A discussion of Professor Vadim Vasilyev’s argument against the local supervenience of the phenomenal on the physical: a presentation of the argument, critical responses to it, and Vasilyev’s replies to the criticism. The compiler is Evgeny Loginov, translation by Artem Iunusov, and editing by Robert and Pamela Howell.

Keywords: local natural supervenience, phenomenal properties, global supervenience, mental causation, causal closure principle, epiphenomenalism.

How to cite this article: Vasilyev V., Olson E., Stoljar D., McGinn C., Frankish K., Dennett D. “Discussion. Local Natural Supervenience and Causation”, Date Palm Compote, 2021, No. 6, pp. 2–27.
introduction note

As far as I know, Vadim Vasilyev's argument against the local natural supervenience of phenomenal mental properties on physical properties is the only analytic-style argument about consciousness and causality that exists in contemporary Russian philosophy. There is some literature on this argument ([Girenok 2013], [Gasparyan 2014a], [Gasparyan 2014b], [Mishura 2014], [Pavlov-Pinus 2014], [Volkov 2015], [Kuznetsov 2016], [Volkov 2018], [Loginov 2019], [Kuznetsov 2020], etc.). Vasilyev's argument can be presented as part of the tradition of anti-physicalist arguments which begins with S. Kripke's 'Identity and Necessity (1971) and T. Nagel's 'What Is It Like to Be a Bat? (1974) and includes great pieces of philosophical work by J. Searle, F. Jackson, J. Levine, D. Chalmers, and many others. In the USSR, materialism was part of the ideology, but despite this, there were researchers who developed their own views that did not fit within the framework of dialectical materialism. In 1961, Anatoly Mickевич (writing under the pseudonym A. Dneprov) offered a story very close to N. Block's and J. Searle's later Chinese Nation and Chinese Room arguments [Dneprov 2018]. Beginning in the early 1970s, David Dubrovsky has been developing an informational theory of consciousness. Beginning in the early 1980s, Alexander Gryaznov, a scholar of the history of modern philosophy, became interested in the analytic philosophy of M. Dummett, P. Strawson, and L. Wittgenstein. Later his interests shifted toward the philosophy of mind. Vadim Vasilyev was his student. Thanks to Gryaznov, Vasilyev became acquainted with Chalmers' book, 'The Conscious Mind' which led him into the world of analytic philosophy. Drawing on his background in the history of modern philosophy, Vasilyev made historical reconstructions of the ideas of J. Searle, D. Dennett, and D. Chalmers. He also described in hitherto unknown detail Wittgenstein's visit to the USSR. In the field of metaphilosophy, Vadim Vasilyev developed his own version of conceptual analysis, which he called argumentative phenomenology, a methodology based on clarifying the connection between the basic cognitive attitudes inherent in human nature. In the field of ontology, Vasilyev has defended two original positions: local interactionism and ultracompatibilism. Local interactionism is qualia interaction which is compatible with the causal closure principle. Ultracompatibilism is a position according to which free will itself (freedom of action plus rational choice) is compatible with the causal closure of the physical because desires are treated as conditions of realizing physical causation. Vasilyev's research in the field of analytic philosophy is presented in his books “The Hard Problem of Consciousness” (2009), “Consciousness and Things” (2014), “In Defense of Classical Compatibilism. An Essay on Free Will” (2017) and in a large number of articles.

I am very grateful to Anton Kuznetsov for his help in organizing this discussion, to its participants, and to those who made it possible: translator Artem Iunusov and editors Robert and Pamela Howell.

Evgeny Loginov

REFERENCES


an argument against the local supervenience of the phenomenal on the physical

Vadim V. Vasilyev, Lomonosov Moscow State University

Abstract: In this paper I discuss a problem with the local natural supervenience of phenomenal mental properties on physical properties. I propose an argument, based on the principle of the variety of causes and on the thesis of the general veracity of human memory, that leads to the conclusion that such supervenience does not occur. I then consider objections to this argument and show how it helps to solve the problem of mental causation.

One part of the mind–body problem is the question of whether we can consider the relation between mental and physical properties as one of supervenience. By physical properties I will mean those properties that can be described in terms of modern physics and neuroscience. As for mental properties, I will be concerned mostly with their internal, phenomenal aspect. In what follows, when I speak of mental properties and states I will by default mean the phenomenal aspects of such properties. By the phenomenal aspect of a mental state I understand that aspect that seems to be accessed exclusively from the first-person perspective.

