Abstract:

Functionalism about mind has been strikingly successful; it is the going paradigm of cognitive science. But qualia appear to resist functional treatment. Intentionalism, the view that qualia are identical to (or supervene on) certain intentional contents, promises an account of qualia that is compatible with functionalism. I set aside traditional worries about intentionalism (e.g., inverted spectrum, knowledge, and explanatory gap arguments). I argue that traditional intentionalism cannot account for the non-contingent relationship between pain or pleasure qualia and behavior. I offer a new intentionalist account which avoids this problem: pain is a self-imperative to avoid certain represented features, and pleasure is a self-imperative to seek out certain represented features. Pain and pleasure are thus necessarily action-guiding. My account explains why we seek the pleasurable and avoid the painful, and it explains how pain and pleasure lead fairly directly to the appropriate sorts of actions.

Defending a Broader Intentionalism

Functionalism about the mind succeeds remarkably in explaining a diverse range of mental phenomena. Functionalism promises to explain, for instance, what beliefs and desires are, how these lead to action, how sensory information leads to belief revision, etc. Moreover, functionalism meshes well with enterprises in cognitive science: it has been a fruitful paradigm, and suggests many avenues for future research. Nevertheless, many philosophers think that functionalism is at best an incomplete theory of the mind. For one mental phenomenon, the phenomenon of *qualia*, is not amenable to functionalist explanation.[i]

There certainly exist functionalist-friendly theories of qualia. *Intentionalism*, the view that qualia are reducible to certain kinds of intentional contents, fits nicely into a functionalist picture. But traditional intentionalist approaches face serious objections, such as spectrum inversion arguments, the knowledge argument, and explanatory gap arguments.[ii] Intentionalists have attempted to respond to all of these concerns, and the jury is still out on whether these responses succeed. In this paper, however, I wish to bracket these familiar objections. I think that there is a rather serious problem for traditional intentionalism that has been overlooked: it cannot properly account for certain features of pleasure and pain qualia. In this paper, I aim to press two related objections against the traditional intentionalist account of pain and pleasure qualia. Then I will offer a modified intentionalist account of such qualia which avoids these objections.

I. Narrow Intentionalism and Its Troubles

Traditionally, intentionalists have been *representationalists*.[iii] Representationalism is the subset of intentionalism that holds that qualia are identical to (or perhaps merely supervene on) certain *representational* contents. All representational content is intentional, but not all intentional content is representational. Representational contents must be true or false; intentional contents need not be. To wit, sentences like "Please pass the salt" and "Where did I put my keys?" have intentional, but not representational (or not *merely* representational), content. Representationalists think that, besides representational contents, no extra intentional resources are required to make sense of qualia. I think pleasure and pain qualia are more felicitously accommodated with the help of some non-representational content. More on this below.

A traditional representationalist account of pain and pleasure qualia – the sort of account that I do not endorse – usually looks something like this. Consider the sharp, slicing pain of a paper cut. This pain has certain features which can be captured representationally; typically pain qualia involve representations of *damage* to the organism. Different varieties of pain feel different because of the different types of damage which get represented. These different representations of damage, which are identical with pain qualia, cause the individual in pain to respond differently. The dull pain of a headache involves one sort of damage representation, which leads to behaviors like avoiding light and seeking rest. The sharp pain of the paper cut, meanwhile, leads to different behaviors: treating the injured area gingerly, for instance. Each type of pain qualia, says the traditional representationalist, is identical to some type of damage representation.

The traditional representationalist analyzes pleasure qualia similarly, in terms of representations of positive bodily changes. Take the delicious taste of a hot fudge sundae. The qualia of the yumminess of the sundae are merely representations of the sundae producing a positive change on the tongue and throat. Other pleasures (like the pleasure of getting a back massage) feel different from the pleasure of eating sundaes because they represent different sorts of bodily changes; the sundae-eating pleasure and the back-massage pleasure take place in different regions of the body, and the changes that get tracked are also different. When one eats a sundae, certain changes in temperature, pressure, and taste-bud stimulation are represented, whereas when one receives a back massage, kneading of the muscles and similar changes are represented.[iv]

Such accounts face two grave problems. The first I will dub the Problem of Function. The representationalist says that tokens of pain are identical to the particular damage representations instantiated on the relevant occasion; tokens of pleasure are identical to the particular benefit representations instantiated. The representationalist will also want to say that the roles that pain and pleasure play in our cognitive economy – the fact that we tend to avoid painful things and to seek out pleasurable ones – can be fully explained by the representational contents of pain and pleasure. But it is not at all clear why representations of damage would reliably cause us to avoid the damaging things. Nor is it clear why representations of benefit would reliably cause us to seek out the beneficial things.

