

“An Argument Against the Mind Being a Physical Mechanism”
by John M. DePoe, Western Michigan University

In his important paper, “The Conceivability of Mechanism,” Norman Malcolm argues against a mechanistic account of the mind.¹ Malcolm contends that a completely mechanistic account of the mind allows no place for “purposive explanations” to bring about events. Consequently, if the mind is merely a neurophysiological mechanism, no person ever performs an action for reasons the person holds. In other words, a wholly mechanistic account of the mind leaves no room for rational explanations of events. Just as it is incorrect to claim that the mechanistic actions of mousetraps, thermostats, and fax machines are rational, so it would be false to say that the mind operates rationally, if it is another physical mechanism of the same sort. Malcolm finds a mechanistic account of the mind inconceivable because it is impossible for someone to assert that the mind does not act on the basis of reasons, which is itself asserted on the grounds that one has good reasons to believe it.

If Malcolm’s argument is sound, no mechanistic account of the mind can be asserted as true. In what follows, I will develop Malcolm’s argument against mechanism and defend it against some of its most prominent objections.

An Argument Against Mechanism

Before arguing that affirming a mechanistic account of the mind is untenable, it is important to specify what is meant by a mechanistic concept of the mind. A mechanistic account of the mind is what would be the case for a strictly physical world (hereafter, just “physical

¹ Norman Malcolm, “The Conceivability of Mechanism,” *Philosophical Review* 77 (1968): 45-72.

world”). A physical world is, for my purposes, one that possesses three characteristics.² The first trait of a physical world is that it accepts the causal closure of the physical domain. As Jaegwon Kim has noted, the causal closure of the physical domain is a central tenet of physicalism. Rejecting the causal closure principle would spell disaster for physicalists, as Kim notes, because the denial of this principle would permit explanations and causes where non-physical entities bring about physical events.³

A second defining characteristic of a physical world is that all causes within it must be *physical causes* and all explanations must be *physical explanations*. This follows from the causal closure of the physical domain. If something nonphysical could cause or explain something in a world, this would violate the causal closure of the physical domain. But what should count as a physical cause or explanation? Some may be inclined to say that physical causes and explanations are precisely those that follow the laws of physics or some unified scientific theory. This approach to define physical causes, however, is not likely to offer much help due to the dilemma made famous by Carl Hempel.⁴ Hempel’s dilemma states that if physical causes are identified with a scientific theory, it must refer either to contemporary science or to some ideal completed science. The first horn of the dilemma is unacceptable because most physicalists would concede that current science is false.⁵ The second horn may provide a sufficient hope for the physicalist to “keep the faith,” but this appeal is too vague to be useful as a definition of physicalism.

² These traits come from William Hasker’s analysis of the why the physical isn’t closed. See William Hasker, *The Emergent Self* (Ithaca, NY: Cornell University Press, 1999), ch. 3.

³ Jaegwon Kim, “The Myth of Non-Reductive Materialism,” in *Supervenience and Mind* (New York: Cambridge University Press, 1993), 280.

⁴ Carl Hempel, “The Function of General Laws in History,” *Journal of Philosophy* 39, no. 2 (1942): 35-48.

⁵ Some think this is obviously true since contemporary science affirms both quantum theory and string theory, which are incompatible with each other.

Rather than try to explain physical causes in terms of science, there is another approach that attempts to distinguish the nature of physical causes. William Hasker takes this approach by providing a useful criterion of physical causes and explanations: “All physical causation and physical explanation must be mechanistic.”⁶ But what does it mean to be mechanistic? The view that the world is literally a machine that runs like Newtonian clockwork is not applicable for contemporary science (nor is it likely to fit future models of science). So, what is meant by “mechanistic”? The answer to this question is explained in my next characterization of physicalism.

The third trait of physicalism gives a suitable definition for “mechanistic.” Hasker proposes the following: “Mechanistic causation and mechanistic explanation are fundamentally nonteleological.”⁷ This definition of mechanism seems fitting because it does not commit one to any particular view of physics, and it seems flexible enough to account for any robust physicalism. The idea is that mechanism implies that the proximate causes are “blind” to the ends they bring about.⁸ Rather than acting for future purposes and goals, mechanistic causes follow non-purposive antecedent conditions that fix their respective outcomes.

