

CAUSAL ROLE OF PHENOMENAL CONSCIOUSNESS

ROBERTO HORÁCIO DE SÁ PEREIRA

Federal University of Rio de Janeiro / CNPq BRAZIL

robertohsp@gmail.com

Abstract. My account of the causal role of consciousness in a physical world is modeled on Dretske’s celebrated explanation of the causal role of beliefs (something that Dretske himself never offered). First, behavior must be understood as a (broadly individuated) process that begins with some external stimulus causing some neurological event C, and ends with causing a bodily movement M (e.g., the Kennedy assassination is a process that begins with Oswald pulling the trigger at 12:30pm CST on November 23 in 1963 in Dallas, Texas, but only ends half an hour later when Kennedy is pronounced dead at 1 pm CST). The internal neurological event C causes bodily movement M, but only by virtue of being recruited by natural selection to represent the instantiation of some external property F when properly stimulated under normal circumstances. But the reason why C causes M lies in the fact that C represents the instantiation of the external property F. I withdraw my hand from a hot surface because the activation of nociceptive specific neurons in my parietal lobe (together with the activation of neuronal patterns in my motor cortices) was recruited by natural selection to represent the tissue damage in my hand. The activation of nociceptive specific neurons *causes* my hand to withdraw *but for the reason that it represents* tissue data at the time that I felt pain in my hand.

Keywords: phenomenal causation • structuring causes • triggering causes • structuring behavior

RECEIVED: 12/05/2021

REVISED: 15/03/2022

ACCEPTED: 25/05/2022

1. Preliminaries

As I understand it here, physicalism about sensory experience is the claim that conscious properties supervene on or are identical with physical properties. That said, any possible world that is a minimal microphysical duplicate of the real actual world is also a phenomenal duplicate. Anti-physicalism consists in the rejection of the claim of a metaphysical necessity: there are natural laws that connect the physical with the phenomenal, but these laws are not metaphysically necessary; they are valid only in the actual world and all nearby worlds. Accordingly, what people today call panpsychism or protopanpsychism are not forms of physicalism in the way that I am proposing, since they posit the existence of phenomenal micro-quiddities that are irreducible to the physical properties that science reveals.



The central problem that the anti-physicalist has faced since Descartes is the problem of mental causation. If the physical domain is causally closed, and if mental properties are irreducible to physical ones, the only way out is to appeal to overdetermination: a child was healed of an illness because he took the proper medication, but also because the prayers offered for his recovery were heard by God.

Mutatis mutandis, if the physical domain is causally closed, and if phenomenal and conscious properties are irreducible to physical ones, the only way out is to appeal to overdetermination. However, as Kim has emphasized, if overdetermination is not impossible, it is quite an implausible claim: “no event can have more than one sufficient cause occurring at a given time — *unless it is a genuine case of causal overdetermination*” (Kim 2005, p.42, emphasis added). Thus, the anti-physicalist faces Kim’s dilemma: if he wants to avoid epiphenomenalism, to show that consciousness makes a difference in the physical world, he must embrace some implausible form of overdetermination; and vice-versa, if he wants to avoid some implausible form of overdetermination, he must embrace epiphenomenalism, namely that consciousness is casually idle.

The essay is composed of the following sections. The next one is devoted to showing how the charge of epiphenomenalism emerges considering Dretske’s case of the upper-register supplication of a coloratura soprano shattering a wine glass (see Dretske 1988, p.79.). In the following section, we consider, alternatively, Schaffer’s sergeant-major case (Schaffer 2000, p.175) and Won’s variant sergeant-sergeant case (Won 2014, p.218) as models of overdetermination to rescue anti-physicalism. Schaffer’s sergeant-major case is not a case of genuine overdetermination: the major’s command preempts the sergeant’s. In contrast, Won’s is a genuine case of overdetermination. However, it is useless as a model for conscious causation.

In the following section, I will eliminate the main obstacle to understanding phenomenal causality in physicalist terms, namely, the so-called explanatory gap. To be sure, sometimes physicalists about sensory experience hold that brute metaphysical necessitation is not strong enough for capturing their view. In addition, we should provide some further explanation of how physical properties *ground* phenomenal ones. Nonetheless, along the same line of Papineau’s reasoning (1993), I hold that identities dispense with explanations: they are brute fact.

