Identity, Autonomy, and Eliminativism

Gregory Johnson

Mississippi State University

gregory.johnson@msstate.edu

words count (excluding notes, captions [160], and references [477]): 8,216

abstract word count: 129

Polger and Shapiro (2016) present a novel defense of the mind-brain type-identity theory. One apparent consequence of this, or any, type-identity theory is that psychological explanations are redundant and can, with no loss of explanatory power, be eliminated. Polger and Shapiro reject eliminativism and argue that, despite mind-brain identity, psychology is still autonomous. This autonomy is achieved, they claim, because psychological explanations are genuine causal explanations. I review how their definition of autonomy differs from other definitions (Antony & Levine 1997; Fodor 1974, 1997; Richardson 1979). I then argue that, given type-identity and their definition of autonomy, either psychology is not autonomous, in which case the eliminativist conclusion is warranted, or it is, but the autonomy is too weak to block the charge that mind-brain identity leads to eliminativism.

Identity, Autonomy, and Eliminativism

words count (excluding notes, captions [160], and references [477]): 8,216

abstract word count: 129

Polger and Shapiro (2016) present a novel defense of the mind-brain type-identity theory. One apparent consequence of this, or any, type-identity theory is that psychological explanations are redundant and can, with no loss of explanatory power, be eliminated. Polger and Shapiro reject eliminativism and argue that, despite mind-brain identity, psychology is still autonomous. This autonomy is achieved, they claim, because psychological explanations are genuine causal explanations. I review how their definition of autonomy differs from other definitions (Antony & Levine 1997; Fodor 1974, 1997; Richardson 1979). I then argue that, given type-identity and their definition of autonomy, either psychology is not autonomous, in which case the eliminativist conclusion is warranted, or it is, but the autonomy is too weak to block the charge that mind-brain identity leads to eliminativism.

1. Introduction

In *The Multiple Realization Book* (2016), Thomas Polger and Lawrence Shapiro argue that (a) actual cases of multiple realization are much less pervasive than has generally been believed and (b) the psychological and brain sciences rely on identities between psychological and neurobiological processes. As an example of the latter, consider memory, which, as is well known, is not a single kind but rather a collection of different capacities—episodic memory,

semantic memory, procedural memory, short-term memory, and so forth. In general terms, Polger and Shapiro describe the development of a theory of one of these types of memory this way:

> The cognitive psychologist offers a theory of memory that the neuroscientist refines with observations about dissociations. The structure of memory reveals itself through a combination of evidence involving psychological experimentation and neuroscientific manipulations. Thus, a theory of memory develops out of the close interplay of psychological and neuroscientific progress. (pp. 168 – 169)

They also approvingly quote Bechtel and McCauley who say that

> Identity claims typically play a *heuristic* role in science. Scientists adopt them as hypotheses in the course of empirical investigation to guide subsequent inquiry—rather than settling on them merely as the results of such inquiry. (p. 169)

And then, to that observation, Polger and Shapiro later add,

> We like to think that the discovery of mind-brain identities by the deployment of these heuristics has the justificatory effect of vindicating the heuristic assumptions and, as it were, discharging their provisional or heuristic status. (p. 181)

This is a compelling way to motivate their account, and we will return to it. The larger part of *The Book*, however, focuses on responding to the evidence and empirically based arguments for the multiple realization thesis—put forward first by Putnam and Fodor and later by others (e.g., Von Melchner et al. 2000; Aizawa & Gillett 2011). This part of their project will not be our focus here, although to give a more complete picture, I will note that they allow that there may be cases in which mental states and processes are multiply realized. But those would be the exceptions rather than the rule (passim [e.g., p. 34]). The

result, then, is that Polger and Shapiro provide a defense of a mind-brain type-identity theory.

My interest, however, is not with the substance of Polger and Shapiro's identity theory but with a consequent of any mind-brain type-identity theory. Namely, if mind-brain identity is true and if we have a complete neurobiological account of our cognitive capacities, then it seems that we can dispatch with psychology and just use neurobiological explanations. As Polger and Shapiro explain, given identity and a complete neurobiological account, an additional psychological explanation would appear to be either redundant or false (p. 200). Although, as we will see, this is an appearance that they reject.

The principle that "there cannot be two theories that explain any one phenomenon *x* unless one of the theories is redundant" Polger and Shapiro call *explanatory exclusion* (p. 201), and, to avoid the threat of eliminativism, they must justify the rejection of this principle. Since they equate explanatory legitimacy—i.e., the rejection of explanatory exclusion—with autonomy, their task is to demonstrate that psychology (or psychological explanations) are autonomous.[1] I am skeptical that they succeed, and I will explain my reasons for that skepticism in sections 3 and 4. My sympathies are with eliminativism and, for our purposes, I will grant Polger and Shapiro their identity theory. The stronger conclusion I will draw is that Polger and Shapiro do not have grounds for claiming that, if the identity theory is correct, psychology is autonomous. Thus, they can't reject the

[1] I'll grant that establishing autonomy is sufficient to block the explanatory exclusion principle, although some intermediate steps are needed to show that having autonomy leads to the rejection of this principle. Rejecting the explanatory exclusion principle entails 'there can be two theories that explain any one phenomenon *x* and it's not the case that one of the theories is redundant'. Thus, in this context, autonomy means that, although there are two theories explaining the one phenomenon *x*, the two explanations are not redundant.

explanatory exclusion principle, and the identity theory leads to eliminativism. Alternatively, if we allow that psychology is autonomous in some weaker sense, that weaker sense will not give us grounds for rejecting the explanatory exclusion principle and so the identity theory still leads to eliminativism.

2. Autonomy

2.1

Polger and Shapiro define *actual autonomy* as follows: "psychological explanations gain their legitimacy when they explain effects by citing their actual causes" (p. 210), and to support this definition, they invoke Woodward's interventionist account of causal explanation (2003).[2] I will discuss Woodward's theory in more detail in the next section, but, for now, Polger and Shapiro's synopsis will suffice: "Explanations cite causes, and an event C causes an event E if, roughly, changes to C make a difference to E" (p. 207). They then add, "the modal force of the counterfactual claim—viz., that the intervention on C *would* make a difference to E—is backed by actual invariances relating C and E" (p. 207). Relating this to psychology, they say, "according to us, psychological explanations gain their legitimacy from their identification of actual causal invariances" (p. 209), and "we believe that causal invariances hold between mental states or processes and physical and mental effects; and those are the basis for explanation in the mind and cognitive sciences" (p. 208). Therefore, since on their view psychological explanations

[2] It is not exactly clear why they refer to *actual causes*—by which they presumably mean, token-causes. On Woodward's account, type-causation and token-causation are related, although the analysis of each is somewhat different. Given that at least some psychological explanations are general, it seems as though Polger and Shapiro should want their definition of autonomy to be about type-causation. In any event, in sections 3 and 4, I will refer to what seems most relevant, which will hew to the interventionist account of type-causation.

are causal explanations, by their definition of actual autonomy, those explanations are autonomous (p. 210).