Imagine the following individual, whom I dub Francis. Francis seeks out many experiences that we would find painful. She avidly pursues paper cuts, headaches, broken bones, and singed fingers. When queried about her strange behavior, Francis doesn't say

that she wants to punish herself or that she believes in a life of denial; rather, she claims to enjoy greatly the various sensations she undergoes. “The feeling of a paper cut,” Francis might say, “is a sharp sort of feeling that is quite delightful. I enjoy it almost as much as the dull pounding of headaches.” Francis’s sensations when she cuts herself might, that is, have the same representational content as my paper-cutting experiences, but one is tempted to think that there is something radically phenomenologically different for Francis. Paper cuts, though they may be represented as self-damage, are fun for her! Since all Francis’s paper-cut sensations might apparently have exactly the same representational content of my paper-cut sensations, but Francis isn’t in pain, the representational account of pain provided above seems mistaken.

Lest one be tempted to think that it isn’t possible for Francis to exist, I’ll point to real-life cases which lack the drama of Francis’s situation but share the features relevant for our purposes. I dislike crunchy vegetables *because* they are crunchy; fresh carrots don’t taste good to me. Others, though, love carrots for precisely the reason that I can’t stand them: they delight in the crunchy texture. Similar situations may exist with music. Consider Jason and Sonia, two highly trained musicians. Jason plays viola professionally, while Sonia plays electric guitar for a rock band. Jason might hear Sonia’s music and despise it, *even if* his representation of that music is very similar to Sonia’s. Jason might hate the music because of its chord structure, while Sonia might love her music for just that chord structure. So why couldn’t there be someone who loves paper cuts, headaches, broken bones, and singed fingers for having precisely the qualities that cause most of us to find them painful?

Here is one reply which the traditional representationalist might try. Perhaps the fact that we are disposed to avoid painful things isn't due solely to the representational content of pain. Rather, it's that content *conjoined* with some other contingent but nearly universal features of human psychology that does the explaining. Such features might include some combination of desires to avoid self-damage and reflexes to avoid self-damage. Perhaps in situations where deliberation is involved, it's the desire to avoid damage to oneself which, when conjoined with the representational content of the pain, does the explanatory work. In situations where one reflexively avoids pain, on the other hand, it's the reflex to avoid self-damage, plus the representational content of the pain, which does the explaining.

This response strikes me as inadequate, for it has a strange consequence. Recall that, on the traditional representationalist view, the reason that a chocolate sundae tastes good is *just* that the organism represents the sundae as being beneficial. And, according to the response under consideration, the explanation of the fact that I seek out chocolate sundaes is that fact *plus* the fact that I have a reflex or desire to get things that are beneficial to me. Suppose, then, that I didn't have such a reflex or desire. Suppose that I had a reflex or desire to seek out things that harmed me, and to avoid things that were beneficial to me. Then, according to the traditional representationalist, I would no longer seek out sundaes, *but they would taste just as good to me*. In actuality, then, it's just a remarkable coincidence that I desire things that make me feel good – the fact that they feel good does not do much work in explaining my behavior. But one wants to say that the reason I seek out sundaes is that they taste good to me, full stop. There isn't any need to posit an additional reflex or desire to seek out what tastes good! Call this the Problem

of Explanation (but don't confuse this with Explanatory Gap problems, which are quite different).[1]

I have argued that there is a dilemma for the traditional representationalist about pleasure and pain qualia: she will face either the Problem of Function or the Problem of Explanation. But intentionalism can be salvaged if we can find some *intentional* but *non-representational* difference between Francis's paper-cut qualia and mine. It is to this task that I now turn.

II. A Broader Intentionalism

Imperatives are intentional items. We tell one another to do things – I may say "Shut the door" – and these imperatives have intentional content; in this case, the content is, roughly, that the door be shut by the person to whom I issued the imperative. But the intentional content of an imperative is not representational, since imperatives are not true or false. To be sure, an imperative may be satisfied or not: if you shut the door, my imperative to you gets satisfied, and if not, my imperative remains unsatisfied. But one should not confuse the notion of satisfaction with the notion of truth.

The explanation for this difference between representations and imperatives lies in the direction of fit between the intentional content and the world. For something to be a representation, its intentional content must have content-to-world direction of fit; that is, a representation is the sort of thing that tries (or is designed, or evolved, or …) to track conditions in the world. For something to be an imperative, by contrast, its intentional

[1] The traditional representationalist may want to alter his position to deal with this worry. He can say that sundaes taste good to me because they are represented as beneficial *and* I have a desire for them. Depending on one's account of desires, this modified position may shade into my own view.

content must have world-to-content direction of fit; that is, an imperative is the sort of thing that guides one to change the world to match its content.