However, the absence of a gap between phenomenal properties and physical properties is not enough to provide a reasonable account for the causal role of consciousness in the physical world. I use Dretske’s soprano case again to show that Papineau’s qualitative view also fails to account for the causal role of consciousness in a physical world. If consciousness is an intrinsic property of the brain that is only contingently connected to representational properties, it is hard to see how consciousness finds its way out to the world outside the skull.

In the last section, I present my own account for the causal role of conscious-

ness in the physical world. My model is Dretske's account for the causal role of basic beliefs (Dretske 1988, p.79–95). To begin with, we must understand behavior as a *process* that starts with some external stimulus causing some neurological event C, and ends up with a bodily movement M (Kennedy's assassination is a process that starts with Oswald pulling the trigger at 12:30pm CST on November 23 in 1963 in Dallas, but ends when Kennedy is pronounced dead at 1:00pm CST, at the hospital). The internal event C triggers the bodily movement M in virtue of being recruited by selection to represent the instantiation of some external property F when appropriately stimulated. Roughly, I consciously pull my hand from a overheated surface because nociceptive specific neurons in my parietal lobe were recruited to represent the tissue damage in my hand in virtue of the painfulness in my hand.

2. Epiphenomenalism

Suppose that, in an opera, the upper-register supplication of a soprano accompanied by the statement "I love you" makes a wine glass shatter.¹ As a matter of fact, the soprano's singing is an event constituted by a sequence of sounds with various physical properties: a certain pitch, amplitude, etc. (they constitute what people usually call the *vehicle* of content). Yet, the sequence of sounds also has a semantic property, that is, the property *that I love you* (the content itself). Now, just as this semantic property is not identical with any of the physical properties in the causal sequence of sounds, nor does it even supervene on them, the following counterfactual is true: even if what the soprano was singing had a different meaning, the glass would shatter in the same way, *ceteris paribus*. The semantic property of the sung sentence is causally idle (see Dretske 1988, p.79; Tye 2021, p.94).

If we assume anti-physicalism (phenomenal properties do not even supervene on physical properties), it is fair to claim that phenomenal properties stand for the meaning of the sentence "I love you" *in the same contingent way* that physical properties stand for the physical makeup of the sequence of sounds. Now, but as the relationship between one and the other properties is contingent, the same sound sequence can enunciate different meanings in the same way that the same neuronal property can convey quite different phenomenal properties. What does this show us? If phenomenal properties do not even supervene on physical properties, Dretske's analogy shows that phenomenal properties are causally idle.

Suppose now that I consciously pull my hand from a hot surface. In this model, my pain has a physical makeup (say the stimulation of the nociceptive specific neurons of the parietal lobe) and its phenomenal or conscious character is like the semantic property of the sentence: "I love you." What follow from this? The answer is that the (phenomenal) painfulness of my pain is causally irrelevant with respect to my

pulling my hand away from the overheated surface. *What is making me do that is just the stimulation of the nociceptive specific neurons.* And this is what we should expect, if we assume that the phenomenal painfulness is neither identical with any of the physical properties of the neural event nor supervenes on them. The following counterfactual seems true: *I would have pulled my hand from the overheated surface even if I had not been in pain.* All that is required is the stimulation of the nociceptive specific neurons. Indeed, that is what Chalmers's zombies scenario tells us (Chalmers 2010). It is metaphysically possible that there are possible worlds of physical and functional replicas of humans that pull their hands from some overheated surface without, however, feeling pain.

Here is not the place to discuss whether if this is conceivable, the zombies scenarios or possible worlds are also possible. Be that as it may. Recently, Tye (2021) complained that this common objection is confused. He argues as follows:

If it is true that in the nearest possible world to the actual world in which C [e.g. painfulness] is missing, E [I pull my hand from an overheated surface] is missing too, that suffices (in the absence of a defeater) to show that C causes E. It is not necessary that in all other metaphysically possible worlds in which C is missing, E be missing too. (2021, p.98)

What grounds a psychophysical law such as pain in my hand causes me to pull my hand from some overheated surface is only the possibility of nearest worlds to the actual one in which the following counterfactual is true: *if I was feeling pain in my hand, I would not pull my hand from the overheated surface.* To be sure, natural laws connect the physical with the phenomenal, but these laws are not metaphysically necessary since worlds with different laws are metaphysically possible. Still, if the pain in my hand does supervene on the stimulation of nociceptive specific neurons in my parietal lobe, and moreover, this neurological event is sufficient to cause me to pull my hand away, it seems that this neurological event is doing all causal jobs. The painfulness in my hand is causally idle in the face of my pulling my hand from the overheated surface.

