Phenomenology and the Cognitive Sciences
https://doi.org/10.1007/s11097-024-09978-7

Resolving the evolutionary paradox of consciousness

Brendan P. Zietsch[1]

Accepted: 25 March 2024

Abstract

Evolutionary fitness threats and rewards are associated with subjectively unpleasant and pleasant sensations, respectively. Initially, these correlations appear explainable via adaptation by natural selection. But here I analyse the major metaphysical perspectives on consciousness – physicalism, dualism, and panpsychism – and conclude that none help to understand the adaptive-seeming correlations via adaptation. I also argue that a recently proposed explanation, the phenomenal powers view, has major problems that mean it cannot explain the adaptive-seeming correlations via adaptation either. So the mystery – call it the evolutionary paradox of consciousness – remains. Some have used this mystery to argue for non-naturalistic (e.g. theistic) explanations. But I propose a naturalistic, non-adaptive explanation of the adaptive-seeming correlations: namely, 'sensational associative learning' during development. In this perspective, pairing of particular sensations with unconditioned stimuli – fitness rewards or threats – cause the sensations themselves to come to be interpreted as good or bad, respectively. Sensations, like colours, that are not reliably paired with either fitness rewards or threats remain largely unvalenced. Sensational associative learning also provides explanations for adaptive-seeming structural aspects of sensations, such as the observation that sounds of different pitch are experienced as ordinal in correspondence to their wavelengths while the same is not true of colours of different hue. The sensational associative learning perspective appears compatible with physicalism, panpsychism, and dualism (though not epiphenomenalism).

Keywords Evolution · Psychophysical harmony · Adaptationism · Phenomenal powers · Qualia · Subjective experience

✉ Brendan P. Zietsch
zietsch@psy.uq.edu.au

[1] Centre for Psychology and Evolution, School of Psychology, University of Queensland, Brisbane, Australia

Published online: 01 April 2024

1 Introduction

Phenomenal consciousness (or, simply, consciousness) is experience. As Nagel (1974) put it, for a thing or state to be conscious means there is something it is like to be that thing or in that state. One of the great mysteries of science and philosophy is how and why experience arises from non-experiential constituents. I will not solve this 'hard problem of consciousness' (Chalmers, 1995) here. My target is a different but related question: why does consciousness feel the ways that it does? More specifically, why do the characters (feels) of different sensations[1] tend to align with what seems evolutionarily adaptive? The observation of these adaptive-seeming correlations was memorably put by James (1890):

> *It is a well−known fact that pleasures are generally associated with beneficial, pains with detrimental, experiences. All the fundamental vital processes illustrate this law. Starvation, suffocation, privation of food, drink and sleep, work when exhausted, burns, wounds, inflammation, the effects of poison, are as disagreeable as filling the hungry stomach, enjoying rest and sleep after fatigue, exercise after rest, and a sound skin and unbroken bones at all times, are pleasant. Mr. Spencer and others have suggested that these coincidences are due, not to any pre−established harmony, but to the mere action of natural selection which would certainly kill off in the long−run any breed of creatures to whom the fundamentally noxious experience seemed enjoyable.*

Here, James used the observation to argue against epiphenomenalism (or automaton-theory, as he called it): the metaphysical perspective whereby consciousness is not efficacious (i.e. does not affect the physical world). Instead, James takes the alternative, 'common sense' position that these correlations between the valence of sensations and their fitness consequences are explainable by the action of natural selection on efficacious consciousness.

This position seems sensible, even obvious, given that consciousness as we know it (i.e. as in a human mind) is a biological phenomenon, and evolution is the unifying theory of biology. But careful evolutionary analysis reveals a mysterious paradox: it is not only epiphenomenalism that fails to explain these adaptive-seeming correlations. It turns out that none of the existing metaphysical perspectives on consciousness can easily explain the adaptive-seeming correlations via natural selection. It is this paradox that I aim here to elucidate (following from others: Corabi, 2015; Mørch, 2017; Robinson, 2014), and to which I propose a resolution.

[1] By sensations I mean the same as what Tye (2021) defines as qualia: "(1) Perceptual experiences, for example, experiences of the sort involved in seeing green, hearing loud trumpets, tasting liquorice, smelling the sea air, handling a piece of fur. (2) Bodily sensations, for example, feeling a twinge of pain, feeling an itch, feeling hungry, having a stomach ache, feeling hot, feeling dizzy. Think here also of experiences such as those present during orgasm or while running flat-out. (3) Felt reactions or passions or emotions, for example, feeling delight, lust, fear, love, feeling grief, jealousy, regret. (4) Felt moods, for example, feeling elated, depressed, calm, bored, tense, miserable." I avoid the term qualia because it is sometimes defined or interpreted in limiting ways (e.g. as nonphysical) that would be confusing in this article (Tye, 2021).

I proceed as follows. In § 2 I expand on why consciousness seems adaptive – this goes beyond the alignment of pleasant and unpleasant sensations with fitness rewards and threats, to structural aspects of sensations. In § 3 I discuss how the major metaphysical perspectives on consciousness, physicalism, dualism, and panpsychism, might help understand the adaptive-seeming correlations. I conclude that they do not help. In § 4 I consider Mørch's (2017) phenomenal powers view, which aims to explain the adaptive-seeming correlations by proposing that sensations have intrinsic causal powers such that unpleasant sensations metaphysically necessitate attempts at avoidance and pleasant sensations metaphysically necessitate attempts at approach. I outline major problems with this view and conclude that it cannot explain the adaptive-seeming correlations. In § 5 I propose a non-adaptationist explanation of the adaptive-seeming correlations, involving associative learning. In § 6 I briefly analyse how such an explanation fits (or does not fit) with the major metaphysical perspectives on consciousness, and in § 7 I wrap up.

2 The characters and structures of sensations *seem* adaptive

Adaptations are features that have arisen through natural selection because they have helped organisms survive and pass on their genes (i.e. they have increased organisms' fitness). The character and structure of an adaptation is connected to its function: for example, the shape of a bird's wings and the feathers they wear, combined with the behavioural adaptation of moving them in particular patterns, enable fitness-enhancing flight. Many of our sensations seem like adaptations in the same way, given their subjective characters and the fitness-relevant situations in which they arise.

When we go too long without food we experience a certain unpleasant sensation. Looking over a cliff edge gives another bad sensation, touching a hotplate yet another. We get unpleasant sensations from smelling something noxious, hearing unexplained sounds in the dark, being socially or sexually rejected, and failing in front of a large audience. All these situations would have posed threats to fitness during evolution, whether by starvation, falling, burning, poisoning, predation, ostracization, reputation damage, and so on. On the other side of the coin, situations that might have benefited fitness during evolution tend to generate various sorts of pleasant sensations: for example, eating an energy-rich meal, having sex, winning a public contest, being safe and dry in a stormy night, receiving a kindness, and so on.