Notably, this commitment to autonomy comes with the rejection of distinctness. Based on the purported identity of psychological states or events and neurobiological states or events, Polger and Shapiro deny that psychological and neurobiological processes are distinct, while, of course, maintaining that psychological explanations are autonomous (pp. 197, 206). That autonomy and distinctness are often treated as similar (if not the same) seems not to be an issue—although even Shapiro elsewhere states that "most philosophers already regard autonomy as a kind of distinctness" and "I shall continue to think of autonomy and distinctness as conveying the same idea, more or less" (2017, n3). Since Polger and Shapiro do not elaborate on how we can have autonomy without distinctness, this is an issue that we will set aside, although I will note that one might think that the minimum requirement for autonomy is that there are two of whatever is the basis for the autonomy and those two are distinct in the relevant way.[3] If psychological states and brain states are not distinct in any way at all, then psychological explanations will fail to meet even this minimum requirement.

[3] Consider, for instance, *autonomy* in a different context: Thomas Hill's example of the deferential wife. Hill writes,

> She buys the clothes *he* prefers, invites the guests *he* wants to entertain, and makes love whenever *he* is in the mood. She willingly moves to a new city in order for him to have a more attractive job, counting her own friendships and geographical preferences insignificant by comparison. (1973, p. 89)

The deferential wife has preferences, but she is still not an autonomous agent. She lacks autonomy because her preferences are what they are because of her husband's preferences. Of course, a wife and husband may share many of the same preferences, and one may acquire preferences because of the other. But the deferential wife's preferences are not, even in principle, distinct from her husband's. We get autonomy when her preferences are, at least in principle if not in fact, distinct from her husband's.

In any case, theirs is a novel definition of autonomy, and it will be useful to consider some alternatives. The one that they discuss and reject is Louis Antony and Joseph Levine's *essential autonomy*:

> Psychological properties must be admitted in order to capture real and important regularities in the organization of the world. The metaphysical principle suggested here is that a property is real (or autonomous) just in case it is *essentially* invoked in the characterization of a regularity. (Antony & Levine 1997, p. 91)

In this context, "*essentially* invoked" means that the regularity (or phenomenon) cannot be fully explained without that entity or property. So, for instance, let's call language comprehension a regularity. A psychological process such as shallow parsing is real on this view insofar it must be invoked to fully explain language comprehension. Or, flipped around, if we try to give a neurobiological explanation of language comprehension, then according to Antony and Levine, even if that explanation is complete, we will be missing something about this psychological ability. As they say, "it is impossible to capture all the nomological regularities there are without invoking mental properties, and no disjunction of physical predicates can replace a mental predicate without a loss in explanatory power" (1997 p. 94; see also Polger & Shapiro, pp. 202 – 203).

In a similar vein, Jerry Fodor says that "a law or theory that figures in bona fide empirical explanations, but that is not reducible to a law or theory of physics, is ipso facto *autonomous*" (1997, p. 149). And, as we know, Fodor maintains that psychology cannot be reduced to neurobiology because of the alleged multiple realizability of mental states. He writes,

> If psychology is reducible to neurology, then for every psychological natural kind predicate there is a co-extensive neurological natural kind predicate, and the generalization which states this co-extension is a law. . . .
>
> . . . [But] what seems increasingly clear is that, even if there is such a co-extension, it cannot be lawlike. For, it seems increasingly likely that there are nomologically possible systems other than organisms (namely, automata) which satisfy natural kind predicates in psychology, and which satisfy no neurological predicates at all. (1974, pp. 104 – 105)

In short, the multiple realization of mental states, or even just their potential multiple realization, makes psychology independent from neurobiology.

Although both of these definitions of autonomy are incompatible with an identity theory, they give us a clear picture of how and why psychology would be independent. Both articulate the idea that psychological phenomena can only be fully explained in psychological terms. Trying to offer such explanations using the resources of neurobiology will, according to essential autonomy, leave something out or, according to irreducibility autonomy, simply be wrong. Of course, Polger and Shapiro do, or must, reject both definitions of autonomy. About essential autonomy, they say at one point,

> That conception of the "autonomy" of psychology is very onerous. It makes psychological explanation of psychological phenomena mandatory—there can be no neuroscientific, behaviorist, information-theoretic, or ecological explanation. It's psychology or bust. (p. 205)

The central point of essential autonomy is, for them, a seemingly obvious objection to it. But they also cannot accept essential autonomy simply because, on their view, psychological states are not essential. Most, or perhaps all, cognitive abilities can be explained with neurobiological

states and processes. Relatedly, Polger and Shapiro do believe something along the lines of "for every psychological natural kind predicate there is a co-extensive neurological natural kind predicate, and the generalization which states this co-extension is a law." Thus, on their account, psychology is reducible to neurobiology, and Fodor's definition gets no traction. Insofar as both essential autonomy and irreducibility autonomy are incompatible with identity, by either definition, identity leads to eliminativism. That is, if mental states, properties or events are type-identical to neurobiological states, properties or events, then neither of these definitions of autonomy apply and, in the absence of any other definition, by the explanatory exclusion principle, we get the elimination of psychological explanations.

A definition of autonomy that may not conflict with type-identity is what Robert Richardson calls *de facto* autonomy (1979). He writes,

> What is important here is the conviction … that psychology can give an acceptable characterization of the phenomena it studies without being dependent upon descriptive categories derived from the physical sciences—or, for that matter, any other science at all. A psychological taxonomy, in short, is to be based on psychological rather than physiological or physical principles. (1979, p. 538)

The autonomy here is methodological: psychology is autonomous insofar as psychological explanations can be constructed without any knowledge, awareness, or interest in neurobiological processes. But as a consequence, methodological independence leads to explanatory independence. Or, even if an interest is taken in some neurobiological evidence, psychology can still, "provide *in its own terms* descriptions of its own phenomena and its domain" (pp. 555, italics added).

We will examine two aspects of *de facto* autonomy. First, right after he defines it, Richardson adds,

> Whether there might be physiological equivalents for psychological categories and concepts is not here at issue, and therefore the reducibility or irreducibility of psychology is also not at issue. The only issue here is whether psychology can give a proprietary characterization of its domain. (p. 538)

But Richardson does take it that psychological states are multiply realizable. (Although contra Fodor, Richardson argues that multiple realization is not a barrier to the reducibility of psychology to neurobiology.) And the alleged multiple realizability of mental states has an important role. If psychological states and processes are multiply realized (and if there might also be multiple higher-level functions of neurobiological states), then, Richardson argues, we will find regularities that can only easily be stated and predictions that can only easily be made at the higher level. As he concludes,

> Thus, though our reductions may be transitive, increasing complexity limits their usefulness and intelligibility. Methodological independence [but not principled independence] becomes the only recourse. (p. 556)

Shortly thereafter, he adds, "the most any of this leads to is an acceptance of the indispensability of higher level theories and not their principled independence," where *principled independence* is different than the mere methodological independence that he sees (p. 556). Psychology will not, however, be indispensable, or at least not for the same reason, if multiple realizability is replaced with type-identity.