The term "imperative" invites a misunderstanding which I need to avoid. Typically, we think of imperatives as being part of language, a language that includes conversational presuppositions, implicatures, and a host of other complexities. We do not think of, say, plants as delivering any imperatives, even though plants do send signals to their buds to open. (Similarly we do not think of plants as having beliefs, even though parts of the plant track changes in temperature, pressure, etc.) I need a stripped-down notion of imperatives, one somewhere between the full-blown imperatives found in natural language and the proto-imperatives found in plant signalling. For the remainder of the paper, I will use the term *imperative* in a non-standard way, and anyone who objects to the use of this term may supply a more innocuous one. By *imperative*, I will refer to things that have intentional content with world-to-content direction of fit. I will not presuppose any of the other rich features that occur in natural language. There are, however, conditions that an item must meet to count as being intentional at all, and imperatives must meet those conditions; it is quite likely that plant signals do not meet those conditions.

As a first pass, I propose that pain is just a self-imperative to avoid whatever results in certain specified qualia, while pleasure is just a self-imperative to do whatever results in certain specified qualia. (Call this the *self-imperative* account of qualia.) For instance, suppose I enjoy eating my stir-fry dinner. I savor the saltiness of the soy sauce, the texture of the broccoli, and the flavor of the mushrooms. This whole experience involves qualia. According to an intentionalist account, the qualia of the taste of the soy

sauce, the broccoli, and the mushrooms can be explained fully in terms of the features *F1, F2, F3, ...,* and *Fn* that my brain attributes to the food: a certain texture, a given amount of saltiness, and so on. Let's say that feature *F3* is the taste of the soy sauce, feature *F41* is the texture of the broccoli, and feature *F908* is the flavor of the mushrooms. I propose that the pleasure involved in my consumption of the meal consists in the following intentional feature of my brain: a self-imperative to do whatever results in features *F3, F41,* and *F908*. Every pleasure involves such self-imperatives which refer to particular features. Similarly, the experience of a paper cut might involve a representation of the paper cut with features *G1, G2, ..., G312*. This experience is painful for me because my brain also sends a self-imperative to avoid whatever results in some of these features – perhaps features *G3, G47*, and *G288*. That self-imperative is what the painfulness of the sensation consists in.

A qualification is in order. No representationalist holds that every representation is a quale; the thesis is only that every quale is nothing more than a *certain kind* of representation. I naturally must extend this point to my intentionalist account. Not every self-imperative is a pleasure or pain quale, but every pleasure or pain quale is a self-imperative of a certain kind. In this paper, I will not try to specify what that certain kind is, for that is a project for another paper.[v]

There are some details left to be explained in the self-imperative account. There are plenty of sensations that are pleasant in some respects and painful in other respects; maybe (though I don't know) there are some sensations that are pleasant and painful in *precisely the same* respect. An example of the former: I might continue to eat a sumptuous meal even after I am full. The eating is pleasant, in that I enjoy the taste of the

food; it is painful, in that I simultaneously develop a stomach ache. On my account, this experience is pleasantly painful because my brain sends a self-imperative to seek out rich food and another self-imperative to avoid making myself too full. Perhaps it is even possible for there to be an experience that is pleasant and painful *in the same respect*. If this is a genuine possibility, then, on my account, the brain would send a self-imperative to seek out and to avoid whatever results in one and the same feature.

Of course, pleasures and pains are not all equal. Being stabbed in the arm hurts considerably more than being injected with a flu vaccine; having an orgasm is typically much more pleasurable than eating pumpkin pie. But some pleasures and pains are difficult to compare (as utilitarians frequently discover to their chagrin). I do not know which is more pleasant for me: listening to a beautiful song for the first time or winning a close game of chess. Nor do I know whether I would rather run 5 miles as hard as I can or have a low-grade fever for a day. The self-imperative account can explain both of these facts. For self-imperatives can have different strengths, and these strengths may not always be comparable. The self-imperative in the case in which I am stabbed in the arm is much stronger than the self-imperative in the flu-vaccine case; the self-imperative involved in the orgasm case is also typically much stronger than the self-imperative in the pumpkin pie case. But there may be two self-imperatives such that neither imperative *dominates* the other. That is, in some situations, the first imperative will override the second, and in other situations, the second imperative will override the first.

III. Problems Revisited

Now let's return to the puzzles about pain and pleasure and see how well the self-imperative account handles them. I suggested earlier that traditional representationalism, which analyzes pain in terms of damage representations and pleasure in terms of positive bodily change representations, ends up in a dilemma: it succumbs either to the Problem of Function or the Problem of Explanation. But the self-imperative account avoids both horns. Let's start with the Problem of Function, that is, the problem of explaining why organisms *seek out* pleasurable experiences and *avoid* painful ones. It is quite clear why, if the imperative account is correct, an organism would avoid painful stimuli and seek out pleasurable stimuli: because pain just is a self-imperative to avoid whatever results in a specified set of qualia, and pleasure just is a self-imperative to seek out whatever results in a specified set of qualia. A self-imperative is a causally efficacious feature of the mind, and is a functionally kosher item. One of functionalism's traditional strengths has been its ability to link the mind to behavior and observable states of the organism – features which can be studied in familiar scientific ways – and the self-imperative account extends that strength to the study of qualia.

