THE ABSENTMINDED PROFESSOR

Suppose that absences can enter into causal relations. The view naturally falls out of counterfactual theories of causation (Lewis, 1973 & 2000): if the gardener had watered the plant, it wouldn't have died, and so the death counterfactually depends on the gardener's failure to water it, an absence. But in addition, aside from any particular causal theory, to deny absence causation is to be forced into an embarrassing revisionism about a wide range of scientific and commonsense causal claims. For instance, you are forced to deny that shooting someone in the heart can cause their death, given the role in the process played by absences, such as the absence of oxygenated blood in the brain (Schaffer, 2000 & 2004). So at least for now, let's work with absence causation as a premise.

Next, suppose that functionalism is the correct account of the mind, including consciousness (Lewis, 1972; Dennett, 2001). Perhaps pain is functionally defined as roughly the higher-order property of having some lower-order property or another whose instances are typically caused by tissue damage and cause winces and groans. In humans, pain is realized by a neural property occupying this functional role—firing C-fibers, according to philosophers' yore. But of course the great attraction of functionalism is that it allows for multiple realizability (Putnam, 1967; Fodor, 1974). Perhaps in Martians, who have an entirely different biology, pain is realized by inflating D-tubes rather than firing C-fibers (Lewis, 1980). And perhaps in artificial intelligence devices of the near future, it will be realized by something plastic or metal.

Put the two premises together and see what follows. Absences can enter into causal relations. Therefore, absence properties can occupy functional roles. Therefore, absence properties can realize mental properties like pain. Functionalism says you can make a mind out of almost *anything*—brains, silicon parts, even beer cans (Searle, 1980). I say, why not a mind made out of *nothing*? That is, a mind where all the realizers of mental properties are absence properties occupying the right functional roles.

You might envision it this way. There are familiar thought experiments in which you start with a normally functioning human brain and then gradually replace its biological parts with artificial computer parts that play the same causal role (Searle, 1992; Lycan, 1996; Chalmers, 1996). Let's follow their lead, but use absences in lieu of computer parts. Take my normally functioning brain and replace my C-fibers with a chamber that sometimes fills up with Dr. Pepper and sometimes empties out. Wire it to the rest of my brain so that the absences of Dr. Pepper are typically caused by tissue damage and cause winces and groans. Then the property of having an absence of Dr. Pepper in my head chamber is what realizes pain in me.

Next, carry out the same type of procedure for the neural realizers of all the other mental properties I instantiate, so that by the end all of my mental properties are realized by absence properties. The end product will be functionally indiscernible from the present (biologically normal) me, and so will instantiate all the same mental properties I do, but will not have a brain. To bypass questions of personal identity, and whether *I* could survive this procedure, call the end product the *Absentminded Professor*. (We assume that it will be allowed to keep my job.)

Functionalists often embrace the token identity theory even as they reject the type identity theory on the grounds of multiple realizability (Fodor, 1974). If we adopt that stance here, every particular mental event the Absentminded Professor undergoes will be token

identical with some absence or another. Outdoing Sartre (1969): when the Absentminded Professor is overwhelmed by a feeling of nausea upon staring at the gnarled root of a chestnut tree in the park, it is not just that the nausea *discloses* a kind of nothingness in the world, but rather that the nausea itself *is* a kind of nothingness—it is an absence of Coca-Cola from a certain head chamber. When you gaze into the abyss, your visual experience itself *is* an abyss (Nietzsche, 1966). That is, it is token identical with an abyss (absence) of some soft drink or another. Or at least this is true if you are the Absentminded Professor.

The scenario is strange. But is it more than strange, does it raise novel philosophical puzzles? I can see two. First, it seems to generate a new form of *explanatory gap* (Levine, 1983). The old, familiar gap concerns how *physical entities* could give rise to consciousness. The new gap we encounter here has to do with how *nothings* could give rise to consciousness. I notice the coffee mug before me on the desk. It is empty, and so contains an absence of Dr. Pepper. How could *this* be the stuff of consciousness, I wonder, staring into my empty mug. This seems even more magical than thinking that firing C-fibers or other neural properties could be. At any rate, regardless of where the most magic is located, it is not obvious in advance that a solution to the first, familiar explanator gap for physical properties will generalize and work as a solution to this new explanatory gap for absence properties. In that case, we have a new problem for functionalists to solve.

Pausing briefly here, Block (1978) famously argues that functionalism absurdly entails that the economy of Bolivia would be conscious if, hypothetically, its constituent parts were arranged so that they enter into a pattern of causal relations mirroring the pattern that our mental states enter into. We can set up an analog for absences. There is some distant region of outer space that contains no asteroids at the present time *t*. This absence counterfactually depends upon—and so, we assume, is caused by—an adjacent region of space containing no asteroids at the slightly earlier time *t*-1 (if there had been an asteroid in this second region at *t*-1, traveling in the right direction, then there would have been an asteroid in the first region at *t*).

