Consciousness and the Representational Relation: Why Experience Can't Be Objectified

Thomas W. Clark

Institute for Behavioral Health, Schneider Institutes for Health Policy, Heller School for Social Policy and Management, Brandeis University

Keywords: consciousness, qualia, physicalism, representation, phenomenal experience, objectivity, content

Please note: This is a pre-publication draft for the Frontiers series on neo-naturalist approaches to consciousness that may change given reviewer feedback. Comments welcome via the contact page at Naturalism.org.

Abstract

Humans are physical beings that are also conscious, but physicalism struggles to locate consciousness in the natural world described by science. The world appears to conscious creatures in terms of experienced sensory qualities – qualia – but science doesn't find qualities in that world, only physical objects and properties. I argue that the failure to locate consciousness in spacetime is a function of our necessarily *representational relation* to reality as knowers: we won't discover the terms in which reality is represented by us in the world as it appears in those terms. Other instances of this failure are uncontroversial: we don't expect to find concepts, numbers or propositions as locatable entities or properties in the world they participate in describing. Rather, we understand that they are mind-dependent representational terms or tools – basic elements of cognitive content – that we deploy in characterizing reality. Philosophers sometimes suggest that mathematical entities have objective ontological status: they exist in a mind-independent abstract Platonic realm. Likewise, physicalists who are realists about consciousness generally assume its objectivity: experience must be something identical with physical processes or properties, perhaps the intrinsic nature of the physical, or perhaps some micro-physical, neural, or emergent property. I argue that this assumption wrongly reifies consciousness; it expects to find qualitative representational content in the physical world as characterized using such content. Instead, we should grant that conscious experience constitutes a mind-dependent, subjective, *representational* reality for cognitive systems such as ourselves, and that the physical world given in experience and in science is a *represented* objective reality. The former, since it exists only for conscious subjects, won't be found as an entity in the latter. I suggest that naturalistic approaches to explaining consciousness should acknowledge the representational relation and the non-objectivity of experience, and be constrained by evidence that consciousness accompanies certain sorts of behavior-controlling representational functions carried out by complex, physically-instantiated mind-systems. I evaluate a variety of current hypotheses about consciousness on that basis, and suggest that a science of representation could help explain why, perhaps as matter of representational necessity, experience arises as a natural but not objectively discoverable phenomenon.

1. Introduction

On a naturalistic view of ourselves, we are entirely physical beings who are also conscious, but thus far there is no consensus on the nature of consciousness. For the naturalist (as opposed to supernaturalist), the explanation of phenomenal experience would ideally be consistent with the current scientific understanding of the world, take into account the relevant research, not appeal to empirically unmotivated claims about the nature of reality, and most definitely not traffic in ectoplasmic entities such as souls or spirits.

The central difficulty, of course, is that the defining feature of conscious experience – the qualitative "what it is like" or phenomenal character of tasting a mango, seeing a red rose, or dreaming about a blue lake – is not available to intersubjective observation or measurement (Gamez, 2014). If it were, there would be no problem of consciousness, nor of other minds, for instance whether fish feel pain. Pain would be out there in public, and we'd know that they either do or don't suffer when hooked. But pain isn't public, unlike whatever its neural or otherwise physical correlates might be, whether in fish, fowl, or us. And so it is with all experiences: they are only available to, only exist for, individual conscious subjects. If, as eliminativists and illusionists about consciousness claim (Dennett, 2016; Frankish, 2016), there are no experiences but only *seem* to be, the problem of explaining these morally significant and apparently private seemings would replace the problem of consciousness, but I suspect would prove as perplexing.[1]

Despite its subjective nature, many naturalistic approaches to explaining (not eliminating) consciousness in effect hypothesize that experience is "out there" in some sense, that it's an objective, physically-embodied phenomenon. It's just that we don't yet know precisely the physical states, processes or properties with which it is identical. Optimistic physicalists about consciousness, perhaps the majority among naturalists working on the problem, suppose that given a few more decades of work on identifying the physical and functional correlates of experience, consciousness will have found its place in the material world as described by science, in particular, cognitive neuroscience as it studies the mammalian brain.

The hope for such an outcome is understandable since consciousness presents an embarrassment for physicalism and, since most naturalists are physicalists, for naturalism. Since there's no canonical philo-scientific[2] account of how consciousness fits into the natural world, anti-naturalists will sometimes seize on the "hard problem" of consciousness as a defeater for naturalism (Moreland, 2009, 16-40). Absent such an account, appeals to something extra-physical, god-like, or occult are more likely to gain traction.

The basic problem of consciousness for physicalists is how to objectify something that by its nature seems essentially resistant to objectification (Skrbina, 2009b; Howell, 2013)[3]. Objectification establishes a phenomenon as resident in spacetime, something potentially detectable by any suitably placed observer, perhaps aided by instruments and measuring devices. Theories aiming to objectify consciousness range from reductive identity theories that equate

[1] As Frankish (2016) himself acknowledges, the problem of explaining the illusion of phenomenology replaces the "hard problem" of explaining consciousness.

[2] "Philo-scientific" is my neologism to designate collaboration between philosophers and scientists, for example when investigating the nature of consciousness, the interpretation of quantum mechanics, why there's something rather than nothing, and the nature and scope of science.

[3] "Given mechanistic physicalism, mind must be shoe-horned into the world. But it refuses to fit." Skrbina (2009b), 364.

experiences with physically-realized states or functions of cognitive systems; non-reductive, property dualist theories in which mental properties, including consciousness, supervene on material or functional states; panpsychist theories that hold phenomenology to be a fundamental property of matter; theories involving quantum or magnetic fields as instantiators of consciousness; and radically enactivist theories that identify conscious states with ordinary physical objects (Manzotti, 2011).

Should one of these theories win the day, future scientists might confidently point to certain neural or physical goings-on and say "That just *is* your pain. We can see it right there." As things stand now, the obvious rejoinder is that the essential nature of pain is to be felt or undergone. Pain, euphoria, and other sensations, emotions and thoughts can't literally be seen or observed, that is, measured, detected, or otherwise perceived as entities; only their physical correlates can be observed (Clark, 2005; Goff, 2016, 94-5). We can call this the *privacy constraint* on consciousness. But perhaps the final theory, maybe some sort of eliminativism or illusionism, will show the privacy constraint to be fatally misguided. Consciousness will be understood to be intersubjectively available just as are brains: there really are no categorically private phenomenal episodes, there only *seem* to be (Dennett, 2016). Alternatively, if phenomenal states aren't eliminated by the final theory, perhaps everything about conscious experience, including its qualitative character, will be shown identical to certain observable physical processes or properties. Consciousness will not have been eliminated, but it will have been publicized and objectified by the final theory.

Although we can't conclusively discount these possibilities, I'd suggest we not hold our collective breath. The existential privacy of an experience, that it exists only for the mind undergoing it, isn't likely going away. My prediction is that (fortunately) you won't ever be in a position to see my pain, or know its qualitative character in physically specifiable terms. Instead, I suggest we ask why conscious experience isn't objectifiable even though as conscious creatures we are physically objective. The answer I propose has to do with what I'll call the *representational relation*: the world is only *known* by systems using content-bearing representations. We are one species, one type, of such systems. Conscious experience is a type of representational content – *qualitative, phenomenal* content – and the world appears to each of us as a conscious subject in terms of that content. The reason we don't find conscious experience in the world, the reason we can't objectify it, is because as a rule we don't and won't find representational content in the world as modeled by it. We only find the physical objects and phenomena characterized *in terms of* such content, including the physically-instantiated content vehicles. Theories that suppose we can objectify consciousness, putting it in the public domain, are thus barking up the wrong tree.

Still, the puzzle remains of why, as current evidence strongly suggests, only certain sorts of physically and functionally-specified cognitive *systems*, running certain sorts of representational processes in service to behavior control and system integrity, end up hosting conscious experience (Clark, 2005, 52-5). We can perhaps narrow down the naturalistic explanatory possibilities in light of the realization that we don't, and won't, discover consciousness to be among the occupants of the world as it appears to us, either in experience or as described by science. This suggests that naturalism may not be equivalent to, or require the truth of, physicalism, but that a broader representationalist approach, including a mature science of representation, might work to naturalize consciousness.

