

THE NEW NEW MYSTERIANISM

DEREK BALL

A familiar style of argument against physicalism proceeds from the claim that consciousness cannot be explained in microphysical terms to the ontological conclusion that consciousness is a non-physical phenomenon. This argument has reached its fullest flower in the work of David Chalmers (1996; 1999; 2003; 2007; see also Jackson (2004); Chalmers and Jackson (2001)). Chalmers holds that reductive explanation of the sort the physicalist requires depends on *armchair deducibility*: if physicalism is true, then it is in principle possible to infer the phenomenal truths from the microphysical truths on the basis of conceptual competence and armchair reasoning.¹ But he maintains that the existence of the explanatory gap shows that the phenomenal truths are not so inferable, and thus that physicalism is false.

Unlike most physicalists (e.g., Tye (1995); Loar (1997); Block and Stalnaker (1999); Byrne (1999)), I am prepared to grant that physicalism is committed to in principle armchair deducibility of the type that Chalmers describes. Unlike others (e.g. McGinn (1991); Stoljar (2006)), I do not maintain that the source of the explanatory gap is truths that we do not know (although the view I will suggest is compatible with this claim.) Instead, I will argue that our *reasoning abilities* may be limited in such a way as to make us (in a strong sense that I make precise below) unable to infer the phenomenal truths from the microphysical truths. This possibility shows that Chalmers's style of argument is invalid: an armchair physicalist can maintain that the some set of reasoning abilities would suffice to infer the phenomenal truths from microphysics, while admitting that there may be an unbridgeable explanatory gap for beings with our abilities.

More dogmatically: the explanatory gap cannot tell against physicalism. The phenomenal truths may be deducible *in principle* though *we* cannot deduce them; and our inability may be an essential consequence of our current intellectual nature. Perhaps the physicalist must hold that god could bridge the gap, but she need not grant that we

Date: 1/12/08.

¹Chalmers claims that such an inference may require among its premises the indexical truths and a 'that's all' truth (Chalmers and Jackson, 2001). For the purposes of this paper, I will set these aside.

could bridge it; it is bridgeable in principle, but not bridgeable with the tools we have, not even in principle.

I begin by showing that the undecidability results proved by Church, Turing, Gödel, and others pose a *prima facie* problem for the claim that physicalism is committed to armchair deducibility. I then argue that this problem can be resisted only by admitting that some truths cannot be known by armchair reasoning by thinkers with limited intellectual resources, but that this admission undermines Chalmers's anti-physicalist argument.

1. NEW EXPLANATORY GAPS

A *Turing machine* is an idealized computing device.² It consists of an infinite one-dimensional *tape* and a *read-write head*. The tape is divided into *cells*, each of which contains either a '0' or a '1'. The head is controlled by a set of instructions called a *program*. At any time, the machine is in a particular state (i.e., a particular step of the program). In a state, it scans a single cell; on the basis of the content of that cell, it can move one cell to the left or right, or write a symbol on the current cell.

Turing machines were designed to model a certain sort of human reasoning. A problem is effectively computable if and only if could be solved in finite time by a human clerk working with pencil and paper according to a set of logical rules that can be applied mechanically. *Turing's thesis* is that a problem is effectively computable if and only if it can be solved by a Turing machine (Turing, 1936). It is widely accepted on the grounds that (i) no counterexamples have been found, and (ii) a large number of other attempts to formalize the idea of effective computability are provably equivalent to Turing machines.

When a Turing machine goes into a state for which it has no explicit instructions, it halts. But not all machines halt. The Turing machines can be put into a 1-1 correspondence with the natural numbers. Given the number of an arbitrary Turing machine, is it always possible to determine whether that machine halts? It can be proved that no Turing machine can determine this. Given the Turing-Church thesis, this means that the halting problem is not effectively computable: it could not be solved by a thinker whose intellectual abilities were limited to the sort of reasoning envisioned by Turing.

Turing machines are a particularly apt example for the physicalist because they are in principle physically realizable. Perhaps there can

²Turing machines were first devised by Turing (1936). I will abstract away from the formal details. For an introduction, see Boolos et al. (2002).

be no Turing machine in the actual world, but there could be in some recognizably physical world: for example, in a Newtonian world. In such a world, that machine T fails to halt would be an ordinary macro-physical truth. But it could not be inferred from the microphysical truths by a thinker whose intellectual abilities are limited to the effectively computable. Given Chalmers's view that reductive explanation requires armchair deducibility, such a thinker would face an explanatory gap between the microphysical truths and the halting truths.³

It is simply a matter of logical fact that there are some truths that some thinkers could not infer from the microphysical truths on the basis of armchair reasoning. Since these truths pose no problem for physicalism, no physicalist is committed to the claim that every truth is so inferable, at least not by every thinker. The strongest claim that can be pinned on the armchair physicalist is that *some* set of reasoning abilities (possibly much more powerful than our own) would permit the inference.