The supervenience relation can be considered in a number of ways: one can speak of a global supervenience and a local one, a logical supervenience and a natural one, and so forth. In this paper I will touch upon global supervenience only in passing. I am also not interested here in the logical supervenience of the mental on the physical. I agree with David Chalmers’ arguments against such a supervenience [Chalmers 1996]. So I will deal with the natural (or nomological [McLaughlin, Bennett 2018]) local supervenience of phenomenal properties. Following Jaegwon Kim’s definitions [Kim 2005, 55], by such a supervenience I will understand a relation of phenomenal and physical properties of an organism such that any phenomenal property \( M \) instantiated by the organism at \( t \) presupposes that that organism has an internal physical property \( P \) at \( t \) in such a way that duplication of \( P \) guarantees duplication of \( M \). (By the internal physical properties of an organism I will mean those of its physical properties that can be established by local observation of that organism.) Despite \( P \) and \( M \) being correlated, we can distinctly conceive of the occurrence of \( P \) in the absence of \( M \). (The impossibility of a conceptual dissociation of \( M \) from \( P \) would mean that one of them is logically supervenient on the other.)

In the contemporary literature on philosophy of mind there is something of a consensus regarding the local natural supervenience of phenomenal mental properties on physical ones. The majority of researchers seem convinced that such a supervenience exists. To avoid any possible confusion I must note that in the case of local supervenience of the mental \( \text{simpliciter} \) on the physical, the situation is different: already Donald Davidson, to whom we largely owe the term ‘supervenience’ in its modern use [Kim 2017], has denied the existence of a supervenience of this kind [Davidson 2001, 62]. His reason for the denial was the fact, demonstrated by Hilary Putnam and others, that certain mental states (for example beliefs) are constituted partly by external physical factors. (Davidson, however, admitted the existence of supervenience in another sense that we can ignore for present purposes, see [Shagrir 2011].) Many philosophers have agreed with him. But this is an externalist aspect of the problem of local supervenience, and denying supervenience in this sense does not imply that physically identical organisms could have different phenomenal experiences.

The picture is completely different when we pass to the internalist aspect of the problem, the one dealing with \( \text{phenomenal} \) mental properties. Thus, for instance, after examining the cases of an apparent absence of the supervenience of the mental on the physical because of the considerations described above, J. Kim specified the kind of supervenience that does obtain anyway: ‘Every internal psychological state of an organism is supervenient on its synchronous internal physical state’ [Kim 1982, 59]. The examples that he then offers
show that by ‘internal psychological states’ he means precisely the phenomenal states or properties that I am concerned with. Chalmers was equally clear on this point, noting in a similar context that ‘consciousness is naturally supervenient on physical properties’ [Chalmers 1996, 58]. This kind of supervenience was also admitted by Colin McGinn [McGinn 1996, 52] and John Searle. Searle was particularly emphatic: ‘conscious states are totally dependent, or supervenient, on brain states’ [Searle 2004, p. 148]. One could easily multiply examples.

It is interesting, however, that the almost universal agreement on this point has led to something of a deficit of detailed accounts of why such a supervenience is supposed to be the case at all [cf. Lewis 1985, 167]. Many probably consider it to be a position almost without alternatives (because its rejection seems to detach the phenomenal from its physical basis) and therefore self-evident. In what follows I will, however, attempt to show not only that there is nothing self-evident about such a supervenience but also that there is no local supervenience of the mental on the physical at all. I will proceed as follows. First I will set out my argument (it is sometimes called the ‘causal trajectories argument’, but I prefer to call it the ‘accidental twins argument’). Then I will add some clarifications and consider a few objections. Finally, I will discuss the implications of denying local supervenience of the mental on the physical for the problem of mental causation and for the critique of epiphenomenalism. I will also show how we can avoid detaching the phenomenal from the physical even if local supervenience of the mental on the physical does not obtain.

