

Faith and Philosophy: Journal of the Society of Christian Philosophers

Volume 26 | Issue 5

Article 4

12-1-2009

"The Hard Problem of Consciousness" and Two Arguments for Interactionism

Vadim V. Vasilyev

Follow this and additional works at: <https://place.asburyseminary.edu/faithandphilosophy>

Recommended Citation

Vasilyev, Vadim V. (2009) ""The Hard Problem of Consciousness" and Two Arguments for Interactionism," *Faith and Philosophy: Journal of the Society of Christian Philosophers*: Vol. 26 : Iss. 5 , Article 4.
Available at: <https://place.asburyseminary.edu/faithandphilosophy/vol26/iss5/4>

This Article is brought to you for free and open access by the Journals at ePLACE: preserving, learning, and creative exchange. It has been accepted for inclusion in Faith and Philosophy: Journal of the Society of Christian Philosophers by an authorized editor of ePLACE: preserving, learning, and creative exchange.

“THE HARD PROBLEM OF CONSCIOUSNESS” AND TWO ARGUMENTS FOR INTERACTIONISM

Vadim V. Vasilyev

The paper begins with a restatement of Chalmers’s “hard problem of consciousness.” It is suggested that an interactionist approach is one of the possible solutions of this problem. Some fresh arguments against the identity theory and epiphenomenalism as main rivals of interactionism are developed. One of these arguments has among its corollaries a denial of local supervenience, although not of the causal closure principle. As a result of these considerations a version of “local interactionism” (compatible with causal closure) is proposed. It is argued that local interactionism may offer a fruitful path for resolving the “hard problem.”

In 1994 the more or less routine discussions on the mind-body problem were thrown into an uproar by David Chalmers. In his talk at the University of Arizona conference “Toward a Science of Consciousness,” he made a distinction between “easy problems of consciousness,” dealing scientifically with psychological mechanisms, and “the hard problem of consciousness.”¹ The hard problem of consciousness is the question: Why and how does brain activity give rise to consciousness?² One can bracket the “how” part of the question and give its deepest “why” part a more neutral expression: Why is brain activity accompanied by subjective experience?³ There have been many responses to this question, but now, after fifteen years of discussion, there has emerged a kind of consensus: while the hard problem looks like a philosophical question, in fact it is hardly possible to solve this problem by philosophical conceptual means, and if it can be solved at all, the solution will come from the side of neuroscience, the progress of which strikes our imagination.⁴

¹David told me in a personal communication that, as far as he remembered, he began using this distinction in his Seminar on Consciousness at Washington University “in late 1993.”

²Cf. D. J. Chalmers, “Facing up to the Problem of Consciousness,” *Journal of Consciousness Studies* 2 (1995), pp. 200–219.

³See D. J. Chalmers, *Consciousness and Cognition* (unpublished, 1990), available at <http://consc.net/papers/c-and-c.html>.

⁴Not everyone would agree that we have such a consensus. Chalmers, for example, believes it would be better to avoid such sociological remarks. Maybe we should ask experimental philosophers indeed. Still, I believe, most analytic philosophers these days would give the kind of answer I have sketched (as is certainly suggested by their preoccupation with experimental data in talking about consciousness). Pessimists, like C. McGinn in *The*



I think, however, that this is a rather strange consensus, because it is possible to give quite a promising philosophical explanation of the fact that brain activity is accompanied by subjective experience with qualia. According to this explanation, the brain simply could not work, as it usually works, without subjective experience—it could not produce behavior of the sort that normal human beings demonstrate in their ordinary lives.