When you think about this case, and appreciate how widely this sort of result will generalize, it seems inevitable that there must presently exist a great many intricate networks of absences entering into a wide range of causal relations with one another (cf. Menzies, 2004, on "profligate" absence causation).[1] Presumably, at least some small subset of these networks enter into causal patterns mirroring those of our mental states—just as a matter of chance, it seems inevitable that some networks of absences will happen to match the patterns of our minds. But then, forget about Block's merely *hypothetical* Bolivian economy. I suggest that there must *actually* exist regions of outer space that are presently conscious, because the absence properties instantiated in such regions realize mental properties, functionally defined. Outer space is haunted by conscious ghosts, if you will.

Conscious, disembodied space ghosts can serve as a way to make vivid the new explanatory gap I am describing. But they also can do more. Chalmers (1996: 247), who of course is no physicalist, endorses a *principle of organizational invariance*, according to which

[1] Just in case further illustration is needed: the absence of an asteroid from a region of outer space at time *t* (the one mentioned above in the text) goes on to cause the absence of a crater on a nearby planet at time *t*+1—if an asteroid had existed, it would have collided with the planet, leaving a crater. The absence of a crater on this planet at *t*+1 goes on to cause the absence of a lake in the years ahead. And on and on.

entities that have the right functional organization are guaranteed to be conscious, not as a matter of metaphysical necessity (as physicalists might hold) but as a matter of natural law. He goes on to acknowledge that a consequence of his view is that the Bolivian economy would be conscious in the right circumstances, but he bites this bullet by emphasizing how unlikely it is for the constituent parts of the economy actually to enter into the right causal patterns. It would take "the most outrageous coincidence," he says (Chalmers, 1996: 252).

However successful this may be as a response to the Bolivian economy problem, it will not work as a response to space ghosts. Again, given just how many absences there are[2] and how easily they enter into counterfactual dependence relations, it seems inevitable that the organizational invariance principle will entail that there are many *actually existing* conscious space ghosts. And so if space ghosts constitute a reductio of something, they are a reductio even of non-physicalist views like Chalmers'.

Moving on, the second philosophical puzzle that I see here arises because absences seem to be neither physical nor physically realized entities. To illustrate, consider that the actual world contains an absence of unicorns. This absence does not itself seem to be physical, since there could be an entirely nonphysical world that also has an absence of unicorns—a world with but a single bit of vibrating ectoplasm, say. Is the absence of unicorns at least physically realized then? No candidate realizer seems promising. For, it is generally thought that the instantiation of a realizer property must *entail* the instantiation of its realizee (McLaughlin, 2007; Shoemaker, 2011; Tiehen, 2014a). But, no actual physical truth entails the truth that our world contains no unicorns, as demonstrated by the fact that there is a possible world that contains a duplicate of every actual physical entity that we have here in the actual world, but then in addition contains a unicorn (and so does *not* contain an absence of unicorns). If such a world is possible, then the totality of physical truths must not entail that the actual world is unicorn-less. And this means the absence of unicorns is physically unrealized, assuming that realization requires entailment.

This sort of point is familiar from discussions of how to formulate physicalism (Chalmers, 2012; Tiehen, 2018). Physicalism must not be understood as the thesis that the totality of physical truths entail all truths, because the physical truths alone do not entail negative truths (e.g., truths about absences). So instead the standard maneuver is to formulate physicalism as the thesis that the physical truths, *taken together with* a negative "that's-all" truth, serve as the entailment base (Tiehen, 2014b). If we add to that entailment base indexical truths as well (a point we will pass over here without discussion), we can formulate physicalism as the thesis that all truths are entailed by the set of truths consisting of the physical truths, the indexical truths, and a that's-all truth (Chalmers, 2012).

Now, physicalists have largely made their peace with the point that negative truths are not generally entailed by physical truths. But what about once we start building minds out of absences, minds with mental properties that apparently are neither physical nor physically

[2] Consider a modal argument. (P1): The absence of Dr. Pepper in my coffee mug could continue to exist even if my mug were filled with Coca-Cola. (P2): The absence of Coca-Cola in my coffee mug could *not* continue to exist even if my mug were filled with Coca-Cola. (C): The absence of Dr. Pepper in my coffee mug ≠ the absence of Coca-Cola in my coffee mug. If this argument is sound, the style of reasoning is obviously going to generalize widely, and so our ontology of absences will be quite abundant.

realized? This seems like trouble; this seems like the kind of thing physicalists should want to reject. Perhaps then physicalists should be more worried about absences than they have been up to now, something that becomes fully clear only when we imagine minds made out of absences.

To start wrapping things up, what are the potential responses to these puzzles? Given that our argument has proceeded from two premises, the most obvious options are to reject one of those two. First, we could deny that absences enter into causal relations after all. There are familiar accounts of causation that do this already, like Dowe's (2000) conserved quantity theory. The gardener who fails to water the plant transmits no conserved physical quantity to it, and so does not cause the plant's death, says Dowe. If we were to accept such a theory, we could then deny that absence properties occupy functional roles, and so deny that they can realize mental properties.