Malcolm’s argument claims that in order to be rational, a person must act for “purposive reasons.” Since physicalism requires (A) the causal closure of the physical, (B) the totality of all causes and explanations must be physical, and (C) mechanistic explanations are fundamentally non-purposive—it naturally follows that no proximate action or explanation could be performed

⁶ Hasker, *Emergent Self*, 62.

⁷ *Ibid.*, 63.

⁸ Hasker is careful to say that this only commits one to say that proximate causes are non-purposive. Some mechanistically caused events can be explained teleologically. For example, the thermostat turns the furnace off and on *in order to* keep the room at a constant temperature. Yet, the thermostat’s proximate cause for turning off and on (e.g., a strip of metal, cooled by the ambient air, became bent in such a way to close the electrical circuit) is non-teleological.

by a person for purposive reasons. Furthermore, if someone came to believe that his mind operated as a physical mechanism, he would be forced to confess that he did not come to that belief from rational causes. Therefore, the doctrine that the mind is a physical mechanism entails an inconsistency when anyone asserts that it is rational to believe it is true. Note Malcolm's concern:

The absurdity lies in the human act of asserting the doctrine. The occurrence of this act of assertion is inconsistent with the content of the assertion. The mere proposition that mechanism is true is not self-contradictory. But the conjunctive proposition, "Mechanism is true and someone asserts it to be true," *is* self-contradictory. Thus anyone's assertion that mechanism is true is necessarily false. The assertion implies its own falsity by virtue of providing a counterexample to what is asserted.⁹

This argument has been extended and slightly amended by several contemporary philosophers.¹⁰ Malcolm's argument originally focused on the problem of mechanism to explain human behavior. The amended forms of Malcolm's argument, however, focus on the implications his argument has for the rational nature of the mind. Hasker explains what follows if the mind is a physical mechanism: "In a physicalistic world, principles of sound reasoning have no relevance determining what actually happens."¹¹ Consequently, "On the assumption of the causal closure of the physical, no one ever accepts a belief because it is supported by good reasons. To say that this constitutes a serious problem for physicalism seems an understatement."¹² On pain of unassertability, then, physicalists must reject one of its core

⁹ Malcolm, "Conceivability of Mechanism," 67-68.

¹⁰ For example, William Hasker, "The Transcendental Refutation of Determinism," *Southern Journal of Philosophy* 11 (1973): 175-83; Lynne Rudder Baker, *Saving Belief: A Critique of Physicalism* (Princeton, NJ: Princeton University Press, 1989), ch. 7; Victor Reppert, "Eliminative Materialism, Cognitive Suicide, and Begging the Question," *Metaphilosophy* 23, no. 4 (1992): 378-92; Hasker, *Emergent Self*, ch. 3; Victor Reppert, "Causal Closure, Mechanism and Rational Inference," *Philosophia Christi* 3 (2001): 473-84.

¹¹ Hasker, *Emergent Self*, 71.

¹² *Ibid.*, 68.

principles and allow for non-physical causes¹³ to bring about rational actions and beliefs. Since physicalism undermines the grounds for its own assertability, physicalism contains a problem of self-reference. Although physicalism could still be true, its truth comes with the steep price of undermining the rationality of believing physicalism itself.

The physicalist is presented with a dilemma. If physicalism is true, then the belief that physicalism is true is unassertable on rational grounds. On the other hand, if people hold some beliefs (including the belief that physicalism is true) on the basis of rational reasons, then physicalism is unassertable. The physicalist's options are not attractive: either accept irrationality or reject physicalism. I should note that strictly speaking this is not a proof that physicalism is false.¹⁴ However, this provides grounds for dismissing physicalism since this shows if physicalism is true, then it is irrational to believe that physicalism is true.

Objections

This application of Malcolm's argument to the mind, leads me to conclude that the mind is more than a physical mechanism. In what follows, I will respond to three objections that I foresee as the most plausible ways out of this argument for physicalists. I intend to show that these routes are "dead ends" for physicalists. Although, I suspect these points of disagreement might simply unveil further fundamental disagreements that physicalists and advocates of Malcolm's argument maintain. If the latter is so, I suspect that my project will require more arguments to establish principles that I have presumed are fundamental.

¹³ As will become evident later in this paper, the kinds of causes that I believe are needed are agent causes. Stay tuned for more on this later in the paper.

