# Disproportional Mental Causation

## By Justin Tiehen

Gilmore is in agony.  His toothache keeps getting worse, and now he has broken down and started to cry.  If nonreductive physicalism is true there must be some physical realizer of Gilmore's pain; call the type of realizer '$P_1$'.  Then, letting $M$ be the proposition that Gilmore is in pain, $P_1$ that he is in a $P_1$ state, and $E$ (for effect) that he cries, consider the following counterfactual.

   [Cry]: $(M \& \sim P_1) > E$.

That is, if Gilmore had been in pain but not $P_1$, he still would have cried.

Here is one way this could be true.  Imagine that pain is multiply realized, with $P_1$ and $P_{101}$ being its two nomologically possible physical realizers.  Imagine next that there is a physical law that $P_{101}$ states causally necessitate crying.  Then, putting things in terms of the standard Stalnaker-Lewis analysis of counterfactuals, the closest world where Gilmore is in pain but not $P_1$ will be a world where his pain is realized by $P_{101}$.  Assuming that the physical law just mentioned obtains in this closest world, Gilmore will cry there.  Thus, [Cry] is true.

A number of philosophers think that counterfactuals like [Cry] are crucially important in accounting for mental causation, and more specifically in solving the causal exclusion problem facing nonreductive physicalists.[1]  Especially influential on this front has been Stephen Yablo, who in a series of well known papers has defended an account of mental causation based on such counterfactuals.[2]  In this paper I argue that such an

---

[1] Kim (1998) provides the classic presentation of the problem.
[2] See Yablo (1992), (1997), and (2003).  Other authors who appeal to counterfactuals relevantly like [Cry] include LePore and Loewer (1987), Loewer (2001), Mills (1996), Bennett (2003), and Bealer (2007).

approach is misguided. In Sections I-V, I argue that accounts of mental causation that appeal to counterfactuals like [Cry] are susceptible to counterexamples of a certain sort: counterexamples involving what I call *disproportional* mental causation. In Section VI, I argue that the core problem with an account like Yablo's is not its appeal to counterfactuals per se, but rather its guiding idea that mental and physical states *causally compete*. Any nonreductive view that posits such causal competition will be susceptible to counterexamples involving disproportional mental causation. Finally, in Section VII, I show that there are alternative nonreductive approaches to mental causation that do not posit causal competition and which thus are able to account for disproportional mental causation. This point is developed into a novel argument in favor of such non-competitive approaches.

**I**

We begin by reviewing Yablo's account of mental causation, which comes in two parts. First part: Yablo claims that the realization relation that obtains between mental and physical states is either identical or at least very similar to the determination relation that obtains between determinables and determinates.[3] On this view, the sense in which scarlet and crimson are different ways for an object to be red is either identical to or at least very much like the sense in which being in $P_1$ and being in $P_{101}$ are different ways for the subject in our example to be in pain. If, as Yablo further contends, determinables and their determinates do not compete for causal *influence* – which Yablo understands as "encompassing everything from causal relevance to causal sufficiency"[4] – then this initial move of construing realization in terms of determination promises to go some way by

---

[3] In his (1992) he claims that realization is determination, while in his (1997) Yablo weakens it to the claim of strong similarity.
[4] Yablo (1992: 274).

itself toward dissolving the exclusion problem.[5] For the sake of my argument I am

willing to grant Yablo this component of his view. I will assume for now that realization

just is determination.

The second part of the view is Yablo's claim that causes generally are

*proportional* to their effects, meaning, inter alia, they do not incorporate detail irrelevant

to those effects. This idea is captured with the following principle.

> [PP]: A state D incorporates detail that is irrelevant with respect to an effect E,
> and so does not cause E, if there is some state C such that C is a determinable of
> D and the following counterfactual is true: $(C \& \sim D) > E$.

To see [PP] in action, consider Sophie the pigeon who pecks whenever she sees

red.[6] When presented with a scarlet triangle, Sophie pecks. Is the triangle's being scarlet

properly regarded as causing Sophie's pecking? No it is not, says [PP]. For red is a

determinable of scarlet, and had the triangle been red without being scarlet – for instance,

had it been crimson – Sophie still would have pecked. The intuition you are invited to

have here is that it is the triangle's being red *and not* its being scarlet which causes the

pecking.[7] The Yablo view, then, is what while determinables and determinates do not

compete over causal influence – again, understood as encompassing causal relevance and

sufficiency – they *do* compete when it comes to causation itself. In the present case, the

triangle's being red wins the causal competition; its being scarlet loses.

The same dynamic is alleged to arise in the mental/physical case. Reconsidering

Gilmore's pain, and granting again that realization is determination, [PP] entails that if

[Cry] is true then Gilmore's $P_1$ state incorporates detail irrelevant to his crying, and so is

---

[5] Although see Gillett and Rives (2005), who raise exclusion worries for determinables.
[6] Yablo (1992: 257).
[7] On Yablo's view the triangle's being scarlet is causally sufficient for Sophie's pecking, but does not cause the pecking. There are serious questions about how exactly Yablo understands the relation between causation and causal sufficiency, but I will not press them here.

not properly regarded as causing his crying.[8] This makes room for the possibility that what causes Gilmore's crying is his pain, just as in the Sophie case what causes the pecking is the triangle's being red rather than its being scarlet. Again, the intuition you are invited to have here is that the truth of [Cry] shows that the crucial thing for causing Gilmore's crying is the pain itself, not the particular way it happens to be realized.

When a counterfactual like [Cry] is false, on the other hand, the account entails that it is the underlying physical realizer which causes the effect in question, not the mental state being realized. Suppose there is a $P_1$-detector pointed at Gilmore while he undergoes his pain, and let $B$ be the (true) proposition that the detector beeps, registering the presence of a $P_1$ state. Then the following counterfactual is false:

[Beep]: $(M \mathbin{\&} \sim P_1) > B$.