The degree of this alignment is not one that can be easily accounted for by chance. If we take (as a crude illustration) only the thirteen example situations I just mentioned and assume the accompanying sensations could each be pleasant, unpleasant, or neutral with equal probability, the chance of them aligning such that the sensations accompanying fitness rewards all felt pleasant and the sensations accompanying fitness threats all felt unpleasant would be one in 1,594,323 (i.e. 1/3 to the power of 13). The probability that all our valenced sensations would align as well as they do with ancestral fitness contingencies by coincidence would be vanishingly small.

Not all sensations feel pleasant or unpleasant. Many simple colours or sounds, for example, are simply neutral in valence, though their individual characters are just as distinctive nonetheless. But sensations of these sorts also exhibit adaptive-

seeming correlations, with respect to their structure. For example, experiences of hue and pitch both involve distinguishing different wavelengths; blue light (~470 nm) is associated with an experience different from green light (~530 nm), and middle C (~1.3 m) is experienced differently from middle D (~1.2 m). But the structure of how the experiences of different hues relate to each other differs from that for different pitches. Middle C and D (and any other pitches) are experienced as ordinal with respect to each other – anyone can arrange simple sounds from lowest to highest wavelength based purely on the character of the sensations, how high or low they 'feel' – whereas the same does not apply to colours: without learned knowledge, a person cannot typically arrange different colours in order of their wavelength based on how they 'feel'[2]. This difference in ordinality between sensations of hue and pitch can be seen to match fitness contingencies too: the wavelength of sound correlates ordinally with the size (and thus with the fitness threat) of the object, predator, or competitor making the sound, whereas the wavelength of light does not correlate ordinally with anything of adaptive significance.

Overall, given that the alignment between the character and structure of sensations and their ancestral fitness contingencies *seems* adaptive and is vanishingly unlikely to have come about by chance, the obvious conclusion is that they *are* adaptive and accordingly came about via adaptation by natural selection. This is where James's (1890) analysis of the issue rested. But when we start to consider *how* the character of sensations could be relevant to survival and reproduction, the position seems far less clear.

Take the aforementioned hotplate example. It may seem to us that the pain sensation we experience upon touching a hot surface causes us to withdraw our hand and minimise bodily damage. But the withdrawal reflex occurs *before* any pain is experienced – by the time the sensation of pain arrives, the damage minimisation function has already been achieved. How, then, did the pain sensation's particular character contribute to fitness? This temporal precedence does not apply to deliberate behaviours, of course, but the more general point remains: any behavioural response to a harmful stimulus is presumed to occur through the transmission of neural signals, albeit via far more complex routes in a deliberate response compared with a reflexive response. What difference, then, does the particular character of the associated subjective sensation make, as long as appropriate behaviours occur through the cause and effect of neural transmissions? Without an obvious answer to this, it is unclear how the characters of sensations could be adaptations, and thus how adaptation could explain the adaptive-seeming correlations.

As we will see, explaining how the character of sensations could be naturally selected, and thus how the adaptive-seeming correlations could be explained via adaptation, is difficult no matter which of the major metaphysical perspectives one takes on the nature of consciousness in terms of its relation to the physical world. But the different perspectives – physicalism, dualism, and panpsychism – present different sets of difficulties, and I will discuss these for each perspective next.

[2] For example, the wavelength order of purple, red, and green is not usually obvious without thinking of one's knowledge of the light spectrum.

3 How might the metaphysical perspectives on consciousness help to understand the adaptive-seeming correlations?

3.1 Physicalism

Physicalism is the idea that everything is physical: a closed system of matter-energy and physical forces (gravitation, electromagnetism, and the strong and weak interactions) operating in space-time[3]. All causation is physical too: an organism's behaviour can in principle be entirely explained by tracing backwards in time through a cause-and-effect chain of purely physical laws and events (e.g. neuronal firings, body movements). Subjective experience (i.e. consciousness) does not feature in the laws of this closed system in any way – not as a cause or effect or substance or property. Indeed, it is not obvious how subjective experience fits in a physicalist perspective at all. Different physicalist perspectives attempt to reconcile this problem in various ways.

For example, mind-brain identity theorists posit that an experience *is* a brain state – not that my brain state produces, causes, gives rise to, or has the non-physical property of, my experience of joy, but that they are identical, one and the same thing. Everything that is true of the experience of joy, for example, must be true of the physical brain processes. Illusionists, on the other hand, posit that I (and they) are in the grip of an *illusion* of consciousness – my experience of redness does not really exist, I just mistakenly believe that it exists (e.g. Dennett, 2018; Frankish, 2016; Humphrey, 2020)[4]. Various problems with physicalist perspectives have been raised – especially that they seem to deny the very phenomenon in question. Formal arguments against physicalism include the knowledge argument (Jackson, 1982), the conceivability (or zombie) argument (Chalmers, 1996), and the explanatory gap argument (Levine, 1983). There has been much debate about the validity of these arguments and of physicalism itself, but the details are beyond the scope of this article.

How does physicalism relate to the adaptive-seeming correlations? At first, mind-brain identity theory might seem a promising avenue for understanding them. If fitness-benefiting neural processes are adaptations, then the subset of these that are identical to conscious states must also be adaptations. Calling the conscious states adaptations, which mind-brain identity theory trivially enables, may seem helpful, because adaptations are normally explainable by natural selection. But mind-brain identity theory (given natural selection) does not, *a priori*, predict or specify or make likely any particular relations or pattern of associations between fitness contingencies and the characters of accompanying sensations (e.g. adaptive-seeming correlations). For example, the identities represented by the dashed lines (which do not yield adaptive-seeming correlations) in Fig. 1 are, *a priori*, equally consistent with physicalism, given natural selection, as are the identities represented by the solid lines (i.e. which

[3] I omit versions of physicalism in which 'physical' encompasses anything causally relevant for matter-energy, even if it is beyond our current conceptions of physics, since that stretches the definition beyond any usefulness here. That is, per Chalmers (2015), I use physicalism to refer to 'narrow physicalism'.

[4] By denying the reality of consciousness, illusionism eliminates many of the quandaries its existence raises, including the evolutionary paradox described here, so I do not consider it further.