The argument that I would like to offer here against the natural local supervenience of phenomenal mental properties on the physical relies on two main premises that may seem quite obvious. (Shorter versions of this argument have been published before; see for example [Vasilyev 2009, 520-521].) The first premise states that it is really possible for qualitatively identical events to have qualitatively different causes (I will call this the variety-of-causes claim). I understand real possibility as accordance with the natural laws of our world. (Real possibility can also be expressed by appeal to those possible worlds whose natural laws are indistinguishable from the laws of our world: something really possible in our world would then exist in one of these possible worlds.) The more commonly used term for this is ‘physical possibility’, but I consider this phrase to be quite misleading because one can in principle speak of this kind of possibility in the case of mental events as well (the other well-known term for it, ‘nomological possibility’, is, however, equally acceptable).

Real possibility is distinguished from logical possibility, for which a simple conceivable of events and their combinations is sufficient. One can hardly deny the logical possibility of qualitatively different causes of identical events, but I think we can admit the real possibility of such causes as well. To support this claim one can refer to actual examples (the fact that those examples are actual would indicate that we are dealing here with a real possibility). Let us consider the classical example of billiard balls. The event in question would be a movement of a billiard ball. It is hard to deny that the billiard ball can move due to a variety of causes ranging from the movement of another billiard ball to the table falling over. As the event we chose as an example does not seem to be extraordinary, it seems quite reasonable to suppose that we can expand the scope of the variety-of-causes claim to include other events as well.

The second premise is that one’s memory reflects one’s past. Memory is of course a complex phenomenon, but here I am thinking specifically of so-called episodic memory. It produces in us images of past events. These images are private and can be directly accessed exclusively from the first-person perspective. They are therefore not only mental but also phenomenal in their nature. The statement that one’s memory reflects one’s past should not be understood as saying either that it reflects the past perfectly (we know that not to be the case) or that it does so necessarily. The point of the statement is rather that a difference in one’s past tends to lead to a difference in one’s memories.
Let us now combine the two premises. If one uses the first premise in the general sense that I have given it, it should hold for a wide variety of events including, among others, those that are taking place in my brain. Let us take the state of a thing at t to be an instantiation of the thing’s internal properties at t, with no particular stress being given to the procedural aspects of this instantiation. If one now understands an event as an instantiation of a property at a certain moment in time [Kim 2005, 19], and if one considers only the events that instantiate the internal properties of a thing and bring that thing into a certain state, then it turns out that at this very moment my brain is a result of the instantiation of a large number of internal physical properties. According to the first premise, the instantiation of this set of properties could have been caused by instantiations of different sets of properties at preceding moments in time. In other words, my brain could have reached its current state by different trajectories. There is thus a real possibility that it and the organism of which it is a part could have had a different causal history — that is, a different past. But, according to the second premise, a different past tends to produce different memories. Therefore, while being physically the same, I could have had different phenomenal mental states. This point can also be made by appeal to twins: an exact twin of me (in the sense of a person having exactly the same internal physical properties) with a different causal history could have had phenomenal states qualitatively different from mine. So, an instantiation of a qualitatively identical set of local internal physical properties does not guarantee an instantiation of a qualitatively identical set of correspondent phenomenal mental properties. Given that real possibility is in question here (both premises were formulated in terms of this kind of possibility), we can infer that there is no local natural supervenience of the phenomenal on the physical.

I will now present the argument in a more formal way.

(1) It is really possible for all qualitatively identical events to have qualitatively different causes.

(1a) Because any person’s brain states (or organism states) at a particular moment in time can be regarded as (complex) events, it is really possible for them to be caused by qualitatively different causes—that is, it is really possible for these brains and organisms to reach their current states by different causal paths.

(2) Human episodic memory tends to reflect the past of the organisms that have it and therefore their causal histories or paths.

(5) Therefore identical organisms with different causal histories would tend to have different memories, which implies that it is really possible for these organisms to have different memories.

Now, episodic memories are an instance of phenomenal mental states. But the local natural supervenience of the mental on the physical excludes the real possibility of physically identical organisms having different mental states. Therefore this argument shows that there is no such supervenience.

Critics of this argument can target either its premises or the claim that its conclusion follows from them.

The second premise seems to be the most secure one, and I will not here deal with arguments against it in any detail. It is indeed possible to come up with such arguments: for instance, one could argue that memory is non-phenomenal (or ‘non-qualitative’) in its nature [Volkov 2018, 70]. If so, episodic memory would not reflect our past simply because it would not exist, as it is phenomenal in its nature. But denial of the existence of episodic memory seems to me so counterintuitive that I allow myself to discard outright this kind of criticism.

