The Evolutionary Argument for Phenomenal Powers

Hedda Hassel Mørch

Forthcoming in *Philosophical Perspectives*

1 Introduction

Epiphenomenalism is the view that phenomenal properties – which characterize *what it is like*, or how it feels, for a subject to be in conscious states – have no physical effects. One of the earliest arguments against epiphenomenalism is the evolutionary argument (James 1890/1981; Eccles and Popper 1977; Popper 1978), which starts from the following problem: why is pain correlated with stimuli detrimental to survival and reproduction – such as suffocation, hunger and burning? And why is pleasure correlated with stimuli beneficial to survival and reproduction – such as eating and breathing? According to the argument, the fact that we have these particular correlations and not other ones must have an evolutionary explanation. But given epiphenomenalism, differences in phenomenal properties could not cause differences in fitness, so natural selection would not be expected to favor these correlations over any other ones. Epiphenomenalism thus renders these correlations an inexplicable coincidence, and should therefore be rejected.

The evolutionary argument has been widely criticized and few have deemed it cogent (Broad 1925; Jackson 1982; Robinson 2007; Corabi 2014). In this paper, I will consider previous and potential criticisms and conclude some of them are indeed fatal to the argument if it is understood, as it traditionally has been, as an argument for any standard version of non-epiphenomenalism such as physicalism and interactionism. I will then offer a new and improved version of the argument, as an argument for a particular non-epiphenomenalist view, which I will call the phenomenal powers view. This is the view that phenomenal properties *produce* and thereby (metaphysically)

necessitate their effects in virtue of how they feel, or in virtue of their intrinsic, phenomenal character alone – along the lines of C. B. Martin and John Heil's powerful qualities view (Martin and Heil 1999; Heil 2003). I will argue that the phenomenal powers view explains the correlations given natural selection far better than any other view. It follows that if (and only if) understood as an argument for the phenomenal powers view, the evolutionary argument is far stronger than it is usually thought to be.

The phenomenal powers view is logically and *prima facie* compatible with any non-epiphenomenalist view, including physicalism, interactionist dualism and Russellian monism (Russell 1927; Strawson 2006; Alter and Nagasawa 2012; Chalmers 2013). However, upon closer examination, we will see that it might not be fully compatible with physicalism in view of some further considerations. This may partially explain why the phenomenal powers view has so far been overlooked as a conclusion of the evolutionary argument.

2 Background and Overview

The classic version of the evolutionary argument is due to William James:

> There is … [a] set of facts which seem explicable on the supposition that consciousness has causal efficacy. *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. An animal

> that should take pleasure in a feeling of suffocation would, if that pleasure were efficacious enough to make him immerse his head in water, enjoy a longevity of four or five minutes. But if pleasures and pains have no efficacy, one does not see … why the most noxious acts, such as burning, might not give thrills of delight, and the most necessary ones, such as breathing, cause agony. (James 1890/1981: 146, emphasis original)

Another influential version of the argument is due to Karl Popper:

> … the theory of natural selection constitutes a strong argument against Huxley's theory of the one-sided action of body on mind and for the mutual interaction of mind and body … If Huxley had been right, mind would be useless. But then, it could not have evolved, no doubt over long periods of time, by natural selection. (Popper 1978: 350)

Popper (building on Eccles and Popper 1977) appears to be offering a deductive argument, which can be reconstructed as follows:

1. Phenomenal properties evolved.
2. If epiphenomenalism is true, phenomenal properties are useless.
3. Useless features do not evolve.
4. Therefore, epiphenomenalism is false.

The main weakness of this argument, anticipated by C. D. Broad (1925) and later pointed out by Frank Jackson (1982), is that its third premise is false: useless features do evolve. Sometimes useless features are correlated with useful, adaptive features as by-products, or "spandrels" (as Gould and Lewontin called them). Phenomenal properties could still have evolved as inert by-products of adaptive physical properties.

James, on the other hand, appears to offer an abductive argument, i.e., an inference to the best explanation. The explanandum is the following correlations:

- Pain is correlated with stimuli detrimental to survival and reproduction, such as suffocation, starvation and burning.
- Pleasure is correlated with stimuli beneficial to survival and reproduction, such as air and food.

In what follows, I will refer to these facts as "the correlations", and regard them as pertaining to *phenomenal* pain and pleasure (as opposed to, e.g., merely bodily, neurological, representational or functional states). I will understand the terms pain and pleasure broadly, so that pain includes all kinds of phenomenal properties with clear negative hedonic value, such as feelings of discomfort associated with, e.g., suffocation, and pleasure includes all kinds of phenomenal properties with clear positive hedonic value, such as feelings of comfort associated with, e.g., rest after fatigue.

James implies that the best explanation for the correlations is that the theory of natural selection is true and epiphenomenalism is false. His reasoning can be interpreted roughly as follows. If epiphenomenalism is false, pain causes creatures to avoid correlated stimuli and pleasure causes creatures to pursue correlated stimuli. Any creature for whom pain happens (by mutation or some other source of biological variation) to be correlated with beneficial stimuli will then be caused to avoid beneficial stimuli, and therefore be selected against, and that any creature for whom pain happens to be correlated with harmful stimuli will be caused to avoid harmful stimuli, and therefore be selected for – and *mutatis mutandis* for pleasure. Non-epiphenomenalism thereby explains the correlations in the sense of predicting them, i.e., increasing their probability or rendering them more to be expected. In contrast, epiphenomenalism is also *compatible* with the correlations, but gives them lower probability than non-epiphenomenalism. If pain and pleasure do not cause behavior, they could still have ended up as by-products of physical properties that do, but there

would be no reason to expect this. It is equally likely that they would have ended up as by-products of some other physical properties.

Now, prediction or probability-raising is arguably not the only virtue of a good explanation – the best explanations are not only predictive, but also have other virtues such as simplicity, elegance and so on. But epiphenomenalism seems at best equal to non-epiphenomenalism with respect to other explanatory virtues. Therefore, inference to the best explanation of the correlations seems to favor non-epiphenomenalism overall, given its predictive advantage.

As an abductive argument, the evolutionary argument seems more compelling than in its simple, deductive form. But it is nevertheless subject to criticism. First of all, one might ask: do the correlations require an explanation in the first place – of the kind suggested by James? One might think the correlations require no explanation, i.e., are best regarded as a matter of chance or coincidence. Or, that they are best explained by appeal to anthropic reasoning – as some claim is the case for the problem of fine-tuning for life. In section 3, I will argue that this is not the case for the correlations – they really do call for, or at least suggest, an explanation of a non-anthropic kind.

Second of all, one might ask: do non-epiphenomenalist views, such as physicalism and interactionism, really explain the correlations better than epiphenomenalism, given natural selection? Critics such as William Robinson (2007) and Joseph Corabi (2014) have argued that they actually do not. Their criticism can be paraphrased as follows. As noted, epiphenomenalism could explain the correlations given natural selection by positing one-way psychophysical laws according to which pain is a by-product of physical properties that cause avoidance in particular, and pleasure is a by-product of physical properties that cause pursuit in particular. But this does not seem like a very good explanation, roughly because these one-way psychophysical laws seem

as inexplicable or improbable as the correlations themselves. However, in order to explain the correlations given natural selection, it seems non-epiphenomenalist views must also posit specific principles: interactionism must posit specific two-way psychophysical laws according to which pain causes avoidance and pleasure causes pursuit in particular (i.e., as opposed to other kinds of behavior or non-behavioral effects), whereas physicalism must posit specific psychophysical identities according to which pain is identical with physical properties that cause avoidance, and pleasure is identical with physical properties that cause pursuit in particular. According to the criticism, these posited laws or identities seem just as inexplicable or improbable as epiphenomenalism's one-way laws.