Had Gilmore's pain been realized by anything other than a $P_1$ state – for instance, had it had been realized by a $P_{101}$ state – the $P_1$-detector would not have beeped. On Yablo's account the falsity of [Beep] shows that it is the $P_1$ state and not the pain which causes the beeping. In this case at least, the amount of detail the $P_1$ state incorporates is just right.[9]

Let's summarize. On Yablo's view mental states do indeed causally compete with their physical realizers, at least when it comes to causation itself. No worries though. Mental epiphenomenalism is avoided because mental states often *win* their competitions. The way we assess which side wins a causal competition is by appealing to proportionality considerations, which involves evaluating counterfactuals like [Cry] and [Beep]. When such counterfactuals are true, as [Cry] is, the winner of the competition is the mental state – Gilmore's pain and not his $P_1$ state causes his crying. When such

---

[8] The $P_1$ state is still causally sufficient for the crying. See the preceding note.
[9] Compare Yablo's (1992: 277-278) discussion of the epiphenomenalist neuroscientists.

counterfactuals are false, as [Beep] is, the winner is the physical realizer – Gilmore's $P_1$ state and not his pain causes the detector to beep.

**II**

Sometimes critics of this general sort of counterfactual-based approach complain that the counterfactuals in question could be true even while mental states were causally inert. In our example these critics would argue that even if [Cry] is true this does not establish that Gilmore's pain causes him to cry. Instead, what the truth of [Cry] might reflect is just that if Gilmore's pain had been physically realized in some other way, that alternative physical realizer would have caused him to cry, causally excluding his pain. In short, these critics charge that the approach fails to come to grips with the true depth of the exclusion problem.[10]

My objection runs in roughly the opposite direction. What I will argue is that there are cases in which the relevant counterfactual is false and yet we nevertheless have compelling reason to say that there is mental causation. In a sense, my charge is that Yablo takes the exclusion problem too seriously; he concedes too much to those opponents of nonreductive physicalism who push the problem. Getting well ahead of myself, I believe that Yablo's mistake is to concede that there is at least some domain in which causal competition takes place between mental states and their physical realizers. Nonreductive physicalists should insist there is no form of causal competition at all.

For the purpose of constructing a counterexample to Yablo's account, let's suppose that the laws of nature are such that there are exactly four nomologically possible physical realizers of pain: $P_1$, $P_2$, $P_{101}$, and $P_{102}$. The subscripted numerals here are meant to track physical similarity: $P_1$ and $P_2$ are very similar to one another, while each is quite

---

[10] See for instance Leiter and Miller (1994), who object to LePore and Loewer (1987) along these lines.

different from $P_{101}$ and $P_{102}$. To flesh out the story we can imagine that $P_1$ and $P_2$ are physical realizers found only in human beings, while $P_{101}$ and $P_{102}$ are found only in Martians. Next suppose that while $P_1$, $P_{101}$, and $P_{102}$ states all causally necessitate crying, $P_2$ states do not. In fact, we can even add, it is nomologically impossible for a being in a $P_2$ state to cry. Again to flesh out the story, we can imagine that people with $P_2$ realizers behaviorally manifest their pains in various ways. They wince, they scream, they gnash their teeth, and so on. But never do they cry.[11]

Given this setup, [Cry] is false. Gilmore is a human being, and so if he had been in pain but not $P_1$ his pain would have been realized by a $P_2$ state.[12] It is nomologically impossible for subjects in $P_2$ to cry though. Therefore, if Gilmore had been in pain but not $P_1$ he would not have cried.

Even though [Cry] is false, however, we still have compelling reasons to say that Gilmore's pain causes his crying in the scenario set out. Sections III-V are devoted to spelling out these reasons in detail, but as a first pass the guiding idea can be put as follows. Gilmore's actual pain is realized by $P_1$, not by $P_2$. But then, I say, the causal status of his pain should not turn on how things go with $P_2$. It should turn only on how things go with $P_1$-reazlied pains. Since $P_1$ causally necessitates crying, we should conclude that $P_1$-reazlied pains, like Gilmore's, cause crying. If I can defend this view

---

[11] To make things most realistic, we could suppose not that $P_1$, $P_{101}$, and $P_{102}$ states by themselves causally necessitate crying, but rather that there is some background condition BC such that any one of these states taken together with BC necessitates crying. By extension we could then suppose not that $P_2$ by itself is nomologically incompatible with crying, but rather that it taken in conjunction with BC is. So, people with $P_2$ realizers do sometimes cry, but never when BC obtains – $P_2$ is incompatible with crying *given* BC. For ease of exposition I leave out explicit reference to background conditions in the text, but I invite you to read them into my discussion if it makes the example more compelling.

[12] The present argument could make do with the weaker claim that if Gilmore had been in pain but not $P_1$ he *might* have been in $P_2$. On the standard analysis of counterfactuals this would be enough to ensure the falsity of [Cry]. I make the stronger claim in the text because it allows me to simplify my presentation and because I find the stronger claim not implausible. See note 20 though.

successfully, I will have established a counterexample to Yablo's account: Gilmore's

pain causes his crying even though [Cry] is false, and thus even though it is not

proportional to his crying.

**III**

There are two possible strategies a defender of Yablo could adopt in trying to

resist my line of attack.  The first is to argue that I have failed to describe a scenario in

which [Cry] really would be false; the second is to concede that [Cry] is false but contend

that this is no embarrassment to Yablo's view since Gilmore's pain does not cause his

crying.  In this section I will address the first strategy.