2. The representational relation

It is a commonplace that as knowers we deploy various sorts of representations in negotiating our contact with the world. We use *concepts* to categorize and characterize objects and events, and *numbers and other quantitative abstractions* to count them and specify their physical attributes and relationships. Our conceptual and quantitative descriptions thus participate in modeling reality, whether in everyday life (learning a new job) or in science (investigating the nature of dark matter). The world is represented by us conceptually and quantitatively in the "manifest image" of ordinary human discourse and the "scientific image" of physical theory (Sellars, 1962). Both of these are indeed *images*, models, not the unmediated, mind-independent world itself, which presumably exists whether or not it's being modeled. That we are in a representational relation to reality seems an unavoidable condition of our being limited, situated creatures with particular perspectives on the world, whether individual or collective. This means that reality – the world – only appears to knowers such as ourselves in terms of our representations; it never appears undressed, so to speak, but always arrayed in perspectivally conditioned models.[4] Still, the models we humans deploy generally include the very plausible assumption of objective physical realism: there exists a mind-independent world of material entities and processes, of phenomena that appear to us as having spatio-temporal properties as given both in science and everyday experience. Among those phenomena are composite, complex, and integrated systems that constitute minds – mind-systems – at least some of which, like ourselves, are conscious.

On the face of it (always a good place to start), consciousness too seems to mediate our contact with reality. All of what we know and interact with – persons, objects, events, situations, even our own body – appears to each of us in terms of phenomenal (that is, qualitatively rendered) experiences. There's usually a non-coincidental and behaviorally crucial correspondence between the structure of our waking experience and what's the case in our immediate surroundings, a correspondence underwritten by causal interaction with the environment via our sensory-perceptual modalities; we can thus fairly say that our experience is *informative* (Chalmers, 2006). Indeed, most of the time we unreflectively take the world as given in experience simply to be the physical world as it is in itself, directly presented to us. The colors, textures, smells, tastes and acoustic characteristics of objects we perceive are taken as *their* properties, not our subjective take on them. But we can infer, on the basis of dreams, hallucinations, afterimages, optical illusions and other sensory and perceptual artefacts, that experience is a selective and fallible individual-level model of what's outside the head. Consciousness looks to be a *representation*, not a window on the unrepresented world itself. Subtract your experience, as for instance in a dreamless sleep, and you subtract the world as represented in experience, as well as the phenomenal subject to whom the experienced world is presented.[5]

[4] I call this thesis about the representational relation "epistemic perspectivalism" (Clark, 2010); it seems a close cousin to the "model-dependent realism" proposed by Hawking and Mlodinow (2010).

[5] The sense of being a subject to whom the world is given in experience is itself part of the overall experiential gestalt, see Metzinger 2003, chapter 6 and Clark 2005, 49-52.

This shouldn't be construed as saying that in having experience we somehow see the experiential model instead of the world (Rockwell, 2013);[6] that is, we shouldn't suppose we're in an observational relation to consciousness (Clark, 2005). We can avert our gaze and otherwise perceptually distance ourselves from physical objects, but cannot divorce ourselves from the experience in terms of which they appear and disappear for us. The representational relation of consciousness to the world is for us to subjectively *consist of* an experiential world-model or representational interface (Revonsuo, 2015), what Thomas Metzinger calls an "ego-tunnel" (Metzinger, 2009) that is modulated and constrained by our direct, physical contact with the world itself. We have developing theories of such (sub-personal) contact, e.g., predictive coding (Seth et al. 2012; Clark, 2013): impinging stimuli activate sensory modalities that feed into neural networks whose job it is to inform the brain's current multi-modal world-model, thus minimizing prediction errors in service to behavior control. The continually updated mappings and covariances between the world and brain – the representational relation – allow for successful action and system maintenance, given the nature and needs of the organism (Kanwisher, 2001; Dehaene and Changeux, 2011; Sterling and Laughlin, 2015).

The representational processes associated with *consciousness* can be identified by contrasting the neural networks activated when performing tasks only possible when conscious (e.g., complex learning, planning) to those networks subserving behavior that can be handled unconsciously (e.g., habitual or automatized behaviors) (Baars, 1997; Clark, 2005, 52-55). The format of the representational elements, neural or otherwise, of course need not resemble what in the world (including the body) they represent; they only need to reliably co-vary with the objects and events our survival depends on knowing – the world's behaviorally relevant structure. Nothing in any of the foregoing suggests that representation can't ultimately be cashed out as a fully natural phenomenon.

What we don't yet have, however, is a canonical, accepted account of *why* conscious experience is apparently associated with only some (not all) of these system-regulatory, behavior-controlling, and informationally rich processes. Nevertheless, we can see that experience, since it closely correlates with neural networks that carry information about the world, *itself* tracks the world (including the system itself), at least when we're awake and in perceptual contact with our body and environment; this is why we're justified in taking our waking experience very seriously as a guide to behavior. Still, we can grasp that consciousness is a world-*model* since the full blown technicolor experience of 3-D reality can be present for us when we're asleep and having dreams, especially lucid dreams in which we realize (have the explicit conscious understanding and thought) that we are asleep *and* conscious (Blackmore, 2006, 137-48; Metzinger, 2009, 139-148; Revonsuo, 2015). When lucidly dreaming, the representational nature of experience – that it's an interface – is a directly presented fact about our current experience. We realize during the dream (and after awakening) that we don't need the actual world as fed to us in online waking perception for us to experience a world, albeit not one that necessarily corresponds to that of waking experience. The brain just needs to be doing the right thing for a world, and the subject who experiences it, to appear in terms of experience. This would seem to support a system-internal (internalist) view of consciousness, although externalists and enactivists maintain that

[6] "I don't perceive the world in front of me because my retina makes a copy of the world and presents it to my brain." Rockwell (2013), 234.

dreams are not the evidence they seem (e.g., Noe, 2010), and some claim that consciousness is in no sense representational (Travis, 2004; Manzotti, 2011; Hutto and Myin, 2012).

That experience carries information about the world is not necessarily to say that one's (waking) phenomenal world-model plays a causal role in 3rd person, scientific accounts of behavior (Oakley and Halligan, 2017). Such accounts can only traffic in intersubjectively observable entities, structures and processes – "observables," for short. Since (as I will argue below) phenomenal content – conscious experience – isn't an observable, it is logically barred from playing a role in such explanations (Clark, 2013a), which will involve various neuro-muscular goings-on and ascriptions of intentional content (beliefs, desires, etc.) based on overt behavior (Dennett, 1987). However, this doesn't make consciousness *epiphenomenal* with respect to behavior, since the charge of epiphenomenalism can only fairly be brought against entities in the physical world that might have played a causal role, but do not (Clark, 2013b). The human appendix, on the (now contested) assumption it has no significant biological function, would fairly be described as epiphenomenal with respect to maintaining one's bodily health. But since conscious experience doesn't appear in the physical, objective world in which behavior occurs, it is not in a position to be epiphenomenal with respect to behavior.

3. The reality of phenomenal content

Despite the apparent reality of consciousness as constituting our subjective world-model, one approach to solving the problem of how it fits *into* the world as thus modeled – what we justifiably think of as the spatio-temporal physical world – is simply to deny the existence of phenomenal experience itself. A good example is Daniel Dennett's most recent denial (Dennett, 2016; Dennett, 2017, 335-70), in which he uses the experience of afterimages to (as he would put it) "pump the intuition" that phenomenology is illusory. Having induced in us an afterimage of a red stripe (by looking at an image of a green- and white-striped American flag), he challenges us to locate the afterimage: where is it? Well, we're not going to find it poking around in the brain; it certainly isn't out there in the world; and according to Dennett it doesn't exist in experience either, even though we might think so. All that *really* exists is a complex set of neural spike trains that carries the representational content *red*, but there is no quality of red that we experience when having the afterimage. Dennett agrees that qualities such as red are real, but only as properties of intersubjectively observable, locatable physical objects, e.g., an American flag that has red stripes (Dennett 2016, 71), or apples, which can really be red.