In section 4, I will consider this criticism in more detail. I will first consider prediction or probability-raising, and conclude that no *standard* version of non-epiphenomenalism increases the probability of the correlations relative to corresponding versions of epiphenomenalism. Mainstream versions of non-epiphenomenalism tend to (at least implicitly) deny the phenomenal powers view, so by a standard version of any non-epiphenomenalist view, I will mean any version that denies the phenomenal powers view (as will be discussed, the phenomenal powers view is at least logically and *prima facie* compatible with all general non-epiphenomenalist views, but not entailed by either).

I will then consider other explanatory virtues. I will note that standard views could increase the probability of the correlations by positing specific psychophysical laws or identities, but would thereby sacrifice what I will call *generality* and *ultimacy*, i.e., they would make the explanation as complex as the explanandum and give rise to explanatory regress. Therefore, no standard view seems to explain the correlations very well when explanatory virtues besides prediction are taken into account.

In section 5, I introduce the phenomenal powers view, the view that phenomenal properties produce their effects in virtue of how they feel, or their phenomenal character. As noted, this can be regarded as a version of the powerful qualities view (Martin and Heil 1999; Heil 2003; Strawson 2008), which is in turn a version of dispositional essentialism (Shoemaker 1980; Mumford 2004; Bird 2007). I will argue that the phenomenal powers view predicts the correlations as well as specific versions of the standard views, but without itself positing any specific principles (i.e., principles that explicitly link pain and pleasure to particular powers). It thereby explains the correlations on a more general basis and without generating explanatory regress, i.e., while preserving the virtues of generality and ultimacy. I will also argue that other versions of dispositional essentialism do not share all these explanatory virtues. Inference to the best explanation therefore favors the phenomenal powers view alone.

Finally, in section 6, I will consider how the view fits with physicalism, interactionism, as well as Russellian monism. Note that, for simplicity, I will set Russellian monism aside until the final section because there is no reason to believe that standard Russellian monism (i.e., versions that deny the phenomenal powers view) explains the correlations better given natural selection than standard interactionism or physicalism.

3 Chance and Anthropic Explanations

The evolutionary argument, as posed by James, presupposes that the correlations require an explanation – of the sort that increases their probability relative to chance. But why could the correlations not just be a matter of chance?

An intuitive response is that the correlations would be a highly fortunate coincidence, i.e., one that we should be happy or thankful for. If the correlations were different, but our behavior remained

roughly the same (which would be required for the correlations to still evolve given natural selection), life would be at worst terrible and at best absurd. Imagine, for example, that pain were correlated with beneficial stimuli such air and food, but we still found ourselves pursuing them. Or, that pain were correlated with ubiquitous neutral stimuli, such as seeing colors or hearing ambient noise, but we still found ourselves not avoiding it despite the painfulness. It seems highly fortunate that we do not find ourselves in any of these situations. And intuitively, it might seem fortunate facts require explanation.

But it is not clear that this intuition is valid in general. Why would fortunate facts be less likely to result from chance than neutral or unfortunate ones? The more fundamental intuition seems to be that fortunate facts are more likely than non-fortunate ones to be brought about by an *agent*, because we know that agents often have both a preference for and a capacity to bring about facts that are fortunate in some sense. Fortunate facts may therefore suggest an explanation specifically in terms of intentional agency, but they may not suggest any other kind of explanation. For example, if someone wins the lottery 10 times in a row, this suggests cheating, i.e., that an intentional agent for whom this outcome would be fortunate deliberately did something to cause it. But if the cheating hypothesis can be independently ruled out, it seems reasonable to conclude that the outcome was a matter of chance, insofar as there is no non-intentional, purely mechanistic process that could be expected to tend towards the particular lottery numbers selected by this player.

Or, consider the problem of fine-tuning for life. According to this problem, it is highly improbable, given current physics, that the fundamental physical constants are such as to allow for life to exist. If some physical constants were different, the kinds of structures necessary for life could not arise, and the range of physically possible constants is enormous. This has prompted scientists and

philosophers alike to look for an explanation that reduces the improbability. The fact that the existence of life strikes us as fortunate seems to motivate theistic proposals in particular, according to which a divine agent which also finds life valuable fine-tuned the constants in order to bring about life. Yet, as Roger White (2007) has argued, if theism and other hypotheses that posit an intentional bias for what we perceive as valuable are set aside, it is not so clear that fine-tuning for life suggests a non-chancy explanation – because there is no reason to expect that any non-intentional process should be biased toward life, similarly to how there is no reason to expect a non-intentional bias for particular lottery numbers. Given atheism, then, there may not be any strong reason to expect that life did not arise by pure chance.

The problem of the correlations could, in a similar way, be regarded as a problem of *hedonic* fine-tuning: why are the psychophysical laws such as to allow for the correlations to evolve? Furthermore, the evolutionary argument also sets aside theism – or what James refers to above as "pre-established harmony" – as a potential explanation of the correlations. In this paper, I will follow James in assuming that this is justified – at least as long as an adequate non-theistic alternative explanation is available, as I will go on to argue is the case for the correlations.[1] But if theistic explanations – the only kind of intentional explanation that seems applicable to the correlations – are set aside, the fact that the correlations are fortunate no longer seems to suggest that they have an explanation.

[1] This can be justified by appeal to further problems with theism, such as lack of parsimony, the problem of evil (which could be regarded as either logically incompatible with theism or as weighty counterevidence to it), inherent metaphysical problems, and so on. However, those who do not find theism inherently problematic, and also see it as a good explanation of the correlations, are welcome to read this paper as supporting the more limited claim that inference to the best explanation for the correlations does not *univocally* support theism, but rather a disjunction of theism and the phenomenal powers view, i.e., the non-theistic explanation that I will go on to defend.

But the correlations nevertheless suggest an explanation simply because they are fairly improbable given chance. Given chance, it seems just as probable that the correlations would be inverted as that they would be roughly the same. Additionally, one could imagine scenarios where the correlations are different but not quite inverted, for example, that pain is neutral and pleasure is harmful, or pain is beneficial and pleasure is neutral, or both pain and pleasure are neutral.[2] It seems these scenarios should also be given some probability. The probability of the correlations being different would then add up to more (perhaps far more) than 0.5 (this would not render the actual correlations as improbable as fine-tuning for life, of course, but they would still improbable enough to motivate looking explanations that would render them less so). Furthermore, even if there is no non-intentional, non-theistic bias for life, there could still be a non-intentional bias for the correlations, such as some form of non-epiphenomenalism. If some form of non-epiphenomenalism renders the correlations more probable than chance, this would support the hypothesis and discredit the chance hypothesis.