Second, and returning to the primary issue, Richardson points out that, to maintain *de facto* autonomy, the properties that are ascribed to a mental state (or used to define it) must be

purely functional and so, presumably, psychological. If they were not purely functional, then psychology would be dependent on some other science and *de facto* autonomy would be violated. As noted earlier, however, Polger and Shapiro assert that psychological explanations do depend on neurobiological evidence. Sticking with memory as our example, they say,

> Indeed, the study of memory contains many examples where neuroscientists introduce new distinctions among kinds of memory processes when they discover them to be implemented in different brain processes. . . . And the reason is plain: Neural differences in memory systems typically manifest themselves in psychological differences. (p. 100)

It could be that psychologists take these sorts of insights from neurobiological research but then offer their final explanations in psychological language only. No neurobiological terms are needed, and maybe we've maintained *de facto* autonomy. Richardson accepts this basic picture, but the example he uses is more minimal: the complexity of the neurobiological system constrains the possible descriptions of any higher-level functional system. He says,

> Since our taxonomies at the two levels are supposed to be disjoint (one structural and one functional), and thus at cross purposes, the only constraint on psychological laws is that they not demand more complexity than the organism supposedly realizing those laws manifests. (p. 557, see also p. 535)

This seems innocuous, and similarly, we might use neurobiological evidence to inform us that memory abilities that are localized in different parts of the brain are, in fact, different capacities. (Or maybe we can't do this without violating *de facto* autonomy. Richardson isn't precise about what is allowable when evidence is "gained indirectly from such contact" between the two disciplines [p. 557].) But while we might be able to give a satisfactory account of, say, declarative memory using only psychological language (even when we use a little bit of

neurobiological evidence to shape the explanation), it would be odd if we prevented ourselves from pointing out—especially if we take identity to be correct—that the encoding of declarative memories is localized in the hippocampus just so that we would not violate *de facto* autonomy.[4] (For Richardson, there is no oddness since, if psychological states or abilities are multiply realizable, the structural details won't be part of the most precise definition of the state or the ability.)

That oddness aside, maintaining *de facto* autonomy seems not to be possible with Polger and Shapiro's identity theory. When discussing pain, Polger and Shapiro say,

> All animals that experience pain do so in part because they have nociceptors. Octopuses and other mollusks have nociceptors, but nociception by itself is generally considered insufficient for experiencing pain. …
>
> … If we have confidence that octopuses experience pains this must be because they exhibit various pain-like behaviors, *and*, *e.g., because they have nervous systems that include nociceptors*. Their behavior discriminates among noxious stimuli in just the ways that we would expect given the kinds of nociceptors that they have—for instance, octopuses lack thermal nociceptors and, unsurprisingly, show no aversion to cold; but they do have mechanoreceptors and, again unsurprisingly, display an aversion to damaging mechanical stimuli. (p. 113, italics added)

Here we find that the neurobiological evidence does not just indirectly guide the psychological explanation. The presence of mechanoreceptors and absence of thermal receptors is part of the

[4] I won't explore it here, but even if there was an identity theory that did not violate *de facto* autonomy, that still might not give us a reason to reject the explanatory exclusion principle.

account of pain. Consequently, by including the presence or absence of certain kinds of neurobiological details in the description of a psychological ability we have violated *de facto* autonomy.[5]

So, *de facto* autonomy, like essential autonomy and irreducibility autonomy, is inconsistent with type-identity, or at least, inconsistent with Polger and Shapiro's type-identity theory. It just remains to be seen whether Polger and Shapiro's own definition of autonomy fares differently and is sufficient to block the explanatory exclusion principle.

2.2

As discussed at the beginning of section 2.1, Polger and Shapiro's definition of autonomy relies on psychological states participating in causal interactions, and while they say that there are invariances between mental states and other mental and physical states, they don't provide a defense of that claim. At first glance, doing so doesn't seem to be especially difficult, however. As Woodward points out,

> On an interventionist conception of cause … all that is required for changes in a mental state M_1 to cause changes in a second mental state M_2 (or in behavior B) is that it be true that under some intervention that changes M_1, M_2 (or B) will change. Common sense certainly supposes that episodes like these are very widespread. (2008, p. 231)

[5] Nociceptors and pain is a specific example showing that Polger and Shapiro take neurobiological details to be part of the description of a mental state. This more general statement also appears to indicate that they reject *de facto* autonomy:

> The best overall model of psychology and neuroscientific processes make substantial and important use of identities. In that model (or, as is likely, set of models) important psychological process kinds are identified with brain process kinds, *and those identifications are explanatory*. (p. 35, italics added)

But mental causation has been challenged—notably by Kim's causal exclusion argument (1989, 1998, 2005)—and a cursory account of mental causation does nothing to dispel that challenge.

Elsewhere, however, Shapiro provides a response to Kim's challenge, and examining that response will be useful for our analysis of his and Polger's definition of autonomy (Shapiro 2010, 2012, 2017, Sober and Shapiro 2007). He argues that, when we employ Woodward's account of causation, we find that mental states, qua their functional characterization, are causally efficacious. In summary,

> From the perspective of a manipulationist way of identifying causes (for the definitive statement, see Woodward [2003]), Shapiro and Sober (2007) and Shapiro (2010) have argued that both macro-properties and the micro-properties on which they supervene count as causes. Leaving aside the details, the thrust of the manipulationist response to causal exclusion arguments is to note that the metaphysical necessity that connects macro-properties to their supervenience bases makes impossible an intervention on the former without an intervention on the latter. (Shapiro, 2017, p. 1053)

In the next section, I will review Shapiro's argument and explain two ways in which it fails to support his and Polger's claim that, once mind-brain identity is established, psychology is still autonomous. I will follow that, in section 4, by examining a weaker sense of autonomy. It may be that this weaker sense of autonomy cannot be rejected outright, but, as we will see, it cannot block eliminativism.

3. Mental states and causation

Kim's causal exclusion argument (1998, pp. 37 – 46) begins with the assumption that mental states supervene on brain states. Next, suppose that a mental state, *M*, causes another mental

state, *M**. Since *M* is a mental state, it supervenes on a brain state, *P*, and likewise, *M** supervenes on a brain state, *P**. So, if *M* causes *M**, then it must do so in virtue of causing *P**. Consequently, it looks like *P** has two independent and sufficient causes: *M* and *P*. (See figure 1.) Kim rejects the possibility that both *M* and *P* are independent causes of *P** and suggests instead that *P* is the only cause of *P**. Mental states, as causes of other mental states and of brain states, are excluded from the story. Although Kim does not set the problem up probabilistically, Shapiro succinctly summarizes Kim's position this way: "He appears to think that if $\Pr(P^*|P \,\&\, M) = \Pr(P^*|P \,\&\, \sim M)$, it follows that *M* is not a cause of *P**" (2010, p. 599). In other words, if *P** is just as likely to occur without *M* as with it, then *M* cannot be the cause of *P**.