But of course the afterimage is really red too (we all agree about its color) even though it can't be found anywhere; that it appears red is thus not an illusion. Even though the afterimage itself is illusory, the *red* is indeed real, just not locatable. This means we can't reasonably discount the reality of the subjective qualitative content that individuates our experiences, whether these experiences involve afterimages, hallucinations, dreams, or physical objects. Dennett and other eliminativists and illusionists (e.g., Churchland, 1988, 43-49; Frankish, 2016) claim that our intuition about the reality of such content involves a big mistake: there is no such content, there only *seems* to be. But since we uncontroversially characterize real physical objects in terms of such content, this approach won't wash, at least not obviously. Calling phenomenal content into question automatically calls into question the apparent qualities, e.g., colors, of physical objects, and we probably don't want to say that it's an illusion that a *real physical apple* appears red. That the afterimage and its qualitative character can't be located anywhere simply highlights the problem of how to incorporate phenomenal experience into a naturalistic worldview.

Looking at a red apple over there, we naturally want to say the red *too* is over there. But is it? What Dennett's afterimage example actually shows, I suggest, is that the phenomenal content in terms of which we characterize physical objects in our experience, unlike the objects themselves, *isn't anywhere*. Of course, we commonsensically take the redness of the apple as the apple's own mind-independent objective physical property, and some philosophers (color realists or objectivists) argue that colors are indeed such properties, for instance relational properties such as surface reflectances interacting with impinging light (e.g., Tye, 2000, 145-170). But as the afterimage example suggests, as much as colors are used to pick out and characterize physical objects, and thus are their *represented* properties (properties we represent them as having), they are in the first instance unlocatable experiential contents. There really is a locatable object with objectively specifiable properties (mass, surface reflectance, etc.), but it appears to us in conscious experience in terms of qualitative content that isn't itself anywhere.

A good naturalist-physicalist, Dennett wants to banish what he takes to be the spooky, anti-naturalist dualism of unlocatable, thus apparently non-physical, phenomenology. But my suggestion is that we need not be irrealists about the qualitative content of conscious experience to avoid anti-naturalism. If we consider the nature of representation, as I will below, the fact that we can't locate such (real) content in the world it helps to characterize actually makes sense. What we *should* be skeptical of are qualia as Dennett defines them: internal mental objects or properties that partake of an immaterial medium, something like the sense-data of yore, to which we have privileged epistemic access. If we are not, as noted above (and by Dennett 2017, 363-4), in an observational, epistemic relation to consciousness, then the qualitative contents of experience aren't things we observe via introspection – they aren't things at all, least of all mental substances that exist objectively alongside neural goings-on.

This means, *pace* David Chalmers, that we don't have a perspective on, or view of, experienced qualities (e.g., the sweetness of sugar as I experience it) such that they constitute a set of first-person facts or data that science can't capture (Chalmers, 2004); rather we have individual perspectives *on the world* (Clark, 2005). We are in an epistemic relation to, and thus know facts about objects (e.g., that sugar is sweet), not the basic, irreducible qualitative terms in which objects are characterized in consciousness. The terms themselves – sweetness, redness, painfulness – are not themselves facts assessible for truth or falsity (Metzinger, 2003, 170). After all, the representational relation must bottom out somewhere: we only need reliable co-variance between the model and the world to have actionable knowledge about the world. What we don't need are private factual truths about the basic qualitative elements of phenomenal experience itself, which would only threaten a representational, epistemic regress. Following common practice, I'll refer to the various phenomenal qualities that comprise experience as "qualia," hoping to have headed off any presumption of substance dualism. But I also hope to have motivated *qualia realism* (Goff, 2016): the physical world only reliably appears real to us in all its extended, massive, colorful glory in terms of real phenomenal content.

4. Objectification

The *prima facie* privacy of conscious experience, that my pain exists only for me, contrasts with the public accessibility and observability of physical objects. Short of radical skepticism about the external world, there's little dispute that objects that appear to me (the apple) can also appear to you, were you awake and suitably situated. To confirm we've observed the same object, we compare descriptions (red, round, on the kitchen table in a blue bowl), and if the descriptions

match sufficiently, we can agree that what we've seen (or touched, heard, tasted or smelled) is mind-independently real, a physical object resident in spacetime. Each of us consciously encounters the apple via our subjective experience of seeing it, the qualitative elements of which, e.g., red, are neither true nor false. But the intersubjective agreement of qualitative descriptions supports the judgment that the apple is *truly there*, not just a matter of my (private) experience, or yours. The apple therefore counts as a physical object in good standing, not something we'd suspect of being illusory or entirely mind-dependent, even though its appearance to each of us may differ in some respects. To objectify something – to represent it as having mind-independent existential status as a physical phenomenon – is thus to show it as actually or potentially intersubjectively accessible (observable, perceivable, measurable).

Before science steps in, the physical objectivity of things like apples gets certified by intersubjective consensus on descriptions couched largely in the qualitative terms of ordinary everyday sensory experience: we generally agree about the apple's redness, roundness, surface texture, firmness, and perhaps smell and taste, all properties we commonsensically attribute to it (even though, as noted above, they are *represented* properties that participate in our phenomenal world-model). The apple is concretely, tangibly present: it has physical substance – "stuff" – as given in our experience. The apple thus achieves its objective, intersubjectively available status as what I'll call a *folk-physical* object. After science gets involved, however, the descriptions include quantitative parameters that make no mention of experienced qualities: the pH level of the apple's acidity, the relative concentration of certain esters, its rest mass, reflectance properties, etc. Eventually, the scientific description of the apple as a physical object might leave behind any qualitatively rendered, experienced component at all – it has become a *science-physical* object.

We can think of this transition, from descriptions of objects in terms of qualities (folk-physical) to entirely quantitative descriptions (science-physical), as a matter of increasing objectification. The phenomenal qualities in folk descriptions of objects can of course vary from individual to individual, depending on their perceptual capacities. Color-blind individuals may not be able to reliably discriminate a red from a green apple, and if you've lost your sense of smell, cider will taste merely sweet, lacking its characteristic appley-ness that would allow you to discriminate it from orange juice (texture and mouth feel aside). The descriptions afforded by science aren't prone to such variability since the objects they pick out are reliably identified as having quantifiable properties according to reproducible measurements that in principle anyone or any*thing* could carry out, including aliens and AIs. Because we ordinarily think of reality as having its own mind-independent nature, descriptions that leave behind experienced, mind-dependent qualities in favor of quantifiable characteristics are in that sense more objective.

We should not forget, however, that the maximally objective, quantified specification of the world – the spatio-temporal world as described by science – is still a representationally encoded *model*, not the unrepresented world itself. The unrepresented self-nature of reality, what we intuitively believe exists independently of our representations of it, is still at an epistemic remove, even though we ourselves participate in that reality. At its best, what the scientific model can afford us is a predictively successful and explanatorily consistent *structurally isomorphic rendition of reality*, one couched in terms of physical parameters and constants (Ladyman and Ross, 2007). Unrepresented self-natures are necessarily left out of the maximally objective picture of the world since the epistemic interface between knowers and known is always in place. Nevertheless, the model (unlike the qualia of experience) is assessible for

accuracy in terms of its predictive and explanatory adequacy, e.g., as applied to climate change; if it passes muster, it should be taken very seriously as a behavior-guiding proxy for the mind-independent reality in which are embedded. We have no better factual basis for action.

5. The non-objectivity of representational content

Having maximally objectified the world, science seems to stumble when it comes to consciousness. There is simply nothing in the world, as described by our best, most objective account of it, that answers to what we think of as the hallmark of consciousness: qualitative phenomenology. If there were, the problem of consciousness would not arise. However, if we grant the reality of phenomenal content, we naturally want to objectify it: locate it in the physical world as given to us in our best scientific models. That we can't locate it affronts our explanatory ambition to unify, for instance under physicalism, all that we hold to be real.