Now, we should keep in mind that there are three ways in which one might hold that the brain could not work, as it usually works, without subjective experience. The first way would occur if subjective experiences were a condition of the very existence or efficacy of the physical events making up brain activity. This is conceivable, for example, if qualia are the substantial basis of physical properties. The attempt to solve the hard problem in such a way was made by Chalmers in his 1996 book *The Conscious Mind*, but his recent papers seem to abandon this approach. And it is understandable why: it is a highly speculative road, based on a bizarre ontological model, lacking arguments in its favor. Among other things, it is not clear why physical properties need any basis. Chalmers referred to Bertrand Russell's idea that our knowledge about matter is a knowledge of relations only, and relations presuppose the things that relate to each other. But no one has yet proven that physical reality as we know it is just a complex of relations. It is no surprise that almost nobody among major philosophers likes this solution.⁵

The second way in which subjective experience might be necessary for brain activity derives from the identity theory. According to this well-known theory, originating in works of U. Place, H. Feigl, and J. J. C. Smart, so-called subjective experience is identical with physical processes in the brain, and thereby is necessary for its normal activity. Since its presentation fifty years ago this theory has been sharply criticized from many angles. I am quite convinced by that criticism, especially by objections to this theory proposed by S. Kripke, D. Chalmers, and M. McKinsey.⁶ To their semantic arguments I would add another one, of a different kind: the basic proposition of the identity theory, if we understand this proposition as a statement about the ontological identity of qualia and physical processes in the brain, does not allow verification, in contrast, say, to the thesis of the identity of the Morning Star and the Evening Star. It might be possible to verify a thesis of a correlation

Mysterious Flame: Conscious Minds in a Material World (New York: Basic Books, 1999), and conceptualists, like Chalmers, do not represent the mainstream view.

⁵One of the exceptions is Gregg Rosenberg, who defends a similar view in his *A Place for Consciousness: Probing the Deep Structure of the Natural World* (New York: Oxford University Press, 2004). See also E. Holman, "Panpsychism, Physicalism, Neutral Monism and the Russellian Theory of Mind," *Journal of Consciousness Studies* 15.5 (2008), pp. 48–67.

⁶S. A. Kripke, *Naming and Necessity* (Cambridge, MA: Harvard University Press, 1980); D. J. Chalmers, "The Two-dimensional Argument against Materialism," *The Character of Consciousness* (New York: Oxford University Press, forthcoming); M. McKinsey, "Refutation of Qualia-physicalism," *Situating Semantics: Essays on the Philosophy of John Perry*, ed. M. O'Rourke and C. Washington (Cambridge, MA: MIT Press, 2007), pp. 469–498.

between physical processes and some qualitative experiences, but proponents of identity theory (except perhaps Ullin Place) make a distinction between correlation and identity. If the identity theory cannot be verified, it should be rejected. And the common suggestion that the identity theory is an example of “inference to the best explanation” is of no help. Such inferences can make sense only when it is possible to verify them. Otherwise they lead to absurdity. Let us imagine, for example, that we have found a curious anomaly: two clocks of a different construction in different parts of the world are in exact synchrony with each other. We have no idea how this is possible. So what? Should we infer to the “best explanation,” saying that these two clocks are in reality just one clock, that is, identical? But we cannot even imagine how they could be a single clock, except in a figurative sense. As we have no method of verification of the “clocks identity theory,” this theory should be unacceptable, as it surely is. But the very same considerations can be applied to the mind-brain identity theory.⁷

Thus, if we still want to solve the hard problem of consciousness by appeal to the necessity of qualitative mental properties or states for our brain activity, we seem to have no choice but to turn to the third way in which this necessity might arise. According to this way, qualia are ontologically distinct from physical properties but have a real influence on physical processes in the brain. This is none other than interactionism (but not necessarily substance interactionism; I think we have more reasons to accept a kind of emergent interactionism). Interactionism is not a very popular position among analytic philosophers, and the reason is fear that it leads to violating the causal closure principle, which, as many believe, is a foundation of a contemporary physics. I hope that we can avoid such violation, and I will return to this topic at the end of the paper.

However, now I want to note that, if we set aside the identity theory, it looks like the only serious alternative to the preceding kind of interactionism is epiphenomenalism, a theory according to which our qualitative mental states are ontologically distinct from physical reality but have no influence on physical processes. In recent times there have been a few attempts to destroy epiphenomenalism. The main goal of the present paper is to survey these attempts briefly, to show that they don't provide conclusive arguments against epiphenomenalism, and to supplement them with two other arguments. One of these two arguments is only relatively new, although I hope to strengthen it here. But the other argument is quite new and, I hope, conclusive. In the course of offering this second argument, I

⁷Common criticism of the verification principle misses the point in our case. It seems to be true that when we talk about some abstract propositions, the verification principle does not work, as indicated by S. Soames, *Philosophical Analysis in the Twentieth Century*, Vol. 1: *The Dawn of Analysis* (Princeton: Princeton University Press, 2003), pp. 289–291. But in our case the proposition in question is a factual one, and no one, as far as I know, has managed to show that this principle does not work with propositions of such a kind.

will also give reasons for accepting the interactionist position about qualia that I have mentioned above.