One might think the correlations could still be explained in other ways. Some think the problem of fine-tuning for life can be resolved by appeal to a multiverse hypothesis combined with anthropic reasoning. The multiverse hypothesis says that every physically possible universe including a fine-tuned one has a high probability of existing as one out of many in a multiverse. The anthropic principle says that our universe is necessarily one of the fine-tuned ones, because if it were not we would not exist, but a universe where we do not exist would not be *our* universe. It is, of course, highly controversial for a number of reasons whether this would really constitute a good explanation of fine-tuning for life. But assuming it could, one might think the correlations

[2] These scenarios can be filled in in such a way that all kinds of stimuli are correlated with phenomenal properties by adding that various kinds of hedonically neutral phenomenal properties (such as color experiences) are correlated with harmful or beneficial stimuli.

could be explained in an analogous way. However, with respect to the correlations, the anthropic principle would in any case not apply – because in any universe where the correlations are different there would still necessarily be living, phenomenally conscious creatures around to experience them. Hence, the only candidate for a non-theistic explanation of the correlations seems to be some form of non-epiphenomenalism.

4 Standard Non-Epiphenomenalism

Assuming theism, the multiverse hypothesis and further non-intentional bias hypotheses can be set aside, the question is whether non-epiphenomenalism explains the correlations any better than epiphenomenalism given natural selection. I will first consider whether standard versions of physicalism and interactionism render the correlations more probable than epiphenomenalism given natural selection. I will then consider whether any other explanatory virtues separate them.[3]

As noted, critics of the abductive version of the evolutionary argument have denied that non-epiphenomenalism contributes to a better evolutionary explanation. Corabi (2014) puts this criticism explicitly in terms of prediction or probability-raising. One reason epiphenomenalism seems like a bad explanation is that it only predicts that pain and pleasure are inert by-products of *some* causally efficacious physical properties, as opposed to physical properties that cause avoidance or pursuit in particular. But, as Corabi points out, interactionism and physicalism are no different from epiphenomenalism in this regard. Interactionism only predicts that there are *some*

[3] The relationship between purely probabilistic, or Bayesian, reasoning and abductive reasoning is contentious. In this paper, I focus on the explanatory virtues of prediction, generality and ultimacy, and leave it open how precisely generality and ultimacy relate to prediction or probability-raising (which according to Bayesian reasoning is the only feature relevant for confirmation). However, one way they could be related is that other explanatory virtues are relevant in a Bayesian framework by influencing the prior probability of a hypothesis. For example, generality might increase the prior probability of a hypothesis because general hypotheses are disjunctions of more specific ones. With ultimacy, one might think a belief that ontological explanatory connections cannot descend infinitely should (at least in some cases) increase the probability of explanatory hypotheses that demonstratively do not do so.

two-way psychophysical laws, but it does not predict that there are laws according to which pain causes avoidance in particular, and pleasure causes pursuit in particular, as opposed to other kinds of effects. Physicalism predicts that phenomenal properties are identical to, or constituted by, *some* physical properties, but it does not predict that pain is identical to a physical property that causes avoidance, and that pleasure is identical to a physical property that causes pursuit, as opposed to other physical properties with other effects. Therefore, the probability of the correlations is not any higher given physicalism or interactionism than it is given epiphenomenalism.

Physicalists might respond that identities are metaphysically necessary. Therefore, even though general physicalism does not increase the *epistemic* probability of the correlations, it increases their *metaphysical* probability. Corabi retorts that since inference to the best explanation is an epistemic matter, non-epiphenomenalist hypotheses cannot be supported by an inference to the best explanation for the correlations unless they render them less epistemically contingent (2014: 220). This seems reasonable. Consider the hypothesis of necessitarian determinism – the view that everything is determined according to some metaphysically necessary, deterministic laws and metaphysically necessary initial conditions. This hypothesis raises the metaphysical probability of any actual event to 1, but does not increase the epistemic probability (i.e., enable the prediction) of any particular event at all. If we want to reject, as seems reasonable, that necessitarian determinism is strongly supported by an inference to the best explanation for everything, we need to say that only epistemic probability is relevant to such inferences.

A further problem is that necessitarian versions of either epiphenomenalist or interactionist dualism, according to which there are metaphysically necessary (one-way or two-way) psychophysical laws, will increase the purely metaphysical probability of the correlations as much as physicalism. Therefore, physicalism may not have an advantage even with respect to purely

metaphysical probability (unless necessitarianism about laws can somehow be dismissed as incoherent, but there is no obvious case for this). [4]

There is one general version of physicalism, analytic functionalism, given which the epistemic probability of the correlations would be close to 1. According to this view, it is analytically true that phenomenal properties are identical to (or otherwise associated with) their actual causal roles. Nowadays, analytic functionalism is widely regarded as deeply problematic and the majority of physicalists hold that psychophysical identities are a posteriori, not analytic and thus a priori. But would the evolutionary argument nevertheless lend analytic functionalism some support? It would be odd to employ the evolutionary argument, construed as an inference to the best explanation, in support of analytic functionalism, because if psychophysical identities were analytic the correlations should not strike us a puzzling and fortunate and in need of explanation in the first place – insofar as we fully grasp the concepts of pain and pleasure (and the background theory of natural selection). In general, it seems analytic views are best defended by direct appeal to conceptual analysis rather than indirectly by appeal to their explanatory value.[5] Therefore, analytic functionalism is not among the non-epiphenomenalist hypotheses that would be well supported by the argument, and can be set aside as a possible conclusion of it.

[4] Necessitarianism about laws is often based on dispositional essentialism. As will be discussed, the phenomenal powers view is a version of dispositional essentialism and therefore a necessitarian view, but unlike other necessitarian views this also increases the *epistemic* probability of the correlations (or so I will argue below).

[5] Analytic functionalists might say that although the psychofunctional identities or correlations are in principle a priori discoverable analytic truths, they still are not obvious, but only discoverable upon thorough reflection. Still, it would be odd if they were so hard to discover that they are better defended indirectly by appeal to their explanatory value than by appeal to instructions for how to directly discover them a priori by conceptual analysis. If a priori conceptual truths are so hard to discover that they are better defended indirectly, this indicates that they are not really a priori conceptual truths after all (at least when it comes to relatively simple alleged truths about psychofunctional correlations—with complex mathematical truths it might be different). Appeal to anything like the evolutionary argument therefore seems self-undermining for analytic functionalism.

Note that this is not to say that the evolutionary argument constitutes an argument *against* analytic functionalism, or that it can be categorically set aside as a possible challenge to it. By presupposing that the correlations are puzzling, the evolutionary argument also presupposes that analytic functionalism is false, but it does not provide any new reason to think so. If a new conceptual analysis were to come along to convincingly demonstrate an analytic connection between phenomenal and functional concepts, the evolutionary argument would be undermined. But it seems highly plausible that most people in fact have phenomenal concepts that are logically distinct from functional concepts.