¹⁴ I would like to thank John Dilworth for pointing this out to me in an earlier draft of this paper, which attempted to argue for the more strong thesis that physicalism is false. I see that my previous conclusion cannot be maintained on the grounds I am arguing in this paper, and I am grateful that John forced me to see this.

Objection #1—Global External Justification

Malcolm's argument might be averted if one accepts a global form of external justification. After all, Malcolm's argument arises because a person is asserting something is true without the grounds for doing so. Perhaps by adopting global external justification, one could still assert something to be true without having access to the appropriate internal cognitive reasons that justify the assertion.

The problem with this objection is that it is far too implausible. For it is not sufficient merely to vindicate external justification in order to rebut Malcolm's argument. In order to achieve this end, one would have to maintain that *all knowledge* is warranted through purposeless physical mechanisms, which are not internally introspectable by anyone. Malcolm's argument is consistent with a mitigated view of external justification. It is not compatible, however, with an epistemology that allows even one instance of internal justification. In order to substantiate the radical form of externalism needed to overturn my application of Malcolm's argument, one would have to vindicate global external justification.

This attempt to overturn Malcolm's argument fails, though, because of numerous and independently conclusive obstacles—two of which I will rehearse here. First, if global externalism is true, this would go contrary to the actual experience one has in the practice of reasoning. We often believe that we draw inferences on the basis of reasons of which we are aware that constitute grounds for reaching those conclusions. Typical forms of externalism claim that this sort of internal reasoning accounts for some, but not all, of the ways we are warranted in acquiring knowledge. Global externalism, on the other hand, must accept that none of our

knowledge can be acquired from internal inferences of this sort. It is this extreme thesis that flies contrary to the actual experiences we have when drawing inferences.

More significantly, global externalism seems guilty of the charge of vicious circularity. On what basis is one supposed to evaluate whether a purposeless external mechanism is reliable? The only answer available to the global externalist is the very external process, which is in question. This circularity is problematic and obviously so. One attempt to overcome this difficulty will appeal to the theory of evolution to ground the rational inferences of a mechanical mind, which is the objection I shall address next.

Objection #2—Evolution as the Grounds for Rationality

Another way to reject Malcolm's argument is to argue that evolution has brought about rational cognitive faculties through non-purposive means. The idea is that creatures with cognitive mechanisms that did not reliably produce true beliefs would be more likely to die and not pass on their genes. Creatures with reliable cognitive mechanisms, on the other hand, would be more likely to survive and pass on their genes. Therefore, it is compatible, *contra* Malcolm, that the mind could be a physical mechanism that generates true beliefs.

But appealing to evolution to validate our cognitive faculties has been challenged by Alvin Plantinga's renowned evolutionary argument against naturalism (EAAN).¹⁵ EAAN presses evolutionary theory to answer the following question: Why should we expect purposeless evolutionary processes to provide the necessary means for giving humans reliable cognitive

¹⁵ Alvin Plantinga, *Warrant and Proper Function* (Oxford: Oxford University Press, 1993), ch. 12; idem, *Warranted Christian Belief* (Oxford: Oxford University Press, 2000), 227–40. See also James Beilby, ed., *Naturalism Defeated?: Essays on Plantinga's Evolutionary Argument Against Naturalism* (Ithaca, NY: Cornell University Press, 2003).

faculties? “If our cognitive faculties have originated as Dawkins thinks,” writes Plantinga, “then their ultimate purpose or function (if they *have* a purpose or function) will be something like *survival* (of individual, species, gene, or genotype); but then it seems initially doubtful that among their functions—ultimate, proximate, or otherwise—would be the production of true beliefs.”¹⁶ Plantinga notes Patricia Churchland’s description of evolution’s process for selection:

Boiled down to essentials, a nervous system enables the organism to succeed in the four F’s: feeding, fleeing, fighting and reproducing. The principle chore of nervous systems is to get the body parts where they should be in order that the organism may survive. . . . Improvements in sensorimotor control confer an evolutionary advantage: a fancier style of representing is advantageous *so long as it is geared to the organism’s way of life and enhances the organism’s chances of survival*. Truth, whatever that is, definitely takes the hindmost.¹⁷

Even Darwin himself worried about the epistemic ramifications of his evolutionary theory:

With me the horrid doubt always arises whether the convictions of man’s mind, which has been developed from the mind of lower animals, are of any value or at all trustworthy. Would any one trust in the convictions of a monkey’s mind, if there are any convictions in such a mind?¹⁸

Plantinga exploits “Darwin’s doubt” in EAAN by figuring that given (R) that our cognitive faculties are reliable, what is the probability of the following three propositions being true: that (N) metaphysical naturalism, (E) humans have cognitive faculties that have arisen through evolution, and (C) we have cognitive faculties of this sort that produce beliefs of this kind. Using Bayes’s Theorem, Plantinga expresses EAAN this way:¹⁹

$$P(N \& E \& C | R) = \frac{P(N \& E \& C) \times P(R | N \& E \& C)}{P(R)}$$

¹⁶ Plantinga, *Warrant and Proper Function*, 218.