1

If we try to sum up the current arguments against epiphenomenalism we notice that they can be divided into two groups: (1) arguments from common sense, and (2) more special philosophical arguments. The argument from common sense usually takes this form: ordinary experience shows us that our qualitative mental states, such as pains or desires,⁸ play an important role in the production of our behavior.⁹ If, say, I want a drink of water, then it seems obvious that my subsequent behavior through which I satisfy this desire is in some sense determined by it. However, the epiphenomenalist can easily avoid this objection by noting that our ordinary experience shows only a correlation between mental states and behavior. But correlation is not yet causation. It is possible that the real causes of my behavior are some neuronal processes, hidden from ordinary experience. They produce not only the behavior, but epiphenomenal mental states as well. And, as real causes are not directly available to us, we take these epiphenomenal states for the real causes of the behavior. But this is just an illusion. And it is not common sense which produces this illusion. Rather it is produced by philosophers who misinterpret the common-sense attitude.

For a long time the main special philosophical argument against epiphenomenalism was an argument from evolutionary theory.¹⁰ It goes as follows: if qualitative mental states, or consciousness, were not causally efficacious, then they could not play an adaptive role and be under pressure of natural selection; and so they could not evolve at all. This argument, however, presupposes that all enduring phenotypic traits have a positive adaptive value and cannot be adaptively neutral. But this presupposition does not follow from the theory of evolution. Let us suppose that trait A has a positive adaptive value but that the mutation as a result of which it originates produces another neutral trait B as well. In such a case we would have a neutral enduring phenotypic trait. And consciousness very well could have been such a neutral trait, connected with adaptive behavioral patterns. So it seems this argument also fails.

In recent decades there has also emerged an epistemological argument against epiphenomenalism:¹¹ if consciousness does not affect behavior,

⁸It goes without saying that desires can be interpreted as behavioral dispositions. I believe, however, that they (as well as other conscious mental states) also have a qualitative side. Cf. J. R. Searle, *Mind: A Brief Introduction* (New York: Oxford University Press, 2004), p. 134.

⁹See, for example, R. Swinburne, *The Evolution of the Soul* (New York: Oxford University Press, 1997), p. 1.

¹⁰In recent times this Jamesian argument has been pushed forward by William Hasker, *The Emergent Self* (Ithaca: Cornell University Press).

¹¹It is not clear who invented this kind of argument. Recently, it has been vigorously defended by A. Elitzur. See, for example, A. Elitzur, "Consciousness Makes a Difference: A

then our reasonings about consciousness do not depend on its existence. The knowledge expressed in such reasonings is not derived from consciousness and its properties. Correspondingly, this knowledge is not a knowledge about consciousness. And if this is so, then, assuming the truth of epiphenomenalism, we can reasonably doubt that we can know anything about consciousness. Or, rather, if epiphenomenalism is true, then our consciousness should be something totally unknowable. But, as a matter of fact, we know something about consciousness, and this fact means that consciousness must have a real influence on our behavior. Defenders of epiphenomenalism (such as D. Chalmers),¹² however, respond¹³ that this argument is based on the so-called causal theory of knowledge: in order to know something about any object I should have been affected by it. As consciousness is causally impotent it is impossible to know about it. But why should we believe in the universality of the causal theory of knowledge? How can we exclude the possibility that some things could be known immediately? And if we cannot eliminate that possibility, as well as the possibility that consciousness is such an object of immediate knowledge,¹⁴ then the epistemological argument against epiphenomenalism loses its force.