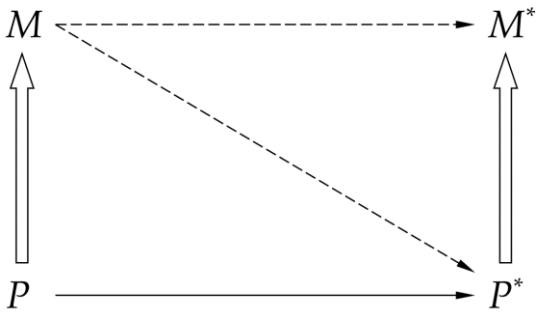

Figure 1 A brain state *P* is the cause of second brain state *P**. *M* and *M** are mental states that supervene on *P* and *P** respectively. If *M*, by being a cause of *M**, is also a cause of *P**, then *P** has two independent causes, *M* and *P*.

In response, Shapiro suggests that we use Woodward's interventionist account of causation to think through the problem in a more empirical way. According to Woodward,

> *X* causes *Y* if and only if there are background circumstances *B* such that if some (single) intervention that changes the value of *X* (and no other variable) were to occur in *B*, then *Y* would change. (Woodward 2008, p. 222; see also Woodward 2003, p. 59)

So, one variable, *X*, is a cause of another variable *Y* when an intervention on *X*—with all other variables being held constant—changes *Y*. For instance, if we want to determine whether yeast causes the transformation of a sugar such as glucose into carbon dioxide and ethanol during fermentation, then we would keep all of the other conditions, namely the presence of glucose, unchanged and intervene by introducing the yeast—that is, changing the value of that variable from *absent* to *present*. That would change the amount of carbon dioxide and ethanol that is produced (from none to some positive amount) and indicate that the presence of yeast causes the production of carbon dioxide and ethanol. (Conversely, we could withhold the yeast and observe the change to the amount of carbon dioxide and ethanol relative to when the yeast was present.)

If we apply the same procedure to an alleged case of mental causation, we would not hold *P* fixed (i.e., let it operate or keep it present) while manipulating *M* (i.e., preventing *M* from occurring). As Shapiro points out, if *M* supervenes on *P*, then having *P* happen but not *M* is impossible. Rather, to test whether *M* is a cause of *P**, the cause of *M* should be held fixed and *M* should be prevented from happening—which will also, of course, prevent *P* from happening. (See figure 2, where P_0 is the variable that must be held fixed.) Shapiro explains,

> But what happens when P_0 is held fixed and *M* is wiggled? Because changing *M* is impossible without simultaneously changing *M*'s supervenience base *P*, and because *P* is a cause of *P**, a change in *M does* result in a change in *P**. *This is evidence that M is a cause of P**. (2010, p. 601)

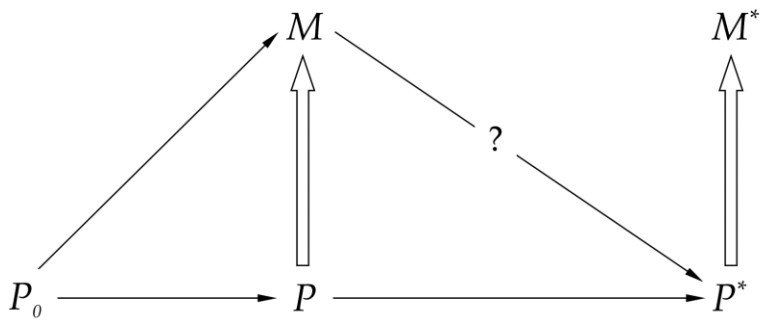

Figure 2 In the empirical test that Shapiro proposes, P_0, which is the cause of *M*, is not changed while *M* is prevented from happening. If that intervention on *M* results in a change to *P**, then, Shapiro suggests, *M* is a cause of *P**.

Even if we grant Shapiro his conclusion, the picture that we get for psychological and neurobiological states is, it seems, exactly the one that he and Polger don't want. The central premise of his argument is that "changing *M* is impossible without simultaneously changing *M*'s supervenience base *P*," which means that his contention that *M* causes *P** is based on there really only being one variable that is the cause and, consequently, only one causal process. Shapiro, in fact, agrees. As he and Sober say,

> We can grant that a functional property has no causal powers beyond those of its realizer. However, this does not mean that the causal relations into which the functional property enters are unreal. This conclusion requires an additional assumption—precisely the one that Kim makes—that functional properties have real causal powers only when they have powers *additional* to the ones possessed by their realizers. But an

> assumption like this requires that the functionalist accept some kind of magic. (2010, p. 245)

But if there is only one causal process in this picture, then it can't be a case of autonomy where that autonomy is based on causation. If, as Shapiro and Sober say, "a functional property has no causal powers beyond those of the its realizer," then the cause is either the neurobiological one or a cause that is shared by the functional and structural properties. There is no cause, however, that can be attributed only to (or independently to) the functional property. Maybe we can maintain that *M* is, in some sense, a cause, but this will not establish autonomy if *M*'s causal powers are the same as *P*'s and cannot, even in principle, be distinct from *P*'s. On the contrary, when autonomy is based on having causal powers, arguing that "a functional property has no causal powers beyond those of its realizer" seems to imply that this is a case where the functional description is not autonomous.

A second problem with Shapiro's argument is that it involves a misapplication of Woodward's interventionist account of causation. Giving a rough characterization of *intervention*, Woodward says,

> an intervention on some variable *X* with respect to some second variable *Y* is a causal process that changes the value of *X* in an appropriately exogenous way, so that if a change in the value of *Y* occurs, it occurs *only in virtue of the change in the value of X and not through some other causal route*. (Woodward 2003, p. 94, italics added)

Such a definition rules out cases, like the one diagrammed in figure 3, where a single intervention has a causal effect on two of the possible causes of *Y*. If an intervention triggers multiple possible causes, then we obviously will not be able to establish that *X* is the cause of *Y*. With respect to the question of whether *M* causes *P**, a successful intervention on *M*

requires that any change to *P** is only because of the change to *M*. It cannot be because of the change to *P*, which, as is shown in figure 1, is on separate causal route (that is, it's not between *M* and *P**). But, as already noted, Shapiro emphasizes that "changing *M* is impossible without simultaneously changing *M*'s supervenience base *P*." Thus, any intervention will be one that changes both *M* and *P*, and that means that it is impossible to intervene on *M*, at least according to Woodward's definition. Consequently, we are unable to show that *M* is a cause of *P**.[6]

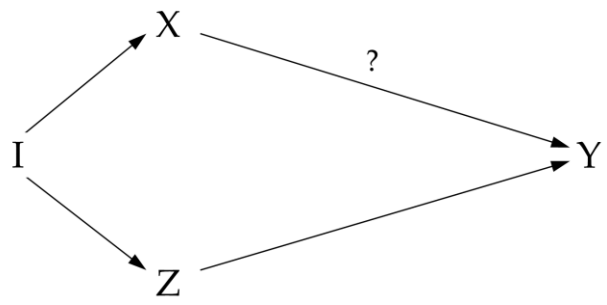

Figure 3 This scenario is ruled out by Woodward's conditions I3 and I5 for a variable to be an intervention. The intervention (*I*) on *X*—which is supposed to show that *X* is a cause of *Y*—also directly causes *Z*, and *Z* is a cause of *Y* but is not on the path from *X* to *Y*. Figure adapted from Woodward (2003, p. 101).

