SUPERPROPORTIONALITY AND MIND-BODY RELATIONS[1]

S. Yablo, MIT

I. Introduction

If you want to be unburdened of the problem of epiphenomenalism, then the following has worked for me and (who knows?) it may work for you as well.

First though a reminder of what epiphenomenalism is and why it is indeed a problem. A passage from Ray Bradbury's novel Dandelion Wine sets the issue up nicely. Early one morning, Bradbury writes, twelve year old Douglas Spaulding climbed

> the dark spiral stairs to his grandparents' cupola [to] perform his ritual magic...He pointed a finger...A sprinkle of windows came suddenly alight miles off in dawn country… "Grandma and Greatgrandma, fry hot cakes!" The warm scent of fried batter rose in the drafty halls..."Mom, Dad, Tom, wake up." Clock alarms tinkled faintly. The courthouse clock boomed. Birds leaped from trees like a net thrown by his hand, singing. Douglas, conducting an orchestra, pointed to the eastern sky. The sun began to rise...Yes sir, he thought, everyone jumps, everyone runs when I yell...

A lot of things take place when Douglas yells, but do they happen because he yells? This is certainly Douglas's view, and it may be true in the story as well. But suppose

we were reading Bradbury's narrative as a report on actual events. Then we would regard Douglas as seriously deluded.

Why? The clock's booming, to take that example, is already causally guaranteed, quite apart from Douglas's activities, by the cogwheel's turning, and the clapper's swinging. So what Douglas does is an irrelevant add-on making no contribution whatever to the effect.

Notice the principle we are relying on here, sometimes called the exclusion principle: if an outcome is causally guaranteed by factors distinct from X, then X is causally irrelevant to that outcome.

The question for us is: what does this principle say about outcomes that Douglas intuitively does have some control over, like his finger's turning toward the eastern sky? This outcome occurs when Douglas decides to point east, but does it occur because of his decision?

Surprisingly the answer seems to be no. Like any physical event, the motion of Douglas's finger is causally determined by its physical antecedents. By the exclusion principle, then, nothing distinct from those antecedents -- such as Douglas's decision -- can be relevant to the finger motion. Douglas's decision to move his finger plays no causal role in his finger's subsequently moving! Douglas is no less deluded when he claims credit for the motion of his finger than when he credits himself with making the sun rise.

II. BELOW

That is one threat to mental causation: I call it the threat from below (BELOW for short) because it has beliefs and desires pushed aside by neural states at a "lower level of description," and, many would say, a lower and prior level of being.

How are we supposed to defend ourselves against BELOW? I am going to suggest an answer that may seem laughably simple-minded. The answer is that this talk of neural states as occupying a "lower and prior level of being," is not to be taken literally. It's only a metaphor.

It had better be only a metaphor, because it seems to me that your typical epiphenomenalist is right: if neural states really were prior, that could not help but make the psychological states posterior and dependent. And if the psychological depends on the neural then there is only one possible conclusion -- drawn already by Thomas Huxley over a century ago. Here is what Huxley said:

> [assuming that] all states of consciousness ... are immediately caused by molecular changes of the brain-substance ... our mental conditions are simply the symbols in consciousness of the changes which take place automatically in the organism...the feeling we call volition is not the cause of a voluntary act,

> but the symbol of that state of the brain which is the immediate cause of that act.[2]

This looks bad, but it also shows the way forward. The picture of mental states as depending on their physical concomitants leaves them out of the causal loop. So, we should reject that picture. The key is not to let them depend.

But what is the alternative? Evidently we've got to let mental states stand in some other and more intimate relation to their physical so-called underpinnings.

The most intimate relation of all is of course identity. But to call mental states identical to their physical bases essentially just runs away from the traditional problem, namely, how to arrange for mental states conceived as different from physical states manage nevertheless to exercise causal influence over physical states?