One way to advance the first strategy would be to raise broadly functionalist

worries about whether $P_2$ could qualify as a physical realizer of pain given that it is

nomologically impossible for subjects in $P_2$ states to cry.  According to functionalism, a

physical state is a pain realizer only if it occupies the causal role associated with pain.  If

a capacity to cause crying is an essential part of that role, then $P_2$'s inability on this front

disqualifies it as a pain realizer.  If this is right, then the Gilmore scenario we have

imagined is impossible upon closer inspection.[13]

In response, I note that we can assume that $P_2$ does an otherwise perfect job at

occupying pain's causal role.  $P_2$ states are causally necessitated by tissue damage, they

causally necessitate wincing, screaming, teeth gnashing, and so on.  The only blemish on

their causal role résumé is the part about the crying.  If $P_2$ does such a near-perfect job at

---

[13] Fodor (1991: 25) briefly considers this sort of objection in response to an argument made by Schiffer (1991), but ultimately sets it aside.  According to Fodor, a state like $P_2$ could qualify as a pain realizer, despite its inability to cause crying, if it has something sufficiently in common with other pain realizers. As I'm about to argue, this holds in our example.

occupying pain's functional role, it is utterly implausible to disqualify it as a realizer of

pain on functionalist grounds.

In case this implausibility needs to be drawn out a bit, consider an intuition pump.

Imagine I come up with an evil new medical experiment I want to run on human beings.

The good news is that the experiment promises to increase our medical knowledge ever

so slightly; the bad news is that it is excruciatingly painful for typical subjects to

participate in.  Now, as it turns out, in one tenth of all human beings the closest thing

there is to a perfect realizer of pain's functional role is the merely near-perfect $P_2$.  Just to

be clear: this bit of the story begs no question against proponents of the present

functionalist-inspired line, for that line does not entail that a state with $P_2$'s causal profile

is impossible, nor does it entail that such a state could not be found in human beings.

Rather, what the view in question entails is that if a tenth of the population were to have

$P_2$ states, then that tenth would not really undergo pain, since they would have no

physical state perfectly realizing pain's functional role.

Given the functionalist-inspired view in question, then, it would seem there could

be no strong moral objection to me performing my medical experiments on the tenth of

the population with $P_2$ states.  For, although the tenth will wince and scream and gnash

their teeth during the experiment; although they will hide from us to try to avoid

participating in it, and then later beg and plead with us after we find them; although they

will act like typical pained human beings in almost all respects, they will not really be in

pain.  They cannot really be in pain, according to this functionalist-inspired view, since

they do not cry.  This *alone* is grounds enough to deny that their $P_2$ states are pain

realizers, and thus to deny that they feel pain. They do not cry, so we can start the medical experiments with an easy conscience!

Thankfully, most functionalists are not committed to such a crazy view. Following David Lewis, many functionalists appreciate that the psychological theory used to functionally define mental states (i.e., the theory functionalists Ramsify in generating their functional definitions) might not be perfectly realized, but that if it's close enough – if, for instance, the disjunction of the conjunctions of most of the clauses of the theory is true – we should still say that the states defined by the theory obtain (i.e., that the mental terms of the theory refer). This is what we find in our present scenario with $P_2$: the vast majority of the clauses of the defining psychological theory will be true of people with $P_2$ states, with the only false clause being that pain causes crying in them. Adopting terminology from Lewis, what we should say in such a case is that since $P_2$ does such a near-perfect job occupying pain's functional role, it counts as a *near-realizer* of pain, and thus as a realizer-simpliciter of pain.[14] On this very familiar and most plausible version of functionalism, beings in $P_2$ states are in pain, just as my objection to Yablo supposes.

Moving on, here is another way one might argue that I have failed to describe a scenario in which [Cry] is false. Even if one grants that $P_2$ qualifies as a pain realizer, one might contend that it is a comparatively unusual realizer given that it does not causally necessitate crying. From there, one could then argue that the nearest worlds in which Gilmore is in pain but not $P_1$ are not worlds where his pain is realized by a $P_2$ state. Those worlds are far away. The closest worlds where his pain is alternatively realized are worlds where it is realized by $P_{101}$ or $P_{102}$, since these are the more typical realizers. If so,

---

[14] See Lewis's (1970: 432) discussion of near-realization.

then [Cry] is true since $P_{101}$ and $P_{102}$ states causally necessitate crying. In short, the idea is that the unusualness of $P_2$ as a physical realizer of pain trumps its physical similarity to $P_1$ when it comes to determining the proximity of worlds for the sake of evaluating [Cry].

In response, I note that we can simply stipulate that $P_2$ is no more unusual as a realizer of pain than any of the other realizers are. We can stipulate that it is nomologically impossible for $P_1$ states to causally necessitate wincing, for $P_{101}$ states to necessitate screaming, and for $P_{102}$ states to necessitate teeth-gnashing. Each of these realizers is then just like $P_2$ in doing a merely near-perfect job occupying pain's associated causal role. If so, then $P_2$ isn't a comparatively unusual realizer of pain after all, and the present line of defending Yablo fails to get off the ground.

I cannot think of any other promising argument for denying that [Cry] is false in the scenario described, so at this point let's shift to consider the second strategy that a defender of the Yablo account could adopt. She could concede that [Cry] is false but contend that this poses no problem for the account because the correct verdict in the Gilmore case is that Gilmore's pain does not cause him to cry. I have two independent arguments against this suggestion.

**IV**

First, consider Counterpart Gilmore, an intrinsic duplicate of Gilmore's who inhabits a different possible world. Counterpart Gilmore's world is as much like Gilmore's as possible except that there $P_2$ states *do* causally necessitate crying. Suppose then that, like Gilmore, Counterpart Gilmore suffers a pain, his pain is realized by a $P_1$ state, and he cries. When evaluated with respect to Counterpart Gilmore, [Cry] will thus be true, not false. Had Counterpart Gilmore been in pain but not $P_1$, his pain would have

been realized by $P_2$, and $P_2$ states causally necessitate crying at Counterpart Gilmore's world. Therefore, had Counterpart Gilmore been in pain but not $P_1$, he would have cried.

