The Rigidity of Pain

December 30, 2008

"The standard way for materialists to avoid Kripke's argument is to deny that the argument applies to the theory in question," reports Polger (1999, p. 1), "typically on the grounds that the theory does not specify identities using rigid designators. This response may be used by functionalists, but is not readily available to the identity theorist." We concur with Polger's diagnosis, although we will employ a different strategy to solve Kripke's (1980) problem. Our approach will be similar to David Papineau's (2000). We shall accept that there is a basic psychological level at which pains and, say, C-fiber firings appear to us to be distinct occurrings, but we shall argue that these intuitions are inadequate: "it seems to me that there is room for physicalists to respond to [Kripke']s analysis by admitting that they are less than fully committed to mind-brain identities at an intuitive psychological level, while continuing to insist on such identities at a theoretical level." (Papineau, 2000, p. 4)

The point of our undertaking is not to argue in favor of physicalism, although we do support this doctrine. The point is to show that the existence of the Kripkean intuitions of contingent identity does not entail that they are true. By doing this, we plan to undercut Kripke's argument as an argument against physicalism.

Before we begin, let us specify what we take Kripke's argument to be against. We take it to be an argument against type-type identity, against token-token identity, and even against functionalism. Type-type identity is

charged with not accounting for the intuition that pain could have turned out to be something else than CFF (as we shall call C-fiber firings from now on). Token-token identity is charged with not accounting for the intuition that any particular CFF could have happened to not be a pain. Functionalism is charged with making 'being a pain' a contingent property, although it should be essential. (Kripke, 1980, pp. 144-55)

The concept of pain is not functional

We subscribe to Kripke's view that the meaning of 'pain' cannot be elucidated in functional terms. This is not because we think functionalism is wrong across the board – on the contrary, it seems to us that functionalism is usually right, except when it comes to certain basic mental contents. Neither does this mean that we accept Kripke's claim that 'painful' is an essential property. It doesn't matter, for our purposes, whether pains are essentially painful, nor does it matter whether anything at all is an essential property of something. What matters is that painful events have a certain component which can be abstracted from them, and which does not appear to be definable in terms of causal roles.

What we have in mind is "that something" that is a common component of all experiences of pain, *as well as* of memories of pain, anticipations of pain, etc. If my tooth hurts, this event has certain causes and effects, and perhaps being-in-pain can be defined in terms of them. If, now, I remember how my tooth hurt yesterday, remembrance will have different causes and effects, and perhaps remembering-pain can be defined in terms of *these* causes and effects. But events of the two kinds have something in common, something that we shall call a *neural marker*. This neural marker is what we have in mind when we talk about "the way pain feels." And it would make scarce sense, if any, to claim that the way pain feels is a causal role: what would it mean for something to play the role of the way pain feels? The way pain

feels is not a role, it is an actor.

pain#

pain# (read *pain-sharp*, no pun intended) is the common component of pain-like states. It is the neural marker which we identify by saying that it feels like pain, the one which is contained in our pain experiences, and which we can also call to mind retrospectively, without incurring additional pain. It is our assumption that the neural marker is unique, that there is only one kind of neural state that *marks* pain, while there are many kinds of complex mental states that involve this marker. Why would we make this assumption? After all, there are at least two reasons why one might assume pain# to be non-unique:

- pains admit degrees, and how can one and the same thing have different degrees?
- pains are not all of a kind, so why should we have just pain# instead of sharp-pain#, dull-pain#, etc.?

Our answers to these objections are based on the fact that, although we take pain# to be a single mental content, we do not take it to be a single object, or a single event, or a single state-of-affairs, etc. The unicity of pain# is akin to the unicity of the name "Kripke": it has many tokens, but it is only one name.

This appears to conflict with our claim that nothing can play the role of the way pain feels: it seems that the tokens of pain# might play the role of pain#. But they don't play the role of the way pain feels, they just feel like pain. The way pain feels is the way they are, and nothing that is not a token of pain# feels like pain. A functionalist can, indeed, talk about states of pain within systems that don't have pain#, as long as these systems use some other marker in its place, say pain*#. We do not object to this talk,

but we underline that pain*# wouldn't feel "like pain." It would feel "like pain*." This doesn't mean that the systems would be "in pain*." They would be "in pain," just like we are, but they would be in pain by feeling pain*. This is possible because "being in pain" is a functional concept, which involves a flaccid definite description that picks out whatever the marker of pain is in the mind we are talking about.