The first premise is more questionable. It can be criticized for vagueness (exactly what notion of cause is implied in it?) or for an unwarranted extension of the scope of the variety-of-causes claim so that it covers any event whatsoever. One can also attempt to demonstrate that this claim is false altogether: certain philosophers (Hume for instance) hold that identical events must have identical causes.

One could overcome all these objections if one could reformulate the first premise without using the notion of cause. Here is one way to do so: it is really possible that
identical local events are preceded by non-identical local events. This claim may be easier to defend than the first premise, and when combined with the second premise it leads equally well to the conclusion of the absence of local supervenience of the mental on the physical. Yet it is less clear intuitively. Therefore it is better to keep the notion of cause and, following the classical definition of John Mackie [Mackie 1974], to clarify it by saying that a cause is understood here as one of the necessary (‘non-redundant’) parts of a sufficient condition of an event’s occurrence. This definition fits well with our pretheoretical conception of cause and does not exclude the possibility of there being various causes of identical (in a qualitative sense) effects, as it does not exclude the possibility of there being various sufficient conditions of an event. The claim that a variety of causes is impossible presupposes a different notion of cause that would be difficult to reconcile with our pretheoretical intuitions. As for the charge of an unwarranted extension of the scope of the general variety-of-causes claim so that it covers all events and thus turns into some kind of variety-of-causes principle (without such an extension the argument might indeed fail), one can try to reinforce this charge in the following way. Let us take the event ‘ball bouncing off the floor’ and consider its causes. It is clear that the causes of such an event may vary—that is, that particular tokens of this event-type can be brought about by qualitatively different causes. Let us, however, give a more specific description of this event, such as ‘ball bouncing off the floor by two meters’. Intuitively, the number of possible causes of the event would then decrease. Our intuition might deceive us here as it deceives us in telling us there have to be more natural numbers than there are even natural numbers, but let us suppose that in this particular case it is correct. One can then argue that this tendency of the variety of causes to decrease given an increase in the specificity of the description of an event should hold generally, so that in the case of a completely specific event no variety of causes is possible.

We can respond to this objection by partially agreeing with it. If we are talking about a completely specific event, then it can be admitted that its tokens should have qualitatively identical causes. However, one can argue that a completely specific event must include in its description the state of the whole world at the moment of the occurrence of that event (provided that the notion ‘the moment’ can be meaningfully applied to the whole world). Otherwise, its tokens could be specified still further by adding to their descriptions all the details, left out originally, that pertain to the states of the world at the moments of the occurrence of those tokens. Our argument was, however, supposed to deal with local events, and the logic of the objection does not allow one to conclude that it is impossible for the causes of such events to vary. The objection might, however, lead us to change the wording of our argument. It is better to word the first premise as ‘All the qualitatively identical local events,’ and so forth.

Let us now turn to the question of whether the conclusion does indeed follow from the two proposed premises. One possible problem here is that the argument intends to establish the claim that different causal histories of physical systems such as a brain can be represented in different memories, even if at a given time these systems are physically identical. So for the argument to work it has to be the case that this claim does not contain a logical contradiction. Yet it might be argued that it could contain such a contradiction, namely this will be the case if one accepts the once quite popular view that phenomenal states of conscious experience are identical with physical states of the brain. If this view is correct, it is logically necessary that different memories differ from each other on a physical level so that it would be contradictory to speak of the qualitative identity of brains or organisms with different memories. So it might seem that the proposed argument relies on an additional implicit assumption: that mental and physical properties are not identical. One way to deal with this difficulty would be just to agree to expand the argument by adding this assumption as a third premise. The problem, however, is that most philosophers will surely say that such an expansion will render the argument epistemically uninteresting, because the question of the identity of the mental and the physical is itself far from being solved.

Luckily, this objection can be met. The redundancy of
the third premise is easiest to show with the help of an analogy. Let us imagine that every day we see from our window the following picture: A plump man approaches a building across the road, enters it, and then after a while a slim man exits the building. In the evening the slim man returns to the building, after which the plump man leaves it again. The difference in their appearances makes us assume that these are different people who work in shifts, one after another. We can imagine, however, that someone could try to accuse us of a logical fallacy here. Might it not be the same man disguised in different ways? If so, in concluding that these are different people, we would find ourselves caught in a logical contradiction. The conclusion that these are different people holds only if these people are indeed different. In drawing such a conclusion, we implicitly presuppose the very thing that we infer. Such an accusation would of course in fact be unwarranted. We do not know whether they are the same person, but the differences in our observations leads us to the hypothesis that they are different.