3. Overdetermination

The question now is whether the anti-physicalist can avoid phenomenal epiphenomenalism by embracing overdetermination. The usual way of characterizing overdetermination is to say that an event is overdetermined when it has two distinct causes such that each without the other would still have caused the event. The question that arises is whether there would be overdetermination within the scope of phenomenal causation. According to Won, Lewis's analysis (1986) suggests the following picture. C1 and C2 overdetermine an event E "and" *iff*:

1. Both c1 and c2 are causes of E.
2. If c1 had occurred without c2, E would still have occurred.
3. If c2 had occurred without c1, E would still have occurred.
4. If neither c1 nor c2 had occurred, E would not have occurred. (2014, p.209)

Consider the following case that

Schaffer presents (2000, p.175). Imagine that a major and a sergeant stand before the same troops. Both soldiers simultaneously shout “charge!” and the troops charge. Yet, if the major had not shouted, the troops would still have charged because they would have obeyed the sergeant’s order. Do we have a case of genuine causal overdetermination? The intuitive answer is no, since it is the major’s shouting that causes the troops to charge *because orders coming from higher-ranking soldiers trump those of lower-ranking ones*.

Now consider Won’s (2014) variant case. Two sergeants stand before the troops. As before, both sergeants shout “advance!” at the same time, and the troops advance. Now suppose that there was a major standing there too. He was about to shout “retreat!” But the major hears the two sergeants shouting “advance!” and he remains silent. But let us assume, in addition, that if only one of the sergeants had ordered the advance, the major would have shouted “retreat!” which would have caused the soldiers to retreat. Won illustrates the case as follows (see Figure 1).

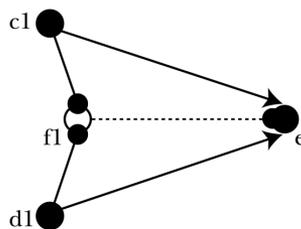


Figure 1: “Here c1 and d1 represent the two sergeants’ orders to advance; f1 the major’s order to retreat; e the soldiers’ advancing. In this case, e is overdetermined, but c1 and d1 do not meet the condition that either alone would have caused e, and there is not a time before e’s occurrence at which there are two causes that meet that condition.” (Won 2014, p.218–219)

In contrast to Schaffer’s case, in Won’s variant case, both C1 and D1 cause the troops to retreat. However, is that a genuine case of overdetermination? Certainly not in the light of Lewis’s counterfactual analysis: conditions 2 and 3 are not met, since if either C1 or D1 had not shouted “advance!” the major would have shouted “retreat!” and E would not have happened.

Be that as it may, my question is not whether Won's case is a genuine case of overdetermination or whether Lewis's counterfactual analysis fails to capture genuine cases of overdetermination. The key question is: are those putative cases of overdetermination of any use as a model to rescue the anti-physicalist account of the charge of epiphenomenalism? Let us consider Schaffer's case again. If this example is to be of any use as a model to the anti-physicalist, we must think of the *brains-state-tokens* that underlie *phenomenal-state-tokens* as different from the *phenomenal-state-tokens*, just as we think of the sergeant as different from the major.

The analogy works as follows. Suppose that I am playing soccer when I suffer a strain in the adductor muscle in my thigh. I feel a terrible pain. By way of analogy, think of my pain as the major's command "charge!" and of the simultaneous neural state (say the stimulation of nociceptive specific neurons) underlying my pain as being like the sergeant's shout "charge!" My pain causes me to put ice on my thigh in the same way as the major's command "charge!" causes the troops to charge *by virtue of being an order from the senior officer*. Now, if the anti-physicalist were right, even without stimulation of nociceptive specific neurons in the parietal lobe, but feeling pain, I would still have asked for ice (like the troops would obey the major's order if the sergeant remained silent). Does this model work?