Humor me for a minute while I ask: what is the next most intimate relation after identity? Last time I looked, it was the relation between a thing and its parts. But which relation do I mean, for there seem to be at least two relations (or types of relation) with some claim to be considered relations between part and whole.

On the one hand we've got extensive part/whole, the relation in which, say, the Battle of the Bulge stands to WW II, or my hand stands to my body. Very very roughly, A is an extensive part of B iff it is what you get when B is confined to just certain spatiotemporal positions.

Secondly though there's intensive part/whole. This is the relation in which Socrates' drinking the hemlock, say, stands to his guzzling the hemlock; or someone's driving home on a certain occasion stands to her speeding home. A is an intensive part of B iff B is what you get when A is confined to just certain possible worlds. (Alternative way of putting it: extensive wholes exceed their parts in size, intensive wholes exceed their parts in strength.[3])

The proposal as you have probably guessed is that if mental states are in the intensive sense parts of their physical concomitants, this very much blunts the force of the threat from below. You can see this by looking at the principal examples we have of intensive part/whole relations: the relation that individual conjuncts bear to their conjunction, and the relation that determinables bear to their determinates.

So -- suppose a pigeon is trained to peck at red, round patches. No one is going to say that since the patch's being red and round was sufficient for her pecking, its redness was irrelevant! Or, second example, let the pigeon be trained to peck at crimson patches of any shape. Assuming that the patch's crimsonness was sufficient for her pecking, does anyone seriously want to conclude from this that its redness made no causal difference?!

III. Proportionality

After all this, is anything still left of the threat from BELOW?

Maybe there is. One could argue that as between, say, my pain, and the underlying brain state, the "real" cause of my wincing is the brain state -- not because the pain is preempted by the brain state (wholes don't preempt their parts) but just because, well, we don't want to have two causes, and the pain offers no advantages to compensate us for the sheer kookiness of nominating a non-physical event as the cause of a physical one.

The answer to this is, who says the mental states offers no advantages! It's a general principle of causation that we want causes to be as far as possible proportional to their effects: that is, to be required by them -- nothing less would have done -- and enough for them -- nothing more was needed. To put it precisely, say that

> one would-be cause x screens off another y from effect e iff, had x occurred without y, e would still have occurred.

Then the proportionality principle is this:

> c causes e only if (i) c is not screened off by any of its parts, and (ii) c screens off whatever it is part of.

To the extent that the pain screens my brain state off from my wincing -- to the extent that my wincing would still have occurred had the pain been differently implemented at the neural level -- my headache is more proportional to the wincing and hence a better candidate for the role of cause.

IV. WITHIN

Now things begin to get really tricky. The proportionality principle, having just turned back the threat from BELOW, appears to propel us straight into the arms of another and equally serious threat: the threat from WITHIN.

The target this time is intentional mental states: states like belief and desire individuated in terms of truth or satisfaction conditions. Hilary Putnam taught us that truth conditions can vary between internally indiscernible agents (e.g. me and my doppelganger on Twin Earth). It follows that intentional states are extrinsic, or not wholly a matter of what goes on within the thinker's skin.[4] Add to this that it is intrinsic states that determine causal powers -- as Fred Dretske puts it,

> you can change [extrinsic states], remove them, or imagine them to be different in various respects, without ever changing the causal powers of the object or person that is in this extrinsic condition --

and you see the problem (quoting now Jaegwon Kim):

> how can extrinsic facts about A, depending as they do on factors that are spatially and temporally remote from A, help explain A's current behavior? Surely what explains, causally explains, A's raising her arm or pushing a button are intrinsic facts about A.[5]

Example: I desire water and extend my hand. But of course Twin Me, who desires not water but twater, would have done the same in my circumstances[6] -- as indeed would anyone intrinsically just like me. So intentional states, like brain states, are overloaded with unneeded detail. The only difference is that this time the unneeded detail is "without" rather than "below."