General versions of physical and interactionism thereby seem no better off with respect to prediction or probability-raising than general versions of epiphenomenalism. But one might also consider specific versions of the views. Physicalists can point to the more specific hypothesis that pain and pleasure are (a posteriori) identical, not to some physical property or other, but to precisely those physical properties that future science will reveal as neurological causes of avoidance or pursuit behavior respectively – or to these particular causal roles themselves, as per (a posteriori, non-analytic) functionalism. Interactionists can point to the more specific hypothesis that pain and pleasure respectively cause avoidance and pursuit behavior in particular. These specific hypotheses would greatly increase the probability of the correlations. But this leads to the obvious problem that epiphenomenalists can make the exact same move. Epiphenomenalists can point to the specific hypothesis that pain and pleasure are by-products of particular physical properties that cause avoidance or pursuit respectively, which would render the correlations just as probable as specific versions of non-epiphenomenalism.

Another problem is that these specific hypotheses are not very explanatory if we take other explanatory virtues into account. As mentioned, prediction is arguably not the only explanatory

virtue. We tend to think that the best explanations are *general*, i.e., that they allow specific facts to be derived from general principles. Hypotheses that list specific psychophysical identities or laws are no more general than the correlations they are supposed to explain, because the number of facts that need to be individually stated in the explanation is equal to the number of correlations to be explained. We also tend to think the best explanations are the more *ultimate* ones, i.e., those that do not give rise to a regress of further explanatory questions. The specific versions of non-epiphenomenalism give rise to further questions such as: *why* these particular psychophysical laws, or *why* these particular psychophysical identities? In this way, they do not seem fully satisfactory.

One might object, on behalf of physicalism, that identities are primitive relations that cannot in principle be further explained. Robinson (2007: 30) responds that psychophysical identities seem inexplicable or unintelligible in a way other identities (such as Mark Twain=Samuel Clemens, or water=H_2O) do not. This point has also been defended by David Chalmers (2003: 113). Therefore, physicalism should at least explain why psychophysical identities, unlike other purely physical identities, are *not* further explicable, or why they intuitively appear to require an explanation even though they actually do not. This is something many physicalists acknowledge – the so-called phenomenal concept strategy (Loar 1997; Papineau 2002) is an attempt to explain why psychophysical identities appear inexplicable.[6]

In summary, non-epiphenomenalism and epiphenomenalism seem to be on an equal explanatory footing. General versions of physicalism and interactionism do not predict the correlations.

[6] According to Chalmers (2007), the phenomenal concept strategy gives rise to an explanatory regress about the phenomenal concepts themselves. However, if this and other important objections to the phenomenal concept strategy can be answered, physicalism together with the phenomenal concept strategy could perhaps be regarded as just as explanatory as the phenomenal powers view with respect to the virtue of ultimacy. But specific versions of physicalism (without the phenomenal powers view) would still be worse off than the phenomenal powers view with respect to generality (as I will argue below).

Specific versions that do predict them are matched by equally specific versions of epiphenomenalism, and lack the virtues of generality and ultimacy.

The evolutionary argument may therefore seem powerless against epiphenomenalism, as seems to be the conclusion of most philosophers who have considered it since James and Popper. I will now argue that it is actually far from powerless – as there remains one particular non-epiphenomenalist view, which critics have so far overlooked, which explains the correlations significantly better than all the views discussed so far: the phenomenal powers view.

5 The Phenomenal Powers View

The phenomenal powers view is the view that phenomenal properties are intrinsically powerful, which is to say that they *produce* or *bring about* their effects, or *make* them happen, in virtue of their intrinsic character alone. Production should be understood a defeasible necessary connection: causes that produce their effects metaphysically necessitate their effects *ceteris absentibus,* that is, in the absence of interference from other powerful causes. Furthermore, the view takes the intrinsic character of phenomenal properties to consist in their phenomenal, qualitative character, i.e., what it is like or how it feels for a subject to experience them. According to the view, then, pain *makes* subjects who experience it try to avoid it simply in virtue of feeling bad, and pleasure makes subjects try to pursue it simply in virtue of feeling good.

The phenomenal powers view is clearly distinct from the regularity theory (Hume 1739; Lewis 1973) and the governing laws view (Armstrong 1978) of causation as applied to phenomenal properties. The regularity theory claims that causes do not necessitate their effects but are rather merely contingently followed by them. The governing laws view takes causes to necessitate their

effects, but in virtue of external irreducible laws or relations, not in virtue of their intrinsic character.

It is also distinct from other forms of realism about causal powers as applied to phenomenal properties. The main version of realism about causal powers is dispositional essentialism. Dispositional essentialism takes dispositions to be irreducible and roughly equivalent to powers as defined above. Dispositional essentialism also takes properties to be essentially dispositional. For example, the essence of the property of mass would consist in a power to resist acceleration, attract other entities with mass, and so on. To say that a massive object does not have these powers would be to say that it is not massive after all.

There are two main versions of dispositional essentialism, *dispositional monism* and *the identity view* of the categorical and the dispositional, also known as *the powerful qualities view*. According to dispositional monism (Mumford 2004; Bird 2007), properties are pure powers, which is roughly to say that there is nothing more to them than their capacity to bring about effects. According to the identity view (Martin and Heil 1999; Heil 2003; Strawson 2008), properties are necessarily both dispositional and categorical (where "categorical" can be read as "not purely dispositional"). Their categorical aspect is standardly described as qualitative (I will therefore use the terms categorical and qualitative interchangeably).

According to the phenomenal powers view, phenomenal properties are not purely dispositional. Rather, their essence consists in what it is like for a subject to experience them, which can be captured by a direct phenomenal concept such as "feeling like *this*", when pointing to an occurrent first-person experience. It cannot be captured in dispositional terms, such as "the property of

making subjects try to avoid it". In this way, phenomenal properties are more like powerful qualities.[7]

In what follows, I will argue that the evolutionary argument supports that the phenomenal powers view is true for at least pain and pleasure. It should be noted that it could also be true for other, and perhaps all, phenomenal properties. In particular, emotional phenomenal properties, such as anger or joy, may appear to make agents act in particular ways in virtue of their phenomenal character similarly to pain and pleasure.[8] But for other kinds of phenomenal properties, such as purely sensory phenomenal properties, it is not as clear how the view could apply. Colors, for example, at least do not appear to have any *strong* motivational power in virtue of their phenomenal character. But one might think their phenomenal powers are just less noticeable, perhaps because their powers are very weak compared to the powers of pain, pleasure, emotions and so on, which they are almost always experienced together with. But for the purposes of the evolutionary argument one can leave it open how far the phenomenal powers view could be extended.

It should also be noted that the phenomenal powers view does not imply that subjects *always* avoid pain and pursue pleasure, only that they do so in the absence of interference from other motives, i.e., *ceteris absentibus.* We often endure or pursue pain, and avoid pleasure for many kinds of interfering motives or reasons. For example, someone might endure pain because they believe it

[7] Proponents of the powerful qualities view are generally not committed to an a priori connection between particular qualities and powers (or particular qualitative and powerful aspects of properties). It is therefore useful to think of the phenomenal powers view as a distinct version of the powerful qualities view, rather than as equivalent to the powerful qualities view simply applied to phenomenal properties.

