsequent subjective awareness with qualitative character that differs from the content properties of either of the earlier perceptions. But since the subsequent subjective awareness varies with the mental properties of the two earlier perceptions, it's natural to assume that each earlier perception had some qualitative character to begin with, though it was not conscious.

One could avoid unconscious mental qualities by rejecting the generally accepted view that perceiving occurs unconsciously. Thus Ian Phillips (2018) denies unconscious perception in part because whether one reports seeing is affected by the internal criterion that determines for one whether there is conscious seeing. If one's criterion, conceived in signal-detection terms, is too demanding, one may inaccurately report not seeing a stimulus. Arbitrary factors can affect one's criterion. So Phillips argues that even when subjects report not seeing, if task performance is above chance the seeing is conscious. And if performance is not above chance, there's no evidence for any seeing at all. There is then no room for unconscious seeing.

But consciousness is literally constituted by subjective awareness. So whatever affects an individual's subjective impression of seeing, however arbitrarily, typically also affects what is and is not conscious for that individual at that time. Indeed, it's wholly unclear what consciousness could be apart from that subjective impression of awareness. Noise can affect one's ability to report. But that aside, a report of no subjective impression of seeing reliably establishes the absence of conscious seeing. Above-chance performance then shows that seeing did occur, despite its not being conscious.

Others, though not denying unconscious perceiving altogether, make it unreasonably difficult to establish. Megan Peters and Hakwan Lau (2015; see also Peters 2017) take conscious amodal hunches to constitute perceiving, and thus to be conscious seeing. In addition, they require discrimination, not mere detection, for perception to occur, ruling out much unconscious perceiving from the other side. And they reject inference to the best explanation from priming results. But it's not clear what warrants these strictures apart from an independently unsupported skepticism about unconscious perceiving,

It can happen that stimuli A and B are indistinguishable and also B and C, but A and C are readily distinguishable. This is the so-called phenomenal sorites. But how can

that happen if, as on QS theory, we individuate mental qualities by perceptual indistinguishability?

Mental qualities are individuated by appeal to their role in fine-grained perceptual discrimination, and which mental quality a particular stimulus elicits can depend on which other stimuli it's being discriminated from. So B can elicit one mental quality when compared with A and a distinct one when compared with C. Perceptual indistinguishability is not transitive, as many have noted (Goodman 1951; Morrison 2015). QS theory nicely explains why.

VI. Misrepresentation and the Origin of Consciousness

On a HO theory, the property of a state's being conscious is distinct from the state, and so independent of that state's other mental properties. So a HO theory allows that one could be subjectively aware of being in a state with mental properties different from any state that one is actually in. Subjective awareness would then misrepresent the state one is in, even if sometimes only slightly.

It's likely that no other currently popular theory accommodates this possibility. On a global-workspace theory, for example, a state is conscious if its content is available for suitable downstream processing. A state with one content might somehow result in the downstream availability of some other content. But it's not clear why on such a theory that would make that state conscious. And that aside, such availability is not itself a way of representing a state. So global-workspace theory seems not to allow subjective awareness to misrepresent what state one is actually in. But HO theories do.

Advocates of intrinsic consciousness maintain that subjective awareness cannot misrepresent. So they insist that accommodating that possibility by itself decisively undermines HO theories. But that's plainly not so. If one were convinced that such HO misrepresentation cannot occur, one could still endorse a HO theory and simply stipulate that such misrepresentation isn't possible.

It's in any case plain that subjective awareness does sometimes misrepresent what mental state one is in. A dramatic case occurs in change blindness. Even when subjects don't consciously see a change, priming effects and forced-choice guessing can show that the change that was consciously missed was nonetheless seen unconsciously (Fernandez-Duque & Thornton, 2000; replicated by Laloyaux et al 2006). Subjects unconsciously see the change even though subjective awareness continues to register the pre-change stimulus (Odegaard et al, 2022, slides 16-20). Since visual cortex will often register the post-change stimulus, subjective awareness misrepresents the visual state the subject is in.