2. INACCESSIBILITY

One moral of this story is that what one can infer from a given set of premises depends on the resources available to one's reasoning. A Turing machine cannot solve the halting problem. Moreover, this limitation is difficult to overcome: no finite increase in the amount of time, memory, or vocabulary available to the machine makes a difference (Boolos et al. (2002), esp. p. 94). No quantitative improvement in a Turing machine's resources can close the gap. The only machines that can solve the halting problem are *qualitatively* different from Turing machines in some respect: for example, machines that can complete infinitely many computations, and machines (like Turing's O-machines) that have access to additional primitive computational operations. (See

³In conversation, Michael Tye objected that this new explanatory gap depends on ignoring some physical truths: although the halting truth could not be deduced by a thinker who knows all the microphysical truths *at a particular time*, they could easily be known by a thinker who knew all the microphysical truths about the *future*. But there are physicalistically acceptable properties to which this style of objection is irrelevant. For example, imagine a world in which there are infinitely many space-time points. These points could be put in a 1-1 correspondence with the Turing machines. Each point would have either the property of corresponding to a Turing machine that halts, or the property of corresponding to a Turing machine that does not halt, but the truth as to which of these properties a given point has could not be deduced from the microphysical truths given the resources of a Turing machine.

Copeland and Sylvan (1999) and Copeland (2002) for discussion of such ‘hypercomputers’.)

Say that a proposition P is *inaccessible* to a thinker T if P is inferable from T ’s evidence given some set of reasoning abilities, but T cannot infer P without undergoing a qualitative change in her reasoning abilities. For example, the proposition that a particular Turing machine halts might be inaccessible to thinkers whose reasoning abilities are limited to the Turing machine computable.

Perhaps our reasoning abilities exceed the Turing limit; still, it is very plausible that some propositions are inaccessible to us. Wherever there are inaccessible propositions, there will be new explanatory gaps. Could the explanatory gap about consciousness be one of these?

Surely it is possible that there be a creature for whom the truths about consciousness would be inaccessible. For example, imagine a being whose mind is modular in the sense of Fodor (1983). Such a mind would consist of a central system responsible for cognition and reasoning, plus computationally isolated modules that provide inputs to the central system. It is possible that the modules would exceed the central system in computational complexity. If this were the case, then some truths about the modules would be inaccessible for such a being: she could come to know the microphysical truths about the modules, but she could not infer the higher-level truths because of the computational limitation on her central system. If consciousness were an aspect of the relevant modules, then the truths about consciousness could be inaccessible.

It is an empirical question whether we are creatures like this. If we are, then there would be an explanatory gap, despite the truth of physicalism. There is no way to determine from the armchair that we have the ability to understand the relevant aspects of the brain. Antiphysicalist arguments like Chalmers’s are thus invalid in the absence of empirical information. To my knowledge, such empirical information is not at this point available.⁴

⁴It might be suggested that this proposal leaves us with no reason to believe physicalism. How could we know that physicalism is true, if we cannot show how the phenomenal truths can be inferred from the microphysical truths? (This is one of the major concerns of Jackson (1998).) I cannot treat this issue in detail here, but there seem to a number of potential arguments in favor of physicalism that do not turn on the sort of inference at issue: for example, arguments based on the causal closure of the physical.

3. CONSCIOUSNESS

It might be thought that the explanatory gap about consciousness is more serious than the new explanatory gaps I have described. It is not merely that we do not see how the gap can be closed; rather, we can see that the gap cannot be closed. But the arguments for the explanatory gap do not justify this claim. Two sorts of argument are typically used to motivate the gap. The first appeals to the claim that simply adding more physical truths of the sort with which we are familiar would not explain consciousness (e.g. Jackson (1997); Nagel (1997)). The position that I have sketched can simply admit this in the finite case: coming to know finitely many more physical truths, or performing finitely many inferences of the type with which we are familiar on the basis of these truths, will never close the gap. And we are familiar from the case of Turing machines that admitting infinitely many truths or qualitatively different sorts of reasoning can change the conclusions that are apriori accessible from a given set of premises. We can imagine a being whose apriori reasoning abilities are limited to the Turing machine computable, unable to solve the halting problem, finding it puzzling and mysterious how ‘more of the same’ could lead to a solution. We know that such a being ought to draw no metaphysical conclusions from such puzzlement. Why should our puzzlement in the face of the mind-body problem be any different?

The second sort of defense of the explanatory gap purports to show that some feature of the nature of physical truths precludes them from explaining phenomenal truths. Thus Chalmers (2003, §3.1) holds that physical accounts explain at most structure and function, but that structure and function cannot explain consciousness. But in the current context, this argument is question-begging. It is true that we do not know how a physical account could explain consciousness. But this does not entail that such explanation is impossible. A dualist whose powers of reasoning were limited to the Turing machine computable might advance the following argument with equal force:

- (1) The physical truths explain at most whether a Turing machine halts after finitely many computations.
- (2) Whether a Turing machine halts after finitely many computations does not explain whether that Turing machine halts (*tout court*).
- (3) Therefore, the physical truths cannot explain the Turing machine halting truths.