Making the point more precise, we are interested in the third condition in Woodward's definition of an *intervention variable* (*I*)—where the intervention is intended to determine whether *X* is a cause of *Y*:

[6] See Baumgartner (2010) for a similar argument.

> I3. Any directed path from *I* to *Y* goes through *X*. That is, *I* does not directly cause *Y* and is not a cause of any causes of *Y* that are distinct from *X* except, of course, for those causes of *Y*, if any, that are built into the *I*-*X*-*Y* connection itself; that is, except for (a) any causes of *Y* that are effects of *X* (i.e., variables that are causally between *X* and *Y*) and (b) any causes of *Y* that are between *I* and *X* and have no effect on *Y* independently of *X*. (2003, p. 98)[7]

We also have Woodward's fifth condition for an intervention variable, which covers similar ground:

> I5. *I* does not alter the relationship between *Y* and any of its causes *Z* that are not on any directed path (should such a path exist) from *X* to *Y* (2003, p. 99).[8]

I3 and I5, then, rule out the scenario shown in figure 3. In the case diagrammed in figure 1, any intervention on mental state *M* will, necessarily, also change *P*, but according to I5, the intervention on *M* cannot alter the relationship between *P** and *P* since *P* is not on the direct path from *M* to *P**. Moreover, since there is a direct path from *I* to *P** that does not go through *M* (that is, there is the path $I \rightarrow P \rightarrow P^*$), I3 also violated. (And notice that nothing changes if we replace figure 1 with figure 2. The intervention is still on *M*, and *P* is still not on the path from *M* to *P**.) Shapiro is correct that *M* cannot change without changing *P*, but that just means that

[7] Instead of I3, Baumgartner (2010) takes I4 to be the relevant condition for moving a similar argument forward. That doesn't seem to be correct, though. Condition I4 addresses a situation where the intervention variable and *P* are somehow correlated. But whatever the intervention variable might be—P_0 or M_0, or something else—it, presumably, is going to be independent of *P*.

[8] The fifth condition is an optional addition to the four conditions for being an intervention variable. For cases where I5 might be invoked, Woodward believes that I3 will suffice, but he supplies I5 as alternative.

intervening on *M* is impossible. The obvious solution is not to treat *M* and *P* as separate variables, which means giving up the claim that psychology is autonomous.

Polger and Shapiro don't have to be textualists, of course. They may want to reject this aspect of Woodward's account, and perhaps they do. About the general treatment of causal explanation that they are after, they say,

> The account that we have in mind is a difference-making theory of causal explanation. … Explanations cite causes, and an event C causes an event E if, roughly, changes to C make a difference to E. The important feature of the view on which we depend is that C's being a difference maker for E is entirely compatible with their also being another event, D, that is also a difference maker for E. (p. 207)

Of course, an event can have different causes at different times, but it seems likely that Polger and Shapiro mean that C and D are both causes of the same E—as *M* and *P* are both causes of the same *P** (or the same *M**). That's not consistent with Woodward's theory, but so be it. Still, in the general case, the reason for ruling out a scenario like the one in figure 3 is, as mentioned already, clear enough. Using the interventionist framework, we can't determine whether *X* is a cause of *Y* if there is also a causal path from *Z* to *Y* and the intervention causes both *X* and *Z*. So, relaxing I3 and I5 to achieve autonomy for psychology is a move that the eliminativist is justified in rejecting.[9]

A more promising solution is the following. The notion of a *variable set*—the set of variables that describe the system—is a central part of Woodward's account, and one way of

[9] I have been focusing on *M causes P**. At least some of the time, though, actual autonomy is about explanations with the form *M causes M**. But, since *P* is still not on the causal route from *M* to *M**, *M causes M** will encounter the same problem as *M causes P**.

making *M* in *M causes M** a genuine cause might be to exclude *P* and *P** from the analysis so that an intervention on *M* does not violate I3 and I5. Do we have to include *P* and *P** in the variable set? Woodward says, "the choice of variables in a representation [of causal relationships] reflects those possibilities that we are willing to 'take seriously'" (2003, p. 56) where those possibilities are determined by, perhaps among other concerns, the probability of an event occurring, moral requirements, expectations, custom, how controllable an event is, and an investigator's interests and purposes (2003, pp. 88 – 89).[10]

In light of those criteria, it may be fair to sometimes omit brain states when investigating whether a mental state is a cause of another mental state or a behavior. That, apparently, is the attitude that Woodward is expressing in the quote in section 2.2.[11] But there are at least two problems with this solution. The more straightforward one is that, given mind-brain identity, it's hard to see how there can be any principled reason for excluding *P* and *P** from the analysis when a serious attempt is being made to establish a causal explanation. What Polger and Shapiro need is a principle like essential autonomy or irreducibility autonomy, which would give them grounds for holding that the neurobiological states can't (or shouldn't) be included in the analysis. But, given their identity theory, that's not the sort of principle that they are in a position to provide. (To sharpen the point, essential autonomy or irreducibility autonomy would give them grounds for limiting what's included in the variable set. They, however, want to limit what

[10] For reference, Woodward's examination of "serious possibilities" occurs primarily, although not exclusively, while discussing causation by omission—e.g., a doctor fails to administer a lifesaving antibiotic to her patient versus a person who is not a doctor and doesn't know the patient but likewise fails to administer the antibiotic (2003, pp. 79 – 80, 86 – 91).

[11] Polger, Shapiro, and Stern (2018) defend a similar position, although there they do not press for an identity theory. They allow that mental states are multiply realized (p. 53) and maintain the distinctness of mental states and brain states (p. 52).

is in the variable set in order to achieve autonomy.) The second problem is that, while Woodward's definition of a cause is relative to a variable set, his definition of an intervention is not (2003, pp. 59, 98). Hence, the precise problem for Polger and Shapiro is that it is not possible to intervene on *M* when *M* and *P* are treated as separate variables, and that is not solved by circumscribing the variable set to suit their needs.

In this section, we've found that even prior to examining the details of Woodward's definition of an intervention, his account doesn't provide grounds for the claim that psychological states are independent causes. Then, when we look at the conditions that must be satisfied for an intervention, we find that an intervention that is adequate for establishing that *X* is a cause of *Y* is not possible when *M* and *P* are treated as separate variables. Furthermore, excluding brain states from the variable set provides no help. Those problems notwithstanding, there may be a weaker notion of autonomy that fits with what Polger and Shapiro have in mind, and this is a possibility that we will examine next.