[8] Perhaps emotions have unpleasant and pleasant dimensions such that they qualify as forms of pain and pleasure in the broad sense defined above, i.e., as phenomenal properties with clear negative or positive hedonic value. But one might also think emotions have further motivational dimensions not shared with pain and pleasure. The former view would be closer to reductive hedonism, and the latter view would be more pluralistic. The phenomenal powers view is compatible with both (insofar as the non-hedonic motivational dimensions of emotions are also phenomenal).

leads to less pain in the future (as when cleaning a wound), or because it leads to a greater pleasure at the same time (as in masochism) or because they believe it is morally appropriate (as when accepting punishment). But in such cases, it is not the pain that causes people pursue or endure it, but rather the interfering motive. If there are no interfering motives, it seems pain always causes subjects try to avoid it – and the same for pleasure.[9]

I will now argue that the phenomenal powers view, conjoined with the theory of natural selection, predicts the correlations (5.1), on a general (5.2) and ultimate (5.3) basis. It thereby explains the correlations far better than any other view of phenomenal causation.

5.1 Prediction

The phenomenal powers view seems to predict, i.e., increase the (epistemic) probability of, the correlations given (1) knowledge of how pain and pleasure feels and (2) the theory of natural selection. This can be supported by a thought experiment.

Imagine someone, call her Maya, who has never experienced pain because she suffers from congenital insensitivity to pain, i.e., a medical condition of that renders patients incapable of experiencing it. She has also grown up very isolated, so she does not have much information about

[9] One might object there are also people who do not avoid pain in the apparent absence of any interfering motives, namely people who suffer from the medical condition pain asymbolia (Grahek 2007). For these patients, pain simply does not seem to bother them. If asymbolic pain and normal pain feel the same, i.e., have identical phenomenal character, it would strongly indicate that the phenomenal powers view is false. However, the literature on pain asymbolia indicates that asymbolic pain and normal pain do not feel the same. According to Grahek's analysis, what pain asymbolia shows is that normal pain has two components, one affective and one sensory. Furthermore, both components are phenomenal, as evidenced by the fact that it seems they can each be experienced in isolation from the other. Not only can the sensory component be experienced without the affective component, as in pain asymbolia, the affective and motivating component can be experienced without the sensory, as a feeling of "pure unpleasantness" that defies further specification (Grahek 2007: 108-111). Pain asymbolia therefore seems consistent with the view that *phenomenologically normal* or *affective* pain makes people avoid it in virtue of how it feels, or in virtue of how its affective part feels. In view of this, and in order to keep things simple, I will stipulative that, throughout the paper, I use the term *pain* to refer to *phenomenologically normal* or *affective pain*. Phenomenologically normal or affective pain can be defined demonstratively in terms of how the sensory and affective parts feel together. It should and need not be defined as "any phenomenal property that makes creatures try to avoid it", or anything else that would reduce the phenomenal powers view to a form of functionalism.

it: she knows that pain is an experience correlated with some bodily states, but she does not know that it causes or is correlated with avoidance. One day, Maya's insensitivity to pain is cured, and she then experiences pain for the first time – say she steps on nail and has an experience of intense, sharp pain. It seems she would then be able to immediately predict that this is a feeling that will always make her, and any other subjects who experience it, try to avoid it (unless they have another motive not to). She would not need to experience pain several times, and observe subsequent avoidance behavior in herself and others, and apply inductive reasoning. Rather, she could make the prediction without induction, on the basis of a single experience of how it feels.

Furthermore, to make such a prediction, Maya would have to assume (implicitly or explicitly) the phenomenal powers view. Knowing how pain feels, it is hard to see how it could *make* any subject do anything else than try to avoid it *solely in virtue of feeling like that*. Try to imagine the pain that Maya would experience from stepping on the nail. The quality of such an experience can only be described as intrinsically repulsive. Could this quality – alone, in and of itself and in the absence of interfering motives – *make* someone who experiences it try to pursue more of it, remain indifferent to it, or otherwise do anything else than try to avoid it? This seems very hard to conceive of. In contrast, it seems quite (at least *prima facie*[10]) conceivable that pain is merely contingently *followed* by something else than avoidance, as per the regularity theory of causation. Or, that pain necessitates something else in virtue of an external governing law. It is perhaps also conceivable that pain makes subjects do something else than avoid it in virtue of a *pure* power which has nothing to do with any qualitative phenomenal character. But assuming that pain has causal power in virtue of its phenomenal character alone, as per the phenomenal powers view, and is conceived

[10] If the phenomenal powers view is true, then arguably it could only be *prima facie*, but not *ideally* conceivable that the view is false, because arguably, if pain is a powerful quality, then it is necessarily a powerful quality, and it should not be ideally conceivable that necessary truths are false.

of in terms of its phenomenal character (i.e., under a qualitative, phenomenal concept), it seems inconceivable that pain has a different power.

One might object that this connection could just be based on a contingent psychological association between pain and its effects. This association could be habitual and learned, or it could be innate – selected for by evolution because it was somehow useful for our ancestors to be able to predict the effects of pain and pleasure prior to repeated experience. However, no other known psychological associations or expectations render it altogether inconceivable that the associated causes and effects come apart. For example, think of the (arguably innate) expectation that there is no (non-gravitational) action at a distance. The idea of action a distance may be counterintuitive, seem highly implausible and so on, but we can still conceive of it if we try. It might be retorted that the psychological association between pain and its effects could be stronger than other psychological associations (perhaps for evolutionary reasons, given how avoiding pain is highly important to our survival). But this hypothesis does not fit well with the fact that it seems easily (*prima facie*[11]) conceivable that pain has different effects on the assumption that the phenomenal powers view is false (or in other words, the fact that it is possible to be puzzled by the correlations in the absence of the phenomenal powers view, which should not be the case if it were psychologically inconceivable that pain had different powers).

In the same way as it enables prediction of the powers and hence the effects of pain, the phenomenal powers view would also enable prediction of the effects of pleasure. Given this, the theory of natural selection can further predict the correlations. If pain causes avoidance, the theory

[11] As discussed in the previous footnote, if the phenomenal powers view is true, it should perhaps only be *prima facie* not ideally conceivable that it is false and consequently that pain can come apart from avoidance. But the fact that it is easily *prima facie* conceivable that pain comes apart from avoidance, still speaks against the hypothesis that they are strongly psychologically associated, because such an association should presumably prevent *prima facie* conceivability.

of natural selection predicts that any creature for whom stimuli detrimental to survival and reproduction causes pain will then be caused to avoid such stimuli, and hence be selected for. Otherwise, they will be selected against. And if pleasure causes pursuit or attraction, any creature from whom stimuli conducive to survival and reproduction causes pleasure will then be caused to pursue this stimuli, and hence be selected for. Otherwise, they will be selected against.

Strictly speaking, the phenomenal powers view might only predict that pain makes creatures *try* to avoid it, and pleasure makes creatures *try* to pursue it. It might not predict that tryings tend to lead to success, i.e., actual avoidance and pursuing behavior (it seems conceivable that they do not, even in the absence of interference). Perhaps the theory of natural selection predicts that creatures whose tryings tend to be successful will be selected for. Otherwise, the principle that tryings tend to be successful (or as successful as the actual correlations imply) could be presupposed as a general background condition. No other candidate explanation of the correlations predicts this principle either, so this would not detract from the explanatory advantage of the phenomenal powers view.