Yablo's account thus entails there is a deep causal difference between Gilmore's pain and Counterpart Gilmore's pain. Since [Cry] is true when evaluated with respect to Counterpart Gilmore but false when evaluated with respect to Gilmore, it follows that Counterpart Gilmore's pain is proportional to crying while Gilmore's pain is not. In turn it follows that Counterpart Gilmore's pain causes his crying but Gilmore's pain does not.

It is implausible, however, that the two pains could causally differ in any important way. The only difference between them is that one takes place in a world where $P_2$ states causally necessitate crying while the other does not. How could this difference make a difference? After all, neither Gilmore nor Counterpart Gilmore is in a $P_2$ state. We can further suppose that neither Gilmore nor Counterpart Gilmore will ever in their lives be in a $P_2$ state, or even have the slightest causal interaction with one. $P_2$ states are as utterly unconnected to Gilmore and Counterpart Gilmore as anything in their worlds are. If so, I cannot see how a causal difference between the two pains could be grounded in this sole, seemingly irrelevant difference in the laws between their worlds.

This argument turns on a comparison of intrinsic duplicates from worlds governed by different laws of nature. This might spark concern: we don't generally expect intrinsic duplicates to be causally alike if the laws at their worlds differ. The concern is misplaced. For one thing, I emphasize again that the difference in laws has to do with $P_2$ states only, and these laws seem irrelevant to Gilmore and his Counterpart. For another, it is possible to recast the preceding argument in epistemic terms that eliminate entirely the comparison across worlds.

In the epistemic version of the argument we imagine that we presently know that the pain realizers $P_1$, $P_{101}$, and $P_{102}$ all causally necessitate crying. We also presently know that there is a fourth realizer, $P_2$. What we do not know yet is whether $P_2$ causally necessitates crying. Given this state of knowledge, the question once again is whether Gilmore's $P_1$-realized pain causes his crying.

Yablo's account entails that in order to answer this question, we need to learn that which we presently do not know: whether $P_2$ causally necessitates crying. To learn that it does would be to learn that [Cry] is true and thus (according to Yablo) that Gilmore's pain causes his crying. To learn that it does not would be to learn that [Cry] is false and thus (according to Yablo) that Gilmore's pain does not cause his crying. In order to figure out whether there is mental causation in the Gilmore case, then, we need to set Gilmore himself aside and go study $P_2$ states, see what they causally necessitate.

This cannot be right. Given that $P_2$ is as utterly unconnected to Gilmore as anything in the world is, I say that this cannot be the way to come to know whether there is mental causation in Gilmore's case. If my verdict is correct then Yablo's account must be wrong. This epistemic argument is a recognizable variation on the Counterpart Gilmore argument, but, again, avoids comparison across worlds with different laws. I return to the Counterpart Gilmore version, though, in order to frame my final points for this section.

So far I have aimed to establish that Gilmore's pain could not differ in causal status from Counterpart Gilmore's pain. I now add the further claim that both should be regarded as having positive causal status – that is, as causing crying. For, if mental causation ever takes place, then surely it takes place with Counterpart Gilmore's pain.

Any viable account of nonreductive mental causation will agree with this conclusion, including Yablo's. Since the two pains cannot causally differ, it thus follows that mental causation takes place with Gilmore's pain as well. And so, despite the falsity of [Cry], we have compelling reason to hold that Gilmore's pain causes his crying.

**V**

The second argument for the causal efficacy of Gilmore's pain requires some stage setting. There clearly is no strict, exceptionless law linking pain to crying at Gilmore's world, given that not all pains are accompanied by crying there – $P_2$-realized pains are not. In principle, a Yablo defender could try to use the absence of a strict law to argue against the causal efficacy of Gilmore's pain. This would not be a promising line to take, however. For, it is widely accepted that in the actual world there are no strict, exceptionless psychological laws.[15] At best, it is held, there may be ceteris paribus laws.

Now, perhaps the most influential account of ceteris paribus psychological laws is due to Jerry Fodor.[16] For the sake of the argument that follows I don't need to endorse everything Fodor says, or even commit myself to there being such things as ceteris paribus laws. Perhaps whatever true ceteris paribus psychological generalizations there are should be assigned a status short of full-blown lawhood.[17] The one element of Fodor's account I'll be using is his view of what goes on "beneath" a ceteris paribus psychological generalization, at the underlying level of physical realizers.

There is no strict law linking pain and crying at Gilmore's world, but perhaps we can say that pained subjects cry there, ceteris paribus. That is, they cry provided that

---

[15] Davidson (1970) is the classic defense of this thesis, but many philosophers' accept Davidson's conclusion without accepting his reasoning.

[16] Fodor (1974). See also the further discussion in Fodor (1987) and (1991).

[17] For arguments against there being any ceteris paribus laws see for instance Schiffer (1991) and Earman, Roberts, and Smith (2002).

their pain is not realized by a $P_2$ state. $P_2$-realized pains constitute an "absolute exception" to the ceteris paribus generalization that crying accompanies pain.[18] Fodor's view is that the ceteris paribus nature of psychological generalizations is generally to be explained at the underlying physical level in terms of realization-specific absolute exceptions of this very sort. It is precisely because there are such absolute exceptions that the true psychological generalizations there are hold only ceteris paribus rather than strictly.[19] If Fodor is right, then there is nothing unusual or artificial about the Gilmore case we have constructed. Take any actual psychological generalization you please and there will be some absolute exception to it, relevantly like $P_2$-realized pains.