The fundamental trouble here is that, intuitively, it seems that the stimulation of nociceptive specific neurons of the parietal lobe is what is truly making me ask for ice. If the pain in the adductor muscle in my thigh caused me to ask for ice, it is because it supervenes on or is identical with the stimulation of the nociceptive specific neurons in my parietal lobe. Thus, the analogy with Schaffer's case breaks down: the painfulness does not stand for the sergeant or the stimulation of the nociceptive specific neurons for the major. Schaffer's case is useless as a model to rescue anti-physicalism from the charge of overdetermination.

4. Bodily Movement

In an attempt to circumvent the charge of overdetermination, the anti-physicalist might introduce Dretske's distinction between behavior and bodily movements. According to Dretske, bodily movements should not be confused with behavior: "The former is an event, movement, something that *happens to* (say) a paw. The second, I shall argue, is a piece of behavior, possibly an action, something the rat *does*" (1988, p.15, original emphasis). Now the anti-physicalist could appeal to Dretske's distinction between behavior and bodily movement as an attempt to overcome the charge of overdetermination. The reason is the following: for now, it seems, we have different effects for different causes. Behavior is a process while bodily movement is an event. What causes behavior as a process is what Dretske calls "structuring causes." In

contrast, what causes bodily movement as an event is what Dretske calls “triggering causes” (see Dretske, 1988, p.x). This requires a little explanation.

Let us consider one of Dretske’s own examples (see 1988, p.21): the tragedy that occurred on November 23, 1963, in Dallas, Texas, USA, when Kennedy was brutally assassinated by Lee Oswald. It is indisputable that the fact that Lee Oswald pulled the trigger of his rifle is what killed J.F. Kennedy in Dallas. However, the fact that Oswald pulled the trigger does not explain *why* or *for what reason* he murdered J.F.K. Oswald’s bodily movement is an *event* that happens at 12:30pm CST on November 23 in 1963 in Dallas. In contrast, Kennedy’s assassination is a *process* that starts with Oswald pulling the trigger at 12:30pm CST on November 23 in 1963 in Dallas, but only ends when Kennedy is pronounced dead at 1:00pm CST.

Thus, the charge of overdetermination is overcome; after all, we have distinct effects, namely Kennedy being shot in the head (an event) and Kennedy’s assassination (a process). But we also have distinct causes: one physical, where Oswald pulls the trigger of his rifle (triggering cause) and another phenomenal (structuring cause), say Oswald’s hatred for Kennedy. Nonetheless, the problem of phenomenal causation of behavior still stands. First, even if behavior as a *process* is not identical with a bodily movement as an *event*, there is undeniably a close connection between the two. First, *as a process* Kennedy’s assassination incorporates Oswald’s bodily movement: it starts with Oswald pulling the trigger at 12:30pm CST in 1963 in Dallas. It is fair to claim that Kennedy would not have been assassinated on November 23, 1963, in Dallas if Oswald had not pulled the trigger at 12:30pm CST in 1963 in Dallas.

Let us assume the well-known conspiracy theory: J.F. Kennedy was assassinated at the behest of the Mafia, the Cubans, and the CIA. The Mafia wanted revenge on the Kennedys because they had breached an agreement between them. The Cubans wanted revenge on J.F. Kennedy because he did not give them more support in the Bay of Pigs invasion in Cuba. And the CIA considered J.F.K a moderate politician in the face of the Soviet threat at the time. As the conspiracy theory continues, Oswald did not act alone: there was a second sniper, who fired from the grass next to Kennedy’s car. Furthermore, this shot (the so-called “third shot” theory) hit exactly the same spot on Kennedy’s skull at the same time (just like two stones thrown at the same time hitting the same place break a window). The question is: do we now have a genuine case of causal overdetermination? However, to assume that such overdetermination exists is *ad hoc* and unnecessarily complicated. There is no more motivation for introducing it here than there is to try to save the hypothesis that both divine fury and SARS COV-2 simultaneously cause COVID-19. The simple fact is that the scientific explanation of the COVID-19 pandemic displaces the folk explanation that appeals to divine fury. The same is true in the case of phenomenal causation, if phenomenal properties are taken to be irreducibly nonphysical.