2

Given the failure of the preceding arguments against epiphenomenalism, we might at this point conclude that epiphenomenalism is immune to any conceptual objections.¹⁵ But I don't think so. And now I am going to provide such objections. I begin with another argument from common sense. One of our common-sense attitudes is a belief in the reality of other minds. Is this attitude compatible with epiphenomenalism? I infer the existence of other minds on the basis of the similarity of my behavior to the behavior of other beings. My behavior is accompanied by subjective experience, and, if I see a similar behavior, I conclude that the being which behaves

Reluctant Dualist's Confession," *Irreducibly Conscious*, ed. A. Batthyany, D. Constant, and A. Elitzur, (Heidelberg: Universitätsverlag Winter, forthcoming), where he sums up his ideas on this topic.

¹²Chalmers is not an epiphenomenalist. Still, he believes that epiphenomenalism can be saved from criticism and that it has no crucial logical flaws. See Chalmers, *The Conscious Mind*, p. 160.

¹³*Ibid.*, p. 196.

¹⁴Robert Kirk, *Zombies and Consciousness* (Oxford: Clarendon Press, 2005) attempts to show that this line of defence is flawed; I do not think, however, that he succeeds in this attempt. His argument is based on an odd thought experiment about a zombie with pictures on the soles of his feet, which are exactly parallel and analogous to Kirk's visual qualia. Then he concludes that as there obviously could not be an "epistemic intimacy" as regards those pictures, the same is true about qualia. I think this argument is a non-starter, because I see no reasons to accept such an analogy: qualia, in contrast to "sole-pictures," incorporate some relation to a subject.

¹⁵I set aside experimental objections, if there are any: in fact, with Libet's famous data at hand, an epiphenomenalist could easily disarm them. See B. Libet, *Mind Time: The Temporal Factor in Consciousness* (Cambridge, MA: Harvard University Press, 2004).

in such a way also has a subjective experience, that is, consciousness or mind. But epiphenomenalism denies the causal connection between behavior and consciousness. And if it were right in this denial, it seems I would have no reason to admit the existence of other minds. Of course, I would see that the behavior of other beings is similar to my behavior. But so what? They could behave in such a way even without consciousness. Why indeed should I assume they have subjective experience?

Of course, some epiphenomenalists would say that in reality they do not deny any connection between consciousness and behavior. They would insist that the very same processes in the brain which produce behavior produce our epiphenomenal consciousness as well. So while our consciousness has no influence on our behavior, the lack of consciousness would mean the lack of its neuronal basis and some changes in behavior. But on such a view, how could we know that our epiphenomenal consciousness is produced by the very same brain activity that produces our particular behavior? From the position of common sense, we cannot refer to any experimental scientific data. The only way left is to refer to some simplicity considerations. And some partisans of epiphenomenalism in fact appeal to such considerations. They say that a possible world in which my epiphenomenal consciousness is produced by brain activity distinct from the brain activity producing my specifically human behavior, and where only my behavior might be accompanied by subjective experience, while other human and non-human animals might be simply zombies, would be much less uniform than our world where we assume a systematic correlation between a certain kind of behavior and consciousness.

Since simplicity considerations originate in common sense, it seems that even if we accept epiphenomenalism, we can still retain the idea of the existence of other minds. Still, I am sure this is the wrong line of argument to follow. I am not going to deny that simplicity considerations originate in common sense. But my point is that, in fact, the world in which only I am epiphenomenally conscious would be much more simple than the "more uniform" world replete with epiphenomenal minds. Indeed, this second world shows us a classic example of multiplying essences without any necessity.¹⁶ I have no need to postulate the reality of other epiphenomenal minds: it would be a waste of ontological material, as by definition they must be impotent. At best I could adopt the "intentional stance" (à la Dennett) toward other human beings and non-human animals, because it would help to predict their behavior—without making any ontological commitments to actual intentional states in these beings. (We make no ontological commitments to computers' having intentional states of different kinds when, for example, we play chess with them.)

¹⁶Chalmers, *The Conscious Mind*, and W. S. Robinson (in his comprehensive entry on epiphenomenalism, <http://plato.stanford.edu/entries/epiphenomenalism>) discuss this topic but do not pay due attention to this point.