To soften the blow to physicalism (should we favor it), we can note that it isn't just qualia that we don't find in the world they help represent for us. Although numbers are indispensable in quantifying reality, whether talking about cats or black holes, we don't expect to see them sitting in spacetime. Rather, we take them to be mind-dependent representational tools that we use in *descriptions* of objects, not objects themselves. We can count cats and black holes, but not the numbers we use to count them since numbers aren't anywhere to be counted. Equations specify physical states of affairs but are not themselves among the states they specify. We can write down an equation, but what we see on the whiteboard is a concrete physical phenomenon, the vehicle of representation, not the mathematical content itself. It is sometimes claimed that the reality of math is not just a matter of its (indispensable) representational function, but that mathematical objects exist eternally and mind-independently in a numerical Platonic realm. This seems to me a misguided attempt to reify representational content: we don't need to reify numbers in any sense to count them real *as representations*.

The same point applies to concepts and propositions. There are millions of objects that fall under the concept CAR, but the concept won't be found on roadways or anywhere else. Nor will we find the proposition just proposed, only the physical and cultural circumstances that make it true: cars exist as locatable physical objects, concepts do not. Are concepts and the propositions that make use of them therefore unreal? That we so confidently and necessarily speak of them, given our representational proclivities, suggests that even if they aren't locatable, they, like numbers, participate in what is for us an indispensable collective *representational reality.* Although they don't appear in the physical world as we perceive it, we can't transcend or step outside concepts and quantities when constructing our world-models, folk or scientific, although they might change as our models improve. Likewise, following Thomas Metzinger's view (about which more below), I suggest we can't transcend or step outside the phenomenal world-model associated with our biological sensory-perceptual capacities. Experience, therefore, also constitutes a representational reality, albeit personal, not collective. For knowers, representations, whether conceptual, quantitative or qualitative, are untranscendably real as a condition of having knowledge.

Given these considerations, at this point it is tempting (and I won't resist) to adduce what might be a general rule about representational content: that the terms in which we represent (characterize, describe, grasp) the world will not be found among the phenomena of the world they participate in representing. Put another way, content, although real, cannot be objectified, such that we will find it alongside the physical objects and processes that appear to us in terms of

content, whether folk-physical or science-physical. The constituents of our *representational* reality – concepts, propositions, numbers and, finally, qualia – don't and won't appear as objects or substances, whether physical, non-physical, abstract, or otherwise, in the reality they represent as being objectively the case, our *represented* reality. Physical phenomena, including the vehicles of representation itself, are represented by us as being spatial, concrete, extended, etc. using terms (qualia, concepts, numbers) that aren't themselves spatial, concrete, or extended, and thus are not accessible as observables.

The representational relation, by requiring that knowers always operate in terms of some set of representations, makes representational content ubiquitous as we negotiate the world. We commonsensically understand that we deploy concepts, numbers and propositions *as* representations; they aren't possible perceptible objects inhabiting the objective world as we know (represent) it. We can see they are categorizations and characterizations of the (represented) world we use in thought, imagination and public discourse. In contrast, the representational relation is for the most part hidden when it comes to conscious experience. We don't deploy qualia (sweetness, pain, red, etc.) as representational tools in person-level voluntary cognition, rather they simply *are* how the world and the body present themselves in consciousness as (somehow) a function of sub-personal sensory processing. We have no control over the phenomenal feels of qualia, for example how cloves smell and taste, or how pain hurts, rather they constitute us as conscious subjects; they are "irrevocable" (Ramachandran and Hirstein, 1997). This, plus the seamlessly integrated and temporally continuous phenomenal gestalt of moment-to-moment experience (of which qualia are the raw materials), is what makes consciousness an untranscendable representational *reality* – a reality so robust that we ordinarily take the world given in consciousness to be the directly presented world itself, not the content of a representational interface. Experience is, as it's often put, transparent: we don't see *it*, but rather the world *in terms of it*. When we realize, as for instance in lucid dreams, sensory hallucinations and optical illusions, that consciousness is a reality-*model*, we might go looking for its basic qualitative terms – qualia – in the objective world as thus modeled, only to discover they aren't there.

Against the rule I've tendered, that content can't be objectified, it might be objected that we can sometimes specify and even *see* representational content in information-bearing systems. Consider artificial intelligences designed to categorize and respond appropriately to input originating from outside themselves. We could perhaps specify what states in what processors running a face recognition program code for *Tom Clark face*, content that then gets used in labeling me in social media images. Wouldn't that specification just *be* the content itself? But note that it's the successful *performance of the system* as judged by its concrete output (image, text and auditory tokens) that ends up validating the ascription of content; nothing in the processing or processors looks like me such that you can see that content in a particular partition or activity of the system. All you see is processing and processors – the content-bearing vehicles and the resulting behavior of the system as it labels me in images. The same goes should we look for representational content in IBM's Watson, the famed AI champion of *Jeopardy*. We know content exists in the up-and-running system, given its spectacular success in answering questions about the world (it's now being marketed for businesses, including marketing), but only the system itself can prove that the content exists; it doesn't exist as an intersubjectively accessible observable.

The status of neurally-instantiated content in biological mind-systems such as ourselves is a matter of considerable debate (Hutto and Myin, 2012; Shea, 2013) and ongoing research (Bechtel, 2014; Nikbakht et al., 2018), but on the plausible assumption that brains carry content (both unconscious and conscious), the same point applies: we won't see content percolating in the brain, e.g., my belief that the apple is on the table or my experience of pain; we'll just see neurons doing their thing. Neurally-instantiated content, whether at the sub-personal level (e.g., single channel sensory information) or personal (e.g., beliefs, desires) is thus *proprietary* to the (organic) system, not publicly accessible: it exists as a function of the system's self-regulation and behavior control processes. It is thus system-dependent, even though the specific content will often reflect the system's ongoing interaction with its environment.

Cross-species commonality in brain function and structure, for instance among primates and other mammals, suggests that hosting *conscious* content – phenomenal experience – is not just a human prerogative. The point at which a system, natural or artificial, ends up with a full-blown self-in-the-world reality-model awaits a credible theory of consciousness, and consciousness may not be an all or nothing affair.[7] Still, since content isn't an observable, the inference to similarity in content (conscious or not) across species and artificial systems is just that, an inference based on their observable features and behavior (Dennett, 1987). One has to *be* a certain sort of system, the nature of which is under investigation (Dehaene & Changeux 2011; Prinz, 2012; Clark, 2013), for the content to become, for it, the private representational reality of conscious experience.

6. Evaluating hypotheses on consciousness

If conscious experience constitutes a subjective representational reality, what in the *represented, objective* reality afforded by science and philosophy (the philo-scientific image, we might call it) could explain the existence of experience for mind-systems? The conclusion that I've just defended – that phenomenal content is real but can't be objectified – might constrain our (naturalistic) approaches to explaining consciousness. Related constraints used above to motivate that conclusion, and that I don't think are particularly controversial, will also apply. These include the assumption that there exists a mind-independent but physically-characterized reality that includes mind-systems like ourselves; the necessity of there being a representational relation involved in behavior-guiding cognition; the fact that consciousness seems to be a mind-dependent phenomenon; and the fact that minds, as far as we know, are composite, complex systems which model reality in service to self-maintaining behavior. I also hazarded the suggestions that, since we're not in an epistemic relation to conscious experience, qualia aren't first-person facts, and that phenomenal content, because it isn't an observable, can't play a role in scientific explanations of behavior (see Section 7 below). In what follows, I will bring these considerations to bear in assessing the viability of hypotheses about consciousness. My selection of hypotheses to evaluate is necessarily incomplete given the vast landscape of consciousness studies, and the evaluations themselves will be cursory but I hope suggest promising avenues for exploration.[8]

[7] On consciousness as a "graded phenomenon," see Metzinger (2003), 135.
[8] For compendia on hypotheses about consciousness, see Blackmore (2006), Van Gulick (2014), and Bourget and Chalmers (2018).

6.1 Non-starters

If we don't expect to find consciousness in the (represented) physical world, this rules out hypotheses which hold that experience is somehow causally produced or generated by physical states of affairs. Anything that's caused by physical goings-on will itself be physical – a potentially observable and thus objective phenomenon. As Dennett points out, there is no "second transduction" in the brain that produces conscious experience as a further effect of neural processing (Dennett, 1998). Consciousness isn't generated as a measurable output of its correlates, whatever those turn out precisely to be, so there will be no *causal* explanation of qualia forthcoming (Oakley & Halligan, 2017). Epiphenomenalists who suppose that qualia are somehow caused by physical states of affairs, but then play no causal role in behavior (Robinson, 2010) face this problem: there is no evidence for any causal production or generation of mental phenomena by the physical, so the epiphenomenalist worry seems misplaced until it's established that consciousness is indeed objective – something that exists on the same causal playing field as the brain.