¹⁷ Patricia Churchland, quoted in Plantinga, *Warrant and Proper Function*, 218.

¹⁸ Charles Darwin, Letter to William Graham, quoted in Plantinga, *Warrant and Proper Function*, 219.

¹⁹ Plantinga, *Warrant and Proper Function*, 228.

Since we believe our cognitive faculties are, in fact, reliable, $P(R)$ is very near 1. Given the reasonable doubts expressed by Darwin and Churchland, $P(R | N \& E \& C)$ is less than .5. Yet, even with a high estimation of $P(N \& C \& E)$, it follows that $P(N \& E \& C | R)$ is low (below .5). Consequently, Plantinga's EAAN demonstrates that evolution cannot provide a way to assert that the mind is a rational mechanism.²⁰

Another reason to deny that evolution provides a basis for the mind's rationality is that evolution precludes the genesis of the type of rational capacities requisite to refute Malcolm's argument.²¹ In other words, evolution may give rise to mental states, but these mental states would remain impotent to cause anything. I will quickly rehearse some arguments that have been offered by Richard Swinburne and Alvin Plantinga who have show that this epiphenomenal view of mental entities is untenable.

One of Swinburne's arguments shows that mental sensations do, in fact, cause something to occur, namely one's beliefs about those sensations.²² Mental sensations are the primary, if not the only grounds, for believing that one has mental sensations and for justifying the content of one's mental sensations.²³ Since mental sensations can cause something physical to occur (e.g., my utterances about beliefs that I have about mental sensations), epiphenomenalism is false.

²⁰ In *Naturalism Defeated?*, Plantinga's argument is challenged by eleven capable critics. In his response to their criticisms, he writes "EAAN seems to me to emerge unscathed—or if a bit scathed, then at least bloody but unbowed. Indeed, I think the comments enable a strengthened version of the argument...." (205).

²¹ For more on this see Hasker, *Emergent Self*, 77-78; William Hasker, "Theism and Evolutionary Biology," in Philip Quinn and Charles Taliaferro, eds., *A Companion to the Philosophy of Religion* (Oxford: Oxford University Press, 1997), 430-31; Angus Menzies, *Agents Under Fire: Materialism and the Rationality of Science* (New York: Rowman and Littlefield, 2004), ch. 6.

²² Richard Swinburne, *Evolution of the Soul*, 2d. ed. (Oxford: Oxford University Press, 1997), 38-41. Swinburne marshals other arguments against epiphenomenalism elsewhere in this work: 82-84, 100-2, 288-92.

²³ Swinburne uses the example of having an image of an eidetic tiger. When one counts the stripes of her image of tiger, the cause for her knowledge of the tiger is her mental sensation.

Plantinga notes that the probability that one's cognitive faculties are reliable given naturalism will be true either with or without one's mental states having causal efficacy. These options are mutually exclusively and exhaustive. In my characterization of physicalism, I have claimed that physicalism is committed to the causal closure of the physical. Since mental states are not physical, they should not be able to cause anything, so the probability that physicalism is true given that one's mental states can cause one to have beliefs is zero. This leaves the only other alternative which states that physicalism must accept that humans possess cognitive faculties that work without one's mental states having causal efficacy.²⁴ Because physicalism is committed to the causal closure of the physical, at best physicalists can say mental states are epiphenomenal. But it seems obvious that some mental states are causally efficacious, therefore, something is wrong with physicalism.²⁵

So, appealing to evolution does not seem like a viable way to save physicalism from my version of Malcolm's argument.