When a CFF occurs within my brain, its pain-like quality does not rely on playing a causal role. Now it is easy to show that CFFs do have causes and effects. For instance, we could think of complex mental states similar to being-in-pain or to memories-of-pain, which contained something else instead of tokens of pain# – perhaps they would contain tokens of pleasure#. These states would have different effects, thus showing that whether something is a token of pain# is not causally irrelevant. But although pain# has effects, it does not *consist in* having these effects. Being a token of pain# is an intrinsic property. The meaning of 'pain' is not just "whatever causes me to withdraw my hand from the fire, when experienced directly, and whatever causes me to feel relieved, when remembered, etc." If it were just this, then how would we know that these two *whatever*s are the same? How would we establish the identity of the two, if pains didn't have at least one intrinsic trait which our brains can recognize? Of course, we *could* establish the identity by showing that both definite descriptions are satisfied by the same neural states. But we don't – we are aware of the identity directly, without engaging in neural research.

On to the two objections: if every pain involves the presence of one and the same neural marker pain#, then how can pains have varying degrees? The answer is that pain# is present by means of a number of its single-use tokens. Let's assume, for clarity, that each event of one C-fiber firing is one token of pain#. Then a stronger pain will consist in stringing together multiple tokens of pain# – in other words, in multiple CFFs.

Why do states of pain differ qualitatively, if they all rely on the presence of pain#? The answer is that pain# is not all they consist in. If a pain has a more distinct quality then simply being painful, then this is due to the composition of pain# with another mental content, such as dull#.

(It is not important for our discussion whether dull pains are actually composed this way, rather than being tokens of a single mental content dull-pain#. In the latter case, our discussion would simply shift from pain to dull pain.)

The reader may now ask how we distinguish composition from abstraction. When is something a mental content such as pain#, and when is it a mere concept abstracted from mental contents, such as 'sensation'? After all, we claim that a sharp pain is implemented by composing pain# with sharp#, but we wouldn't claim that a sensation of red is composed of sensation# and red#. There is only red# in a sensation of red, and 'sensation' is abstracted from red# and others. We will not spend much time on this, but here is a quick test: if the term behaves more naturally like a mass term, it stands for a neural marker, and if it is more naturally a count term then it stands for an abstraction. "I see red" and "I feel pain" fall within the first category, "I see a color" and "I have a sensation" within the second. Why this should be so is an interesting question. Markers like pain# or red# have the ability to string their tokens together, which is presumably why they behave like mass terms. But concepts like "color" or "sensation" are predicates of many distinct mental contents, and these don't string their tokens together: an olfactory and an auditive sensation will hardly add up to anything. This is why these latter terms behave like count terms.

pain@

We have already talked about red#. red# is a neural marker, a neural state that we instantly recognize when it occurs in our brains, the one we call "what

red feels like." But there is another red in our minds, aside from the feeling of red: the concept of red. Not the concept of a red thing or of red light, but the concept of red#, as it is involved in the sentence "some people can't see the difference between red and orange." Notice that, in order to understand this sentence, one need not bring the feeling of red to one's mind, one need not imagine red. People who have never seen red or orange can understand the sentence, just like we can understand sentences about the sensations of bees, which are sensitive to the ultraviolet spectrum.

We call this concept of red# 'red@.' In general, we suffix a name with '@' when it stands for a mental content that is directed towards some content of its own,[1] and we suffix it with '#' when it names a mental content that is used 'as is,' rather than to stand for something else. red# is the way red feels, red-thing@ is the concept under which red apples fall, and red@ is the concept under which red# falls.[2]

pain@ is to pain# what red@ is to red# – it is a device that our brains use to think about pain#. When we read something about chest pain, we can comprehend it although we don't necessarily pause to imagine the pain – which we may not even be able to do. Something else than pain# enters our thoughts, and this something else is pain@.

hesperus@ and phosphorus@

hesperus@ is the thing in our mind that stands for Hesperus, just like pain@ stands for pain#. phosphorus@ stands for Phosphorus. hesperus@ and phosphorus@ are *equirepresentative*, i.e. they stand for the same thing. They are, however, not identical – or better yet, they weren't identical in the beginning. As long as we don't realize that the two representations stand for the

[1] The clash between the two meanings of 'content' is unfortunate, but harmless.