If beliefs and desires don't cause behavior, what does? Any behavior that beliefs and desires might seem to generate must really be due to some intrinsic surrogate: syntactic states, perhaps, or narrow-content quasi-beliefs, or even brain states.[7] Intentional causes are displaced by factors -- intrinsic surrogate states -- internal to the agent, which gives the threat from WITHIN its name.

V. Superproportionality

I said that WITHIN presents such a threat because it appears to use the very same proportionality principle that we relied on in our response to the threat from BELOW. You can see what I meant by that if you put the objection like this: would-be intentional causes are screened off by internal surrogate states -- I would still have reached out my hand even without desiring water so long as my narrow content state had been just the same -- and this puts intentional causes out of proportion with their seeming effects.

But now wait a minute. The proportionality condition doesn't say that causes can't be screened off at all; it says that causes

can't be screened off by proper parts of themselves. Suppose then that we distinguish two sorts of intentional state.

Thick intentional states are rich in internal detail, so much so that you can strip their extrinsic aspects away and still have enough left to constitute what we've called a surrogate state, e.g., a narrow content state. Thick intentional states include surrogate states as (intensive) parts. Thin intentional states by contrast are subjectively impoverished, so that when you strip their extrinsic aspects away there is not enough left to make up a surrogate state. Thin intentional states do not include surrogate states as parts.

Now, about "thick" beliefs and desires the objector may well be right; to the extent that they have intrinsic surrogate states as parts -- surrogate states that screen them off from the effect -- they're in violation of proportionality. But "thin" attitudes, remember, have no intrinsic parts to speak of, hence none to screen them off. The moral is that while proportionality may indeed make trouble for thick beliefs and desires, thin ones it leaves entirely untouched.

Suppose we focus our attention on "thin" beliefs and desires, that is, ones that don't include the intrinsic surrogates as parts, that is, ones that are relatively impoverished on the subjective side.

Then what you would need to make the objection from WITHIN work is not proportionality -– that doesn't apply for reasons just explained -- but the enormously stronger condition of SUPERproportionality:

> c causes e only if (i) c is not screened off by any other candidate cause, and (ii) c screens off every other candidate cause.

Bertrand Russell seems to have been in the grip of some such principle in his paper "On the Notion of Cause." Because here is what he argued, or provided the materials for arguing:

> c cannot cause a strictly later event e except via some causal intermediary d. But then c is not really enough for e, since it would not have been followed by e but for d's assistance.[8] Nor is it really required, since given d it makes no difference to e whether c occurs or not. So there can be no temporal gap between cause and effect.[9] The only true causation is simultaneous causation.

Russell intended his argument as a reductio of the whole notion of cause. But it works better as a reductio of the super-proportionality principle. The real lesson of Russell's argument is that to insist that causes screen off subsequent events, while not being screened off by them in return, imposes an absurd degree of intimacy on causal relations.

Where does this leave us? No one imagines it makes beliefs and desires epiphenomenal to be screened off by events subsequent to themselves. But many do seem to think it makes them epiphenomenal that they are screened off by intrinsic states. This is interesting because it seems to me that to

count this sort of screening off disqualifying also imposes a disastrous degree of intimacy on causal relations. The only difference is that now the intimacy is of a modal nature rather than a temporal one. Instead of being forced to exist at the same times, c and e are forced to occur at the same or similar worlds.

Here are some crude statistics to suggest what the superproportionalist is up against. Suppose $c_1,\ldots, c_n$ are coincident events each up for the role of causing e. Then c_i causes e, according to superproportionality, only if

> for all c_j, e would still have occurred had c_i occurred in c_j's absence, and for all c_j, e would not have occurred had c_j occurred in c_i's absence.

Call the scenario where none of the c_is passes this test -- where each has its candidacy destroyed by some other -- collective self-destruction. How probable is this scenario?