4. The communicative sense of autonomy

If we are seeking a weaker notion of autonomy than Polger and Shapiro's *actual autonomy*, one option is the idea that as long as we can talk about psychological states causing other psychological states and behavior, then psychology is autonomous. That is to say, insofar as the sentence *M causes M** is meaningful, *M causes M** is autonomous with respect to neurobiology.[12] Call this the *communicative sense of autonomy*. The observation that mental

[12] "The sentence *M causes M** is meaningful" may not be very precise, but, according to Polger and Shapiro's identity theory, the truth of *M causes M** (or of *M* has been caused*) is, or may be, dependent on neurobiological facts. So, claiming that *M causes M** is true, while maintaining independence from the relevant neurobiological

causes are often used to explain mental effects and behavior appears, at some points, to ground Polger and Shapiro's commitment to autonomy, and so this definition is not out of place. For instance, they say,

> Exclusion fears arise when one assumes that the presence of one kind of causal explanation—one that cites neural causes—competes with another kind—one that cites mental causes. However, identification does not create competing explanations. The identification of mental and neural processes does not force one to choose sides. We might explain a given phenomenon by citing causes of either sort. The cognitive sciences tend to cite psychological causes, and the identity of these causes with neural processes does nothing to strip such explanations of their legitimacy. (p. 198)

And later, they add,

> Assume for the sake of argument that explanations can be distinct even if the phenomena that they explain are not—explanations are intensional or hyperintensional. In the interventionist framework there is no obstacle to having more than one causal explanation for a phenomenon. Exclusion principles are simply not part of the view. Consequently, we see no reason to view psychological explanation as in any sense less legitimate than neuroscientific explanation. (p. 209)

We can also stipulate then that the communicative sense of identity is consistent with their identity theory. So, does this give them grounds for rejecting the explanatory exclusion principle and avoiding eliminativism?

---

descriptions, runs into problems. Maintaining instead that "*M causes M*** is, on its own terms, meaningful" hopefully avoids those problems.

While being consistent with mind-brain identity is helpful, the communicative sense of autonomy is clearly a weaker notion of autonomy than essential autonomy or irreducibility autonomy. It is also weaker than Richardson's *de facto* autonomy, although perhaps less clearly so. Polger and Shapiro may want to say that psychology can "provide in its own terms descriptions of its own phenomena and domain," but as we saw in section 2, being able to do so conflicts with how they understand psychology and its relationship with neurobiology. The communicative sense of autonomy, therefore, does not require that psychology can provide complete descriptions of psychological phenomena but only that a speaker can say "*M causes M**" and be understood. Nonetheless, this may be enough to underwrite the autonomy of psychology. At the same time, as we will see in the remainder of this section, it fails to block the threat of eliminativism.

4.1

The neurobiological states *P* and *P** are typically cast as structural descriptions, but that's not exactly correct. They have functional descriptions that are given in neurobiological language. *M* and *M** are additional functional descriptions that are offered on top of, so to speak, the existing neurobiological functional descriptions. We could, then, if we wished, add more functional descriptions above the *M* and *M**; call them *H* and *H**. Let's say that, just as *M* = *P* and *M** = *P**, identity holds for *H* and *M* and for *H** and *M**—and so, also, *H* = *P* and *H** = *P**. Psychological explanations and *H*-explanations are similar insofar as both are purely functional descriptions and both stand in an identity relation with brain states and processes.

That *H* causes *H** (and *H* causes *M** and *P**) is a pointless explanation, but that doesn't matter since Polger and Shapiro reject the explanatory exclusion principle. Given the identities,

if *P* causes *P**, then, at least to those of us familiar with *H*-explanations, *H causes H** expresses a meaningful idea, and the *H*-explanation meets the criteria for being autonomous in the communicative sense. Now, stepping back for a moment, Polger and Shapiro's task is to demonstrate that psychological states or events are causes. If that can be established independently, then we have actual autonomy and we can reject the explanatory exclusion principle. This account of *H*-explanations, however, demonstrates that, when autonomy cannot be obtained by establishing that *H*-explanations cite genuine causes (and cannot be obtained any other way), the autonomy must be established by, first, rejecting the explanatory exclusion principle. In other words, this is the order of events:

(1) Reject the explanatory exclusion principle.

(2) Introduce the *H*-explanation (and stipulate the relevant identities).

(3) The *H*-explanation achieves autonomy in the communicative sense.

Instead of using autonomy to justify the rejection of the explanatory exclusion principle (that is, once we have autonomy, we get to reject the explanatory exclusion principle), the rejection of the explanatory exclusion principle is now prior to autonomy. Moreover, the rejection of the explanatory exclusion principle is not grounded in any way. Presumably, we should have a basis for rejecting the explanatory exclusion principle, and, in this case at least, we do not. Since we can't reject the explanatory exclusion principle, the elimination of *H*-explanations is justified.

Of course, the history of psychological explanations is different than the way that *H causes H** was just introduced. For psychology, as Polger and Shapiro's would have it, the order is something like this:

(4) Prior to establishing identity, assume a strong version of autonomy (e.g., Fodor's or Antony and Levine's, or even Descartes's).

(5) Establish type-identity.

(6) Given type-identity, reject the stronger versions of autonomy, but maintain autonomy in the communicative sense.

(7) Based on (6), reject the explanatory exclusion principle.

Here autonomy is prior to the rejection of the explanatory exclusion principle, but this rejection is tenuous. For one, there's no clear way to establish that we aren't rejecting the explanatory exclusion principle after establishing identity but before settling on autonomy in the communicative sense. But even if (5), (6), and (7) are in the correct order, that (6) precedes (7) just seems to be based on the lingering effects of (4). That is, *M causes M** is meaningful (and not pointless in the way that *H causes H** was) because we initially had a stronger—albeit false, according to Polger and Shapiro—version of autonomy in place.

4.2

A *hyperintensional context* is one where, as Marc Lange explains, "necessarily equivalent terms cannot always be intersubstituted *salva veritate*" (forthcoming), and the claim that 'explanations are hyperintensional' (or as Lange puts it, " '. . . explains . . .' is a hyperintensional context") is the closest that we get to a justification for the communicative sense of autonomy. Polger and Shapiro, quoting Lange, provide this example of necessarily equivalent terms that cannot be intersubstituted *salva veritate* in an explanation:

> Though "being water" and "being composed of H_2O molecules" necessarily designate the same property, they pick it out in different ways . . . Because "being water" and "being composed of H_2O molecules" pick out the same property in different ways, they differ in the information they supply and thus differ in their roles in causal explanations,

> since causal explanations work by supplying contextually relevant information about the explanandum's causal history or, more broadly, about the world's network of causal relations. (p. 214)

The general point that Lange is making is that, given a hyperintensional context, identities can be explanatory (a point to which Polger and Shapiro are also committed, although not for reasons having to do with autonomy or the threat of eliminativism). So, for instance, despite those differences between *being water* and *being composed of H_2O molecules*, according to Lange, this explains why water boils at 373K:

1. Stuff composed of H_2O molecules boils at 373K (under standard conditions).
2. Being water = being composed of H_2O molecules.
3. Therefore, water boils at 373K (under standard conditions).[13]

Hence, in this case at least, an identity, along with information about a causal process, is explanatory. But, with respect to autonomy, the net effect of making the identity explicit is to discharge any presumed independence between how *water* fits into explanations and how *composed of H_2O molecules* does. Once we are informed of the identity, *water* and *composed of H_2O molecules* are intersubstitutable. If they were not, then the explanation for why, under standard conditions, water boils at 373 K wouldn't succeed. Consequently, an explanation that

---

[13] The argument is explanatory, Lange says, because boiling is a molecular process and the first premise provides the needed information about that molecular process. Analogously, this argument explains why *M* causes *P**:

1. *P* causes *P**.
2. *M* = *P*
3. Therefore, *M* causes *P**.