There are other, more esoteric demonstrations of misrepresentation by subjective awareness (Rosenthal 2012). In chromatic flicker fusion, isoluminant colors flickering at high frequency fuse into a different nonflickering conscious color. But fMRI reveals

cortical flickering, which consciousness misrepresents (Jiang et al 2007). Also, in the postdictive effects described above, since perceptions of two successive stimuli combine to generate a subjective awareness with qualitative character distinct from the content properties of either perception, that subjective awareness misrepresents both perceptions. Finally, we seem subjectively to see colors in the visual periphery that peripheral retinal cells arguably can't support (e.g., Knotts et al 2019).

Since subjective awareness does sometimes misrepresent what state one is in, that HO theories accommodate such misrepresentation is not the disadvantage intrinsicists imagine, but rather compelling support, all the more if no other theory accommodates it.

Why then do advocates of intrinsic consciousness insist that subjective awareness cannot misrepresent? Intrinsicism rejects the possibility of reaching beyond the subjective appearances to some underlying mental reality that's responsible for those appearances. So its advocates see no underlying mental reality that the subjective appearances could misrepresent. If the subjective appearances exhaust the mental reality of consciousness, there is no room for misrepresentation. So we must assess this denial of mental reality independent of the subjective appearances.

Consciousness would occur automatically with each qualitative state if it were intrinsic to qualitative character, but not if it consists in distinct HOAs. So we must explain how those HOAs come to occur. Since mental qualities determine perceptual functioning independently of consciousness, HOAs aren't needed for that.

HOAs likely arise in the earliest stages of infant perceptual development, in which qualitative inputs must constantly be calibrated across modalities. The infant must coordinate visual and tactile qualities pertaining to size and shape, visual and auditory qualities pertaining to location, and visual qualities with olfactory and gustatory qualities and with pleasurable and unpleasant qualities.

Because this early cross-modal calibration occurs constantly, the infant will come fairly often to wonder whether qualities from two modalities result from the same external objects. So the infant will come to be aware of those states, and those awarenesses are HOAs. HOAs pertaining to qualitative states initially result from early cross-modal calibration, and their continued frequent occurrence becomes entrenched. We can accordingly predict that qualitative states are conscious in various creatures to the degree that successful perception requires significant cross-modal calibration. HOAs pertaining to purely conceptual states, such as thoughts and desires, require a different, more complicated explanation (Rosenthal, 2005, ch. 10).

Since cross-modal calibration has great utility, one might conclude that a state's being conscious itself confers some distinctive utility. But the utility here attaches just to the calibration, not also to the HOAs. It's unlikely that a state's being conscious does confer significant additional utility (Rosenthal 2008, 2012). A methodological consequence is that we cannot learn about consciousness by investigating its utility.

VII. Appearance and Reality

The claim that qualitative character is intrinsically conscious goes hand in hand with the denial of any distinction between appearance and reality as applied to consciousness. Consciousness is indeed the way our mental lives appear to us subjectively. But intrinsicalists extrapolate from that to the extreme claim that subjective appearance is all there is to consciousness, denying any underlying mental reality of consciousness. Thus Nagel (1974, 444): "The idea of moving from appearance to reality seems to make no sense" for conscious experiences.

If the subjective appearances did exhaust the mental reality of consciousness, we could not theorize about that mental reality or investigate it scientifically. First-person phenomenology would then be the only way to learn about the mental reality of consciousness, which would be closed off behind a wall of subjectivity, separated from anything objective, a conclusion endorsed by many who find Nagel's picture compelling.

But there are, in addition to the subjective appearances, objective mental realities that underlie and implement those subjective appearances. For one thing, there are the qualitative and conceptual states that subjectively appear to us, and which we can theorize about independently of consciousness (Rosenthal 1986, 2010). In addition, there are the mental occurrences that implement the subjective appearances themselves (Rosenthal 2022).