As a basis for calculation, let's say that between the hypothesis that e would have occurred had c_j occurred without c_i, and the hypothesis that it wouldn't have occurred, there is nothing to choose; one candidate cause is a priori as likely to screen another off as not to do so. (This is debatable but never mind; any other estimate only increases the chances of collective self-destruction.) Then the probability of c_i's escaping elimination at the hands of c_j is 1/4 -- for there is half a chance of its being screened off by c_j and half a chance

of its failing to screen c_j off. Assuming that these probabilities are relevantly independent,[10] we can reason as follows:

> the chance of c_i escaping elimination by c_j = 1/4, so the chance of c_i escaping elimination altogether = $(1/4)^{n-1}$, so the chance of c_i being eliminated = $1-(1/4)^{n-1}$, so the chance of each c_i being eliminated = $(1-(1/4)^{n-1})^n$.

This is not a negligible figure, even for small values of n. With two candidate causes, self-destruction is 56% likely; with three it is 82% likely; with four it is 94% likely; and with five it is 98% likely. With six candidate causes there is only one chance in a hundred that some c_i will stave off elimination.[11]

It is true that the "right" candidate cause could beat the odds. But think what "right" has to mean here. A c_i which occurred in the very same worlds as e would not be in any danger. But any departure from this ideal is potentially a departure from superproportionality. For e to occur without benefit of c_i in even a single world w opens c_i up to charges of not being superrequired for e. (What it would take to make the charges stick is a c_j such that w = the closest world to actuality in which c_j occurs in c_i's absence.) Likewise a single world in which c_i occurs without e opens c_i up to charges of not being superenough for e. (Here we would need a c_j such that w = the closest world in which c_i occurs in c_j's

absence.) Superproportionality comes perilously close to the demand that causes be unconditionally necessary and sufficient for their effects -- as close as the pool of candidate causes permits.[12] It thus appears that the threat from WITHIN rests on an overheated conception of proportionality.

VI. After WITHIN

The claim so far is that WITHIN does not show that extrinsically individuated mental states are out of proportion with their putative effects. Can anything be said to clarify how they might actually be proportional with them? Details will have to wait, but let me give three examples/models of how an extrinsic cause might be better proportioned to an effect than the competition.

Unity Model

An extrinsic cause might be needed because it takes something physically outside of c to "unify" the various ways in which c might have taken place. Quine gives a relevant example in "Propositional Objects." Why did the cat jump onto the roof? Presumably because it wanted to get onto the roof, an extrinsic desire if there ever was one. One could try to nominate the corresponding brain state as cause, but the effect would still have occurred even if the desire had occurred by way of a different brain state. As Quine puts it,

> the particular range of possible physiological states, each of which would count as a case of [the cat] wanting to get on that particular roof, is a gerrymandered range of states that could surely not be encapsulated in any manageable anatomical description even if we knew all about cats...Relations to states of affairs,...such as wanting and fearing, afford some very special and seemingly indispensable ways of grouping events in the natural world[13]

An example from my own experience: I get anxious whenever I believe myself to be in Paris. Why attribute my anxiety to the extrinsic belief that I'm in Paris, as opposed to an intrinsic attitude with the content that I'm in a place of wine, baguettes, unfiltered cigarettes, etc. Because, to paraphrase Quine, the particular range of narrow states, each of which would count as a case of my believing I'm in Paris, and hence of my becoming anxious, is a gerrymandered range that could not be encapsulated in any natural narrow specification even if we knew all about my psychology. The ways I might think of myself as being in Paris are just too many, and from an internal point of view too open-ended, to permit an internal rendering of the robust counterfactual relation that obtains between my Paris-beliefs and my getting anxious.

Matching Model

An extrinsic cause might be wanted because the effect depends on my internal state "matching" the environment along a certain dimension, relatively independently of the

precise values either of the matching items assume. If I win the prize on a game show, this might be because my beliefs about snack food are by and large correct. Why not say that I won because I believe that salty snacks make you thirsty, and they do make you thirsty; I believe that most people prefer regular pop to diet pop, and they do prefer regular pop to diet pop; and so on? Such an answer is out of proportion with the effect. For if they had asked different questions about snack foods, or if they had asked the same questions but the truth about salty foods had been different and my beliefs different as well, I would still have won the prize.