So we see that epiphenomenalism concerning my own consciousness leads to the zombification of other people and animals. As this zombification contradicts common sense, epiphenomenalism is incompatible with common sense.

And the conclusion that epiphenomenalism must fail can be strengthened with the help of another argument. This second—special philosophical—argument is based on three premises. Each of these premises looks quite unproblematic. The first premise is that identical (from the qualitative side) events can have different causes. This point is a commonplace. For example, the same rise in stock markets can be caused by different factors; you could break a cup in the same way with your left hand or with your right hand, etc. Speaking abstractly, any event is something like a vector, which can be considered as a sum of other vectors. And it is obvious that different components (in our case different components of a complex cause) can bring about the same resultant. The second premise is also quite obvious, or so it seems. It notes that the memories belonging to human beings reflect or represent their past lives, their individual histories, and, in general, these memories do so correctly. Finally, the third premise directs our attention to the fact that our behavior is correlated with the qualitative mental states we have. For example, if I want to drink some water, I pick up a glass of water, not a glass of wine. Even the epiphenomenalist would not deny a correlation of this kind. Of course, there may be cases in which such correlation does not hold, but, as a rule, it holds.

As far as I can see, these premises are independent of each other. And, let me repeat, I am sure that they would hardly give rise to controversy. It is interesting, however, that combining them leads to the refutation of epiphenomenalism and provides a strong case for interactionism. Indeed, if identical events can (in the sense of real, or “natural,” possibility, and not just of logical possibility) have quite different causes, and if my brain at a given moment of time can be considered as a sum of neural events (all or at least some of which could have had different causes), then it is obvious that its current complex state could have been a result of very different trains of events, very different causal trajectories. But in such a case I would have had a different individual history, and, according to the second premise, other memories, than I actually have. And the contents of our memories are the source of intentional objects (which are qualitative in their nature) of other mental states, like beliefs and desires. Meanwhile, the third premise tells us that our desires and other mental states are correlated with our behavior. Hence, if my current physical state had been the result of a different train of events, providing me with different memories, desires and so on, then, while being (currently) the same physically as I actually am, I would behave differently. And the differences in my behavior would depend on the fact that I would have different mental states. And if this is so, we must admit that my actual behavior also crucially depends on the mental states I currently have. And that fact means that our qualitative mental states have a real impact on our behavior.

When I discussed this argument with some colleagues, they proposed a few objections, which I am going to survey here. Some said that while all the premises are correct, the conclusion may not follow in case of a Putnam-like externalist scenario. Indeed, it is possible to have had a different individual history simply due to some differences in environment, for example. And such differences could be undetectable at the qualitative level of our memories. Then I could have had a different past than I actually have, but still I would have the very same memories. I think this is really possible, but this fact does not affect my argument, because such scenarios do not exclude other scenarios where the differences *are* detectable at the level of perceptions and memories. Such scenarios are also possible, and that is all that I need for my conclusion.

Another objection¹⁷ draws our attention to the fact that, according to my argument, my present mental states depend not only on the physical state of my brain, but also on my past experiences. The worry is that such dependence may involve some kind of a "causal action at a temporal distance," that is, from the past to present, which is hard to swallow. I think, however, that in reality we have no serious reasons to worry. First of all, not every dependence is a causal one. There is such a thing as logical dependence, for instance, and our case can give an example of it. But, what is even more important, we will see that in fact we have no need to deny the supervenience principle at all, that is, we have no need to deny the strict dependence of my present mental states on physical states (only after denying the supervenience principle would we have a need for additional causation). All we need to deny is the local supervenience principle, but this can leave intact the global supervenience principle (I will discuss this later). Of course, such denial can shock somebody and maybe is also hard to swallow, but that is what my argument is about. And a priori it is difficult to understand why the local supervenience principle (same brain—same qualia) is more solid than the reverse principle (same qualia—same brain) which is universally denied since the coming of functionalism. We simply have to reject some philosophical superstitions.

The only way to criticize my argument, I think, is to attack its premises. But as these premises are universally accepted, the burden of proof is on those who want to attack them. And it hardly helps to say that all these premises express theses that allow exclusions, and that scenarios I describe presuppose the very same exclusions. It hardly helps because this thesis is also to be proved, and I see no clear ways to prove it.