For appearances in general, there is always some process or mechanism that generates and implements the appearances, a process or mechanism that's distinct from the appearances themselves. Intrinsicalists might urge that the appearances of consciousness seem subjectively to exhaust its mental reality. But for appearances in general, we never rely solely on the appearances themselves to learn what underlying reality generates and implements them. Why think we can do so for the appearances of consciousness?

Intrinsicalists might insist that the underlying reality for the appearances of consciousness is exclusively neurological, and deny any underlying reality at the distinctively mental level. But consciousness and its appearances are mental phenomena. So it strains credulity that there is no distinctively mental process that generates and implements them. And there's no reason to think that apart from intrinsicalism, which itself lacks independent support.

The appearances of consciousness do not exhaust its mental reality. Still, we can explain why so many think they do, an explanation that again rests on the relocation story sketched in section III. On that relocation story, seeing all of nature as mathematically describable leads to denying that colors and other qualitative properties, as conceived by common sense, can be physical properties. So they get relocated as properties of the mind. But commonsense physical qualities resist mathematical

treatment only if we conceive of them as we are consciously aware of them. So when those properties are relocated, the resulting mental properties also have consciousness built in.

As argued in section III, the relocation story explains the apparent appeal of the otherwise unsupported claim that mental qualities are intrinsically conscious. But it also helps with the issue about mental appearance and reality. If we conceive of the relocated properties as we're conscious of them, there will be nothing to the nature of those properties, so conceived, that we're not conscious of. So one will then see the subjective appearances as exhausting the mental reality of those relocated properties. The relocation story explains why so many are tempted to hold that the subjective appearances of consciousness exhaust its mental reality.

But as stressed earlier, the relocation picture is wholly unwarranted. Consciousness is not built into the properties we perceive. So relocation is not needed for the perception of those properties to square with mathematical science. Nor is relocation needed to explain the subjective appearances that occur when we consciously perceive those properties. All we need is the objective mental reality of consciousness: mental qualities construed independently of consciousness and the higher-order states that implement the subjective appearances (Rosenthal 2022).

Superficial appearances were once seen as determining essences. But by themselves appearances are never enough to fix the relevant reality. This is now universally accepted for understanding all aspects of nature, with qualitative consciousness a conspicuous holdout for some. Accepting this for qualitative consciousness as well is crucial for completing the paradigm shift of modern science.

REFERENCES

Aristotle (2016), de Anima, translated by Christopher Shields, Oxford: Clarendon Press.

Beauny, Arnaud, Adélaïde de Heering, Santiago Muñoz Moldes, Jean-Rémy Martin, Albert de Beir, and Axel Cleeremans (2020), "Unconscious Categorization of Submillisecond Complex Images," PLOS ONE, 15, 8, doi 10.1371/journal.pone.0236467 (August 12): 1-23

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 (1995), "On a Confusion about a Function of Consciousness," The Behavioral and Brain Sciences, 18, 2 (June): 227–247.

Block, Ned (2007), "Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience," Behavioral and Brain Sciences, 30, 5-6 (December 2007): 481-499.

Block, Ned (2015), "The Puzzle of Perceptual Precision," in Open MIND, ed. Thomas Metzinger and Jennifer M. Windt, 1-32, Frankfurt am Main: MIND Group, doi: 10.15502/9783958570726.

Chalmers, David J. (1995), "Facing up to the Problem of Consciousness," Journal of Consciousness Studies, 2 3 (March): 200-219.

Cova, Florian, Maxence Gaillard, and François Kammerer (2021), "Is the Phenomenological Overflow Argument Really Supported by Subjective Reports?", Mind & Language, 36, 3 (June): 422-450.

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.

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.

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

Goodman, Nelson (1951), The Structure of Appearance, Cambridge, Massachusetts: Harvard University Press.

Harman, Gilbert (1990), "The Intrinsic Quality of Experience," Philosophical Perspectives 4: 31-52.