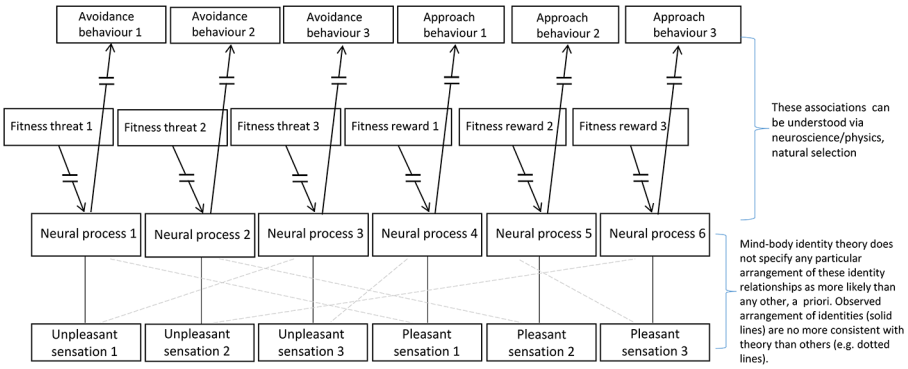

Fig. 1 Mind-brain identity theory does not help to understand the adaptive-seeming correlations. Line breaks denote other neural processes not shown

do yield adaptive-seeming correlations). Therefore, after observing the actual pattern of associations between experiential characters and fitness contingencies (the solid lines), mind-brain identity theory does not receive any support or boost in likelihood of being true. In other words, there is no connection between the truth (or not) of physicalism and the existence (or not) of the adaptive-seeming correlations, so assuming the truth of the former gets us no closer to understanding the latter (Corabi, 2015; Mørch, 2017; Robinson, 2014).

3.2 Dualism

Dualism is the perspective that the body/brain and mind (consciousness) are distinct – that is, consciousness is ontologically irreducible to the physical. Though dualism is often associated with Descartes' position that a mind can exist independent of a body (i.e. Cartesian dualism), most non-theological theorists discuss dualism while assuming that consciousness depends on (or more precisely, supervenes on) a physical substrate such as the brain (Chalmers, 2017b). I will only discuss this naturalistic dualism here.

If the physical and mental are distinct, how do they interact? Epiphenomenalism and interactionism have different answers to this question and therefore have different evolutionary considerations. According to epiphenomenalism, physical events (e.g. brain processes) cause (or have the property of) consciousness, but conscious events do not affect physical events. Compared to physicalism, epiphenomenalism has the advantage of explicitly allowing for consciousness as part of natural reality. Along with physicalism, it also avoids the problem of overdetermination – i.e. physical events being affected by more than what are presumed to be wholly sufficient physical causes[5] – that would arise if consciousness affected physical events.

Some prominent theorists have dismissed epiphenomenalism as inherently inconsistent with evolutionary theory, based on the idea that a functionless feature (inert

[5] Here, and throughout the paper, I assume no randomness – not through any commitment to determinism, but merely to avoid necessitating a tangential qualification at every turn.

consciousness) would disappear or not evolve in the first place (e.g. Popper, 1978; Romanes, 1895). It is true that if conscious states do not affect behaviour or other physical states of the organism, they cannot provide a fitness advantage, and therefore cannot be adaptations. But functionless features can nonetheless evolve and persist. Such features may be byproducts of adaptive features: for example, men's nipples are functionless byproducts of women's nipples, which are adapted for delivering milk to infants. Conscious states could be functionally irrelevant byproducts of the kind of information processing that promotes adaptive behaviour (Broad, 1925; Jackson, 1982; Robinson et al., 2015). Epiphenomenal consciousness is therefore not inherently inconsistent with evolutionary theory. But if conscious states are functionally irrelevant byproducts and not adaptations, then the question remains unanswered as to why the particular characters of these byproducts align so well with ancestral fitness contingencies (Corabi, 2015; Mørch, 2017; Robinson, 2014). That is, epiphenomenalism cannot explain the adaptive-seeming correlations via adaptation, and another explanation is not obvious[6].

The alternative dualistic perspective, interactionism, proposes that conscious states are not only caused by physical events, but also themselves play causal roles. *How* they might do so is generally not specified, and indeed the very notion of physical events caused by conscious events or properties raises difficult problems. In particular, it seems to lead to the aforementioned situation of overdetermination, where physical events are caused by more than what are presumed to be wholly sufficient physical causes. Interactionalists are essentially proposing that physics is not in fact causally closed as is widely assumed (though not known)[7], and that there are undiscovered psychophysical laws that govern how physical arrangements give rise to consciousness and how conscious events can affect physical events.

Interactionism seems to provide a framework that could help us understand the match between the particular character of sensations and their ancestral fitness contingencies. In principle, conscious states could provide fitness advantages to experiencing organisms by affecting their behaviour. Perhaps these consciousness-influenced processes are simpler or more efficient, and thus more likely to evolve, than nonconscious computational processes that would achieve the same outcomes.[8] Thus interactionism is compatible with conscious states as adaptations. But as with

[6] Robinson (2023) suggests an epiphenomenalist explanation that requires supporting the view that "[w]hat "pleasure" refers to in any possible world is the effect in consciousness of [neural events in a reward system] that contribute to continuance or repetition". Defined as such, there is no problem to solve regarding the adaptive-seeming correlations, since the neural events that contribute to continuance or repetition will of course be associated with fitness rewards given natural selection, and pleasure is defined as whatever sensation goes along with those. But in this paper I assume that sensations are defined by their subjective character rather than by the behavioural effects of the associated neural events. For the same reason, I do not discuss analytic functionalism, which defines mental states solely by their functions (e.g. pain is the state that tends to be caused by bodily injury).

[7] Although note that Sturgeon (1998) argues that even if the microphysical (i.e. quantum mechanical) is causally closed, it does not mean that consciousness could not cause macrophysical occurrences (like human behaviours). This qualification applies in later parts of the paper too, but I do not restate it each time.

[8] Worth noting here is that it is not obvious that there are any fitness benefits conferred by consciousness that could not in principle be achieved by non-conscious processing.

mind-body identity physicalism, interactionist dualism (given natural selection) does not, *a priori*, predict or make likely any particular pattern of associations between fitness contingencies and the particular characters of different sensations (e.g. adaptive-seeming correlations). For example, the associations represented by the dashed lines (which do not yield adaptive-seeming correlations) in Fig. 2 are, *a priori*, equally consistent with interactionist dualism, given natural selection, as are the identities represented by the solid lines (i.e. which do yield adaptive-seeming correlations). Therefore, after observing the actual pattern of associations between fitness contingencies and the particular characters of sensations (the solid lines), interactionist dualism does not receive any support or boost in likelihood of being true. As there is no connection between the truth (or not) of interactionist dualism and the existence (or not) of the adaptive-seeming correlations, assuming the truth of the former gets us no closer to understanding the latter (Corabi, 2015; Mørch, 2017; Robinson, 2014).