[2] Strictly speaking, we should call red@ 'red#@.' This, however, wouldn't actually make the exposition easier to follow, which is why we have opted for the more readable variant.

same thing, the representations will remain distinct. This is because our brains automatically assume that wherever there is only one representation, there is at most one thing the representation stands for. This assumption can be defeated under special circumstances like twin confusion comedies, but this shouldn't obscure the point: true or not, the assumption is present in the brain. As long as we don't believe that Hesperus is Phosphorus, yet we entertain both notions, we entertain two distinct representations.

What happens, then, when we discover that the two representations stand for the same thing? Initially, we still have both representations in our minds, which is why we make seemingly incoherent outings like "these two planets are identical," or "those two are just one." The reason why these sentences seem incoherent is that we are counting representations on one hand and things on the other. So far so good.

As long as hesperus@ and phosphorus@ have been distinct, they have been responding separately to the names 'Hesperus' and 'Phosphorus.' When we say that a representation responds to a name, we do not simply mean that it represents the thing that is denoted by the name. What we mean is that the representation contains its own encoding of the name, just like a dog's brain contains an encoding of the dog's name. Whenever I speak the name 'Hesperus' in my mind, hesperus@ steps forward and comes to me, like an old and faithful dog.[3]

When we use a name that is tied to a representation in this intimate way, the name denotes whatever the representation represents. *This law also holds of counterfactuals.* If we say that "If Phosphorus were a star rather

[3]An exhaustive account of this would have to tell a story about contexts. Multiple representations can be tied to the same name. I can know many people called 'John,' and their representations in my mind would all be tied to 'John' – or at least the representations of the Johns whom I know well. And yet, when I speak the name 'John' in my mind, I don't make all these representations come to me. Only the contextually relevant will do so. This suggests that our minds contain distinct models within which names are spoken, and that a name spoken within a model can only be 'heard' by representations inside that model.

than a planet, it would be a bright star," then 'Phosphorus' still denotes whatever phosphorus@ represents. It cannot denote anything else, because phosphorus@ is the representation that automatically, mechanically comes to mind when we say 'Phosphorus.' We have no control over this. 'Phosphorus' *has* to call phosphorus@, at least unless we work on this and retrain our brains. We call a name that is tied to a representation, with Kripke's words (1980), a *rigid designator*. What we mean, though, is not that the name *denotes the same thing* within all counterfactual contexts (although this is typically the case), but that the name is *tied to the same representation*.

Obviously, more is borrowed from Kripke than the words. We share with him the concern for doing justice to the way names function inside counterfactuals. We consider the phenomenon of rigid designation highly important. We do not believe that rigid designation can tell us much if anything about the world, except for the small parts of it we call brains. It can, however, give rise to a class of confusions akin to optical illusions. As the reader already anticipates, we shall argue that Kripke's argument against the identity theory relies on a confusion of this kind.

As we were saying, hesperus@ and phosphorus@ have started out as distinct representations, and they were still distinct shortly after our astronomical discovery. But what happened to them later? They grew into one and the same representation. The process, of course, was gradual and complex, and its details are far from clear. But what matters is that, after a period of restructuring, our brains ended up with only one representation instead of the original two. The two names, on the other hand, did not fuse. They *both* became tied into the unified representation. The new representation answers to both 'Hesperus' and 'Phosphorus,' just like Bertrand Russell answers to both 'Bertrand' and 'Russell.'