There are challenges to this way out, however. One is that our brains seem to make use of absence causation. Russo (2016) argues that the physiological mechanisms by which the brain brings about voluntary behavior involves absences as casual intermediaries (e.g., when motor neurons fire, it *prevents* the shielding of actin sites by tropomyosin). Clark (2016) observes that on predictive processing accounts of perception, the absences of error signals seem to play a causal role. The upshot is that if we go this way, we might end up disqualifying neural properties from occupying functional roles either. The Absentminded Professor is not conscious—but neither am I.

In connection, it is a familiar point that you can build a machine in which the absence of voltage is used as part of the hardware, to encode a binary '0' as an input for a logic gate, say.[3] As the Artificial Intelligence age dawns, suppose we build two different robots. The first does not make use of such hardware at all, it uses only "presences." It has silicon properties that occupy the functional roles of mental properties, and so therefore is conscious. The second robot is input-output equivalent to the first but its hardware exploits voltage absences in the way just described. On the view we are presently entertaining, it would then seem to follow that the second robot will be a "zombie"—it acts just like the first, conscious robot, but it does by using absences, and so therefore is not really conscious. This seems awfully implausible. Maybe this sort of minor hardware detail can matter to consciousness, but I would not want to arrive at this conclusion merely because it allows me to dodge the puzzles raised by the present paper.

Our second premise was functionalism, and so we might reject that instead. If we take this route, we will need to reject even the weak functionalist thesis expressed by Chalmers' organizational invariance principle, lest we be stuck with space ghosts. Given my own functionalist sympathies, I would prefer to find another way. For now, I am stuck with an absence of promising ideas though.

WORKS CITED

Block, N. 1978. "Troubles with Functionalism," in *Perception and Cognition*, ed. W. Savage, pp. 9-61. University of Minnesota Press.

[3] This example was given to me by anonymous twitter user @browserdotsys.

Chalmers, D. J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

Chalmers, D. J. 2012. *Constructing the World*. Oxford: Oxford University Press.

Dennett, D. 2001. "Are We Explaining Consciousness Yet?" *Cognition*, 79.1: 221-237.

Dowe, P. 2000. *Physical Causation*. New York: Cambridge University Press.

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

Lewis, D. 1972. "Psychophysical and Theoretical Identifications." *Australasian Journal of Philosophy*, 50: 249-258.

Lewis, D. K. 1973. "Causation," *Journal of Philosophy*, 70: 556-67.

Lewis, D. K. 1980. "Mad Pain and Martian Pain," in *Readings in the Philosophy of Psychology*, ed. N. Block, pp. 216-222. Harvard University Press.

Lewis, D. K. 2000. "Causation as Influence." *Journal of Philosophy*, 97: 182-197.

Lycan, W. G. 1996. *Consciousness and Experience*. Cambridge, MA: The MIT Press.

McLaughlin, B. P. 2007. Mental causation and shoemaker-realization. *Erkenntnis, 67*, 149–172.

Menzies, P. 2004. "Difference-making in Context," in *Causation and Counterfactuals*, eds. John Collins, Ned Hall, and L. A. Paul, pp. 139-80. Cambridge, MA: MIT Press.

Nietzsche, F. 1966. (1886) *Beyond Good and Evil*. Tr. W. Kaufmann, New York: Vintage.

Orlandi, N. and Lee, G. 2019. "How Radical is Predictive Processing?" in *Andy Clark and His Critics*, eds. M. Colombo, E. Irvine, and M. Stapleton, pp.. Oxford: Oxford University Press.

Putnam, H. 1967. "Psychological Predicates," in *Art, Mind, and Religion*, eds. W. H. Capitan and D. D. Merrill, pp. 37-48. Pittsburgh: University of Pittsburgh Press.

Russo, A. 2016. "Kim's Dilemma: Why Mental Causation is Not Productive," *Synthese*, 193. 7: 2185-2203.

Sartre, J. P. 1969. (1938) *Nausea*. Tr. L. Alexander. New York: New Direction Books.

Schaffer, J. 2000. "Causation by Disconnection," *Philosophy of Science*, 67.2: 285-300.

Schaffer, J. 2004. "Causes Need Not be Physically Connected to their Effects: The Case for Negative Causation," in *Contemporary Debates in Philosophy of Science*, ed. C. R. Hitchcock, pp. 197-216.. Malden, MA: Blackwell**.**

Searle, J. R. 1980. "Minds, Brains, and Programs," *Behavioral and Brain Sciences*, 3: 417-457.

Searle, J. R. 1992. *The Rediscovery of the Mind*. Cambridge, MA: The MIT Press.

Tiehen, J. 2014a. "Subset Realization and the Problem of Property Entailment," *Erkenntnis*, 79.2: 471-480.

Tiehen, J. 2014b. "A Priori Scrutabilty and That's All," *Journal of Philosophy*, 111.12: 649-666.

Tiehen, J. 2018. "Physicalism," *Analysis*, 78.3.1: 537-551.