3.3 Panpsychism

Panpsychism is the idea that consciousness is a fundamental and ubiquitous aspect of the universe. Contemporary perspectives normally hold that some kind of basic physical entities, say quarks, have subjective experience (i.e. there is something it is like to be a quark). In particular, a popular version is Russellian panpsychism, according to which consciousness is the actual 'stuff' each fundamental particle consists of (i.e. its quiddity), as opposed to its dispositions (i.e. how it behaves in relation to other things, which is described by the standard laws of physics). By taking consciousness to be the intrinsic nature of basic physical entities, Russellian panpsychism avoids the questions of how/why/from where consciousness arises, to which physicalism has no answers and to which dualism postulates undiscovered, fundamental psychophysical laws.

However, although it avoids these particular questions, Russellian panpsychism still faces the similar question of how the complex and unified consciousness that we experience could come about, if all that is fundamental is the micro-experiences of quarks or other basic entities. This is known as the combination problem (Chalmers, 2017a). There are two basic approaches to this problem (Chalmers, 2015). One,

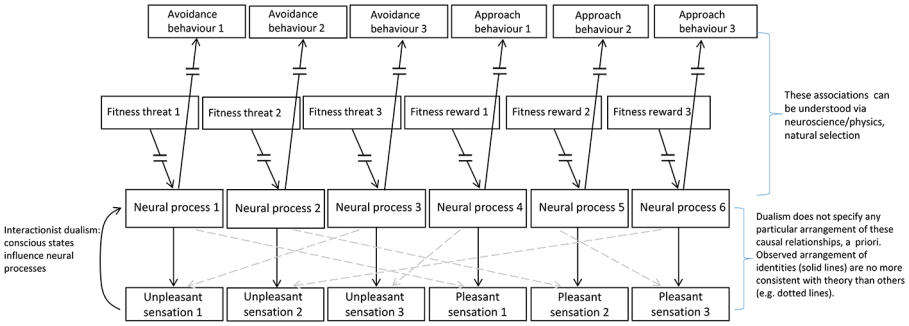

Fig. 2 Standard dualism does not help to understand the adaptive-seeming correlations. Line breaks denote other neural processes not shown

constitutive panpsychism, posits that micro-experiences of basic entities somehow add up to macro-experiences like ours. It is difficult to imagine how such aggregation would work, and Goff (2015) argues that it is implausible. The other approach, non-constitutive or emergent panpsychism, posits that macro-experience is a strongly emergent phenomenon that arises with certain arrangements of micro-experiences. (Strong emergence occurs when processes at 'lower' levels of organisation (in this case, micro-experiences) give rise to something new (in this case, macro-experiences) that is fundamentally undeducible from lower-level facts (Chalmers, 2006).) Emergent panpsychism shares many of the difficulties and advantages of dualism – it needs to posit undiscovered laws by which micro-experiences in certain arrangements yield macro-experiences. On the other hand, one could argue there is less explanatory work to go from micro- to macro-experiences than to go from purely physical entities to macro-experiences, as is the challenge with dualism.

Macro-experiences can be adaptations in Russellian panpsychism, either because they are constituted of micro-experiences, which have causal powers in virtue of being the categorical bases of physical dispositions, or because the macro-experiences are emergent properties that exert downward causation over the lower-level happenings in a similar way to interactionist dualism. In any case, in terms of understanding the adaptive-seeming correlations, Russellian panpsychism does not fare any better than physicalism or dualism. The same arguments as I have made regarding those perspectives apply to Russellian panpsychism's constitutive and emergent versions, respectively. Basically, the theory does not, *a priori*, predict any particular pattern of associations between fitness contingencies and experiential characters, so it does not help to understand the pattern of associations we observe (i.e. the adaptive-seeming correlations).

4 Mørch's phenomenal powers theory

None of the major metaphysical theories of consciousness help us understand the evolutionary paradox of the adaptive-seeming correlations, given natural selection. The reason they do not help is that none of the theories, as they are typically presented, predict any particular arrangement of different experiential characters with respect to ancestral fitness contingencies: for example, that fitness threats should tend to be associated with sensations that feel unpleasant, like pain, disgust, or fear, while fitness rewards should tend to be associated with sensations that feel pleasant, like physical pleasure, pride, or gratitude. As far as the standard theories are concerned, stubbing one's toe might just as well have been associated with the sensation we know as green instead of with the one we know as pain.

But there is a clash with intuition here. The sensation we know as pain seems connected to physical harm in some kind of intrinsic, necessary way, whereas the sensation we know as green does not. Intuitively, it seems the green sensation would not motivate us to avoid it (and the associated physical harm) the way pain does. If the connection between pain and harm – and the connections of other pains and pleasures with fitness threats and rewards – were truly intrinsic and necessary, then perhaps we have the ingredients for a theory that predicts the adaptive correlations?

Mørch's (2017) 'phenomenal powers' perspective aims to provide this theory. Mørch reinforces the aforementioned intuition with a thought experiment involving a girl, Maya, with a congenital insensitivity to pain who suddenly has this condition cured. Soon after, she steps on a nail and for the first time feels pain, and intense pain at that. According to Mørch, Maya would immediately be able to predict that this is a sensation that will always make her, and anyone else who feels the same sensation, try to avoid it. Because she does not have to learn this through repeated associations, but knows it simply from the way it feels, Mørch argues that pain must have intrinsic phenomenal powers (and likewise pleasure, in the opposite direction), and that these phenomenal powers explain the adaptive-seeming correlations. In further support, she argues that it is inconceivable, to anyone that has felt pain, that it would have anything other than a repulsive effect. This perspective is, according to Mørch (2017; p. 311), the only avenue for intelligibly explaining the adaptive-seeming correlations. If that is the case, then the fact of the adaptive-seeming correlations constitutes strong support for a phenomenal powers view. In my view, though, there are major problems with phenomenal powers view, and there is an alternative explanation that does not share these problems. I will discuss the problems now, and in the next section (§ 5) I will describe the alternative explanation.

Key to Mørch's argument for the phenomenal powers theory is the thought experiment involving Maya. But the thought experiment is flawed. First, it does not lend itself to consideration of the pain in and of itself, because the pain is accompanied by the visceral invocation of bodily harm in the form of a nail through the foot. Therefore, it may be bodily harm, rather than the sensation of pain, that we struggle to imagine not finding repulsive. But let's say a modified thought experiment presented the hypothetical intense pain unaccompanied by an image of bodily harm, and let's say we still found it hard to imagine not finding the sensation repulsive. In this case, still, we could only consider the sensation having experienced a lifetime of associations, starting from before we can even remember, between pain and bodily harm. Being informed that Maya has never before felt pain (and thus does not have these associations) does not enable us to unlearn our own deep associations and consider the hypothetical pain sensation afresh. So even this modified thought experiment may be tapping our intuitions about bodily harm, rather than about pain in and of itself. The thought experiment does not clearly demonstrate that someone who has never felt the sensation that normally accompanies bodily harm (i.e. pain) would, on feeling it for the first time, immediately be able to predict that it will always make her, and any other subjects who experience it, try to avoid it.