This is explanatory, according to Lange's account, because it supplies, in the first premise, the appropriate causal process for explaining the presence of *P**, just as a molecular process must be invoked to explain boiling.

includes an identity compels us to give up this recourse to hyperintensionality, and hyperintensional contexts are no longer a justification for autonomy.

Still, Polger and Shapiro's point might be that if we don't invoke an identity (or, at least, if we leave it in the background), then a term such as *water* sometimes figures in explanations and *composed of H_2O molecules* cannot be substituted for *water* in those explanations. For instance, *drink enough water* is an explanation, or a partial explanation, for how to survive when lost in the woods, and it may not be possible to replace *water* with *stuff composed of H_2O molecules* in that explanation without a loss of meaning. Therefore, Polger and Shapiro would surely suggest, *water* is not eliminable.[14] That may be correct, although one reason why *water* won't be eliminated is presumably because, as it is typically used in everyday discourse, it is not indentical to *composed of H_2O molecules*. More to the point, however, even if *water* was always identical to *composed of H_2O molecules* and it still wasn't eliminated, there is no principle that says that it couldn't be.[15] After all, our lexicon can handle terms such as *carbon dioxide*, which,

[14] Whether this is different in some significant way than the case where the identity is made explicit is debatable. Here, we (with knowledge of the identity) know that this is an instance of the substitution of necessarily equivalent terms. The loss of meaning, then, would occur if we made the substitution but didn't share that information with our audience. Given the realities of life, there is the risk of such a loss of meaning sometimes, but that seems to imply only that more people should have more facts at their disposal.

[15] Remember that the issue here is not the one that, for instance, J. J. C. Smart (1959) addressed, which, in two places, he described as follows.

> Any illiterate peasant can talk perfectly well about his after-images, or how things look or feel to him, or about his aches and pains, and yet he may know nothing whatever about neurophysiology. . . Hence the things we are talking about when we describe our sensations cannot be processes in the brain. (p. 146)
>
> [T]he qualities of sensations are something over and above the qualities of brain-processes. That is, it may be possible to get out of asserting the existence of irreducibly psychic processes, but not out of asserting

incidentally, in the 17th and 18th centuries went by other names that have since been replaced. What Polger and Shapiro need is not a principle that states,

> (1) in some explanations, necessarily equivalent terms cannot be intersubstituted *salva veritate*,

but rather one that tells us that

> (2) in explanations that invoke psychological states, events, and processes, necessarily equivalent terms for those states, events, and processes can *never* be intersubstituted *salva veritate*.

But (2) is surely false, and without such a principle, hyperintensionality cannot underwrite the communicative sense of autonomy, much less block eliminativism.

Moreover, (2) illustrates that hyperintensionality is not the right resource to invoke when arguing that psychological explanations are autonomous. Backing up for a moment, *believes* is standard example of a verb that creates a hyperintensional context. Daniel Nolan explains that "'John believes that …' creates a hyperintensional position in a sentence" because, while 'John believes that Jane runs' is true, substituting Jane's online username *Jgrrl* for *Jane* produces a sentence—*John believes that Jgrrl runs*—that is false (2014, p. 151). That much is straightforward, but, of course, it doesn't mean that we can never substitute necessarily equivalent terms into sentences that begin "John believes that." When we can't, as in Nolan's

---

> the existence of irreducibly psychic *properties*. … [T]here must be some properties (for example, that of being a yellow flash) which are logically distinct from those in the physicalist story (p. 148).

These were challenges to mind-brain identity. We, however, are assuming that the type-identity relation is correct, and Polger and Shapiro are not claiming that, once we have a complete neurobiological account of our cognitive abilities, something that needs to be explained will still be left unexplained. Rather, the issue before us is whether, once the neurobiological account is complete, neurobiological language can be effectively used in all explanations.

example, that has implications for our analysis of language, but in a practical sense, the lack of intersubstitutability between *Jane* and *Jrrl* is just a consequence of an identity not being well-known, even though it could be.[16]

4.3

The final problem that I will discuss for the communicative sense of autonomy arises from the explanatory exclusion principle itself. This principle, recall, states that "there cannot be two theories that explain any one phenomenon *x* unless one of the theories is redundant." So far, I have granted that if it can be shown that psychological explanations are autonomous, then we would have sufficient grounds for rejecting the explanatory exclusion principle. But that can't be granted for the communicative sense of autonomy. Here is the situation as it stands.

(1) Given mind-brain type-identity, cognitive abilities can be completely explained with explanations of the form *P causes P**.

(2) The communicative sense of autonomy exists (and maybe only exists) in the context of mind-brain type-identity.

(3) The communicative sense of autonomy tells us that explanations such as *M causes M** are meaningful.

From (2) and (3), it follows that

[16] A cleaner example doesn't involve names that designate a living person (who may be actively trying to thwart intersubstitutability). Consider the following. In 1958, Mississippi State University's official name was changed from *Mississippi A & M* to *Mississippi State University*. Let's say that, in 1959, *John believes that Mississippi A & M is in Starkville* was true while *John believes that Mississippi State University is in Starkville* was false. But, while *Mississippi A & M* and *Mississippi State University* were necessarily equivalent terms that could not, in that context, be intersubstituted *salva veritate*, the term *Mississippi A & M* has still been eliminated.

(4a) there have to be type-identities $M = P$, $M^* = P^*$, and it has to be the case that P causes P^*.

And from (1) and (2), it follows that

(4b) there have to be type-identities $M = P$, $M^* = P^*$, and it has to be the case that, if P causes P^*, then M causes M^*.

(5) Hence, *P causes P** and *M causes M** are explaining the same phenomenon, and, given (4a), (4b), and a standard interpretation of *redundant*, one of the explanations is redundant.

The upshot is that the communicative sense of autonomy simply confirms that there are two theories that are explaining the one phenomenon *x* and one is redundant. As a result, the communicative sense of autonomy is equivalent to the explanatory exclusion principle. At best, it gives us no reason to reject the explanatory exclusion principle, and at worst, we cannot reject the explanatory exclusion principle without rejecting the communicative sense of autonomy.[17]

5. Conclusion

Polger and Shapiro have a dilemma. They have reasons for rejecting the multiple realization thesis and embracing the mind-brain type-identity theory (Polger & Shapiro 2016, chapters 2 – 8; Shapiro 2000; Polger 2004), and they also want to retain cognitive psychology. Of course, they have the option of defending their identity theory and then waiting to see what happens to psychological explanations. Those explanations won't disappear immediately or, perhaps, even

[17] In n1, I pointed out that the rejection of the explanatory exclusion principle means that 'there can be two theories that explain any one phenomenon *x*, and it is not the case that one of the theories is redundant.' Hence, if we reject the explanatory exclusion principle, a sense of autonomy that requires that the two autonomous explanations are redundant (as the communicative sense of autonomy does) will no longer apply.

soon. But, although neurobiological explanations are not yet complete, if we think that the explanatory exclusion principle has force, and they clearly do, then theirs is a dilemma without a simple, or perhaps any, solution.