Likewise, non-reductive physicalists who propose that consciousness somehow emerges from or supervenes on neural processing (or any other presumptive physical base) as an objective but *non-physical* property or feature of such processing must specify what mechanisms and transitions are involved. Such property dualism has it that consciousness can be individuated as a distinct mental aspect of the physical states of affairs with which it is associated, something that doesn't reduce to them but nevertheless exists alongside them in the objective world. In this way, it's possible to maintain that consciousness plays its own behavior-controlling role beyond what strictly physical properties accomplish (which is how it might commonsensically seem). But then the unsolved problem of phenomenal causation arises: how exactly does qualitative experience as something non-physical causally contribute to behavior? In any case, if consciousness can't be objectified, there's no objective non-physical property which could play a behavior controlling role. And indeed, no such property has been discovered: all we have in the objective (represented) world are the physical states and events associated with consciousness.

Reductive accounts that seek to *identify* experience with its observable physical correlates face the difficulty that such an identity would make experience a public object, contravening the privacy constraint. For example, Patricia Churchland, skeptical of any hard problem of consciousness, suggests that qualitative experiences just *are* certain brain states (Blackmore, 2006, 60). In which case, since brain states are completely physical and publicly accessible in principle, so too must be experiences. But however closely you inspect, measure, and schematize the brain states associated with pain, you'll not find pain as an observable spatio-temporal property (Dennett, 1978). You'd have to instantiate (be) those states for pain to exist, and it would then exist only for you, which is not the case for your brain states. Jesse Prinz, a self-described physicalist, has developed an empirically grounded representationalist hypothesis that consciousness (at least in our case) is constituted by neurally-instantiated *attended intermediate representations* (AIRs) in the brain (Prinz, 2012). But in examining the neural AIRs (the vehicles) we wouldn't thereby see or access consciousness (the content). The former are all in the public domain, the latter not. Michael Tye, another physicalist-representationalist, has proposed that consciousness consists of Poised, Abstract, Non-conceptual, and Intentional Content – PANIC (Tye, 2000). Without going into its merits as a brand of representationalism, the underlying physicalist assumption of the PANIC hypothesis is that phenomenal content ultimately reduces to, and thus is identical to, some set of physically-instantiated representational

goings-on – the vehicles. Or if a reduction is not in the offing, then phenomenal content will have objective status as a non-physical property of said vehicles. The first alternative places consciousness in the public domain, thus is ruled out on my view, while the second suffers from the problems confronting any sort of property dualism: explaining how consciousness as a non-physical but objective phenomenon, something irreducible to the physical, emerges from a physically instantiated system and then goes on to play a causal role in behavior.

Susan Pockett proposes another type of phenomenal-physical identity, that consciousness is "identical with certain spatio-temporal patterns in the electro-magnetic field" generated by the brain (Pockett, 2000). Were this the case, in observing and measuring those patterns, we would *per impossibile* be observing and measuring experiences as public objects. This point applies to all hypotheses about consciousness which hold it to be objectively specifiable and locatable, that is, resident in spacetime, even something as intangible (yet physical) as the collapse of the wave function in the brain's micro-tubules (Hameroff, 1998). According to John Searle's biological naturalism,

> Everything that has a real existence has it in a single space/time continuum and the real existence of consciousness is in human and animal brains. Thoughts about your grandmother, for example, are caused by neuron firings and they exist in the brain as a feature of the system at a higher level than that of individual neurons. (Schneider and Velmans, 2017, 327).

Observing that higher-level feature, for instance some sort of neurally-instantiated functional organization, would be to observe the experienced thought, which is precisely what the non-objectivity of consciousness rules out. Although Searle elsewhere acknowledges the subjectivity of experience, its "first person ontology," his insistence that only spatio-temporal entities can be real doesn't sit well with subjectivity.

6.2 Radical objectivisms

Panpsychism in its various forms is among the more extreme approaches to reifying consciousness: the phenomenal is hypothesized to be a completely mind-independent property of the basic constituents of matter, such that experience or perhaps *proto*phenomenology (something quasi-phenomenal that can combine to constitute the phenomenal) might be present just about everywhere in spacetime (Strawson, 2006; Skrbina, 2009a). Thus far there is no empirical support for such proposals, and as noted above the evidence thus far strongly suggests that consciousness is associated with composite mind-systems doing particular representational, behavior-controlling and system-regulatory work. Panpsychism attempts an end run around the question of how qualitative content arises for mind-systems by positing the existence of *objective, system-independent* quanta of qualitative or pre-qualitative subjectivity. But this maneuver is effectively blocked, at least for the time being, by the absence of any proposal for, or evidence of, the way these strictly hypothetical quanta combine to become contentful phenomenal experiences for subjects like us (the so-called combination problem).

What seems a variant of panpsychism, Russellian monism (RM) holds that the structural and dynamical regularities described by physics at the micro-level aren't all there is to the world, but that in addition there exist categorical natures which ground fundamental physical properties, natures that are identical to the qualitative states of experiences (Alter and Nagasawa, 2015). In having experiences, we are thus directly acquainted with the intrinsic, objective, mind-

independent self-nature of the world. This neatly solves both the problem of naturalizing consciousness *and* identifying the unrepresented *concreta* which purportedly lie behind the merely structural, relational and dispositional characterization of the world afforded by physics. But like panpsychism, thus far RM lacks any empirical support. It also runs afoul of the representational relation in supposing that phenomenal experience constitutes direct, non-representational but nevertheless epistemic contact with the (unrepresented) reality of intrinsic self-natures. Such contact may not be in the cards for the project of objectification since, as discussed above, the representational relation can only afford our world-model a structural correspondence with reality (Ladyman and Ross, 2007). Moreover, the idea that physics needs supplementing by knowable intrinsic natures seems a holdover from folk physicalism: that reality must be ultimately concrete, partaking of some kind of stuff or substrate.[9] What's mind-independently real, at bottom (should there be one), may not be under any such obligation.

An even more radical approach to objectifying consciousness is Riccardo Manzotti's externalist "spread mind" hypothesis, which holds that conscious experiences don't represent physical objects, rather they are *identical to* those objects that appear in consciousness:

> …to perceive something does not entail concocting a representation of something, but rather perceiving something means that the something is literally part of the experience of the subject. Whatever the subject sees is identical with a process beginning in the environment and ending in her brain. In turn, the perceived object would be identical to that process. There is no separation between the physical world and the experience of the subject. (Manzotti, 2011, 66)

On the assumption that the physical world is mind-independent, it is ordinarily thought to be intersubjectively accessible: we generally suppose that differently placed observers can observe and have experiences involving the same mind-independent object. However, since we have distinct experiences of an object which may differ depending on our perceptual capacities and location, and since according to Manzotti objects are identical to those different experiences, on his account there is no single, intersubjectively available mind-independent object to which we all have access. The claimed identity of experience and the physical world thus puts the presumptive mind-independent status of reality in doubt. Further, if there is no difference between the physical world and a subject's experience, this means my experience, and yours, ends up being publicly accessible, contravening the privacy constraint. Lastly, for Manzotti experiences can never be informative or misinformative, since they simply *are* what's physically the case, even dreams, hallucinations and visual illusions.[10] All told, the spread mind hypothesis is a tough sell for physicalists holding an internalist, supervenience-based view of consciousness, even more so for representationalists who suppose there's good reason to distinguish between

[9] Harboring this supposition we could call being *stuffy*. A good example is the following passage by philosopher Hedda Hassel Mørch, writing in *Nautilus* (Mørch, 2017); italics in original, underlining added: "What are physical things like *in themselves*, or intrinsically? Some have argued that there is nothing more to particles than their relations, but intuition rebels at this claim. For there to be a relation, there must be two things being related. Otherwise, the relation is empty—a show that goes on without performers, or a castle constructed out of thin air. In other words, physical structure must be realized or implemented by some stuff or substance that is itself not purely structural. Otherwise, there would be no clear difference between physical and mere mathematical structure, or between the concrete universe and a mere abstraction."