But still, might it indeed be inconceivable that pain could make anyone try to do something other than avoid it, in virtue of how it feels? I do not think so. I do not have difficulty imagining pain being something other than repulsive. In fact, I do not even need to imagine it. I frequently bite the inside of my cheek in order to produce a pain. I like the sensation. I can increase the pressure so that the sensation increases in intensity, and at a certain point – just before it would probably damage my cheek tissue – I do not like the sensation anymore. But up until the point near which I risk bodily damage, I like it. And I am not alone. Pain itself, separated from (danger of) bodily harm, is commonly sought out – a phenomenon called 'benign masochism'

(Rozin et al., 2013). If anyone seeks pain in virtue of how it feels, the phenomenal powers view is disconfirmed.

Mørch dismisses masochism generally by assuming it is not the pain itself that is sought in such cases, but an accompanying pleasure (Mørch, 2017; p 303). I do not deny that this may be an accurate account of many instances of masochism – sexual masochism, for example, involves sensation (sexual pleasure) of a different form than the pain that gives rise to it, and it seems a plausible hypothesis that the sexual pleasure is enjoyed while the pain in itself is not. But when I bite my cheek, it is the pain sensation itself that I seek and that I like. The pain sensation does not give rise to another, different sensation, that motivates me to endure the unpleasant pain sensation. If I am correctly describing my own experience – and others I have talked to report similar experiences – then these first-person data are difficult to reconcile with the phenomenal powers view[9].

Another difficulty with the phenomenal powers view is that it is very difficult to see how a sensation (let's stick with pain) could genuinely (metaphysically) necessitate attempts at avoidance. First, there are very many ways to avoid pain (or attempt to avoid it), and the particular way an individual attempts to avoid pain in any given instance will depend on the context: for example, the source of the pain, the available means to address it, the subject's beliefs about the relative pros and cons of those different means, and so on. If my head is in pain, I could take a painkiller or drink some water, if those are available, or I could distract myself by attending to something else; if my hand is in pain, I could take my hand out of the hot water stream, or I could turn the tap off. Each of these different actions (or attempts at them) involve different neural processes and phenomenal experiences. Clearly, a pain sensation cannot metaphysically necessitate all attempts at avoidance, since many are mutually exclusive. And it cannot metaphysically necessitate just one specific kind of attempt at avoidance, since in most instances that one kind of attempt would be inappropriate, ineffective, or nonsensical (e.g. turning a hot tap off when there is no hot tap and no water, because I have stubbed my toe). So what exactly is supposed to be metaphysically necessitated by pain? One or more of many completely different actions (or attempts at such) which each involve different neural processes and phenomenal experiences, the choice of which is contingent on various contextual variables both internal and external to the subject of the pain. The hypothesis that such a contingent relationship is metaphysically necessary (i.e. true in all possible worlds) seems quite contradictory. For example, there are surely possible worlds with pain but without beliefs or means to attempt avoidance, on which the relationship (pain necessitating avoidance attempts) seems to depend.

A further difficulty with the phenomenal powers view stems from the fact that pain sensation and attempts at avoidance have different neural bases in different parts of

[9] Some objections might be raised. One might argue that if I like the cheek-biting sensation and seek it out then it is not pain. But if one defines pain sensation as being disliked and avoided, then the phenomenal powers view would be true by definition and would therefore lack explanatory value. Alternatively, one might argue that I am mischaracterising my cheek-biting experience: that the sensations that I like (low bite force) and dislike (high bite force) are qualitatively different sensations rather than the same sensation at quantitatively different intensities. To this, I can only say I am testing it right now and that is not my experience.

the brain[10]. According to a standard understanding of brain functioning, insofar as a pain sensation leads to one of the possible avoidance attempts (in virtue of how it feels), it would do so via neural pathways. In principle, then, a genius mad scientist could rewire those neural pathways so that a pain sensation led to an attempt at some action other than avoidance. The phenomenal powers view holds that any such alternative wiring is impossible (Mørch, personal communication); but there does not seem any obvious reason to believe in such impossibility, other than the thought experiment and inconceivability claim that I have already challenged. Even if introspection led someone to feel (with his current brain wiring) very strongly as though his pain necessitates that he try to avoid it in virtue of how it feels, he could still acknowledge, without contradiction[11], that if his brain were differently wired then the same pain sensation might then feel as though it necessitates some other action in virtue of how it feels. (As an analogy, the taste of raw tomato is repulsive to me and it feels necessary that I try to avoid it in virtue of how it tastes – it is difficult to imagine otherwise – but I readily acknowledge that if my brain were wired differently I might love and enthusiastically pursue that same taste sensation in virtue of how it feels.)

Given all this, there seems not to be strong reason to believe that pain sensation is intrinsically bad or that it metaphysically necessitates attempts to avoid it in virtue of how it feels; indeed, the hypothesis seems conceptually problematic and at odds with first-person data. If we do not believe this intrinsic badness of pain (or goodness of pleasure) and the metaphysical necessity of connections between pain and avoidance (and between pleasure and approach), then the adaptive-seeming correlations remain unaccounted for. But I earlier foreshadowed an alternative explanation, and I will discuss that next.

5 A non-adaptationist, naturalistic explanation for the adaptive-seeming correlations: sensational associative learning

The difficulties with the phenomenal powers view seem to undermine our only working explanation of the adaptive-seeming correlations (Mørch, 2017; p. 311). From somewhat different angles, other theorists have also argued that adaptation cannot explain what they call "cognitive fine-tuning" (Goff, 2018) or "psychophysical harmony" (Cutter & Crummett, in press; Goff, 2023), which involve the adaptive-seeming correlations discussed in this paper. Those authors propose theism (Cutter & Crummett, in press; Goff, 2018), panagentialism (Goff, 2023), or other solutions

[10] Involved in pain are the somatosensory organs, afferent nerve fibres relaying signals from these to the spinal cord and then to the brain's 'pain matrix', which includes the thalamus and the primary and secondary somatosensory, insular, anterior cingulate, and prefrontal cortices. The brain's pain matrix is activated even during hypnotically induced pain where there is no physical pain input from the body, suggesting the brain activity is sufficient for the experience of pain (Tracey & Mantyh, 2007). Usually, attempts at avoidance would primarily involve the motor cortex, which is key to the planning, control, and execution of voluntary movements.