Objection #3—A Physicalist Account of Agent Causation

Thus far, I've shown that an application of Malcolm's argument shows that physicalism cannot account for rationality primarily because all physical causes are non-purposive. A

²⁴ Another way to state this argument is using the theorem on total probability. Since these options are mutually exclusively and exhaustive, the theorem on total probability can express the $P(R|N)$ this way: $P(R|N) = P(R|N \& C) P(C|N) + P(R|N \& \neg C) P(\neg C|N)$. This idea is that the product on the right-hand side must equal 1 since the causal closure of the physical makes the product on the left-hand side 0.

²⁵ This argument can be extended further, but space is far too limited to do so. In correspondence, Victor Reppert has suggested the following remark that points out the burden for the naturalist: "The point is that in order for the evolutionary argument to even get started you have to naturalize intentionality, you have to naturalize truth, you have to account for mental causation in virtue of content, you need for logical laws to be psychologically relevant, and then if you give the naturalist all that stuff, then you can still argue that evolution won't necessarily pick for truth. If you don't have all the other stuff first, then the evolutionary response 'flunks out completely' (Hasker) because something can't be open to the pressures of natural selection unless it is part of the causal stream of events."

physicalist could object to the requirement that her view entails that all physical causes are non-purposive. She might argue for a physicalist account of agent causation, which would allow for rational causes to bring about events. If agent causation can plausibly be melded with physicalism, it would seem that physicalism does not fall prey to the objections raised in my argument.

Jan Cover and John O’Leary-Hawthorne have presented an argument that highlights the difficulties for physicalism to provide an account of agent causation.²⁶ For agent causes are taken to be enduring substances that persist and are wholly present at a plurality of times. On a physicalist account, the most plausible account of an enduring substance that has agent causal properties (for humans) is to say that human agents are human animals or biological human organisms. Moreover, agent causation holds to a principle that states that the fundamental causal powers of the agent aren’t determined wholly or partly by properties that are extrinsic to the agent.

From the above commitments of physicalism and agent causation, Cover and O’Leary-Hawthorne find the two views an unlikely pair:

In the actual world, Jan agent-causes his decision to stop daydreaming about climbing big mountains and to work on philosophy instead. Suppose the agency theorist identifies Jan with the salient physical organism. Consider now world W where the decision is made by a brain in a vat intrinsically identical with the one Jan enjoys in this world, surrounded by 145 pounds of organism-stuff. In the actual world, the brain does not agent-cause the decision, but in the brain-in-vat world it does. . . . Thus the materialist agency theorist will have to either give up the principle or claim that even in the actual world, agents are brains and not physical organisms.²⁷

²⁶ Jan Cover and John O’Leary-Hawthorne, “Free Agency and Materialism,” in *Faith, Freedom and Rationality*, ed. Jeff Jordan and Daniel Howard-Snyder (Lanham, MD: Rowman & Littlefield, 1996), 47-71.

²⁷ *Ibid.*, 69.

The dilemma Cover and O’Leary-Hawthorne raise shows the improbable compatibility of physicalism and agent causation. Since to deny the principle regarding the causal powers of an agent are wholly intrinsic is to deny agent causation, the physicalist would have to try to make the claim that human agents are brains rather than human organisms. This second option may seem plausible at first blush, but a little reflection demonstrates the philosophical quagmire to which it leads. Should we apply moral predicates of praise and blame to brains, rather than persons? Furthermore, we could continue to divide the parts of the brain up to try to find the “agent” in the brain. Is it the right-hemisphere of the brain that is the agent? For these reasons, it seems awkward, if not altogether irreconcilable, to maintain that physicalism and agent causation are compatible.

Conclusion

My argument has been that accepting (A) the causal closure of the physical, (B) that all causes are physical, and (C) all mechanistic causes are non-purposive, results in a view that undermines the rationality of asserting its own truth. Consequently, one of these principles must be wrong. But since physicalism seems committed to all of these principles, it seems that physicalism must be cast aside. At minimum, I believe this means accepting some kind of theory of agent causation.²⁸ Indeed, Malcolm’s argument seems to show that rationality requires a denial of one, if not all, of the commitments of physicalism.²⁹

²⁸ A recent explication of agent causation is spelled out in Timothy O’Connor, *Persons and Causes: The Metaphysics of Free Will* (New York: Oxford University Press, 2000).

²⁹ I would like to thank John Churchill, John Dilworth, and Victor Reppert for helpful feedback on earlier drafts of this paper, and who bear no responsibility for any errors in reasoning (or otherwise) I may have made.