The self-imperative account also avoids the Problem of Explanation. As I argued above, to avoid the Problem of Function, the traditional representationalist had to concede the possibility of someone who feels pleasure when she eats sundaes, but who *avoids* sundaes because of these pleasure qualia. For the traditional representationalist dodged the Problem of Function by positing independent extra reflexes or desires to avoid the painful and to seek out the pleasurable; this independence of reflexes and desires from pain and pleasure qualia leads straight to the Problem of Explanation. The

self-imperative account doesn't need to posit reflexes or desires to avoid the painful and seek out the pleasurable, so the Problem of Explanation never gets off the ground.

The self-imperative account is also able to handle the case of Francis. Recall that Francis is like the rest of us in most respects, except for one quirk: she enjoys all of the sensations that we find painful. She likes paper cuts, headaches, broken bones, and so on. The traditional representationalist account had some options about what to say about folks like Francis, but none of these options seemed especially appealing. I think that the self-imperative account offers a satisfying story about Francis and her ilk. I will focus on explaining both the differences and the similarities between ordinary people and Francis. For an ordinary person, the sensation of a paper cut includes, not only representations about the type of damage involved, but also the self-imperative to avoid such representations. Francis's paper cut sensations involve the same representations of damage as our paper cut sensations. But her paper cut sensations also include the self-imperative to seek out those sensations. The identity between Francis's damage representations and ours explains the similarities between our paper cut sensations and Francis's paper cut sensations. The difference between Francis's self-imperative and ours explains two dramatic differences between us and Francis: the difference in behaviors (Francis deliberately cuts herself, whereas the rest of us take precautions against being cut) and the difference in what paper cuts feel like (Francis finds them enjoyable, whereas we find them painful).

I conclude that, if one is inclined to be an intentionalist in the first place, one ought to accept a broad intentionalism that includes self-imperatives. Traditional representationalism does not avail itself of much-needed intentionalist resources and thus

cannot satisfactorily explain the link between pain or pleasure qualia and behavior. The self-imperative account faces no such difficulties.

I suspect that self-imperative accounts can be extended beyond pleasure and pain qualia. Several other classes of qualia involve a necessary connection to dispositions to act, and, as I've argued, representationalism is inadequate in such cases. Qualia associated with emotions are the largest class of such cases, but other examples include hunger-qualia, thirst-qualia, and itchiness-qualia. Take, for example, fear-qualia. There is something it is like to feel fear. But, for a creature to count as feeling fear, it must have some disposition to behave in certain ways (e.g., to flee or hide); a creatures that has no such dispositions is not experiencing fear. Likewise, a creature with no disposition to seek or consume food is not experiencing hunger-qualia, and a creature with no disposition to scratch itself isn't experiencing itchy-qualia. A self-imperative account has the resources to explain the necessary connection between such qualia and dispositions to act; a representationalist account does not. I do not have time to defend this claim at length here, but this set of cases suggests that self-imperatives may be an important, widespread, and overlooked feature of qualia.

---

[i] David Chalmers has expressed such a view in Chapter 1 of *The Conscious Mind* (New York: Oxford University Press, 1996).

[ii] For classical expositions of these problems, see, respectively, "The Inverted Spectrum," by Sydney Shoemaker, in *The Journal of Philosophy*, volume LXXIX, No. 7, July 1982, and "Inverted Earth," by Ned Block, in *Philosophical Perspectives*, 4, 1990; "Epiphenomenal Qualia," by Frank Jackson, in *Philosophical Quarterly* 32:127-136, 1982; and "Materialism and Qualia: The Explanatory Gap," by Joseph Levine, in *Pacific Philosophical Quarterly* 64: 354-61, 1983.

[iii] Tye and Lycan are prime examples (see above endnote), and other examples include Fred Dretske (cf. *Naturalizing the Mind*, Cambridge, MA: The MIT Press, 1995) and John McDowell (cf. "The Content of Perceptual Experience," in *Philosophical Quarterly*, 1994).

[iv] In broad outline, Tye's representationalism is of this sort. While other representationalist accounts may differ in the details, the charges that I levy against representationalism apply to them, as well.

[v]For one view about what such conditions might look like, see Michael Tye's *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind* (Cambridge, MA: MIT Press, 1995).