[11] Assuming one's feelings about something don't necessarily correspond to the truth of that thing or one's beliefs about it. For example, I feel as though I have free will, but that doesn't necessarily mean that I actually have free will or believe that I have it.

"that are wildly at odds with naturalism about the mind" (Goff, 2018). Is there an alternative, naturalistic explanation?

I propose such an explanation – one that involves implicit learning of associations between sensations and fitness-relevant stimuli. According to this explanation, unpleasant sensations are associated with fitness threats (and pleasant sensations with fitness rewards) not because natural selection has harnessed intrinsic causal powers of the sensations, but because individuals learn during development to interpret as bad (or good) those sensations that are associated with fitness threats (or rewards). (For the sake of brevity, I will stick mainly with the bad side while explaining this idea, but an equivalent argument applies to the good side.) On this view, interpreting a thing as bad (i.e. to be avoided[12]) after associative learning is a cognitive process that is not itself conscious but can affect one's experience of the thing. "Interpreting as bad" does not itself entail an unpleasant conscious state – for example, one can interpret as bad SARS-CoV-2[13] without feeling bad. But interpreting as bad *one's own sensation* is what we mean when we say that sensation feels unpleasant (or repulsive, or that we dislike it).

Associative learning occurs when an organism perceives contingency relations between paired events (Jozefowiez, 2012). It can happen from just one pairing, as in one-shot learning, as well as incrementally over many pairings (Lee et al., 2015). The earliest studies on associative learning usually focussed on animal behaviour. For example, an animal might receive a fitness-relevant stimulus (i.e. an unconditioned stimulus, e.g. an electric shock) around the same time as a fitness-irrelevant stimulus (i.e. a conditioned stimulus, e.g. a red light). After repeated pairings, the animal comes to avoid the red light even when it is presented alone. Related findings, in humans, extended these effects to subjective valence – a conditioned stimulus comes to be disliked (or liked) if it is paired with a negative (or positive) unconditioned stimulus (Moran et al., 2022; Zanna et al., 1970). The (dis)liking can be revealed by either explicit self-report or implicit tests.

In a nutshell, I am proposing that the sensation of pain is a conditioned stimulus like the red light – except that acquaintance with the sensation is direct rather than mediated through the senses as in the apprehension of the physical red light[14]. To

[12] The principles of evolution by natural selection ensure that, in general, things interpreted as 'to be avoided' are fitness threats – individuals that interpreted fitness rewards as 'to be avoided' would not survive and reproduce well. Evolution by natural selection also determines what things are instinctively interpreted as to be avoided (or to be approached) – that is, what are unconditioned stimuli. Note that interpreting something as bad or good does not require high-level concepts about fitness or harms or benefits – again, information as to what is a fitness threat or reward is encoded into the organism by natural selection. For example, even creatures as simple as honeybees contain the information as to what is a fitness threat (unconditioned stimulus), in that they instinctively avoid it and learn to avoid any other thing (conditioned stimulus) associated with it. (Note that even if one held the view that responses to fitness relevant stimuli were driven by pain and pleasure, a well-adapted organism would still need to contain the information as to what stimuli should produce pain and what stimuli should produce pleasure in the first place, i.e. what are fitness threats and rewards. That is, the pain and pleasure cannot convey to the organism any extra information it does not already have).

[13] The virus, not the experience of being infected with it.

[14] Acquaintance, in the philosophical sense, is a kind of relation in which a subject is directly aware of something, unmediated by anything else (such as the physical processes involved in the senses) (Duncan,

expand: The starting point is that some neural processes correlate with sensations of particular characters (how and why this is true is the hard problem of consciousness, which I have already said I will not solve here). Such a sensation has no intrinsic causal power or even intrinsic valence. Furthermore, the characters of sensations provoked by given stimuli might differ naturally among individuals. For any individual, associative learning, from very early in development[15], causes sensations that accompany the neural processes involved in the detection and avoidance of fitness threats (i.e. unconditioned stimuli) to come to be (1) interpreted as bad, and (2) avoided. Likewise, sensations paired with fitness rewards come to be (1) interpreted as good, and (2) approached. The question of the adaptive-seeming correlations – why fitness threats feel unpleasant and fitness rewards feel pleasant – is thus answered.

This 'sensational associative learning' explanation involves the idea that an overall experience tends to comprise both a sensation and how the sensation is interpreted[16] – that is, the experience of a sensation interpreted in one way feels overall different from the experience of the same sensation interpreted another way. To illustrate the concept, think of looking at a particular face: the visual sensation is the same whether one recognises the face or not, or whether one finds it beautiful or not, but the overall experience differs depending on these interpretations. This example is useful because the interpretation is readily distinguished from the sensation, whereas in the case of pain, the sensation is more easily conflated with its interpretation as bad (because pain sensation so reliably accompanies physical harm). But there is evidence that the sensation of pain can be separated from its interpretation as bad. The most striking case is a condition called pain asymbolia. Individuals with this condition experience the sensation of pain, but they do not interpret it as bad and make no effort to avoid it (Bain, 2014). One might argue that this situation is unique to these disordered individuals – who typically have stroke-related brain damage – and that it says little as to whether pain sensation and badness are separable in healthy individuals[17]. But a simple thought experiment might help healthy individuals intuitively see the separability of pain and its interpretation as bad. Press a fingernail into your palm until you feel a pain sensation. When I do that, I do not interpret it as very bad, and the total experience is only mildly if at all unpleasant. But now imagine you were to feel the same sensation with the same intensity but located unexplainably in your eyeball. I would immediately interpret that sensation as very bad and the total experience would be highly unpleasant. This intuitive contrast in (hypothetical) experiences, one okay and one very unpleasant, based on the same pain sensation in different contexts, makes

2021). One's apprehension of the physical object emitting light of a certain wavelength is mediated through the senses, but one's acquaintance with the sensation of redness is, on this view, direct, i.e. unmediated.

[15] Evidence suggests normal associative learning begins before birth (Kawai, 2010); presumably sensational associative learning would be possible from whatever stage conscious sensations arise, which is unknown.

[16] But I don't claim that experience must always involve interpretation. Raw sensation would still be an experience, though whether that can occur in normal human consciousness or perhaps during meditation or drug-influenced states or in other creatures' minds is an open question.

[17] It nonetheless speaks to the possibility for alternative brain wirings that make pain not necessitate attempts at avoidance, which are impossible according to the phenomenal powers view.

sense if the unpleasantness of pain depends on how the sensation is interpreted, but is harder to explain if pain's unpleasantness is intrinsic.