Tracking Model

An extrinsic cause might be wanted because the effect depends on my internal state "tracking" the environment in a certain way, so that the matching will persist even when the environment changes, and/or my evidential situation changes.

An example of Tim Williamson's[14]: I keep on digging because I know this mine contains gold. Believing, even truly believing, that it contained gold would not have been enough; such a belief might have been inferred from the misinformation that it contained gold precisely here, when in fact the gold was in a completely different place. It seems more than a coincidence that the belief's being inferred from a false lemma both prevents it from constituting knowledge, and gives it less control over my behavior than the corresponding knowledge would have. That my belief is based on a false lemma is the kind

of thing I am liable to discover, in which case I am likely to drop the belief and give up the behavior it rationalizes.

Another example, where it is not the stability of my representations that is enhanced by an extrinsic factor but the stability of their truth. Of the various causes that might be mentioned of my catching a certain ball at time T+1, one is my seeing the ball at time T. Seeing the ball seems a better candidate for the role of cause than something intrinsic, e.g., my having an intrinsic experience as though of a ball at place P, even if we add the fact that P is where the ball really was. Why? One reason has already been mentioned; that the ball was at place P as opposed to P' is less important than my having an accurate representation of its whereabouts whatever they may be. Just as important, though, to see the ball is in part to be well placed to continue to have accurate experiences of its position as it continues to move. This can hardly fail to help with the project of catching it.

Bibliography

Dretske, F. 1988. Explaining Behavior: Reasons in a World of Causes (Cambridge, MA: MIT)

Dretske, F. 1993. "The Nature of Thought," Philosophical Studies 70, 185-199

Fodor, J. 1991. "A Modal Argument for Narrow Content," Journal of Philosophy 88, 5-26

Huxley, T.H. 1911. Method and Results (New York: Appleton)

Kim, J. 1991. "Dretske on How Reasons Explain Behavior," in McLaughlin 1991; reprinted in Kim 1993

Kim, J. 1993. Supervenience and Mind (Cambridge, UK: CUP)

Loar, B. 1985. "Social Content and Psychological Content," in Grimm and Merrill, Contents of Thought (Tucson: University of Arizona Press)

Loewer, B. and G. Rey, 1991. Meaning in Mind: Fodor and his Critics (Oxford: Blackwell)

McLaughlin, B. 1991. Dretske and His Critics (Oxford: Blackwell)

Pettit, P. & McDowell, J. 1986. Subject, Thought, and Context (Oxford: OUP)

Putnam, H. 1975. "The meaning of 'meaning'," in Putnam, Mind, Language, and Reality (Cambridge: CUP)

Quine, W. 1966. "Propositional Objects," in Ontological Relativity and Other Essays (New York: Columbia University Press)

Russell, B. 1917. Mysticism and Logic (London: Allen & Unwin)

Stich, S. 1978. "Autonomous Psychology and the Belief-Desire Thesis," The Monist 61, 573-591

Stich, S. 1980. "Paying the Price for Methodological Solipsism," Behavioral and Brain Sciences 3, 97-8

Stich, S. 1983. From Folk Psychology to Cognitive Science: The Case Against Belief (Cambridge, MA: MIT Press)

Williamson, T. 1998. "The Broadness of the Mental: Some Logical Considerations," Philosophical Perspectives 12, 389-410

Woodfield, A. 1982. Thought and Object (Oxford: OUP)

Yablo, S. 1992a. "Mental Causation," Philosophical Review 101, 245-280

Yablo, S. 1992b. "Cause and Essence," Synthese 93, 403-449

Yablo, S. 1997. "Wide Causation," Philosophical Perspectives 11, 251-281

Yablo, S. 2000. "Seven Habits of Highly Effective Thinkers," in the Proceedings of the 20th World Congress of Philosophy, vol. 9

---

[1] Parts of this paper are taken from Yablo 1992a,b, Yablo 1997, and especially Yablo 2000. Thanks to Tim Williamson for his 1998 and for helpful discussion.