Relating the discussion back to Yablo, I claim that if we deny mental causation in Gilmore's case, then by parity of reasoning we might well be forced to deny that there is any mental causation here in the actual world. To make the argument explicit, let's call a world a *Fodor World* if it satisfies each of the following three conditions: (i) every true psychological causal generalization there holds only ceteris paribus; (ii) every such generalization has an absolute exception, in the sense just spelled out; and (iii) for each mental state falling under a given generalization, the nearest counterfactual world where that mental state is alternatively realized by an absolute exception to the generalization is at least as close as the nearest counterfactual world where the mental state is alternatively realized by a non-exception to the generalization.[20] Intuitively, the idea is that a Fodor

---

[18] The term "absolute exception" is taken from the exchange between Schiffer (1991) and Fodor (1991), who use it to cover cases exactly like Gilmore's.

[19] See Fodor (1974: 108-111). An attractive feature of this view is that by positing absolute exceptions, Fodor is able to explain how it could be that psychological generalizations hold only ceteris paribus while the physical laws governing realizers are all exceptionless.

[20] Condition (iii) is jointly entailed by (i) and (ii), given the following principle: for each mental state M and for each physical realizer $P_i$ of M, the nearest counterfactual world where M is realized by $P_i$ is just as close as the nearest counterfactual world where M is realized by some other physical realizer $P_j$. In other words, the nearest alternative physical realizations of a mental state are all equally far away. I am not sure

World is a place where every putative instance of mental causation is relevantly like the case involving Gilmore's pain. At a Fodor World then every counterfactual relevantly like [Cry] is false. Therefore, just as Yablo's account entails that Gilmore's pain does not cause his crying, it also entails that there is no mental causation at all in a Fodor World.

Two points about Fodor Worlds. First, it would seem to be a wide open empirical question whether the actual world is a Fodor World. I can see no armchair reason to think it is not.[21] Therefore, it is a wide open empirical possibility on Yablo's view that there is no actual mental causation. Second, even if the actual world isn't a Fodor World, it might at least turn out to be rather Fodorish. For instance, it might turn out that while no actual putative instance of mental causation involving a desire is Gilmore-like, every putative instance involving a belief is. Or it might turn out that half of all putative mental causation is Gilmore-like while half is not. These, again, would seem to be wide open empirical possibilities. Therefore, it is a wide open empirical possibility on Yablo's view that there is no belief causation, or that half of all putative mental causation is bogus.

These results are unacceptable. Much of the motivation behind Yablo's account is that it promises to save mental causation as we intuitively understand it, more or less. But now this appears very much up in the air. Whether it does or not depends on the empirical question of whether the actual world is a Fodor World, a possibility that I

---

this principle is correct, but I am not sure it is incorrect either – maybe, when it comes to assessing counterfactuals like [Cry], all alternative physical realizers should be treated equally. The Gilmore case as it has been set out violates the principle since we have assumed that the nearest world where Gilmore's pain is realized by $P_2$ is closer than nearest world where it is realized by either $P_{101}$ or $P_{102}$. However, this assumption has been inessential to the argument: even if the nearest $P_2$-world were merely as close as the nearest $P_{101}$-world and $P_{102}$-world, everything in my argument would go through given the standard analysis of counterfactuals, since this would still be enough to ensure that if Gilmore had been in pain but not in $P_1$, he might have been in $P_2$. See note 12 above. Even if the present principle is to be rejected and (i) and (ii) don't jointly entail (iii), there are other scenarios on which (iii) could be true.

[21] Fodor himself does not explicitly discuss condition (iii). However, he does suggest that both conditions (i) and (ii) hold here in the actual world, and so if (i) and (ii) jointly entail (iii) – as suggested in the last note – he is committed to the actual world being a Fodor World.

otherwise would have thought nonreductive physicalists could be open to. In order to avoid these results we must reject any view that denies that Gilmore's pain causes his crying. We must reject Yablo's view.

The preceding argument suggests that the problem the Gilmore case poses for Yablo's account is potentially quite deep. If all or many actual instances of mental causation are Gilmore-like, then an account of mental causation which yields the wrong verdict on the Gilmore case is doomed to yield the wrong verdict on all or many actual cases of mental causation.

**VI**

From this point on I will take it as established that the preceding arguments give us good reason to reject the Yablo account as it stands. Now, one could agree with this assessment while still thinking there is something importantly right about the appeal to proportionality. A number of philosophers influenced by Yablo have embraced something like the proportionality component of his view while rejecting or at least remaining neutral on his use of counterfactuals.[22] What these philosophers take from Yablo are the ideas that (i) there is indeed causal competition between mental and physical states, and (ii) the way to sort out which side wins such a competition is by looking around at what happens with alternative physical realizations of the putative mental cause. Let me say something about each of these in turn.

Regarding (i), I have been employing the notion of causal competition throughout this paper, but it will be helpful here to clarify it a bit. Let's say that a mental state and

---

[22] This includes Shoemaker (2001) and (2007) and Williamson (2000) and (2005). Shoemaker appeals to proportionality, but explains proportionality in terms of his own subset model of realization rather than counterfactuals. Williamson (2000: 82) cites Yablo and relies on something like proportionality while arguing for the causal efficacy of knowledge, but, as he later (2005) makes explicit, rejects Yablo's counterfactual-based approach to causation.

its physical realizer causally compete just in case there is some important causal dimension along which at most one of the states can have a positive causal status. By "important causal dimension," I mean to cover things like causal relevance, causal sufficiency, and causation itself. By "positive causal status," I mean statuses like that of being causally relevant to a given effect (rather than irrelevant), that of being causally sufficient for a given effect (rather than insufficient), and that of being a cause of a given effect (rather than not a cause).

When Yablo says that mental states and their physical realizers do not compete over causal relevancy or sufficiency, what he means is that it is possible for both states to have a positive status along these causal dimensions – it is possible for both a mental state and its physical realizer to be causally relevant to some effect, or for both to be causally sufficient for that effect. There is another causal dimension along which there is causal competition according to Yablo, however – that of causation itself. It is not possible for both a mental state and its physical realizer to have the positive status of causing a given effect. This is why Yablo appeals to proportionality – to sort out which state wins the causal competition, which state is awarded the positive status of cause.