This conception of pain experience as a composite of the sensation and its interpretation is consonant with evaluativist accounts of pain (Bain, 2017), though they are quite different accounts in general. Evaluative accounts pertain to negative evaluation of bodily damage represented by pain, rather than the pain sensation itself, as is the case in sensational associative learning. And I use 'interpretation' here rather than 'evaluation' because the latter word normally connotes the deliberative weighing of value, whereas the interpreting that I have in mind neither is deliberative nor necessarily relates to value or valence (I will return to this latter point shortly). Further, though I have been referring mainly to pain for convenience, the kind of interpretation I am describing can apply to any hedonic sensations, such as fear, disgust, guilt, shame, sadness, sexual pleasure, joy, pride, gratitude, satisfaction, excitement, and so on. Sensations that tightly co-occur with either (ancestral) fitness threats or rewards are learned to be interpreted as bad or good, respectively, in a way that can seem hard to imagine otherwise. Other sensations may have statistical association with fitness-relevant outcomes, but not in a consistent direction (with regard to threat or reward) – these sensations, such as surprise, excitement, or anticipatory 'nerves', may thus be easier to imagine having hedonic valence in either (or neither) direction. In such cases, the hedonic valence of a particular instance may depend more on real-time interpretation of the circumstances – for example, the sensation of anticipatory nerves might be interpreted negatively before a public speech one has not sufficiently prepared, for example, or positively before a date with an appealing potential partner. Still other sensations (e.g. of colours) may have little or no statistical association with fitness-relevant outcomes, in which case they would not have a valenced interpretation. That the sensational associative learning view easily accommodates ambivalent and neutral sensations is a major advantage over the phenomenal powers view, which struggles to do the same (Mørch, 2017).

Another advantage of the sensational associative learning view is that it can be applied beyond hedonic valence to structural aspects of experience that seem adaptive. As mentioned earlier, higher-frequency pitches 'feel' higher than lower-frequency pitches, whereas higher-frequency colours do not typically feel higher in the same ordinal way. I suggested that this seems adaptive because the wavelength of sound correlates ordinally with the size (and thus fitness threat) of the object or predator making the sound, whereas the wavelength of light does not correlate ordinally with anything of adaptive significance[18]. The implication was that we might have evolved to experience sounds and colours in this discrepant way. But another possibility is that we learn the ordinality of sound experiences of different wavelengths during development, for example because we hear our own voice and control the vibration frequency of our own vocal folds (and hence vocal pitch), and because pitch

[18] Some readers might contend that colour sensations do feel ordinal in the sense that warm-feeling colours (yellows, reds, oranges) have longer wavelengths than cool-feeling colours (blues and greens). I would suggest that the coolness or warmness of the colour feelings is because we implicitly learn the association of the former colours with certain warm things (sunlight and fire) and the latter with certain cool things (shadow and water). I would also suggest that there is only limited correspondence of ordinality of colour feelings with respect to wavelength. Does green feel cooler than blue or violet?

correlates with the size of sound-producing objects, whereas there is nothing similar for colour. That is, sound experiences come to *feel* ordinal because they are correlated to physical generators of which we perceive the ordinality (and control, in the case of our own voices).

To sum up, the sensational associative learning view can account for a wide range of observations that make up the adaptive-seeming correlations. This contribution is in itself important, given that the phenomenal powers view heretofore seemed to be the only naturalistic avenue for explaining these correlations. But more than simply providing one more avenue, I contend that this new explanation does not share the considerable problems of the phenomenal powers view, while at the same time offering other advantages: for example, the associative learning explanation can account for adaptive-seeming structural aspects of consciousness that the phenomenal powers view cannot, and it can also explain why some sensations seem to have a clear valence (e.g. pain) while others have little or no valence (e.g. green) or ambiguous valence (e.g. surprise)[19].

6 How does sensational associative learning fit with metaphysical perspectives of consciousness?

I earlier discussed the compatibility of various metaphysical perspectives with consciousness as an adaptation, since adaptation seemed the most straightforward way to explain the adaptive-seeming correlations. Having come to the conclusion that adaptation appears not to explain the adaptive-seeming correlations after all, and that sensational associative learning might explain them instead, I will now sketch how the latter might fit with the major metaphysical perspectives on consciousness.

A difficulty in considering how sensational associative learning might fit with physicalism is that there is not a well-established, intelligible way that consciousness itself fits with physicalism more broadly. Nevertheless, if we simply equate sensations to brain states, then sensational associative learning implies, in the case of pain for example, that the relevant learned association is between bodily harm and a particular brain state. In standard associative learning, the associated stimuli are perceived by the organism – for example, a red light (conditioned stimulus) and electric shock (unconditioned stimulus). Can an organism perceive its own physical brain states? Physicalists might say yes: to be aware of a conscious experience is to be aware of one's physical brain state. This concept may clash with common intuition, since when I am aware of an experience I have no understanding of what neural transmissions are going on in my brain. Again, this clash is an issue for physicalism more broadly; the sensational associative learning explanation of the adaptive-seeming correlations raises no obvious further issues regarding the physicalist perspective.

Dualism in general presents no obvious conceptual incompatibility with sensational associative learning. In the pain example, the learned association is between bodily harm and pain sensations, the latter of which the organism is directly acquainted

[19] The phenomenal powers view takes all sensations to be intrinsically motivational and causal, so it is not straightforward to explain these apparent differences in motivation or valence.

with. But what of the compatibility of epiphenomenalist and interactionist versions of dualism? The associative learning explanation does not give consciousness intrinsic causal power to motivate behaviour like Mørch's phenomenal powers view does, but that does not mean it fits with epiphenomenalism. In the latter perspective, as it is typically conveyed, behaviour (and the physical world in general) would be the same in the presence or absence of consciousness (Robinson, 2023). But the associative learning explanation, in a dualist framework, relies on immaterial sensations being involved in a learned association, which itself inevitably affects behaviour. That leaves a dualist in the realm of interactionism, in which the existence of phenomenal consciousness makes a difference to the physical world, albeit only indirectly.

Russellian panpsychism attributes consciousness to fundamental particles, but these unchanging micro-experiences pre-date biological evolution. It is the macro-experiences of individual organisms that are relevant with respect to the fit of sensational associative learning with Russellian panpsychism, and in that regard the considerations of its constitutive and emergent versions echo those of the physicalist and dualist perspectives, respectively. For constitutive panpsychism, a macro-experience is an aggregation of the microexperiences of the organism's constituent particles, and while this idea may itself present formidable difficulties (Goff, 2015), it has the advantage over physicalism that the organism will be directly acquainted with its brain state via that phenomenal constitutive relation (thus somewhat resolving the counterintuitive notion of an organism perceiving its own physical brain state). Emergent panpsychism, on the other hand, presents the same considerations as dualism – that is to say, there are no obvious conceptual barriers to fitting it with sensational associative learning[20].