Once the two representations unify, it becomes impossible for us to imagine that Hesperus is not Phosphorus. This is because both names are now tied to the same representation, and we cannot imagine that this represen-

tation is not equirepresentative with itself. hesperus@ and phosphorus@ are now just as identical as Hesperus and Phosphorus, which is why any counterfactual that contains occurrences of hesperus@ is one that contains the same occurrences of phosphorus@. It has become *physically* impossible for us to assume something like "If Hesperus wasn't Phosphorus...," because the thought "Hesperus is not Phosphorus" is prohibited by the laws of physics. The necessity of identity thus stands explained.

Just like hesperus@ and phosphorus@, water@ and h_2o@ have fused, at least in the minds of chemists. A chemist cannot ascribe any meaning to the direction "think of a kind of water that isn't H_2O" (and neither can we). The reason why we cannot do this is not merely that we know that water is identical to H_2O. *Knowledge of identity is not enough* to make distinctness impossible to imagine. We also know that vertebrates are identical to mammals, birds, reptiles, fish, and amphibians taken together,[4] but we have no difficulty imagining more exotic vertebrates from outer space. The reason why we cannot imagine water that is not H_2O is that the very neural tool with which we imagine water is the same with which we imagine H_2O: water@ is identical to h_2o@.

Needless to say, the same goes for heat@ and molecular-motion@.

Why we cannot affirm heat and deny molecular motion at the same time

Kripke claims that 'heat' and 'molecular motion' are rigid designators. We agree, at least once the true nature of heat has been discovered. In the old days, when 'heat' meant "what makes us feel warm," it was a flaccid designator. Now, however, it is a rigid designator, just like 'molecular motion.' And it is impossible to think that "if heat weren't molecular motion, then ...," although we can still think something like "if molecular motion didn't make

[4] Zoology is more complicated, but this shouldn't matter.

us feel warm." But why can we not think a thought of the first kind? Is it because the two terms are rigid designators, therefore they designate the same thing in all possible worlds, providing they designate anything at all? Kripke would have us believe this. But we have strong reasons to propose a different explanation, reasons at which we have hinted at the end of the previous section. The explanation we propose is that the two representations, heat@ and molecular-motion@, have fused and become one, hence the inconsistency of thinking of one while trying to deny the other.

The kripkean explanation of the necessity of identity cannot account for the following Gedankenexperiment, whereas the one we propose can. Imagine you have neighbor, Samuel Langhorne Clemens. You are also an admirer of the writer Mark Twain, with whom you have corresponded in the past. You still have his signature on his letters. Little do you know that he has recently moved to your neighborhood. One day you meet Clemens, and you say "Sam, do you know Mark Twain, the writer?" Clemens explains that he *is* Mark Twain. You don't believe this eccentric pretense at first, because your mental representations twain@ and clemens@ are very far apart, the *a priori* likelihood of them turning out to represent the same object is very low, and the fact that Clemens claims that he is Twain can easily be explained as a joke prompted by your initial question. Seeing that he won't quit, you challenge him to show you his signature. He does so to your satisfaction. You have now established beyond reasonable doubt that Clemens is Twain. You *know* he is Twain. *But you just can't believe it!* clemens@ and twain@ are so different! The effort of bringing them together is so large, that it seems impossible. Yet, in a number of days, the mess will get sorted out. Sooner or later, the fusion of the two representations will be complete. Now it becomes impossible to imagine that Twain is *not* Clemens. The moral of the story? The moment when you get rid of the intuition that Twain might be distinct from Clemens does not coincide with the moment when you make sure that they are the same. On the contrary, even after you know that they

are the same, you still keep saying to yourself "But what if it's not true? I can't believe that Clemens is Twain. I must be making some mistake!" Knowledge of the identity does not come bundled with the intuition that the identity is necessary. This intuition only steps in later, after your mind has reorganized its contents. But if the intuition of non-contingency were due to the mere awareness of identity, coupled with the fact that both names are rigid designators, then situations akin to our Gedankenexperiment should be impossible.