5.2 Generality

The phenomenal powers view does not specify the particular causal powers of any phenomenal properties; it merely says that phenomenal properties produce *some* effects in virtue of their phenomenal character. It thereby predicts the correlations on a more general basis than the correlations themselves. As noted, standard versions of physicalism and dualism can also predict the correlations, but only if they specify one particular identity or law (or power) for each correlation to be explained. The phenomenal powers view consists in a single principle only, from which the specific powers of pain and pleasure follow given knowledge of how they feel. It thereby constitutes a better explanation in the sense of being more general.

One might object that the phenomenal powers view only predicts the correlations given knowledge of how pain and pleasure feel, and since the number of facts about how phenomenal properties feel is equal to the number of correlations that can be explained the explanation is not more general than specific versions of physicalism and dualism after all. But facts about how pain and pleasure feel are not posits of the phenomenal powers view. These facts are directly observable, not part of the explanatory hypothesis. They constitute background knowledge available to all theories. In fact, they are constitutive of the explanandum itself, i.e., the correlations between *phenomenal* pain and pleasure and harmful or beneficial stimuli respectively.

One might think that dispositional monism could also predict the correlations on a general basis, i.e., that someone who assumes that pain and pleasure are *pure powers* will be equally capable of and justified in inferring their effects based on knowledge of how they feel as someone who assumes that they are *powerful qualities.*

Dispositional monism about phenomenal properties comes in an a priori version and an a posteriori version. According to the a priori version, phenomenal concepts (that pick out phenomenal properties in terms of how they feel) are descriptive concepts identical to or analyzable into dispositional concepts (that pick them out in terms of their powers or effects). But this view has the same problem as analytic functionalism, as discussed above. Firstly, it independently seems implausible that phenomenal concepts have this sort of analysis. Secondly, purported analytic truths are not well supported indirectly by inferences to the best explanation (as opposed to directly by appeal to conceptual analysis).

According to the a posteriori version of dispositional monism, phenomenal concepts are demonstrative concepts,[12] distinct from and not analyzable into descriptive dispositional concepts, that directly refer to the same pure powers that dispositional concepts refer to by description. The problem with this view is that if someone like Maya can predict the effects of pain based on a phenomenal concept alone (i.e., the only concept of pain that she possesses), then phenomenal concepts cannot be opaque and uninformative (i.e., blind pointers that reveal nothing about the nature of their referent). Furthermore, the information they reveal cannot be purely dispositional, for the following reason: Someone who only has the phenomenal concept of a phenomenal property can infer its powers, i.e., which dispositional concepts apply to it (as in the Maya case). But someone who only has a dispositional concept of a phenomenal property, i.e., knows that pain is a property with powers such as to make subjects try to avoid it, but has never actually experienced it, cannot infer what it is like, i.e., which phenomenal concepts apply to it.[13] Phenomenal concepts must therefore reveal information that goes beyond the purely dispositional, which implies that phenomenal properties have a non-dispositional aspect.

[12] The reason phenomenal concepts must be demonstrative given a posteriori dispositional monism would be as follows. If phenomenal concepts referred by description, then this description would have to be dispositional or non-dispositional. If phenomenal concepts were (wholly) dispositional, the view would collapse into the a priori version. If phenomenal concepts are (wholly or partially) non-dispositional, they (or their non-dispositional part) would fail to refer, given that dispositional monism implies that phenomenal properties are purely dispositional. These concepts (or their non-dispositional part) should thereby be incapable of conveying any predictive information.

[13] This is supported by thought experiments such as Frank Jackson's famous Mary scenario (Jackson 1982). One might think invoking this claim begs the question against physicalism. But in response to the knowledge argument (based on the Mary scenario), most physicalists acknowledge that phenomenal knowledge cannot be inferred from physical knowledge, including purely dispositional knowledge. They take this to be compatible with physicalism because they deny that new knowledge implies new facts. That is to say, they take Mary's new phenomenal knowledge to be about old physical facts that she already knows, she just conceives of them in a new way—as per some version of the phenomenal concept strategy mentioned above. My claim that purely dispositional knowledge does not imply phenomenal knowledge is therefore fully compatible with physicalism, given the phenomenal concept strategy.

Furthermore, note that the conclusion I presently draw from this claim is incompatible with one form of physicalism, dispositional monist physicalism, but not with other kinds of physicalism. That is to say, I argue that phenomenal prediction (of the kind described in the Maya scenario above) is incompatible with the view that predictive phenomenal concepts refer to *purely dispositional* facts, but not that it is incompatible with the view that they refer to *purely physical* facts of other kinds. However, in section 6 below I will argue that phenomenal prediction (as in the Maya scenario) may be incompatible with physicalism in general in view of some further considerations.

5.3 Ultimacy

Finally, unlike specific versions of standard physicalism and dualism, the phenomenal powers view does not engender an explanatory regress. Specific versions of physicalism and dualism give rise to further explanatory questions, such as "why these particular psychophysical identities?" and "why these particular psychophysical laws?", to which there is no answer – except, for psychophysical identities, the answer that they do not require any explanation even though they intuitively seem to.

Given the phenomenal powers view, the laws "pain makes creatures try to avoid it (*ceteris absentibus*)" and "pleasure makes creatures try to avoid it (*ceteris absentibus*)" are explained as follows: Pain and pleasure have the effects that they do simply because of how they feel. Knowing how pain feels, it is obvious why people try to avoid it – people avoid pain because its phenomenal character is just intrinsically disagreeable and repulsive. Knowing how pleasure feels, it is equally obvious why people try to pursue it – people pursue pleasure because its phenomenal character is just intrinsically enjoyable and attractive. Somebody who genuinely does not understand why subjects avoid pain and pursue pleasure must either not really know how they feel (because they have never experienced them, or fail to vividly remember the experience), or be convinced that their phenomenal character is in fact not the source of their causal power (i.e., reject the phenomenal powers view).

There is no further question as to why something that feels like pain or pleasure makes subjects try to respectively avoid or pursue it, assuming they produce these effects in virtue of how they feel. Such a question can be answered simply by attending to how they feel again (by inducing another experience or by vividly imagining it from memory). There is also no further question as

to why pain and pleasure have this phenomenal character and not some other one. If pain felt different, it would not be pain,[14] and if pleasure felt different it would not be pleasure.

There are, of course, further questions as to why pain and pleasure are correlated with certain physical states and not other ones, but these (insofar as they are relevant to the correlations between phenomenal pain/pleasure and harmful/beneficial stimuli, as opposed to the mind-body problem more generally) will be answered by the evolutionary component of the explanation.

One might object that dispositional monism also ultimately explains the correlations. I have already argued that dispositional monism fails with respect to generality, but given any doubt about this, it is worth considering ultimacy as well. Dispositional monism would explain the laws relating pain and pleasure to behavior in terms of their powers, and claim that there is no further question as to why pain and pleasure have the powers they do, because their nature or essence consists in nothing more than having these powers. If pain had a different power, it would not be pain; if pleasure had a different power, it would not be pleasure.