7 Conclusion

The consciousness we know of is that of evolved biological organisms. But on my analysis, none of the major metaphysical perspectives on consciousness help to understand why the particular characters of conscious sensations *seem* so adaptive, given natural selection – i.e. why fitness rewards feel pleasant, fitness threats feel unpleasant, and structural aspects of consciousness line-up with fitness contingencies. My explanation is that we come, through associative learning in development, to interpret sensations in ways that happen to mirror adaptationist intuitions. On this view, sensations are not intrinsically motivational or valenced and do not have intrinsic causal powers. Nevertheless, sensations may affect the physical world because of our direct acquaintance with them, and because of their involvement in our evolved learning mechanisms. This sensational associative learning explanation seems to be compatible with all the major metaphysical perspectives on consciousness, except epiphenomenalism.

[20] However, one might say that Russellian panpsychism takes consciousness to be intrinsic, whereas on the sensational associative learning view, an overall experience depends not only on the intrinsic nature of a sensation but also on its interpretation (Mørch, personal communication).

Acknowledgements For comments and discussion I am very grateful to Hedda Mørch, Philip Goff, Adam Bulley, Jonathan Redshaw, Natasha Anderson, and anonymous reviewers.

Author contributions Brendan Zietsch was solely responsible for writing the paper.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Data availability There were no data or materials used.

Declarations

Ethical approval was not required for this study as it did not involve any human or animal subjects.

Competing interests The authors have no relevant financial or non-financial interests to disclose.

Informed consent Not applicable.

Statement regarding research involving human participants and/or animals Not applicable.

References

Bain, D. (2014). Pains that don't hurt. *Australasian Journal of Philosophy*, *92*(2), 305–320.

Bain, D. (2017). Evaluativist accounts of pain's unpleasantness. *The Routledge Handbook of Philosophy of Pain* (pp. 40–50). Routledge.

Broad, C. D. (1925). *The mind and its place in nature*. Routledge.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford Paperbacks.

Chalmers, D. J. (2006). Strong and weak emergence. In P. Clayton, & P. Davies (Eds.), *The re-emergence of emergence* (pp. 244–254). Oxford University Press.

Chalmers, D. J. (2015). Panpsychism and panprotopsychism. *Consciousness in the physical world: Perspectives on Russellian monism*, 246–276.

Chalmers, D. J. (2017a). The combination problem for panpsychism. *Panpsychism: Contemporary Perspectives*, *179*, 214.

Chalmers, D. J. (2017b). Naturalistic Dualism. In *The Blackwell Companion to Consciousness* (pp. 363–373).

Corabi, J. (2015). The misuse and failure of the evolutionary argument. *Disputatio*, *6*(39), 199–227.

Cutter, B., & Crummett, D. (in press). *Psychophysical harmony: A New Argument for Theism*. Oxford Studies in Philosophy of Religion.

Dennett, D. C. (2018). Facing up to the hard question of consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1755), 20170342. https://doi.org/10.1098/rstb.2017.0342.

Duncan, M. (2021). Acquaintance. *Philosophy Compass*, *16*(3), e12727. https://doi.org/10.1111/phc3.12727.

Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, *23*(11–12), 11–39.
Goff, P. (2015). Against constitutive Russellian monism. *Consciousness in the physical world: Perspectives on Russellian monism*, 370–400.
Goff, P. (2018). Conscious thought and the cognitive fine-tuning problem. *The Philosophical Quarterly*, *68*(270), 98–122.
Goff, P. (2023). *Why? The purpose of the Universe*. Oxford University Press.
Humphrey, N. (2020). The invention of consciousness. *Topoi*, *39*(1), 13–21. https://doi.org/10.1007/s11245-017-9498-0.
Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly*, *32*, 127–136.
James, W. (1890). *The principles of psychology* (Vol. 1). Macmillan.
Jozefowiez, J. (2012). Associative learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of Learning* (pp. 330–334). Springer US.
Kawai, N. (2010). Towards a new study on associative learning in human fetuses: Fetal associative learning in primates. *Infant and Child Development*, *19*(1), 55–59. https://doi.org/10.1002/icd.654.
Lee, S. W., O'Doherty, J. P., & Shimojo, S. (2015). Neural computations mediating one-shot learning in the human brain. *PLOS Biology*, *13*(4), e1002137. https://doi.org/10.1371/journal.pbio.1002137.
Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, *64*(4), 354–361.
Moran, T., Bar-Anan, Y., & Nudler, Y. (2022). Evaluative Conditioning: Past, Present, and Future. *Annual Review of Psychology, in press*.
Mørch, H. H. (2017). The evolutionary argument for phenomenal powers. *Philosophical Perspectives*, *31*(1), 293–316. https://doi.org/10.1111/phpe.12096.
Nagel, T. (1974). What is it like to be a bat? *Readings in Philosophy of Psychology*, *1*, 159–168.
Popper, K. (1978). Natural selection and the emergence of mind. *Dialectica, 32*(3/4), 339–355. Retrieved from http://www.jstor.org/stable/42970324.
Robinson, W. S. (2014). James's evolutionary argument. *Disputatio*, *6*, 229–237.
Robinson, W. S. (2023). Epiphenomenalism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Vol. Summer 2023 Edition). Retrieved from https://plato.stanford.edu/archives/sum2023/entries/epiphenomenalism/.
Robinson, Z., Maley, C. J., & Piccinini, G. (2015). Is consciousness a Spandrel? *Journal of the American Philosophical Association*, *1*(2), 365–383. https://doi.org/10.1017/apa.2014.10.
Romanes, G. J. (1895). *Mind and motion and monism*. Longmans, Green and Co.
Rozin, P., Guillot, L., Fincher, K., Rozin, A., & Tsukayama, E. (2013). Glad to be sad, and other examples of benign masochism. *Judgment and Decision Making*, *8*(4), 439–447.
Sturgeon, S. (1998). Physicalism and overdetermination. *Mind*, *107*(426), 411–432.
Tracey, I., & Mantyh, P. W. (2007). The cerebral signature for pain perception and its modulation. *Neuron*, *55*(3), 377–391. https://doi.org/10.1016/j.neuron.2007.07.012.
Tye, M. (2021). Qualia. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition ed.).
Zanna, M. P., Kiesler, C. A., & Pilkonis, P. A. (1970). Positive and negative attitudinal affect established by classical conditioning. *Journal of Personality and Social Psychology*, *14*(4), 321.