The notion that this pain could have been something else than a CFF

We cannot conceive a counterfactual starting with "if heat weren't molecular motion." We can, however, say something that begins with "if pains weren't CFFs." One of the reasons for this is fairly obvious: pains aren't CFFs, after all. But now let's suppose that we know of a constant neural correlate of pain-reports. For the sake of familiarity, we will still call it 'CFF', although it isn't C-fiber firing. Could we still hypothesize about non-CFF pains, even if we knew that pains are CFFs? The answers we propose are 'yes,' and 'no.' The correct answer depends on which mental content we would be trying to use: the trick would work with pain@, but not with pain#.

Here is why: First of all, we should remember that pain@ and pain# are different things. pain@ is the concept of pain, just like dog@ is the concept of a dog. dog@ doesn't hunt ducks, and pain@ doesn't hurt. pain# is not a concept. pain# is the way pain feels. pain# is the common component of the states our brains are in, when we perceive or imagine pain. Whether pain# hurts or not depends on where it occurs in our minds: there are mental places where it will go to put us in pain, and there are mental places where it will go to make us imagine pain.

If we were to discover that pains are CFFs, what we would discover is

that tokens of pain# are CFFs. The neural nature of pain@ would be a different business altogether. Let's think a bit about what this would mean, that tokens of pain# are CFFs. It would mean, among others, that when C-fibers fire in our brains, we feel pain. It would also mean that containing CFFs is all there is to feeling pain, just like molecular motion is all there is to heat. This sounds paradoxical, and we believe, with Papineau (2000), that it will always sound paradoxical.

Now, once the identity between pain# and CFF is discovered, we would expect the two concepts to fuse. That is, we would expect pain@ and cff@ to fuse. But would we expect pain# and cff@ to fuse? We would not. One reason why they shouldn't be expected to fuse is that pain# is not the concept of pain: pain@ is the concept. The other reason is that pain# is *read-only*. Unlike concepts, phenomenal states cannot be updated. red# has always felt like red. green# has always felt like green. green# will never go red. If a phenomenal state were to become read-write, this would presumably be cataloged as a neural disorder: pain# would fail to do its job, if it started to feel like bliss. And if it started to carry conceptual contents, such as the ones carried by cff@, then it would presumably break the mental compounds it is a part of, by introducing a concept where a phenomenal state is expected. Finally, if the tokens of pain# are C-fibers firing, then these tokens cannot be updated. They cannot start, upon the discovery of their nature, to contain the concept of C-fibers firing. Either they contained it from the beginning, or they will never contain it, just like a sentence does or does not contain a certain word.

Thus, if we were to discover that pain# is CFF, we would only expect pain@ to fuse with cff@. Are we absolutely sure even of this? No, because it may turn out that the concept of pain@ itself plays a vital biological role, one which prevents it from accepting updates. But then again, this possibility is not germane to our argument, since the argument is not that pain@ and cff@ would fuse sometime in the future. The argument is that pain# and

cff@ will never fuse, and that an introspective grasp of this fact underlies the kripkean intuition that the identity of pains and CFFs will never appear necessary to our imagination.

We will always be able to imagine a pain, and at the same time deny that this image contains any CFFs. But this has nothing to do with whether the pain we imagine is or is not a CFF. The reason why this kripkean intuition is possible is that *all it takes for us to imagine pain is to fire some C-fibers*, and that *this firing of C-fibers does not, and cannot, contain a conceptual representation of itself.* It does, however, contain the feeling of pain: or rather, it *is* the feeling of pain.

Epilogue

We have told a story about the mind, a story that may or may not be accurate. We have reasons to believe that it is accurate, and we have described some of them elsewhere.[5] But the reasons are less important. What is important is that we have provided a plausible account for why we should have kripkean intuitions, even if all there is to our minds is the activity of our brains. Thus, the existence of kripkean intuitions can be explained within the framework of materialist monism.

References

Kripke, Saul: Naming and Necessity. Harvard University Press, 1980

Papineau, David: Kripke's Proof That We Are All Intuitive Dualists. 2000, Available online ⟨URL: `http://sas-space.sas.ac.uk/dspace/bitstream/10065/892/1/D_Papineau_Proof.pdf`⟩

[5] *Reference removed for anonymity.*

Polger, Thomas: Kripke and the Illusion of Contingent Identity. 1999, APA Eastern Division (Boston, MA), Colloquium Paper, Graduate Student Submission