My root objection to Yablo is really an objection to this competitive element of his view. I believe that the proper account of nonreductive mental causation needs to be thoroughly *non-competitive*. It needs to say that there is no causal dimension along which mental states and their physical realizers causally compete in the sense spelled out. I will survey a few non-competitive accounts in the next, concluding section, and show that they are not susceptible to the problems I pose for competitive accounts like Yablo's.

17

In this section, however, my focus will be on (ii), the second idea philosophers have taken from Yablo. I regard the true core of the proportionality component of Yablo's view not to be his particular use of counterfactuals, but rather his idea of settling a causal competition between a mental state and its physical realizer by looking at what happens with alternative realizations of that mental state. Counterfactuals like [Cry] enter the picture because they provide one way of trying to capture this idea for settling causal competitions. Potentially there are other ways one might try to capture the idea instead, however. Perhaps some of these other ways are not as susceptible to counterexample as Yablo's counterfactual-based approach is.

Along these lines, consider the following conveniently unsophisticated proposal. To determine whether a mental state of type M causes an effect of type E, divide M's realizers into two groups: (1) those that do causally necessitate E effects, and (2) those that do not. The putative instance of mental causation is then genuine only if group (1) is bigger than group (2). Applying this proposal to the Gilmore case, $P_1$, $P_{101}$, and $P_{102}$ all belong to group (1) while only $P_2$ belongs to group (2). Therefore, according to the proposal, Gilmore's pain causes his crying – the proper verdict.

What this shows is that it is possible to get the Gilmore case right while still retaining what I'm regarding as the heart of proportionality, the idea of settling causal competitions by looking at what happens with alternative realizations. No doubt, more sophisticated proposals that accomplish this result could be developed as well. Perhaps then the proper conclusion to draw at this point is just that we need a new way of cashing out the guiding idea of proportionality, not that the idea should be jettisoned completely.

18

This is what I now will argue against. I think that the core idea of proportionality needs to be thrown out, regardless of how we cash it out. To make my case I need a new thought experiment, so meet Mullin. Tomorrow is the big logic exam, and Mullin is feeling quite anxious about it. Her anxiety putatively causes her heart to race. Suppose that there are four nomologically possible physical realizers of anxiety: $P_{10}$, which realizes it in human beings (including Mullin); $P_{20}$, which realizes it in Martians; $P_{30}$, which realizes it in Venusians; and $P_{40}$, which realizes it in Jupiterians.[23] Suppose also that while $P_{10}$ states causally necessitate heart rate acceleration, all other physical realizers of anxiety fail to do so. Martians, it turns out, have glowing orbs instead of hearts, and so never undergo heart rate acceleration. Venusians have hearts, but hearts made of iron which always beat at the same constant rate. Jupiterians have hearts physically indistinguishable from our own, but the wiring connecting their hearts to their brains is so different from ours that Jupiterian heart rates decelerate during anxiety.

At Mullin's world then it is only when anxiety is physically realized by $P_{10}$ that heart rate acceleration ensues. The Yablo-style counterfactual is thus false: if Mullin had been anxious but not in a $P_{10}$ state, her heart rate would not have accelerated. More to the present point, however, it seems that no matter how one might try to cash out the idea of proportionality, of taking mental causation to depend on what happens with alternative physical realizations, any proportionality-based account will be forced to say that Mullin's anxiety does not cause her heart rate acceleration, since heart rate acceleration does not occur when anxiety is realized by anything other than a $P_{10}$ state. Reconsider the unsophisticated divide-and-count approach. It entails that Mullin's anxiety does not

---

[23] The similarity of these physical realizers won't matter for the discussion.

cause her heart to race since group (2), which includes $P_{20}$, $P_{30}$, and $P_{40}$, outnumbers

group (1), which includes only $P_{10}$.

Intuitively, this seems like the wrong result. For all we know the real world may

be exactly like Mullin's world. Would learning tomorrow that there are aliens who lack

hearts, or whose iron hearts always beat at a constant rate, or whose differently wired

hearts decelerate during anxiety, put any pressure on us at all to deny what is presently a

matter of commonsense, that in human beings anxiety does cause hearts to race? Surely

not. We might qualify our causal claims about anxiety in light of these findings. Instead

of saying that anxiety causes heart rate acceleration, full stop, it would be better to say

that anxiety causes this effect *in human beings*. But this still would be to assign causal

efficacy to those anxieties that do take place in human beings, and so it would agree with

the commonsense verdict that Mullin's anxiety causes her heart to race.

Even stronger anti-proportionality intuitions are generated if we shift to a case in

which rationality considerations enter in, and then pit rationality against proportionality.

The fact that rationality can oppose proportionality has not been noted in the literature, to

my knowledge, and so I think the following case is especially interesting. After hours of

studying, Mullin begins her logic exam. The first question presents students with a

proposition, *Q*, which deductively entails another proposition, *A*, the sought answer.

Mullin gets the question right. Putatively, her thought of *Q* causes her subsequent

thought of *A*. Suppose there are four nomologically possible physical realizers of

thinking of *Q*: $P_{15}$, which is the realizer of Mullin's thought, $P_{25}$, $P_{35}$, and $P_{45}$. Suppose

next that while $P_{15}$ states causally necessitate thinking of *A*, each of the other physical

realizers causally necessitates thinking of the same wrong answer, ~*A*. We can imagine

the problem as a trick question, where only minds physically built in a certain way avoid being taken in by the trick.