Formally, dispositional monism and the phenomenal powers view give equally ultimate explanations. But intuitively, the phenomenal powers view is more ultimate. There is a widespread intuition that dispositions need categorical grounds, or that powers cannot be pure but must have a qualitative aspect (e.g., Russell 1927; Armstrong 1997; Lowe 2006). Dispositional monists often acknowledge this intuition even though they ultimately reject it. But this leaves them in a similar position as physicalists who posit inexplicable, a posteriori identities (as discussed above). Those

[14] Note that, as discussed in footnote 9 above, by "pain" I mean "phenomenologically normal pain", and hence, "if phenomenologically normal pain felt different, it would not be phenomenologically normal pain." This is to contrast it with asymbolic pain. Asymbolic pain lacks the affective component of normal pain and therefore feels different, but people who experience it still call it pain, so pain asymbolia would be incompatible with this claim without this implicit qualification.

who deny that identities call for further explanation, even though they intuitively seem to, at least owe an explanation of why they actually do not, or of why they appear to need one even though they actually do not (such as the phenomenal concept strategy). Similarly, dispositional monism leaves open the further explanatory question of why dispositions do *not* require categorical aspects or grounds, even though they intuitively appear to – and even though, for phenomenal properties like pain and pleasure, there actually seems to be one available.

5.4 Summary of the Argument

The evolutionary argument for phenomenal powers can be summed up as follows. The phenomenal powers view explains the correlations between pain and harmful stimuli and pleasure and beneficial stimuli – given background knowledge of natural selection and of how pain and pleasure feel – by (1) predicting them, i.e., increasing their (epistemic) probability, (2) on the basis of a single, general principle, and (3) without giving rise to explanatory regress.

No other view explains the correlations while retaining the same explanatory virtues. General versions of standard physicalism, interactionism or epiphenomenalism only predict that pain and pleasure have, or are by-products of physical properties that have, *some* effects, but they do not predict their particular effects. They thereby fail to further predict the correlations. Versions that specify particular psychophysical laws, powers or identities do predict the correlations, but lack generality and give rise to explanatory regress. Dispositional monism may seem to predict the correlations on a general basis, but upon closer examination it cannot account for how phenomenal concepts are more predictive and thereby reveal more information than purely dispositional concepts of pain and pleasure. It is also a less ultimate explanation because it does not explain why dispositions intuitively appear to have or require categorical qualitative grounds or aspects even though, according to dispositional monism, they actually do not.

The correlations call for an explanation that does not appeal to chance because they would be more improbable given chance than given this explanatory hypothesis. They are also not well explained by a multiverse hypothesis combined with anthropic reasoning. The argument also assumes that theism can be set aside (or at least that non-theistic explanations are preferable to theistic ones other things being equal). Given this, by inference to the best explanation, we should accept the phenomenal powers view.

6 Compatibility with Physicalism, Dualism and Russellian Monism

The phenomenal powers view is primarily a view about the nature of phenomenal causation, not about the nature of phenomenal properties more generally. It is therefore logically and *prima facie* compatible with most non-epiphenomenalist views in philosophy of mind, including physicalism and interactionist dualism. The view implies that phenomenal properties are powerful qualities. Physicalists could therefore identify phenomenal properties with physical powerful qualities that ground physical laws (it would only be precluded to identify them with physical pure (i.e., non-qualitative) powers or pure (i.e., non-powerful) qualities). Dualists could identify them with non-physical powerful qualities that ground interactive psychophysical laws.

But on closer examination, the phenomenal powers view may be in some tension with physicalism – because it is not clear whether there are any physical powerful qualities with the same explanatory features as phenomenal powerful qualities, as the view characterizes them. That is to say, even granted (contra dispositional monism and categoricalism) that there are physical powerful qualities, there does not seem to be any physical qualities that by themselves explain and predict physical powers – because any physical qualities seem merely contingently connected to physical powers, as long as they are conceived under qualitative, non-dispositional concepts. For

example, it does not seem predictable based on a single experience (without induction) that, say, all heavy objects fall to the ground at a certain speed, even assuming they have causal powers in virtue of their intrinsic qualities – insofar as heaviness and other related physical qualities are conceived of in purely qualitative terms (e.g., in terms of how they look, feel and on) rather than in terms of their actual, inductively discovered dispositions (i.e., unless heaviness is conceived of as "some quality that disposes objects to fall at a certain speed", or similarly).

In contrast, I have argued that it is predictable based on a single experience of pain that it will always make subjects try to avoid it, assuming it has causal power in virtue of its phenomenal character, even if its phenomenal character is conceived of under a purely qualitative, non-dispositional phenomenal concept (i.e., the quality of feeling like *this*). The kind of intelligible connection that exists between phenomenal qualities and phenomenal powers therefore seems like a kind of connection that is not revealed by physics or the physical sciences, but rather only by first-person experience. If this is correct, then the phenomenal powers view would be incompatible with physicalism, understood as the view that the nature of phenomenal properties can in principle be fully captured by physics or the physical sciences.

It could be speculated that this partially explains why the phenomenal powers view has so far been overlooked as conclusion of the evolutionary argument. In particular, one might worry that given this tension with physicalism, the phenomenal powers view may imply interactionism. Interactionism faces a well-known dilemma between violation of the principle of physical causal closure, a principle which arguably has strong empirical support (Papineau 2001), and systematic overdetermination, which appears highly inelegant and *ad hoc*. Those who see decisive reasons to reject both horns of this dilemma would thereby see decisive reasons to reject the phenomenal powers view. Furthermore, those who see weighty but non-decisive reasons to reject both horns

might find that the explanatory advantage of the phenomenal powers view with respect to the correlations is thereby outweighed by this explanatory disadvantage with respect to mental causation more generally.

But the phenomenal powers view is also compatible with a third form of non-epiphenomenalism, which can arguably also escape interactionism's dilemma of mental causation altogether, namely Russellian monism (Russell 1927; Strawson 2006). According to Russellian monism, physics only reveals dispositional (aspects of) properties, which require categorical grounds (or aspects). Phenomenal – or so-called protophenomenal properties closely analogous with them – which the view takes to be categorical, can therefore serve as the categorical grounds (or aspects) of all physical dispositions. This view is arguably compatible with the principle of physical causal closure without implying overdetermination (Alter and Nagasawa 2012; Chalmers 2013).

Russellian monism does not imply and does not seem to have been explicitly combined with something like the phenomenal powers view. Without the phenomenal powers view, Russellian monism might take phenomenal and/or protophenomenal properties to ground physical dispositions in virtue of entering into Humean regularities, in virtue of being governed by irreducible external laws, or in virtue of (epistemically or ontologically) contingent links between phenomenal qualities and powers.[15] But it could also take phenomenal and/or protophenomenal properties to ground (or serve as the categorical aspects of) physical dispositions in virtue of being intrinsically powerful, in accordance with the phenomenal powers view. If the phenomenal powers

[15] Note that, standard (i.e., non-phenomenal powers) Russellian monism would not be capable of predicting the correlations on a general and ultimate basis, for the same reasons as physicalism and dualism. Therefore, standard Russellian monism does not rival the phenomenal powers view as an explanation of the correlations either.

view is combined with Russellian monism in roughly this way, it has a way of arguably avoiding interactionism's problem of mental causation after all.