Our case is similar. We may not know whether phenomenal mental states are identical with brain states, but we can still infer that they are not—due, among other reasons, to our argument against local supervenience between them. We do not presuppose either the non-identity of the mental and the physical or their identity. But the argument gives us reason to believe that they are non-identical.

Let us suppose that the argument succeeds. Can it then somehow be useful in answering other questions pertaining to the mind-body problem?

The argument’s conclusion can be used as a valuable point in debates on mental causation and in a critique of epiphenomenalism. Epiphenomenalism is the view that phenomenal mental states are produced by the brain and are causally impotent to affect brain processes and an organism’s behavior. This view is regarded by many as inelegant but hard to refute. Therefore any new line of reasoning against this view deserves attention. Our argument can indeed offer a new critique of epiphenomenalism. To flesh it out, we have to add to our argument’s conclusion (which states that there is no local natural supervenience of phenomenal mental properties and states on inner physical properties of the brain) another premise: that people’s behavior tends to be correlated with their mental states, including the phenomenal aspects of their mental states. In other words, different phenomenal mental states tend to be accompanied by different behavior. If I want some juice, then, given a choice between drinks, I will tend to choose the juice. If, however, I want some Coke, I will choose the Coke.

Let us now conceive a pair of physical twins who have reached their identical physical state by different causal paths. (We can conceive them as existing either in our world or in two different worlds with identical natural laws.) As we already know, they can have different memories. Memories can be a basis of other mental states, so differences in the twins’ phenomenal mental states would most likely not be limited just to memories. But different mental states are correlated with different behavior. So the twins would probably behave differently. Given that they are physically identical, the differences in their behavior would have to be due to the difference in their phenomenal mental properties and states. Those states and properties would then not be causally impotent.

Now the proponent of epiphenomenalism can of course insist that, despite the differences in their mental states, these accidental twins will behave identically. The problem with this line of defense, however, is that it might render the correlation between one’s mental states and behavior (which epiphenomenalism does not deny) hard to explain. Indeed, let us assume this correlation to exist in my case as well in the case of most other people. Let us further conceive trillions of our physical twins with different causal histories. On the epiphenomenalist’s hypothesis, these twins would have different mental states but identical behavior. They would thus lack the correlation between the two. But what is to explain the fact that in us as well as in most of the people around us these are in fact correlated? After all, we are no better than the twins, and the causal paths leading to our present states are no more natural than are
Theirs. The fact of this correlation in us would seem to be something like a miracle.

If this critique of epiphenomenalism is successful, then its proponent—an interactionist—is presented with a further well-known problem. I am referring to the danger of violating the causal closure of the physical, which is especially problematic when one rejects the ontological identity between the mental and the physical. (If this identity is the case, then for obvious reasons the causal closure of the physical is not a threat to the causal efficacy of mental states.) If mental states affect our behavior, then the conclusion that physical events can have non-physical causes may seem inevitable. That conclusion in turn is seen by many as undermining the foundations of the experimental sciences of human beings and of nature in general.

However, the arguments above need not cause such concerns. Denying the local natural supervenience of a person’s phenomenal mental properties on the physical properties of that person’s brain does not necessarily imply denying the global natural supervenience of phenomenal mental properties on the physical, where the global natural supervenience is treated as establishing the supervenience relation between phenomenal and physical properties in the entire world. One can deny the local supervenience and admit the global one [Kim 1987]. Indeed, given the claim that physically identical organisms in different worlds with identical natural laws can be bearers of different phenomenal mental states, it is possible to infer the conclusion that physically identical worlds can be bearers of different phenomenal properties only if the worlds in question, containing physically identical but mentally different organisms, do not differ from each other physically in any respect at all. Otherwise one could try to explain the difference of the mental states in those worlds by appeal to physical differences between those worlds, differences that in such a case would prevent the free floating of mental phenomena over their physical bases and could preserve the lawlike connection of physical and phenomenal properties and the global supervenience of the phenomenal properties on the physical ones. In the case of global natural supervenience, physically identical but mentally different organisms can also be conceived to exist within the same world, given that these mental differences would be reproduced after replicating all the physical properties of such worlds. And such differences within the world—just as in previous case, when we considered different worlds—could be explained by reference to physical causes, now somehow connected with different localization of these organisms.