Let me return to the case that I am in pain from a strain in the adductor muscle of

my thigh. First, the anti-physicalist must allow that the property of painfulness might have been missing while the underlying token neural state of having my nociceptive specific neurons stimulated retained all its physical properties. Now, if painfulness had not been tokened but the stimulation of the nociceptive specific neurons had retained all its physical makeup, I still would have made the bodily movement with my tongue and asked for ice. Further, that bodily movement with my tongue would have constituted the very same action of asking for ice (in the same way that Oswald pulling the trigger constitutes in part Kennedy's assassination).

Given this, we are back at Dretske's soprano/shattering glass case again. If it is agreed that the property of singing "I love you" has no causal efficacy with respect to the breaking of the wine glass — for the simple reason that if the property had been missing but the physical makeup of the sound sequence constituting the soprano's singing "I love you" had remained the same, the breaking of the glass would still have occurred — what is the difference in the phenomenal case? On the face of it, the property of painfulness seems to be equally causally idle.

5. Minding the Gap

The first obstacle in the way of a physicalist conception of phenomenal causation is the "explanatory gap." How does the soprano's love for her husband arise? How does consciousness itself arise? According to many philosophers, we cannot provide satisfactory answers to these questions. We face the problem of the explanatory gap: for any physical state presented as the basis for consciousness, we can always ask, "Yes, but why does such a state engender feelings?" And that question has no answer. In fact, there are two explanatory gaps. The first is the one formulated above. The second is the following: why does such and such a physical state generate such and such a feeling? Why, for example, does stimulation of nociceptive specific neurons "engender" pain, but not itching, tickling, love, etc.?

Now, if we consider that phenomenal properties have a physical nature (physicalism), the above supposedly complex questions find simple answers. Consider the phenomenal property P (being in pain). If P has a physical nature, there is some physical property that is identical to Q (say the activation of parietal lobe nociceptive specific neurons). Now, accepting the identity between P and Q, the very question of how P came from Q or how Q gave rise to P makes no sense: if P and Q are identical, one cannot be at the origin of the other! If P and Q are identical, how could P originate from Q? There is no explanatory gap.

To be sure, even some physicalists hold that brute metaphysical necessitation of this sort is not strong enough to capture their view, and they take it to be part and parcel of physicalism that a more powerful relation obtains, for example, that

the physical metaphysically *grounds* the phenomenal, where *grounding* is a relation that has an explanatory dimension beyond the one of metaphysical necessitation. For example, the truth of the conjunction of P & Q is grounded in the truth of P and in the truth of Q. Still, grounding is a non-causal metaphysical and explanatory relation between facts or truths that does not seem to apply to identities between properties or particulars. Given this, as to why the phenomenal would be identical to the physical, there is, in fact, no grounding explanation available.

The point is: is there a problem here? I see no problem at all, since identities between entities to which we directly refer need no explanation. All that is needed is a strong correlation that epistemically supports the counterfactual: P (feeling pain) would not have been instantiated if Q (the property of having nociceptive specific neurons stimulated in the parietal lobe) had not been, and vice versa. True, it is always possible that we are scientifically wrong. That is why “water is H₂O” is metaphysically necessary, but emerges as an empirical discovery. But if such metaphysical identities are correct, it makes no more sense to ask additionally why pain is identical with the stimulation of nociceptive specific neurons, just as it makes no sense to ask why Mark Twain is Samuel Clemens.

However, we cannot deny the existence of an “intuition of distinction” between phenomenal and physical properties: intuitively my pain does not seem to be identical with the stimulation of nociceptive neurons (see Papineau 1993). Most likely, religious beliefs are at the basis of such an intuition. Be that as it may, this intuition of distinction is entirely compatible with the thesis of physicalism. But how are we to understand this intuition? According to David Chalmers’s famous modal argument (2010) the answer is known: it seems that we can imagine the instantiation of the physical property without the instantiation of the phenomenal property. If we can imagine such properties instantiated separately, then they could not be identical or phenomenal properties can supervene on physical ones.