3

So, if the arguments above are correct, epiphenomenalism is wrong, and we should accept interactionism. And then "the hard problem of consciousness" mentioned at the beginning of this paper could have the answer that I sketched above: brain activity is accompanied by subjective

¹⁷Made by Richard Swinburne.

experience *because* without such accompaniment our brains, being the same physically, simply could not work as they usually work, and that's because our mental states do indeed affect our behavior.

But now it is time to think about the cost of this answer. If we admit that mental states have an influence on brain processes, we are in danger of destroying the causal closure principle, that is, the principle according to which every physical event has an immediate physical cause.¹⁸ But why should we be so worried about violation of this principle?

I see at least two reasons to worry. First of all, as I have already mentioned, some philosophers believe that the renunciation of this principle leads to the demolition of the very basis of experimental physics. Secondly, it is rather difficult to deny this principle, as it may be ranked among the fundamental natural beliefs that guide us in our ordinary life.

The first of these reasons is not, I think, very dangerous. For example, David Papineau has shown that until the twentieth century experimental physics did not in fact presuppose the causal closure principle.¹⁹ Papineau is probably right that in the twentieth century this principle became more significant for scientists and philosophers. It is clear, however, that the core of contemporary physics is quantum mechanics, and the relation of quantum mechanics to that principle is notoriously ambiguous, as some interpretations of quantum mechanics allow for the crucial role of the conscious observer in the collapse of the wave function.

The second reason to worry seems to me much more serious. I think that our belief in causal closure can be counted among our natural beliefs. Admittedly, this is far from obvious, and in fact is not a commonplace opinion. Still, we can try to prove it by connecting the causal closure principle with one or another natural belief, such as our belief in the correspondence between the past and the future experiences. To demonstrate this connection we should do some kind of phenomenological analysis. In outline, this analysis looks as follows. As Hume has shown, our reasonings about facts are based on inferences from the past to the future. Such inferences presuppose our belief in the correspondence or "conformity between the future and the past." Looking closely at this belief, we see that it implicitly incorporates the causal closure principle. Indeed, let us suppose that it is possible to infer from the past to the future and deny the causal closure principle at the same time. If this principle is denied we should be ready to accept that there could be situations in which an external physical event B could have no physical correlate (necessarily connected with it) in the previous moment of time. Then, if that previous event A is repeated, we would have no reasons to believe that it will be succeeded by the event B. Note that we could not rely on possible mental causes of B included in A, as they are not among our experiences (and, as we saw in the previous

¹⁸Cf. E. J. Lowe, "Causal Closure Principles and Emergentism," *Philosophy* 75 (2000), pp. 571–585.

¹⁹D. Papineau, *Thinking about Consciousness* (Oxford: Clarendon Press, 2002).

section, we cannot infer to them from the physical characteristics of event A). And, lacking such beliefs, we would not be able to infer from the past to the future. It follows that our belief in the conformity between the future and the past includes belief in the causal closure principle, when we think about physical processes outside our body. And if our belief in the conformity between the future and the past leads us to conclude that the behavior of other people is determined by physical causes, it is reasonable to generalize this conclusion to ourselves. So it is better not to deny the causal closure principle.

I understand, of course, that this brief argument needs to be developed in more detail. But that is not my purpose here. All I want to do is to indicate a possible line of argumentation for the causal closure principle. For the purposes of this paper it suffices to stipulate that this principle is true (and most philosophers would agree with that) and see what happens next. So let us assume that we have no choice but to accept causal closure.

4

Can we reconcile the causal closure principle with the thesis that mental states have an impact on behavior? In order to find an answer to this question let us return to the proof of the causal efficacy of mental states. We saw that we can infer such efficacy on the basis of the assumption that one and the same brain could be accompanied by different sets of mental states, if we assume also the correlation between mental states and behavior. Combining these assumptions gives us the following conclusion: if I consider my brain, then an answer to the question "Why does my brain produce the behavior it produces?" is impossible without taking into account the particular mental states I have: I could have other mental states, and then my brain would produce quite a different behavior. Looking into this argument, we see that it presupposes an important qualification, expressed in the phrase "if I consider my brain, etc." That is, we can prove the efficacy of mental states by the local consideration of a material system we are interested in. In other words, up until now, we have talked about *local interactionism* only.