[10] "It is our direct individual experience that is unerringly true; being one with the external world, it cannot be wrong." – from Manzotti and Parks, 2018, http://www.nybooks.com/daily/2017/11/26/the-pizza-thought-experiment/.

mind-dependent representational content and what that content refers to or tracks, which is often a mind-independent physical state of affairs.

6.3 Better bets

Given my thus far pessimistic (and for reasons of space necessarily superficial and incomplete) assessment of hypotheses concerning consciousness, what approaches might be more congenial to the constraints I have proposed? Stronger candidates would be those that accept the *prima facie* informational content of experiences *and* acknowledge that such content is proprietary, not an observable. It will come as no surprise, then, that I find Integrated Information Theory (IIT) (Tononi & Koch, 2014) to be a promising proposal, since it meets both these criteria in its attempt to account for the essential characteristics of consciousness. IIT wears its representational commitment on its sleeve, since information is ordinarily *about* something, or can be construed to be.[11] Secondly, in IIT the information as qualitatively rendered in experience only exists *for the system*:

> In IIT, information is meant to capture the "differences that make a difference" *from the perspective of the system itself* – and is therefore both causal and intrinsic. These and other features distinguish this "intrinsic" notion of information from the "extrinsic" Shannon notion… (Oizumi et al, 2014, p. 6, emphasis added)

Keeping in mind that by "from the perspective of the system itself" IIT does not mean that the system is in an observational relation to its internal information (which would threaten an epistemic regress), but rather in an existential, *constitutive* relation (that is, of being), we can see that it respects the privacy constraint on consciousness. The central identity thesis of IIT, that "an experience is identical with the maximally irreducible conceptual structure…specified by the mechanisms of a complex in a [system's] state" (Oizumi et al, 2014, 14) means that experience exists only for the instantiating system, not as an observable such as the system itself.

I can't essay a detailed evaluation of IIT here, but simply note that certain of its implications seem intuitively implausible, for example that simple photodiodes host experience, and that perfectly static systems might as well. Such intuitions will have to give way should the theory pan out, but the evidence in hand indicates that experience arises only in conjunction with complex systems that engage in (or can potentially engage in) system-maintaining behavior with respect to their environments. Oizumi et al. and Tononi and Koch (2014) also speak from time to time of consciousness being "generated" by physical systems; but if experience is *identical* to informational content – the maximally irreducible conceptual structure of a system – then it isn't generated as a further effect of being that structure. Another concern is that ITT does not explain why integrated information should feel like anything for the instantiating system, that is, be qualitative, but the theory is yet young.

A rather different take on consciousness, that of James Tartaglia (like myself, a former physicalist), also recognizes that experience isn't found in the world as described by science. Agreeing with Dennett, he says that "…the objective world lacks any evidence for the existence of experience: it is this fact, after all, that generates the problem of consciousness in the first place" (Tartaglia, 2016, 93). But instead of opting for illusionism or eliminativism, Tartaglia

[11] Oizumi et al. (2014) say: "While emphasizing the self-referential nature of concepts and meaning [that is, of the system's informational content], IIT naturally recognizes that in the end most concepts owe their origin to the presence of regularities in the environment, to which they ultimately must refer, albeit only indirectly." (23)

holds that experience is real *and* that it's representational (86-9): we perceive objects "in virtue of having experiences" (88) so subjectively we are in indirect contact with the world, what he calls indirect realism. Each of us embodies a conscious perspective that bequeaths us a world given in terms of experience. But he goes on to argue that experience itself *transcends* that world: it is part of a wider, final context of existence, of a transcendent being that can't ever be grasped as it is in itself (117-9).

I find Tartaglia's "transcendent hypothesis" appealing since it gets so much right about experience, in particular its non-objectivity and representational nature. It also recognizes that ultimate reality cannot be captured in its self-nature precisely because we're always in a representational, perspectival relation to it. But I think it goes wrong in holding that consciousness is mind-*independently* real, that it participates in a transcendent reality. Tartaglia says "…the transcendent reality of experience is not actually caused by an objective world" (166), and so puts aside the question that most nags at naturalists: why, as the evidence suggests, does consciousness only attend the physical brain, doing only particular cognitive things? Even if we renounce *causal* explanations of consciousness, as I have, the obdurate, objective fact remains (a fact that should be included in our collective reality-model) that as far as we know only certain sorts of up-and-running systems host conscious experience; this cries out for a naturalistic explanation. To hold that consciousness is transcendent – a fundamental mind-independent reality – conveniently avoids that question, much in the way that panpsychism avoids it by holding phenomenality to be a particulate fundamental. The transcendent hypothesis, then, seems to depart from naturalism by marginalizing a valid philo-scientific inquiry, that of explaining consciousness as a phenomenon that on its face is neither transcendent nor fundamental, but rather mind-dependent.

Thomas Metzinger, by contrast, is resolutely naturalistic in his ambitious "self-model theory of subjectivity" (Metzinger, 2003, 2006) which was prefigured in earlier sections of this paper. The theory is explicitly representationalist, informed by and predictive of empirical findings, and seeks to explain the central conditions and features of phenomenal consciousness, including experiential transparency.[12] To be conscious is to instantiate a neurally-realized, functionally adaptive, self-in-the-world reality-model or representation, parts of which the system can't recognize as a model. On Metzinger's account, to take something in experience as real is simply not to grasp the content of experience *as a representation*: experience is thus transparent for the system and its content taken as reality (2003, 163-169). We consequently end up as naïve realists about at least some portion of the represented world, in particular ourselves *as selves* (hence the "self-model" theory).

Metzinger argues that the representational nature of consciousness is suppressed for biological systems such as ourselves as a matter of cognitive efficiency. To minimize energy consumption and afford real-time behavioral control, our representational capacities must be limited in their recursive (meta-representational) application, and this limitation can perhaps help explain why we end up hosting qualitative content. The adaptive (and perhaps logically necessary) *closing off* of what would be a paralyzing representational, epistemic regress means that the system will instantiate on the sub-personal level representational content that it can't *further* represent, what

[12] Other features of phenomenality covered by Metzinger include the global availability of information for action control, the immediacy or "nowness" of experience, its coherent self-in-the-world unity, and its hierarchical part-whole integration, dynamicity, and subjective perspectivalness (2003, 107-208).

Metzinger calls "autoepistemic closure" (2003, 131).[13] Such content, in particular that associated with sensory channels such as vision and internal proprioception, perforce becomes cognitively impenetrable, an irreducible basic element of representation itself, that which the system can't specify any component of or take as a representable object (except conceptually of course, as we're now engaged in). But what *is* this if not the essential characteristic of individual, basic qualia: their not-further-characterizable, non-decomposable, hence ineffable phenomenal character, the fact that in consciousness we can't get behind, or into, or away from such things as sweetness, primary red, pain, or any other basic unanalyzable sensory quality? And, according to Metzinger, since we're not in a position to take most qualitative content as representations (in particular that of waking experience), qualities in experience are taken as *object* properties, properties of things construed as mind-independently real.

This possible explanation of proprietary qualitative content as an outcome of being a behavior-controlling representational system, of which the above is just the briefest sketch, is among the most promising features of Metzinger's detailed and theoretically rich approach.[14] Although he holds that consciousness is ultimately a physical, biological phenomenon (2006, 58), it's notable (and laudable on my view) that he offers no causal explanation of phenomenology nor any pat phenomenal-physical identity claim. Instead, he proposes a logical and adaptive *entailment* from the limitations of physically-instantiated representational functions to the existence of phenomenal experience for, and only for, the instantiating system. This, it seems to me, at least gets us in the vicinity of qualia construed as the irreducible elements of a private representational reality.