Now, some philosophers hold that there are constitutive norms of rationality such that a necessary condition on being a thinker at all is that one be fairly rational.[24] There is much to be said for such a view. However, no plausible version of it could be used to rule out as impossible the scenario just described. We can suppose that those beings without $P_{15}$ realizers are as rationally adept as can be but for the one exception that when they think about $Q$, they wrongly infer $\sim A$ instead of $A$. Any version of the constitutive rationality thesis which says that even this is too much irrationality for those beings to qualify as thinkers is absurd.

The scenario is metaphysically possible then. What's more, I don't see any obvious reason to think it is empirically improbable. To begin with, it would be surprising if facts about physical realization *didn't* impose constraints on a thinker's rational acumen, so that certain rationally proper inferences are simply impossible for thinkers built in a given physical way.[25] Assuming this is so, the real question is whether such constraints are uniform across different physical realizations or whether thinkers built one physical way are saddled with *different* rational limitations than thinkers built another. This is an open empirical matter. Human beings make more errors reasoning with modus tollens than they do with modus ponens. Perhaps in Martians the pattern is reversed, and perhaps in addition there is no explanation for this human/Martian difference at the psychological level, but only one at the underlying physical level.

---

[24] Famously, Davidson (1970) assigns this view a central role in his account of mental causation.
[25] It is worth reemphasizing here the point made back in note 11, that the alleged impossibility may be relative to certain background conditions. So, for instance, perhaps the inference in question is one that it sometimes possible, but not given limitations on time, attention, and so on.

As in the anxiety case, any proportionality-based account of mental causation will entail that Mullin's thought of $Q$ does not cause her thought of $A$, and this is true regardless of how proportionality is cashed out. What the present example adds to the discussion is that here, Mullin is performing the way she rationally ought to perform. A thought of the correct answer is precisely what a thought about a logic problem *should* cause. This, I suggest, makes it even more counterintuitive to follow proportionality here and deny causal efficacy to Mullin's thought. If we are forced to deny the causal efficacy of anyone's thought in the example – and, ultimately, I don't think we are – we should deny it for those beings that get the problem wrong, not Mullin who gets it right. After all, irrationality is the sort of thing we would expect from thinkers whose mental states are being causally excluded. Heightened rationality is not.

**VII**

Nonreductive accounts of mental causation can be divided into two camps: those that deny causal competition between mental and physical states, and those that grant some form of competition but contend mental states often win. As we have seen, proportionality based-accounts belong to the latter group. In constructing cases of disproportional mental causation, what we do is stack the deck of the supposed competition as heavily as possible in favor of the physical realizer, so that it must be the winner if there is a winner at all. For instance, if Mullin's thought of $Q$ causally competes with her $P_{15}$ state, then her thought of $Q$ must lose – the case was constructed so as to ensure this result. It is an embarrassment for competitive accounts of mental causation that they are unable to accommodate disproportional mental causation. To avoid this embarrassment, we need to look to noncompetitive accounts instead.

One example of a noncompetitive account is Donald Davidson's.[26] Davidson's view of mental causation is notoriously problematic, but it might still serve as a helpful reference point. According to Davidson, mental event tokens are identical to physical event tokens, while mental types are irreducible to physical types. In addition he holds that causation is a binary extensional relation that holds between events regardless of the properties those events instantiate or exemplify. Given this view of causation, Davidson thinks that an adequate account of mental causation has done its job in full once it has shown mental events are causes. This view thus qualifies as non-competitive by our standards: along the only causal dimension that matters, that of causation itself, it is possible for both mental and physical states to have a positive causal status with respect to a given effect – in fact, given the token identity, it is guaranteed that a mental token has a positive status if and only if physical token does.

Pretend for a moment Davidson's view were satisfactory. Looking back at the examples of disproportional mental causation in this paper, we see that Davidson's account yields the proper verdict each time. Take the logic case. On the Davidson view, Mullin's thought of $Q$ is token identical with her $P_{15}$ state, and so there is no room for causal competition between them, not even – pace Yablo – with respect to causation itself. True, the mental type of thinking of $Q$ is irreducible to the physical type of $P_{15}$; but again, types are irrelevant to causation. If we assume that Mullin's $P_{15}$ state causes her thought of $A$, it follows by the token identity that her thought of $Q$ causes this effect. And this, on the Davidson view, is all that matters for mental causation.

On the view, whatever happens with alternative physical realizations of the thought of $Q$ is irrelevant to the question of whether Mullin's particular thought causes

---

[26] Davidson (1970), and then further defended and elaborated in his (1993).

23

her subsequent thought of *A*. Even if no alternative physical realization of thinking of *Q* causally necessitates thinking of *A*, this does nothing to undermine the proposed token identity between Mullin's thought of *Q* and her $P_{15}$ state, and so it does nothing to undermine her thought's claim to causing her thought of *A*. Davidson gets disproportional mental causation right – he is able to allow for it.

Of course, Davidson's view is not satisfactory as it stands; that it is not is the starting point of much recent work on mental causation. But this undermines our use of it as a reference point only if the problems facing his account are inherent to noncompetitive views generally. And this does not seem to be the case. Davidson's main problem was that he was wrong to think that an adequate account could get by merely with showing that coarse-grained Davidsonian mental events are causes. In addition, any satisfactory account needs to establish the causal efficacy of mental properties. It needs to show that mental events are causes by virtue of their mental properties.[27]

A noncompetitive account that promises to do just this has been developed by David Robb.[28] According to Robb, properties, insofar as they are relevant to causal relations, should be conceived as *tropes*, or abstract particulars. Like Davidson, Robb holds that mental events are token identical with physical events. Going beyond Davidson, he also holds that mental tropes are token identical with physical tropes. The view qualifies as nonreductive because in addition to events and tropes it also recognizes types, conceived as either universals or resemblance classes of tropes. Mental types are distinct from physical types (so understood) for familiar reasons of multiple realizability.