[2] "Animal Automatism" in Huxley 1911, 244; the essay dates from 1874.

[3] Yablo 1992b.

[4] Putnam 1975. Despite our perfect intrinsic similarity, my doppelganger on Twin Earth wants

twater, the colorless drinkable stuff in his environs, while it is water that I desire.

[5] Dretske 1993, 187, with inessential deletions.

[6] There is a considerable tradition of attempting to answer WITHIN by denying this sameness; my Twin, unlike me, would have been reaching for twater. (See the first few papers in Pettit & McDowell 1986, and for criticism Fodor 1991.) I agree that there is something my Twin does that is different from what I do, and vice versa. But I would hate to pin the case against WITHIN on this, for there is something else we do, viz. simply reaching out, that is the same in both cases. I want to argue that WITHIN is wrong even about the behaviors that my Twin and I have in common.

[7] See various papers in Woodfield 1982, especially McGinn's; Loar 1985; Stich 1978, 1980, 1983; Fodor 1991; Dretske 1988 and 1993; and various papers in McLaughlin 1991. Here is Kim's version of the argument: "semantical properties [are] relational, or extrinsic, whereas we expect causative properties involved in behavior production to be nonrelational, intrinsic properties of the organism. If inner states are implicated in behavior causation, it seems that all the causal work is done by their "syntactic" properties, leaving their semantic properties causally idle....How can extrinsic, relational properties be causally efficacious in behavior production?" (1991, 55).

[8] "[T]here must be some finite lapse of time...between cause and effect. This, however, at once raises insuperable difficulties. However short we make the interval ... something may happen during this interval which prevents the expected result. In order to be sure of the expected result, we must know that there is nothing in the environment to interfere with it. But this means that the supposed cause is not, by itself, adequate to insure the effect" (Russell 1917, 136-7).

[99] "[I]f the cause is a process involving change within itself, we shall require...causal relations between its earlier and later parts; moreover it would seem that only the later parts can be relevant to the effect...Thus we shall be led to diminish the duration of the cause without limit, and however much we may diminish it, there will

still remain an earlier part, which might be altered without altering the effect, so that the true cause...will not have been reached" (Russell 1917, 135).

[10] Assuming, that is, that (i) elimination at the hands of one candidate cause is independent of elimination at the hands of another, and that (ii) one candidate cause's being eliminated is independent of another's being eliminated. (ii) is not strictly true since the hypothesis that c_i is eliminated raises the chances that it was eliminated by c_j, which lowers the chances that c_i eliminates c_j in return. (If c_j eliminated c_i by screening it off, then c_i cannot eliminate c_j by failing to be screened off by it, and vice versa.) The formula in the text is close enough to the truth not to matter.

[11] If the power of elimination is reserved to c_i's determinates and determinables, chances of self-destruction are zero until the number of candidate causes hits four. And self-destruction will always be rare, because of the following fact. Using $<$ for the is-a-determinable-of relation, and letting a zigzag be a sequence of c_is such that $c_1 < c_2 > c_3 < c_4 > \ldots$, a set of candidate causes self-destructs only if each of its members is connectable by a zigzag to a circular zigzag of cardinality four or more.

[12] Ordinary proportionality raises similar problems (Yablo 1992b, section 11), but not on anything like the same scale. Technically this is because the chances of finding a c_j screening c_i off (a c_j that c_i fails to screen off) are greatly reduced if we require c_j to exist in all (only) the worlds that c_i exists in. Intuitively it is because a determinate of c_i that screens it off (a determinable of c_i that it fails to screen off) is prima facie a better candidate than c_i for the role of cause. Superproportionality allows c_i to be killed off by its causal inferiors; proportionality keeps c_i alive until something better comes along.

[13] Quine 1966.

[14] Williamson 1998.