7. Phenomenal-physical parallelism

That said, I differ with Metzinger in his supposition that qualia, and the integrated, dynamic phenomenal gestalt of waking consciousness, can play a role in scientific accounts of behavior and the evolution of the neural processes that support consciousness (Metzinger 2006, 54-62). In these accounts, science can only deal in observables or what accepted theories suggest might eventually be observed or detected (e.g., dark matter, Hawking radiation). Since qualia and the experiences constructed out of them are not, and I think never will be, observables, they can't be cited as causal factors alongside or in addition to what their neural correlates accomplish in physical and functional accounts of behavior control.[15] We therefore have to give up on objective phenomenal causation since the phenomenal, although real, doesn't appear in the objective world as represented by science. In the evolutionary cognitive arms race, natural selection undoubtedly selected for the behavior-controlling cognitive functions and underlying neural wetware associated with being conscious, but not for consciousness itself (Rosenthal, 2008; Oakley and Halligan, 2017). However, since we live our subjective lives completely within the ego tunnel of conscious experience, we unsurprisingly take experience to be a causal operator: I eat chocolate

[13] This is somewhat the antithesis of higher-order thought or perception theories of consciousness (Rosenthal, 1997), which propose that a brain state becomes conscious when it becomes the target of yet *further* representational activity by the system.

[14] See Metzinger 2003, pp. 107-208 for his full story on phenomenality, and Clark 2010, part 5 for some supportive commentary that summarizes his major points. See also Clark 2005, parts 6 through 9 for an account of phenomenality that draws on Metzinger and other representationalists.

[15] Note that if conscious states were *identical* to their neural and functional correlates, then they wouldn't play an additional causal role either. We wouldn't need to appeal to consciousness *per se* in scientific explanations of behavior.

because it tastes good to me, right? Well, objectively that taste is nowhere to be found, so science will always default to the neural correlates of the taste in explaining my chocolate habit. But in everyday life I will as a conscious subject continue unperturbed to cite the taste – a phenomenal particular – as the cause. We thus have two sorts of valid, predictive explanations, one involving my experience, another involving its physical correlates.

Objective phenomenal causation would require that experience somehow causally supplements the physical on the objective causal playing field, what we might call objective *explanatory space*. But again, we don't find my taste of chocolate in that space, that which involves physical spacetime. What is more plausible, or at least better governed by the constraints suggested above, is a phenomenal-physical parallelism that involves respective subjective and objective explanatory spaces. Such parallelism respects the reality of both qualia (the private representational reality of experience) and its correlates (the collective *represented* reality of neural processes) without supposing they causally interact, which obviates the problem of phenomenal causation. As suggested above, it could be that some sort of non-causal entailment from physically-instantiated representational goings-on to qualia underwrites the reliable parallelism.[16]

Keeping the phenomenal and physical on separate but correlated tracks also relieves physicalism of the burden of somehow reducing qualitative feels to physical facts. If, as suggested in Section 3, there are no first-person facts about qualia, e.g., no private truths about sweetness, red, or pain, no such reduction is called for. Qualities, experiential and non-objective, are the not-further-specifiable basic terms in which the physical world appears in consciousness, so won't themselves be objects in that world we can point to as facts about it. Rather, folk-physical facts and truths about the world (e.g., that the apple in Section 4 is red) are couched *in terms* of such qualities. Not having to concern itself with phenomenal-physical identity or causation, physicalist science can proceed apace in developing an ever more predictive quantitative and conceptual grasp of reality.

8. Conclusion

As much as I've pleaded the case for phenomenal realism and the non-objectivity of qualia, the widely held naturalistic presumption of physicalism makes the conjunction of these claims at least counterintuitive and perhaps irredeemably obscure. What *physical* sense can be made, after all, of a non-causal entailment from neurally-instantiated representational processes to subjective qualitative content? Physicalists take composition, causation, reduction, and emergence to be primary explanatory relations among phenomena, so of course want to apply them in naturalizing consciousness. Conscious mind-systems are natural, objective, physical phenomena, resident in spacetime, so if consciousness is real must it not also be discoverable there? If consciousness is not a causal production of a physical system, nor identical with at least some parts or properties of it, then it looks as if its association with the system is a mystery. The idea that each person's conscious experience constitutes a private representational reality will of course seem suspect

[16] Accounts of consciousness that identify it with information (e.g., IIT and Pylkkänen's (2017) account suggesting it might be identical with "active information" at the quantum level), face the difficulty of showing how information *per se* could affect physical processes. My recommendation again would be to accept an informational-physical explanatory parallelism, not to suppose that information causally supplements (adds causal power to) the physical processes subserving behavior in physicalist explanations. Both sorts of explanations might well be perspicacious and predictive in their respective explanatory spaces, but combining them might not be workable.

under a regime that requires all of what's real to be objective. But if content, in particular phenomenal content, can't be objectified, then consciousness won't be discoverable in spacetime, and indeed that's the current state of play: experience is a private, not public affair, and naturalistic explanations must somehow respect this (objective) fact about consciousness, at least until it's shown that there is no such fact, or no such thing as phenomenal experience.

The way forward, as I see it, is to continue full steam ahead with the investigation of the neural and functional correlates of consciousness; this will inform a *science of representation* that may lend plausibility to the possible entailment from certain types of representational processes to experience. The nature of that entailment, should it exist, may become clear as the self-model theory and its representationalist competitors and collaborators, e.g., IIT, Prinzian AIRs, and Tye's PANIC account, are refined in light of further research. Understanding the biological and artificial architectures of world-responsive informational systems may show how instantiating a suitably ramified and integrated reality-model makes representational content qualitative for the system – not as a matter of causation or emergence but of representational necessity. We may not find a satisfying *physicalist* account of how consciousness arises in nature, since it doesn't appear in the physical world, but we might find sufficient explanatory resources in representationalism to close the explanatory gap (Levine, 1983).

Such work may end up challenging commonsense intuitions about what counts as real, but naturalistic theories of consciousness will be constrained by empirical evidence and logic, not traffic in unexplained explainers, and will admit their incompleteness until completed. All science and philosophy can ask is that we stay true to epistemic good practice, so if we end up with something initially counterintuitive, so be it. But the end result won't be spooky or supernatural given that it's arrived at under naturalistic evidential and methodological constraints.

We can understand physicalism as a well-intentioned and natural attempt at cognitive unification – a global pronouncement about the nature of reality – but consciousness puts physicalism under considerable pressure.[17] Physicalism forgets, perhaps, that the world as objectified in folk-physicalism and science is a *represented* reality, a world-model, not unrepresented reality. A better, more self-consistent global naturalistic realism will incorporate the indispensable role of representation in the collective world-model itself. Since our world-model represents what we take to be real – our ontology – as thus expanded it will assert the reality of representation, not as a separate substance or property, but as a condition of being knowers. A science of representation might eventually show that physicalism is not the last naturalistic word by highlighting the necessity of the representational relation, understanding how systems host representational content, and perhaps even answering the question of how we physically characterized mind-systems each end up with the subjective, qualitative representational reality of consciousness. If the representational relation puts us as knowers at one remove from unrepresented, mind-independent reality, as natural, physical creatures we ourselves are nevertheless situated in that reality. This means our conscious experience too, even if not

[17] For acknowledgement by physicalists of the limitations of physicalist explanations of consciousness, see Kim, 2005 and Howell, 2013. Kim accepts that qualia can't be functionalized, thus resist any straightforward incorporation into physical science, while Howell argues that phenomenal consciousness, although ultimately physical, can't be objectified, hence we should be "subjective physicalists."

discoverable in spacetime, is fully within the natural world as it engages, through us, in the project of self-knowledge.

Conflict of Interest Statement

The author declares that this manuscript was produced in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

[Acknowledge any reviewer comments and assistance here]

References

Alter, T. and Nagasawa, Y. (2012). What is Russellian monism? *J. Conscious. Stud.* 19, 9-10.

Baars, Bernard J. (1997). "Contrastive phenomenology: a thoroughly empirical approach to consciousness," in *The Nature of Consciousness: Philosophical Debates*, eds N. Block, O. Flanagan and G. Guzeldere (Cambridge: MIT Press), 187-202.

Bechtel, W. (2014). Investigating neural representations: the tale of place cells. *Synthese* 193, 1287–1321. doi: 10.1007/s11229-014-0480-8

Blackmore, S. (2006). *Conversations on Consciousness*. Oxford: Oxford University Press.

Bourget, D. and Chalmers, D. (2018). eds *PhilPapers*. https://philpapers.org/browse/theories-of-consciousness.