Herzog, Michael H, Leila Drissi-Daoudi, and Adrien Doerig (2020), "All in Good Time: Long-Lasting Postdictive Effects Reveal Discrete Perception," Trends in Cognitive Sciences, 24, 10 (October): 826-837.

Irvine, Elizabeth (2014), "Problems and Possibilities for Empirically Informed Philosophy of Mind," in New Waves in Philosophy of Mind, ed. Mark Sprevak and Jesper Kallestrup, London: Palgrave Macmillan, pp. 185-207.

Jiang, Yi, Ke Zhou, and Sheng He (2007), "Human Visual Cortex Responds to Invisible Chromatic Flicker," Nature Neuroscience 10, 5 (May): 657-662.

Kammerer, François (2022), "How Can You Be So Sure? Illusionism and the Obviousness of Phenomenal Consciousness," Philosophical Studies, 179, 9 (September): 2845-2867.

Kentridge, Robert W. (2017), "Sensation and Unconscious Perception," in "Does Unconscious Perception Really Exist? Continuing the ASSC20 Debate," Neuroscience of Consciousness, 3, 1: nix015, pp. 3-5.

Knotts, J. D., Brian Odegaard, Hakwan Lau, and David Rosenthal (2019), "Subjective Inflation: Phenomenology's Get-Rich-Quick Scheme," Current Opinion in Psychology, 29 (October): 49-55.

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

Laloyaux, 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.

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

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

Morrison, John (2015), "Anti-Atomism about Color Representation," Noûs, 49, 1 (March): 94-122.

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.

Norman, Liam J., Kathleen Akins, Charles A. Heywood, and Robert W. Kentridge (2014), "Color Constancy for an Unseen Surface," Current Biology, 24, 23 (December 1): 2822-2826.

Odegaard, Brian, Alan Lee, Isaac Lee, Addison Sans, Leo Ng, Andrew Haun, Dana Chesney, David Rosenthal, and Francis Fallon (2022), "Using Saccade-Induced Change Blindness to Refine Existing Theories of Conscious Perception," PowerPoint presentation at ASSC 25, Amsterdam, July 13, part of the Templeton World Charity Foundation project, Replication and Extension of Crucial John Grimes Experiment: Change Detection during Saccades, grant number TWCF0445, <https://www.davidrosenthal.org/Odegaard-et-al-ASSC-2022-CDDS.pdf>.

Peters, Megan A. K. (2017), "Practical and Theoretical Considerations in Seeking the Neural Correlates of Consciousness," in "Does Unconscious Perception Really Exist? Continuing the ASSC20 Debate," Neuroscience of Consciousness, 3, 1: nix015, pp. 1-3.

- Peters, Megan A. K. and Hakwan Lau, (2015) "Human Observers Have Optimal Introspective Access to Perceptual Processes Even for Visually Masked Stimuli," eLife, 4, e09651.
- Phillips, Ian (2018), "Unconscious Perception Reconsidered," Analytic Philosophy, 59, 4 (December): 471–514.
- 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 (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 (2012), "Higher-order Awareness, Misrepresentation, and Function," Philosophical Transactions of the Royal Society B: Biological Sciences, 367, 1594 (May 19): 1424-1438.
- Rosenthal, David (2016), "Quality Spaces, Relocation, and Grain," in Sellars and His Legacy, ed. James R. O'Shea, Oxford: Oxford University Press, pp. 149-185.
- Rosenthal, David (2019), "Consciousness and Confidence," Neuropsychologia 128 (May): 255-265.
- Rosenthal, David (2022), "Mental Appearance and Mental Reality," in Qualitative Consciousness: Themes from the Philosophy of David Rosenthal, ed. Josh Weisberg, Cambridge: Cambridge University Press, pp. 243-271.
- Sperling, George (1960), "The Information Available in Brief Visual Presentations," Psychological Monographs: General and Applied, 74, 11 (Whole No. 498): 1-29.