In section 3, I argued that, given identity and Woodward's account of causation, psychology is not autonomous when we understand autonomy as *actual autonomy*. That conclusion should not be too surprising. Examining the other definitions of autonomy showed us that the autonomy of psychology and mind-brain type-identity conflict, and there is not an obvious way to resolve that conflict. Weakening the definition of autonomy to something like *communicative autonomy*, as I did in section 4, gives us a definition that is consistent with type-identity but which still fails to help. So, neither the specific definition of autonomy that Polger and Shapiro provide nor a weaker sense of autonomy block the threat of eliminativism.

One further point worth exploring, albeit briefly, is how widely this conclusion applies. It is common to see neurobiological evidence deployed in support of mental or psychological theories. For instance, while defending his Jamesian account of emotion, Jesse Prinz writes,

> There is also anatomical evidence that emotions can be elicited via pathways from early visual structures, such as the pulvinar and superior colliculous, to the amygdala, which instructs other structures to perturb the body (Ledoux 1996; Morris, Öhman, and Dolan 1999). These pathways trigger an emotional bodily response without the mediation of any kind of judgment. (2004, p. 46)

Likewise, Bence Nanay, while arguing that there is a connection between mental imagery and pain, explains,

> In sensory stimulation-driven [visual] perception, the light hits our retina and this triggers perceptual processing (in V1, V2, V4/V8, MT, etc.). If there is perceptual

> processing in these regions that is not triggered by retinal input, we have to talk about mental imagery. Similarly, in sensory stimulation-driven pain perception, the nociceptors are activated and this triggers pain processing in clearly delineated cortical regions (the primary and secondary somatosensory cortices [S1/S2] and the anterior cingulate cortex, among others). And if there is pain processing in these regions that is not triggered by nociceptors, we have to talk about mental imagery. (2017, p. 490)

Then, as an example, he offers the thermal grill illusion:

> This is one of the oldest perceptual illusions involving pain: if the subject touches three bars, the middle one cold and the others warm, she experiences burning pain where the middle bar touches her skin (Craig and Bushnell 1994). This is a clear example of mental imagery in my framework as the activation in S1/S2 is not triggered by nociceptors (in fact, nociceptors are not involved anywhere in the entire process, see Defrin et al. [2002], Marotta et al. [2015]). Nonetheless, the subjects feel pain. Again, mental imagery is not merely partly constitutive of pain (let alone causally involved in pain perception): it is fully constitutive of pain. (2017, p. 491)

Prinz, Nanay, and many others who draw on neurobiological evidence do not explicitly defend a mind-brain identity theory and, if pressed, may be unwilling to do so. Nonetheless, Prinz and Nanay are not referring to token brain events, and given that they do not add any qualifications to their evidence—that, for instance, these are special instances of type-identity—they have, at least tacitly, accepted Polger and Shapiro's contention that "the best overall model of the psychological and brain sciences is one that includes substantial and important explanatory identifications of psychological and cognitive process kinds with neuroscientific process kinds" (p. 35). From there, they only need to accept the related claim that "multiple realization is not as

obvious and ubiquitous as defenders of realization theories have supposed" (p. 35), and they have adopted Polger and Shapiro's version of the identity theory. That then means that they must also be willing to accept eliminativism.

References

Aizawa, K., & Gillett, C. (2011). The Autonomy of Psychology in the Age of Neuroscience. In P. M. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the Sciences* (pp. 202–223). Oxford University Press.

Antony, L. M., & Levine, J. (1997). Reduction with Autonomy. *Philosophical Perspectives,* Vol. 11, Mind, Causation, and World, 83–105.

Baumgartner, M. (2010). Interventionism and Epiphenomenalism. *Canadian Journal of Philosophy*, *40*, 359–383.

Fodor, J. (1974). Special Sciences (Or: The Disunity of Science as a Working Hypothesis). *Synthese*, *28*, 97–115.

Fodor, J. (1997). Special Sciences: Still Autonomous after All these Years. *Noûs*, *31*, 149–163.

Hill, T. E. (1973). Servility and Self-Respect. *The Monist, 57*, 87–104.

Kim, J. (1989). The Myth of Nonreductive Materialism. *Proceedings and Addresses of the American Philosophical Association, 63*, 31–47.

Kim, J. (1998). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.

Kim, J. (2005). *Physicalism, or Something Near Enough*. Princeton, NJ: Princeton University Press.

Lange, M. (forthcoming). Two Kinds of Identity Explanations.

Nanay, B. (2017). Pain and Mental Imagery. *The Monist*, *100*, 485–500.

Nolan, D. (2014). Hyperintensional Metaphysics. *Philosophical Studies*, 171, 149 – 160.

Polger, T. W. (2004). Natural Minds. Cambridge, Mass: MIT Press.

Polger, T. W., & Shapiro, L. A. (2016). *The Multiple Realization Book*. Oxford University Press.

Polger, T. W., Shapiro, L. A., & Stern, R. (2018). In Defense of Interventionist Solutions to Exclusion. *Studies in History and Philosophy of Science Part A*, *68*, 51–57.

Prinz, J. (2004). Embodied Emotions. In R. C. Solomon (Ed.), *Thinking About Feeling: Contemporary Philosophers on Emotions* (pp. 44 – 58). New York: Oxford University Press.

Richardson, R. C. (1979). Functionalism and Reductionism. *Philosophy of Science*, *46*, 533–558.

Shapiro, L. A. & Sober, E. (2007) "Epiphenomenalism: The Dos and the Don'ts", in G. Wolters and P. Machamer, eds., *Thinking About Causes: From Greek Philosophy to Modern Physics*, Pittsburgh: University of Pittsburgh Press.

Shapiro, L. A. (2000). Multiple Realizations. *The Journal of Philosophy*, *97*, 635–654.

Shapiro, L. A. (2010). Lessons from Causal Exclusion. *Philosophy and Phenomenological Research*, *81*, 594–604.

Shapiro, L. A. (2012). Mental Manipulations and the Problem of Causal Exclusion. *Australasian Journal of Philosophy*, *90*, 507–524.

Shapiro, L. A. (2017). Mechanism or Bust? Explanation in Psychology. *The British Journal for the Philosophy of Science*, *68*, 1037–1059.

Smart, J. J. C. (1959). Sensations and Brain Processes. *Philosophical Review*, *68*, 141–56.

von Melchner, L., Pallas, S. L., & Sur, M. (2000). Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature*, *404*, 871–876.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford, New York: Oxford University Press.

Woodward, J. (2008). Mental Causation and Neural Mechanisms. In J. Hohwy & J. Kallestrup (Eds.), *Being Reduced: New Essays on Reduction, Explanation, and Causation* (pp. 218–262). Oxford University Press.