In conclusion, I have argued that the phenomenal powers view – the view that phenomenal properties produce and thereby necessitate their effects in virtue of their intrinsic, phenomenal character – predicts and explains, on a general basis and without generating regress, the correlations between pain and stimuli detrimental to survival and reproduction, and between pleasure and stimuli beneficial to it, given natural selection. Furthermore, it is the only view that seems to do so (assuming theism can be set aside). It is also arguably compatible with physical causal closure (without implying systematic overdetermination), if combined with Russellian monism – which would mean that its explanatory advantage with respect to the correlations would not risk being cancelled out by a failure to explain (elegantly, without overdetermination) the evidence for physical causal closure, as might otherwise be suspected.

I have also compared the evolutionary argument to the problem of fine-tuning for life: why are the fundamental physical constants such as to allow life to arise? The problem of the correlations involves a similar problem of hedonic fine-tuning:[16] why are the psychophysical laws such as to

[16] The problem of hedonic fine-tuning can also be compared with what Philip Goff has called the problem cognitive fine-tuning (Goff 2017). According to this problem, assuming there is such a thing as cognitive phenomenology, it seems like an extraordinary fortunate coincidence that thoughts relate to beliefs, desires, actions and so on in rationally appropriate ways. Goff argues that there is no naturalistic hypothesis that could explain away this apparent coincidence. He suggests three potential non-naturalistic explanations: divine intervention, value-involving laws and the claim that agents have a fundamental capacity to respond to reasons. The problem of hedonic fine-tuning differs from the problem of cognitive fine-tuning, firstly, by not depending on the existence of cognitive phenomenology, secondly, by having a different and more naturalistic solution. Among Goff's non-naturalistic solutions, the phenomenal powers view is closest to the third one, a fundamental capacity to respond to reasons. However, pain and pleasure are traditionally regarded as psychological *causes* rather than reasons, and are indeed often explicitly contrasted with reasons as something that can potentially override them in cases of weakness of will (or *akrasia*). Furthermore, a fundamental capacity to respond to reasons is, according to Goff, "likely to be conceived of as a kind of libertarian free will and/or agent causation" (Goff 2017: 21). The phenomenal powers view does not have these implications.

One might think the phenomenal powers view could thereby potentially offer a fourth, more naturalistic solution to the problem of cognitive fine-tuning. If cognitive phenomenology produces effects in virtue of their phenomenal character, this might predict and explain that it will make creatures act rationally (in the absence of

allow the correlations to evolve? The problem of fine-tuning for life has motivated explanatory hypotheses from theism to anthropic multiverse hypotheses. I have argued that even granted (for the sake of the argument) that fine-tuning for life were best explained in one of these ways, the same would not follow for hedonic fine-tuning. In the case of hedonic fine-tuning, the anthropic principle would not apply, but an adequate non-theistic explanation can still be found in the form of the phenomenal powers view.

interference). However, it is harder to evaluate whether the view could apply to cognitive phenomenology than to pain and pleasure, because cognitive phenomenology is a lot more elusive.

Acknowledgements:

Many thanks to David Chalmers, Brian Cutter, Michael Strevens, Andrew Lee, Brad Saad, Robert Long and participants at the CSMN Work in Progress seminar at the University of Oslo and the CogSci seminar at CUNY Graduate Center for helpful comments and discussion.

This research has received funding from The Research Council of Norway through a FRIPRO Mobility Grant, contract no. 240328. The FRIPRO Mobility grant scheme is co-funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under Marie Curie grant agreement no. 608695.

References

Alter, Torin, and Yujin Nagasawa. 2012. What Is Russellian Monism? *Journal of Consciousness Studies* 19 (9-10): 67-95.

Armstrong, David M. 1978. *A Theory of Universals. Universals and Scientific Realism Vol. II*. Cambridge University Press.

– – –. 1997. *A World of States of Affairs*. Cambridge: Cambridge University Press.

Bird, Alexander. 2007. *Nature's Metaphysics: Laws and Properties*. Oxford: Clarendon Press.

Broad, C. D. 1925. *The Mind and Its Place in Nature*. London: Kegan Paul, Trench, Trubner & Co.

Chalmers, David J. 2003. Consciousness and Its Place in Nature. In *Blackwell Guide to Philosophy of Mind*, eds. S. P. Stich and T. A. Warfield. Malden, MA: Blackwell.

– – –. 2007. Phenomenal Concepts and the Explanatory Gap. In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, eds. T. Alter and S. Walter Oxford University Press.

– – –. 2013. Panpsychism and Panprotopsychism. *The Amherst Lecture in Philosophy* 8 (1-35), http://www.amherstlecture.org/chalmers2013. Reprinted in Brüntrup and Jaskolla 2016.

Corabi, Joseph. 2014. The Misuse and Failure of the Evolutionary Argument. *Disputatio*.

Eccles, John C. , and Karl Popper. 1977. *The Self and Its Brain: An Argument for Interactionism*. Berlin: Springer-Verlag.

Goff, Philip. 2017. Conscious Thought and the Cognitive Fine-Tuning Problem. *The Philosophical Quarterly*.

Grahek, Nikola. 2007. *Feeling Pain and Being in Pain*. Cambridge, MA: MIT Press.

Heil, John. 2003. *From an Ontological Point of View*. Vol. 115. Vol. 1 Oxford University Press.

Hume, David. 1739. *A Treatise of Human Nature*.

Jackson, Frank. 1982. Epiphenomenal Qualia. *Philosophical Quarterly* 32 (April): 127-136.

James, William. 1890/1981. *The Principles of Psychology Vol. 1*. In *The Works of William James*, Vol. 8, eds. F. H. Burkhardt, F. Bowers and I. K. Skrupskelis. Cambridge, MA and London: Harvard University Press.

Lewis, David. 1973. Causation. *Journal of Philosophy* 70 (17): 556-567.

Loar, Brian. 1997. Phenomenal States II. In *The Nature of Consciousness: Philosophical Debates*, eds. N. Block, O. Flanagan and G. Güzeldere The Mit Press.

Lowe, E Jonathan. 2006. *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*. Oxford University Press.

Martin, C. B., and John Heil. 1999. The Ontological Turn. *Midwest Studies in Philosophy* 23 (1): 34-60.

Mumford, Stephen. 2004. *Laws in Nature*. New York: Routledge.

Papineau, David. 2001. The Rise of Physicalism. In *Physicalism and Its Discontents*, eds. C. Gillett and B. Loewer. Cambridge: Cambridge University Press.

– – –. 2002. *Thinking About Consciousness*. Oxford: Clarendon Press.

Popper, Karl. 1978. Natural Selection and the Emergence of Mind. *Dialectica* 32 (3-4): 339-355.

Robinson, William. 2007. Evolution and Epiphenomenalism. *Journal of Consciousness Studies* 14 (11): 27-42.

Russell, Bertrand. 1927. *The Analysis of Matter*. London: Kegan Paul, Trench, Trubner & Co.

Shoemaker, Sydney. 1980. Causality and Properties. In *Time and Cause: Essays Presented to Richard Taylor*, ed. P. van Inwagen. Dordrecht: Reidel.

Strawson, Galen. 2006. Realistic Monism: Why Physicalism Entails Panpsychism. *Journal of Consciousness Studies* 13 (10-11): 3-31.
– – –. 2008. The Identity of the Categorical and the Dispositional. *Analysis* 68 (4): 271-282.
White, Roger. 2007. Does Origins of Life Research Rest on a Mistake? *Noûs* 41 (3): 453-477.