Chalmers, D. (2004). "How can we construct a science of consciousness?" in *The Cognitive Neurosciences III*, ed M. Gazzaniga (Cambridge, MA: MIT Press), 1111-1120.

Chalmers, D. (2006). "The representational character of experience," in *The Future for Philosophy*, ed B. Leiter (Oxford: Oxford University Press), 153-181.

Churchland, P. M. (1988). *Matter and Consciousness*. Cambridge, MA: MIT Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 36, 1-24. doi: 10.1017/S0140525X12000477

Clark, T. (2005). Killing the observer. *J. Conscious. Stud.* 12, 38—59.

Clark, T. (2010). The appearance of reality. *Naturalism.Org*. http://www.naturalism.org/philosophy/consciousness/the-appearance-of-reality

Clark, T. (2013a). "Experience and autonomy: why consciousness does and doesn't matter," in *Exploring the Illusion of Free Will and Moral Responsibility*, ed G. Caruso (Lanham, MD: Lexington Books), 239-254.

Clark, T. (2013b). Respecting privacy: why consciousness isn't even epiphenomenal. *Naturalism.Org*. http://www.naturalism.org/philosophy/consciousness/respecting-privacy

Dehaene, S. and Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200-227. doi: 10.1016/j.neuron.2011.03.018

Dennett, D. C. (1978). Why you can't make a computer that feels pain. *Synthese* 38, 415-456.

Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, MA: M.I.T. Press.

Dennett, D. C. (1998). "The myth of double transduction," in *Toward a Science of Consciousness II, The Second Tucson Discussions and Debates*, eds S. Hameroff, A. Kaszniak, and A. Scott (Cambridge, MA: MIT Press), 97-107.

Dennett, D. C. (2016). Illusionism as the obvious default theory of consciousness. *J. Conscious. Stud.* 23, 65-72.

Dennett, D. C. (2017). *From Bacteria to Bach and Back: The Evolution of Minds.* New York: W.W. Norton & Company.

Frankish, K. (2016). Illusionism as a theory of consciousness. *J. Conscious. Stud.* 23, 11-39.

Gamez, D. (2014). The measurement of consciousness: a framework for the scientific study of consciousness. *Front. Psychol*. 5. doi: 10.3389/fpsyg.2014.00714

Goff, P. (2016). Is realism about consciousness compatible with a scientifically respectable world view? *J. Conscious. Stud.* 23, 83-97.

Hameroff, S. (1998). Quantum computation in brain microtubules? The Penrose–Hameroff 'Orch OR' model of consciousness. *Phil. Trans. R. Soc. Lond.* 356, 1869–1896.

Hawking, S. and Mlodinow, L. (2010). *The Grand Design*. New York: Bantam Books.

Howell, R. J. (2013). *Consciousness and the Limits of Objectivity*. Oxford: Oxford University Press.

Hutto, D. D. and Myin, E. (2012). *Radicalizing Enactivism: Basic Minds Without Content.* Cambridge, MA: MIT Press.

Kanwisher, N. (2001). Neural events and perceptual awareness. *Cognition* 79, 89-113.

Kim, J. (2005). *Physicalism, or Something Near Enough*. Princeton, NJ: Princeton University Press.

Ladyman, J., and Ross, D. (2007). *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.

Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly*, 64, 354–361. doi: 10.1111/j.1468-0114.1983.tb00207.x

Manzotti, R. (2011). The spread mind: is consciousness situated? *Teorema* 30, 55-78.

Manzotti, R. and Parks, T (2018). On Consciousness. *New York Review of Books,* http://www.nybooks.com/topics/on-consciousness/

Metzinger, T. (2003). *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MA: M.I.T. Press.

Metzinger, T. (2009). *The Ego Tunnel: The Science of Mind and the Myth of the Self.* New York: Basic Books.

Mørch, H.H. (2017). Is matter conscious? *Nautilus* 47, http://nautil.us/issue/47/consciousness/is-matter-conscious

Moreland, J. P. (2009). *The Recalcitrant Imago Dei: Human Persons and the Failure of Naturalism*. London: SCM Press.

Nikbakht, N., Tafreshiha, A., Zoccolan, D., Diamond, M. E. (2018). Supralinear and supramodal integration of visual and tactile signals in rats: psychophysics and neuronal mechanisms. *Neuron* 97, 626-639. doi: 10.1016/j.neuron.2018.01.003

Noe, A. (2009). *Out of Our Heads*. New York: Hill and Wang.

Oakley, D., and Halligan, P. (2017). Chasing the rainbow: the non-conscious nature of being. *Front. Psychol*. 14, doi: 10.3389/fpsyg.2017.01924

Oizumi, M., Albantakis, L., Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLOS Computational Biology* 10, doi: 10.1371/journal.pcbi.1003588

Pockett, S. (2000). *The Nature of Consciousness: A Hypothesis*. Lincoln, NE: Writers Club Press.

Prinz, J. J. (2012). *The Conscious Brain: How Attention Engenders Experience.* Oxford: Oxford University Press.

Pylkkänen, P. (2017). "Is there room in quantum ontology for a genuine causal role of consciousness?" in *The Palgrave Handbook of Quantum Models in Social Science: Applications and Grand Challenges*, eds A. Khrennikov, A. and E. Haven (London: Palgrave Macmillan), 293-317.

Ramachandran, V. S. and Hirstein, W. (1997). Three laws of qualia. *J. Conscious. Stud.* 4, 429-58.

Revonsuo, A. (2015). Hard to see the problem? *J. Conscious. Stud.* 22, 52-67.

Robinson, W. S. (2010). "Epiphenomenalism." *Wiley Interdisciplinary Reviews: Cognitive Science* 1, 539–547. doi: 10.1002/wcs.19

Rockwell, T. (2013). Representation and radical empiricism. *Intellectica* 2, 219-240.

Rosenthal, D. M. (1997). "A theory of consciousness," in *The Nature of Consciousness*, eds N. Block, O. Flanagan, and G. Guzeldere (Cambridge, MA: MIT Press), 729-754.

Rosenthal, D. M. (2008). Consciousness and its function. *Neuropsychologia* 46, 829-40. doi: 10.1016/j.neuropsychologia.2007.11.012.

Schneider, S. and Velmans, M. eds (2017). *The Blackwell Companion to Consciousness*. Hoboken: Wiley and Sons, Ltd.

Sellars, W. (1962). "Philosophy and the scientific image of man," in *Science, Perception and Reality* Sellars, W., (London: Routledge & Kegan Paul), 7-43.

Seth, A., Suzuki, K., Critchly, H.D. (2012). An interoceptive predictive coding model of conscious presence, *Front. Psychol.,* 10. doi: 10.3389/fpsyg.2011.00395

Shea, N. (2013). Naturalising representational content. *Philosophy Compass* 8, 496-509. doi:10.1111/phc3.12033

Skrbina, D. (2009a). Ed *Mind that Abides: Panpsychism in the New Millennium*. Philadelphia: John Benjamins.

Skrbina, D. (2009b). "Minds, objects and relations: toward a dual aspect ontology," in *Mind that Abides: Panpsychism in the New Millennium*, ed D. Skrbina (Philadelphia: John Benjamins), 361- 382.

Sterling, P. and Laughlin, S. (2015). *Principles of Neural Design*. Cambridge, MA: M.I.T. Press.

Strawson, G. (2006). Realistic monism: why physicalism entails panpsychism. *J. Conscious. Stud.* 13, 3-31.

Tartaglia, J. (2016). *Philosophy in a Meaningless Life: A System of Nihilism, Consciousness and Reality.* London: Bloomsbury.

Tononi, G. and Koch, C. (2014). Consciousness: here, there but not everywhere. *arXiv.org* arXiv:1405.7089

Travis, C. (2004). The silence of the senses. *Mind* 113, 57-94. doi: 10.1093/mind/113.449.57

Tye, M. (2000). *Consciousness, Color and Content*. Cambridge, MA: MIT Press.

Van Gulick, R. (2014). "Consciousness," in *The Stanford Encyclopedia of Philosophy,* Spring 2104 edition, ed E. Zalta, 34-43.