The physicalist cannot deny that imagination or conceivability is *prima facie* a guide to metaphysical possibility. However, guidelines are revocable: evidence can be adduced that weakens or refutes what our imagination tells us. And the causal considerations adduced in this essay provide just that recalcitrant element to any appeal to the imagination in defense of “the intuition of distinction” between physical and phenomenal properties. When we seem to imagine the instantiation of a physical property — such as the firing of nociceptive specific neurons — without the instantiation of the phenomenal property of feeling pain, what do we imagine? We imagine, “from the outside,” so to speak, the property of the firing of nociceptive specific neurons being instantiated, while we imagine, “from the inside,” the phenomenal property of experiencing pain as absent. This is the thesis that Nagel defended several decades ago in a footnote to his famous article: “When we imagine something sympathetically, we place ourselves in a state of consciousness similar to the thing itself”

(see Nagel 1974, p.445n). This is what Papineau calls the “antipathetic fallacy” (see Papineau 1993, p.177).

Therefore, the imagination cannot show us that the “intuition of distinction” is correct. Furthermore, there are additional abductive reasons to doubt what intuition is telling us when we remember the problem of phenomenal causality that anti-physicalism faces. If the phenomenal realm supervenes on the microphysical realm, a simple and satisfactory explanation of this supervenience is that there are neuronal properties that would be identical to phenomenal properties. If that is the case, the so-called “intuition of distinction” is a misnomer: there is no genuine explanatory gap between the phenomenal and the physical realms.

6. Internal Physical State View

Now, with the exclusion of anti-physicalism there are still two alternatives that must be appreciated as competing forms of physicalism. According to the first of them, phenomenal properties are intrinsic of the brain and are only accidentally connected with relational properties of experience. This is the position defended by David Papineau today (2021). According to Papineau’s qualitative view:

My positive proposal is to identify the conscious properties of sensory experiences with their vehicle properties rather than their representational properties. As I said earlier, I have no doubt that in the actual world all sensory experiences are representations. But it is not what they represent that fixes their conscious character, but how they represent. By way of analogy, take the sentence ‘Paris is south of London’. This has the representational property of being true if and only if Paris is south of London. But it also has the vehicle properties of being written in Times Roman script, in black letters, 12 point, and so on. Note how in a different world (indeed in a different language) a physical state could have just the same vehicle properties yet not that representational property. (2014, p.18)

The phenomenal properties of sensory experience stand for the properties of written marks of a sentence, while the representational property stand for the meaning or the content of the same sentence (2021, p.31). Given this, the key relation between phenomenal and relational/representational properties of sensory experience is as contingent as the relation between the properties of written marks of a sentence on a piece of paper and the representational content of the same sentence. In the case of language, the explanation for why the relation is contingent is quite obvious. For example, *the content* of the sentence “Paris is south of London” is not natural, depending on conventions of the English language. We could state the same with different sentences in English, with different sentences of different languages. And it

is even possible to utter a same sentence (or at least a same word) meaning different things in other languages, related to English.

The analogy with Dretske's case of the soprano shattering the glass is striking. In the same way that the sequence of sounds has a physical makeup, a certain pitch, amplitude, etc., the written sentence also has a physical makeup: written in Times Roman, 12 point, italic, or bold, etc. Moreover, the content and its vehicle, the sentences, stand in the purely contingent relationship. However, in Dretske's case it is tacitly assumed that the mental properties stand for the content of the sung sentence. In contrast, Papineau assumes that phenomenal properties stand for the physical makeup of the written sentences. Given this, in Papineau's model, phenomenal properties are not *prima facie* causally idle; after all, in analogy, the physical makeup of the sound can make a difference in the physical realm such as shattering the wine glass. Is that enough to account for phenomenal causation of behavior?

Let us focus on the core of Papineau's book (2021), namely his "qualitative view": the claim *that* phenomenal properties are intrinsic properties of people, rather than relational properties of experiences. On one hand, *phenomenal properties are narrowly individuated* as properties of *people*. On the other, *content is always broadly individuated* (2021, p.6). The content of experience is up to the outside world, namely to the environment, containing particulars and distal properties. As he likes to put it: "*Consciousness is something that lights up inside the brain*" (p.50 & p.62) when experience takes place; regardless of what is going on outside the brain. E.g., to have the orange-experience is simply to have a certain neural property (p.29–30, p.95–97).