4. CONCLUSION

David Chalmers has proposed a three-fold division of physicalist views. Roughly, type-A physicalists deny that there is an explanatory gap, and thus admit that the microphysical truths are armchair deducible; type-B physicalists admit that there is an explanatory gap, and thus deny armchair deducibility, but hold that there is no ontological gap between the mental and the physical; and type-C physicalists hold that there is a deep epistemic gap between the mental and physical, but that the gap can be closed in principle (Chalmers, 2003). Colin McGinn's (1991) suggestion that we are "cognitively closed" to the solution to the mind-body problem is perhaps the most famous type-C view. On McGinn's proposal, our cognitive constitution permanently precludes us from forming the concepts necessary to understand the psychophysical nexus.

McGinn's view has two striking features. First, it claims that there are truths that we cannot know. Second, it claims that this inability is permanent. A proponent of the view suggested in this paper need accept neither claim. Perhaps we already possess all of the relevant concepts, and nothing other than limitations of time and memory prevents us from knowing all of the microphysical and phenomenal truths. Nonetheless, it could be that our reasoning abilities are not sufficient to bridge the explanatory gap. This deficit could be permanent. But perhaps our inability to understand consciousness can be overcome. Perhaps we can develop the relevant concepts and modes of reasoning; perhaps we can qualitatively increase our own reasoning power (cf. the discussion of Turing's learning machines in Copeland (2002, p. 475)). I have claimed that a gap for us need not be a gap for others. It is equally true that a gap today need not be a gap tomorrow.

Chalmers's main objection to type-C views is that they are "inherently unstable" and prone to collapse into type-A or type-B views (Chalmers, 2003, §7). Perhaps my armchair physicalist is a type-A in sheep's clothing. But I do not see this as a vice: there is no shame in a type-A physicalism with a substantive and principled explanatory gap.⁵

REFERENCES

Block, Ned, Owen Flanagan, and Güven Güzeldere, editors. *Consciousness: Philosophical Debates*. Cambridge, MA: The MIT Press, 1997.

⁵Thanks to David Chalmers, Josh Dever, Janice Dowell, Rob Koons, Cory Juhl, Bryan Pickel, Mark Sainsbury, David Sosa, and Michael Tye for helpful comments.

- Block, Ned, and Robert Stalnaker. "Conceptual Analysis, Dualism, and the Explanatory Gap." *Philosophical Review* 108: (1999) 1–46.
- Boolos, George, John P. Burgess, and Richard C. Jeffrey. *Computability and Logic, Fourth Edition*. New York: Cambridge University Press, 2002.
- Byrne, Alex. "Cosmic Hermeneutics." *Philosophical Perspectives* 13: (1999) 347–383.
- Chalmers, David. *The Conscious Mind*. New York: Oxford University Press, 1996.
- . "Materialism and the Metaphysics of Modality." *Philosophy and Phenomenological Research* 59: (1999) 473–493.
- . "Consciousness and its Place in Nature." In *Blackwell Guide to Philosophy of Mind*, edited by Stephen Stich, and Ted Warfield, Oxford: Blackwell, 2003.
- . "Phenomenal Concepts and the Explanatory Gap." In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, edited by Torin Alter, and Sven Walter, New York: Oxford University Press, 2007.
- Chalmers, David, and Frank Jackson. "Conceptual Analysis and Reductive Explanation." *Philosophical Review* 110: (2001) 315–361.
- Copeland, B. Jack. "Hypercomputation." *Minds and Machines* 12: (2002) 461–502.
- Copeland, B. Jack, and Richard Sylvan. "Beyond the Universal Turing Machine." *Australasian Journal of Philosophy* 77: (1999) 46–67.
- Fodor, Jerry A. *The Modularity of Mind*. Cambridge, MA: The MIT Press, 1983.
- Jackson, Frank. "What Mary Didn't Know." In Block et al. (1997), 567–571.
- . *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. New York: Oxford University Press, 1998.
- . "Postscript." In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, edited by Peter Ludlow, Yugin Nagasawa, and Daniel Stoljar, Cambridge, MA: The MIT Press, 2004, 409–416.
- Loar, Brian. "Phenomenal States." In Block et al. (1997), 597–616.
- McGinn, Colin. "Can We Solve the Mind-Body Problem?" In *The Problem of Consciousness*, Cambridge, MA: Blackwell, 1991, 1–22.
- Nagel, Thomas. "What Is It Like to be a Bat?" In Block et al. (1997), 519–529.
- Stoljar, Daniel. *Ignorance and Imagination*. New York: Oxford University Press, 2006.

Turing, Alan. “On Computable Numbers, with an application to the Entscheidungsproblem.” *Proceedings of the London Mathematical Society* 42: (1936) 230–265.

Tye, Michael. *Ten Problems of Consciousness*. Cambridge, MA: The MIT Press, 1995.