---

[27] Among the first to present this criticism were Honderich (1982), Sosa (1984), and Kim (1984).
[28] Robb (1997) and (2001).

Because types are not relevant to causation, however, this type distinctness does not open the door to causal competition.

On Robb's view, Mullin's thought of $Q$ is the same event as her $P_{15}$ state, and so if the $P_{15}$ state causes the thought of $A$, it follows that the thought of $Q$ causes this effect. But in addition, on Robb's view, the causally relevant physical property exemplified by this event, the $P_{15}$ trope, is token identical with the exemplified mental property, the thought of $Q$ trope, and so by the token identity this mental property is causally relevant. This is the sought result: Mullin's thought of $Q$ causes her thought of $A$, and it does so by virtue of its mental property (trope). The view improves on Davidson's in making mental properties causally relevant, but is like in Davidson's in disallowing causal competition. On Robb's view, it is irrelevant what happens with alternative physical realizations of the thought of $Q$. For again, even if those alternative realizations fail to causally necessitate thoughts of $A$, this does not undermine the proposed token identities, and so it does not undermine the claim that Mullin's thought of $Q$ causes her thought of $A$ by virtue of the former thought's mental property.

Is there any way to develop a noncompetitive account that allows for disproportional mental causation other than by embracing some sort of token identity thesis, be it at the level of events (Davidson), tropes (Robb), or whatever? The key to doing so would be to show that there is some potential mind/body relation other than identity that prevents causal competition from arising. One possibility on this front would be to develop an account that took over the determinate/determinable component of Yablo's view but dropped entirely the proportionality component. The rough idea would be to go further than Yablo by saying that determination gives rise to *no* form of

causal competition, not even with respect to causation. Developing such an account falls outside the scope of the present paper. I mention it here in part because I really do think it might be a promising option, and in part to remind the reader of what I find objectionable in Yablo's account and what I don't. The friend of disproportional mental causation and noncompetitive views can be quite sympathetic to Yablo's comparison of realization to determination.

In this closing section, I have not meant to suggest that noncompetitive accounts are completely without their own unresolved problems. Rather, what I have argued is that such accounts possess a certain virtue: they can allow for disproportional mental causation. This, I have tried to show, is a significant advantage for such accounts over those competitive accounts, like Yablo's, with which they are competing.

WORKS CITED

Bealer, George. 2007. "Mental Causation," *Philosophical Perspectives* 21: 23-54.

Bennett, Karen. 2003. "Why the Exclusion Problem Seems Intractable, and How, Just Maybe, To Tract It," *Noûs* 37: 471-497.

Davidson, Donald. 1970. "Mental Events," reprinted in *Essays on Actions and Events* (1980, Oxford University Press).

Davidson, Donald. 1993. "Thinking Causes," in *Mental Causation*, edited by John Heil and Alfred Mele (1995, Oxford University Press).

Earman, John, John Roberts, and Sheldon Smith. 2002. "Ceteris Paribus Lost," *Erkenntnis* 57: 281-301.

Fodor, Jerry. 1974. "Special Sciences," *Synthese* 28: 77-115.

Fodor, Jerry. 1989. "Making Mind Matter More," reprinted in *A Theory of Content and Other Essays* (1990, MIT Press).

Fodor, Jerry. 1991. "You Can Fool Some of the People All of the Time, Everything Else Being Equal; Hedged Laws and Psychological Explanations," *Mind* 100: 19-34.

Gillett, Carl and Bradley Rives. 2005. "The Non-Existence of Determinables: Or, a World of Absolute Determinates as Default Hypothesis," *Noûs* 39: 438-504.

Honderich, Ted. 1982. "The Argument for Anomalous Monism," *Analysis* 42: 59-64.

Kim, Jaegwon. 1984. "Epiphenomenal and Supervenient Causation," *Midwest Studies in Philosophy* 9: 257-270.

Kim, Jaegwon. 1998. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: Bradford.

LePore, Ernest and Loewer, Barry. 1987. "Mind Matters," *The Journal of Philosophy* 84: 630-642.

Leiter, Brian, and Miller, Alexander. 1994. "Mind Doesn't Matter Yet," *Australasian Journal of Philosophy* 72: 220-228.

Lewis, David. 1970. "How to Define Theoretical Terms," *The Journal of Philosophy* 67: 427-446.

Loewer, Barry. 2001. "Review of *Mind in a Physical World* by Jaegwon Kim," *The Journal of Philosophy* 98: 315-324.

Mills, Eugene. 1996. "Interactionism and Overdetermination," *American Philosophical Quarterly* 33: 105-117.

Robb, David. 1997. "The Properties of Mental Causation," *Philosophical Quarterly* 47: 178-194.

Robb, David. 2001. "Reply to Noordhof on Mental Causation," *Philosophical Quarterly* 51: 90-94.

Schiffer, Stephen. 1991. "Ceteris Paribus Laws," *Mind* 100: 1-17.

Shoemaker, Sydney. 2001. "Realization and Mental Causation," in *Physicalism and Its Discontent*, edited by Carl Gillett and Barry Loewer (2001: Cambridge University Press).

Shoemaker, Sydney. 2007. *Physical Realization*. Oxford: Oxford University Press.

Sosa, Ernest. 1984. "Mind-Body Interaction and Supervenient Causation," *Midwest Studies in Philosophy* 9: 271-281.

Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.

Williamson, Timothy. 2005. "Replies to Commentators," *Philosophy and Phenomenological Research* 70: 468-491.

Yablo, Stephen. 1992. "Mental Causation," *The Philosophical Review* 101: 245-280.

Yablo, Stephen. 1997. "Wide Causation," in James E. Tomberlin, ed. *Philosophical Perspectives* 11: 251-281

Yablo, Stephen. 2003. "Causal Relevance", *Philosophical Issues* 13: 316-328