Now I want to formulate what seems to be the main problem of Papineau's physicalist qualitative view. To be sure, assuming that phenomenal properties are narrowly individuated while content is broadly individuated, and, further, that there is nothing in between, Papineau claims that it is hard to see *how particulars and distal properties make their way into consciousness*. As he puts it: "I don't see how any worldly properties can be *present* in experience" (2021, p.60). My point is that the converse is also true. Again, if phenomenal properties are narrowly individuated, while contents are broadly individuated, and there is nothing in between, *it is also hard to see how phenomenal properties find their way into the outside world of particulars and distal properties*. It is hard to see how consciousness could engage with the outside world.

The problem becomes more acute when we consider that Papineau recognizes only broad contents and, therefore, that the idea of narrowly individuated behavior is out of the question for him: *behavior is always broadly individuated by the broad content that it involves*. Oswald's behavior of murdering Kennedy is broadly individuated, that is, it is individuated by appealing to the Russellian content, which contains *that very particular*, J.F.K, rather than the 35th president of the United States, whosoever he is. Oswald did not murder the 35th president of the United States, because there are possible worlds in which the 35th president of the USA is Richard Nixon.

Oswald did not murder someone in a car in Dallas, Texas. He murdered J.F.K; *that guy straight ahead of him*. But, according to Papineau's qualitative view, as a particular, J.F.K cannot be *present* in Oswald's experience. Again, if consciousness is just something that lights up inside the brain, it is hard to see how Oswald could murder J.F.K.

Let us return to Papineau's analogy between written marks and their representational contents. Now consider an interesting variant case of Dretske's case that Tye introduces. Suppose now that the coloratura soprano is in her dressing room with her adored husband. In addition to that, suppose that she sings playfully to him "I love you." Consequently, he breaks into a big smile (see Tye 2021, p.97). It is true that if the sounds that the soprano produces had not meant *I love you*, her husband would not have broken into a big smile. The behavior is not caused by the sounds she produces, but rather *by the meaning that I love you*.

By way of analogy, let us suppose now that the soprano, instead of singing "I love you," writes the same sentence on a piece of paper and gives it to her beloved husband in her dressing room. Consequently, as before, he breaks into a smile. Again, the question is: what is causing the soprano's husband's behavior of breaking into a big smile? It is certainly not the physical makeup of the marks written on the paper. Rather, what is causing the soprano's husband's behavior is the content/meaning of the written marks on the paper. If she had written the same content in Mandarin, he would not have broken into a big smile (supposing that he did not speak Mandarin). This raises the suspicion that if phenomenal properties are intrinsic properties of the vehicle of content (of sentences, sounds, etc.), which are *only contingently related* with the content itself, they cannot account for a piece of behavior.

This conclusion is reinforced when we remember that behavior is a process in opposition to bodily movements, which are events. Let us suppose just for the sake of argument that the physical makeup of the written marks (the sentence "I love you") is able to make the beloved husband break into a big smile without the intervention of the meaning of the written marks. Still, the soprano's husband's behavior is a process rather than the bodily movement of breaking into a big smile. Now, if conscious properties are internal properties of the brain without any necessary connection to representation of what is going on in the world outside the skull, they cannot account for the process that starts with the soprano's husband reading the written marks ("I love you") on the paper and ends with his breaking into a big smile. Again, meaning is missing.

Finally, Papineau's qualitative view does not fit well with all we know from animal consciousness. To be sure, pain in humans is strongly correlated with the stimulation of nociceptive specific neurons in the parietal lobe. However, cephalopods such as octopuses and squids also feel pain and have more neurons in their tentacles than in their head. In fact, many creatures without a cortex are capable of feeling pain like

we humans. The question is: can we therefore identify “pain” with internal properties of the brain (activation of nociceptive neurons in the parietal lobe)? In regard to the causal role of pain, it seems more natural to assume that in each species a neuronal pathway or something similar is recruited by selection to indicate tissue damage.