This is an important qualification, and if we free ourselves from it and consider my brain in the context of the whole universe, the picture may change drastically. Indeed, couldn't we assume in such a case that my behavior is determined by physical causes after all, only not just by local physical factors in the brain but also by non-local physical factors as well? This assumption could help us to save the causal closure principle.

It is clear that this also is not a solution without costs. Apart from allowing a notion of non-local causality, we again face the question about the role of mental states in producing brain activity: are they epiphenomenal after all?

Before answering this question, I should mention that the solution we are now considering has some obvious merits as well. Most importantly, it can help to save the lawlike relation between the physical and the mental.

When I argued that local interactionism is true, we came to the conclusion that one and the same brain could be accompanied by quite different mental states.²⁰ As I have already mentioned, this conclusion contradicts the local supervenience principle, which establishes an univocal correspondence between the brain and its qualitative mental states. And if the local supervenience principle is wrong, we might doubt whether there exists any lawlike connection between mental states and the brain. The only chance to save such a relation is to show that while the local supervenience principle does not work, this might not affect the global supervenience principle, according to which identical physical worlds are accompanied by the same mental states.²¹ Indeed, from the denial of the local supervenience principle it doesn't follow that the global supervenience principle is also to be denied. For if the global supervenience principle is true, then the falsity of the local supervenience principle means no more than that if my brain were accompanied by a different set of mental states, then the physical world as a whole would not be the same as it is now: there would be some differences in it.

Now it is possible to combine this picture with the causal closure principle. Suppose that non-local (as regards my brain) physical differences of a world in which my brain occurs are accompanied by a different set (compared to the actual world) of mental states that are correlated with a different behavior. Then these non-local physical differences might be considered to be a non-local physical part of the cause of such different behavior. And, correspondingly, the physical differences of the actual world—in comparison with the physical features of that possible world—might be considered to be a non-local physical part of the cause of my behavior in the actual world. Hence, accepting the causal closure principle in combination with local interactionism, we get a confirmation of the global supervenience principle, which in turn can provide us with a reason to believe in the lawlike connection between mental states and their physical bases. It is only that the physical bases now extend beyond the local, physical features of my brain itself.

5

The possibility of such a lawlike connection gives us hope that it will be possible to find a new answer (since, if we believe in causal closure, the straightforward interactionistic answer is wrong) to Chalmers's question: Why is my brain activity accompanied by subjective experience? In order to see how this new answer arises, note that the above discussion raises the following problem: if we retain the causal closure principle and

²⁰J. C. Fisher, "Why Nothing Mental is Just in the Head," *Nous* 41 (2007), pp. 318–334 comes to a similar conclusion, using different arguments and thought experiments. I have published some of my results in V. Vasilyev, "Brain and Consciousness: Exits from the Labyrinth," *Social Sciences* 37 (2006), pp. 51–66.

²¹See J. Kim, *Philosophy of Mind* (Cambridge, MA: Westview Press, 2006), pp. 9–10 for details.

introduce, in the way that we have above, the idea of non-local physical factors in the causation of our behavior, then we may get the impression that our mental states are epiphenomenal after all. And it is hard to explain the existence of epiphenomenal mental states.

However, this problem can be solved. Our situation here is in fact totally different from the one in which we usually talk about epiphenomenalism. If our behavior could be explained by local physical processes in the brain, the existence of mental states would be miraculous. But if we have already proved that our mental states make a difference at the local physical level, and if we then assume that their physical effects still could have had some physical, although non-local, correlates, then we might be inclined to think that it is these non-local correlates that are a kind of epiphenomena. The reason is simple: in general, we have no empirical evidence to believe in non-local causation. But if we want to insist that these non-local correlates are real causes, then, it seems, we should accept the following schema: as non-local causality is not universal, it is not unconditional; and so it is quite plausible that the condition under which a non-local physical correlate might be considered as a cause (or as a component of a complex cause) of some event is precisely the very existence of private mental states accompanying this event within some physical system and correlating with it. In other words, mental states seem to be something like mediators in the realization of non-local physical causation. They are such mediators not in the sense that they are intermediary links between the non-local factors and behavior (this would violate the causal closure principle) but in the sense that they are necessary ontological conditions of the realization of non-local physical causation.²² In such a case the mental states would not be epiphenomenal; they would have causal relevance, if not causal efficacy.