7. The Causal Role of Consciousness

The conclusion that seems to follow from the previous section is that without a necessary connection between neural patterns and the instantiation of distal environmental properties in the world outside the skull, there is no hope of understanding the causal role of consciousness in the physical world. By far the best candidate to necessarily connect what is going on inside the brain and what is going on in the world outside the skull are the so-called representational properties. As we anticipated in the last section, the pain (in my hand) causes me to pull my hand from the overheated surface because it represents the property of painfulness.

However, meanings, contents, and representational properties are abstract entities rather than particulars. Consider Dretske’s case of the soprano again. To be sure, the sung sentence “I love you” has a meaning, a representational content. Still, what makes the glass shatter is not the meaning of the sung sentence, but rather the physical makeup of the sequence of sounds: a high pitch, amplitude, etc. If the sentence had quite a different meaning the glass would still shatter, provided only that the physical makeup of the sung sequence of sounds remained unaltered. The present problem is: how can abstract entities enter into concrete causal relationships?

Recently, Papineau (2021) formulated an argument against representationalism based on this problem, which he calls *the-here-and-now argument*:

- 1) Instantiations of conscious sensory properties constitute concrete facts with causes and effects.
- 2) Instantiations of representational properties constitute abstract facts that cannot feature as causes or effects.
- 3) Conscious sensory properties are not representational properties. (2021, p.72)

To begin with, I question 2). I do not see why instantiations of representational properties must constitute abstract facts that cannot feature as causes or effects. To be sure, what features as causes and effects are particulars rather than abstract facts. Still, particular instantiations of abstract properties are still particulars. Moreover, even if what experience represents is abstract, the representing mental states are certainly particulars that cause bodily movements. That is exactly what Dretske suggests by claiming that a belief state C causes some bodily movement M in virtue of indicating the instantiation of an abstract property F. It is not the meaning or the content F

as abstract entities that are causing the bodily movement M. What directly causes M is the doxastic state C. However, C is only recruited by natural selection to cause M *because of what C indicates*, namely the instantiation of the external property F.

Now, if we replace in Dretske's schema the doxastic state C with a phenomenal state C', we can easily account for the causal role of consciousness in a physical world. The first assumption is that there is a nomological covariation between C and the external property F. Now, as the bodily movement is favorable for the adaptation of the species, the phenomenal conscious state C is "recruited" to represent the instantiation of F. The pain in my hand causes me to pull my hand from the overheated surface, but not because it locally supervenes on the stimulation of the nociceptive specific neurons of my parietal lobe (the qualitative view). Instead, my pain causes me to pull my hand from the overheated surface because the stimulation of the nociceptive specific neurons of my parietal lobe is recruited to *represent* the tissue damage. In this case, however, anti-physicalism is false. If pain does *not locally supervene* on the stimulation of the nociceptive specific neurons of the parietal lobe, *it does globally supervene* on the property of having tissue damage that the nociceptive specific neurons represent.

References

- Chalmers, D. J. 2010. *The Character of Consciousness*. New York, NY: Oxford University Press.
- Dretske, F. 1988. *Explaining Behavior, Reasons in a World of Causes*. Cambridge MA: MIT Press.
- Kim, J. 2005. *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.
- Lewis, D. 1986. Postscript to 'Causation'. In: *Philosophical Papers*, vol. 2, 172–213. Oxford: Oxford University Press.
- Nagel, T. 1974. What is it like to be a bat? *Philosophical Review* **83**(October): 435–50.
- Papineau D. 1993. *Philosophical Naturalism*. Oxford: Blackwell.
- Papineau D. 2014. Sensory Experience and Representational Properties. *Proceedings of the Aristotelian Society* **114**(1): 1-33.
- Papineau D. 2021. *The Metaphysics of Sensory Experience*. New York, NY: Oxford University Press.
- Schaffer, J. 2000. Trumping Preemption. *Journal of Philosophy* **97**: 165–81.
- Tye. M. 2021. *Vagueness and the Evolution of Consciousness: Through the Looking Glass*. New York, NY: Oxford University Press.
- Won, C. 2014. Overdetermination, Counterfactuals, and Mental Causation. *Philosophical Review* **123**(2): 205–229.

Notes

¹I took the example "I love you" from Tye's book (2021, p.94).