If this is the right way to go, then not only can we explain why our brain activity is accompanied by subjective experience, but we can also begin to try explaining how that accompaniment is possible. (Recall that the full version of Chalmers's question includes a request for how it is that brain activity is accompanied by consciousness.) Indeed, since it is very likely that our mental states emerge from brain activity, and if we then take into account that these states must be ontological conditions of non-local physical causation, we can suppose that they are produced by physical systems, which would demonstrate some kind of non-deterministic behavior in the physical systems themselves if we ignore the role of the mental states in the realization of the non-local physical causation. In other words, the mechanisms due to which some physical systems produce mental states must have a relation to the incompleteness, at the physical level, of their local causal patterns. I should note, however, that this scheme does not mean that the existence of mental states is just an odd

²²This conclusion may possibly illustrate what Lowe, "Causal Closure Principles," calls "causation by a mental event of a physical causal fact."

consequence of some physical anomaly. In fact, we saw (see section 2 of the present paper) that the existence of mental states connected with some physical systems provides such systems with an opportunity to take account of their individual histories;²³ and that fact, in turn, can give them a big adaptive advantage.

So our return to the hard problem of consciousness and its possible solution helps us to see that the approach of David Chalmers in *The Conscious Mind* is probably the most fruitful one after all. Let me recall that in that book he considers a solution of the hard problem based on the thesis that mental states are ontological conditions of the realization of physical causation and even of the very existence of physical properties. At the beginning of this paper I said that this idea is rather bizarre. But it looks bizarre only while we use it in the context of a neo-Russellian ontological scheme (in which mental states are something like vehicles of physical properties), as Chalmers did. And from the arguments above, it follows that after some modifications in that scheme this approach may open the way to solving the hard problem. As Ned Block has noticed, to solve the hard problem is to close an “explanatory gap.”²⁴ It is interesting, however, that Block, as well as J. Levine,²⁵ believe there is no difference in explanatory gaps when we try to understand why subjective experience exists and when we try to understand why we have just this subjective experience. But while we certainly cannot find an answer to the second question yet, we are in a much more promising situation in regard to the first one.²⁶

Lomonosov Moscow State University

²³If this is true, then no purely mechanical system can simulate human behavior exactly. Indeed, any such system behaves on the basis of its current physical state only. This fact means that not only epiphenomenalism is wrong; the very same can be said about “conscious inessentialism.” (See O. Flanagan, “Conscious Inessentialism and the Epiphenomenalism Suspicion,” *The Nature of Consciousness: Philosophical Debates*, ed. N. Block, O. Flanagan, and G. Güzeldere [Cambridge, MA: MIT Press, 1997], pp. 357–373.) And maybe John Searle and many others were a bit too over-optimistic in their belief in so-called Weak AI. This over-optimism distracted their attention from the fact that, say, Searle’s famous Chinese Room (which allegedly passes the Turing Test) simply cannot work; it cannot, for example, provide reasonable answers to some indexical questions, like “What time is it now?”

²⁴N. Block, “Comparing the Major Theories of Consciousness,” *The Cognitive Neurosciences IV*, ed. M. Gazzaniga (Cambridge, MA: MIT Press, forthcoming).

²⁵J. Levine, *Purple Haze: The Puzzle of Consciousness* (New York: Oxford University Press, 2001); cf. J. Levine, “Materialism and Qualia: The Explanatory Gap,” *Pacific Philosophical Quarterly* 64 (1983), pp. 354–361.

²⁶I am grateful to Robert Howell whose suggestions were of much help to me in writing this article. I owe also a great debt to David Chalmers, Noam Chomsky, Daniel Dennett, Ned Block, David Armstrong, Richard Swinburne, and Michael McKinsey for their comments in our personal communications.