If this line of argument is correct, then although it is impossible to find exhaustive local physical explanations of a person’s behavior and although such an explanation always requires reference to that person’s phenomenal mental states, these states themselves, and therefore the behavior, can always be explained by reference to some physical properties of the world (even if it is not immediately clear what kind of properties they are). It means that interactionism could be a local one and not violate the causal closure of the physical globally.

REFERENCES


8. Mackie 1974 — J.L. Mackie. The Cement of the Universe: 

University Press, 1996.

10. McLaughlin, Bennett 2018 — B. McLaughlin, K. 
Bennett. “Supervenience”, The Stanford Encyclopedia of 
Philosophy (Winter 2018 Edition), ed. by E.N. Zalta (URL: 
supervenience/)


12. Shagrir 2011 — O. Shagrir. “Supervenience and 
Anomalism are Compatible”, Dialectica, 2011, Vol. 65, 
pp. 241–266.

13. Vasilyev 2009 — V. Vasilyev. “‘The Hard Problem of 
Consciousness’ and Two Arguments for Interactionism”, 

(In Russian.)
vasilyev's argument against psychophysical supervenience

Eric T. Olson

Vadim Vasilyev claims that two people can be physically identical, yet have different mental properties. This is not only possible in the broadly logical sense, he says, but nomologically possible: it’s consistent with the laws of nature. If the Martians were to make an exact physical duplicate of me as I am now, his brain states and activities would be identical to mine. But we may not be mentally alike. A mental discrepancy may be unlikely, like tossing a coin ten times and getting ten heads, but it would not violate any laws of nature.

The claim is not just that my mental states could differ from my duplicate’s in their extrinsic content. It’s widely agreed that my memories of Moscow might correspond in him to states intrinsically identical to mine, but which are not memories of Moscow, or even memories at all, because they weren’t caused by experiences of that city. Vasilyev says more strongly that we could be intrinsically mentally different. My duplicate may not have anything like my memories of Moscow: my Moscow memories might correspond in him to apparent memories of Zanzibar, or to a complete blank.

The argument for this claim has two premises. The first is about causation. It says that intrinsically identical states and events can have intrinsically different causes. (Hereafter all talk of similarity and difference will mean similarity or difference in intrinsic respects.) This is not just logically but nomologically possible. A billiard ball can be caused to move across a certain part of the table at a certain velocity by any of a wide variety of antecedent events: by being struck by another ball, or by a cue stick, or by being pushed by hand or fired from a gun. Likewise, the ball could have its current temperature as a result of cooling or warming, rapidly or slowly, or by having had the same temperature for the past hour.

This applies to human beings as well. A duplicate of me created by the Martians would be physically identical to me. (Or at least he’d be physically identical, when created, to me as I was when the Martians took my measurements.) But our physical states would have different causes: mine would be caused in the usual way, my duplicate’s by that wondrous Martian technology.

That’s the first premise. The second is about memory. It says that our memories are a record of our past histories. They may be an incomplete and unreliable record, but they’re a record all the same. Other things being equal, different pasts tend to produce different memories and similar pasts tend to produce similar ones. The reason your memories are now different from mine is at least largely that different things have happened to us in the past.

I have no doubt that these premises are both true. Vasilyev wants to infer from them that two people could be physically identical, yet have different memories. How? As I understand it, his reasoning is this: It could happen that you and I now have the same physical properties but different causal histories. But different causal histories tend to produce different memories. So it’s likely, given our different pasts, that we’ll now have different memories, despite there being no current physical difference between us. It’s therefore possible for us to be physically identical and at the same time to have different memories: there can be an intrinsic mental difference between two people without any intrinsic physical difference.

This is all correct until the last step. The argument appears to have this form: A certain state of affairs can hold in certain circumstances; a second state of affairs not only can hold in those circumstances, but is likely to hold; thus, both states of affairs can hold at once.

This argument form is not logically valid. The reason is that two states of affairs can each be possible, yet not be possible together. It commits the same fallacy as this one: It’s possible that P; it’s possible that Q; thus, it’s possible that P and Q. There can be a world where P is true and a world where Q is true but no world where both are, giving the argument true premises and a false conclusion. We can see this by substituting ‘not-P’ for ‘Q’. It may be possible, given current weather conditions, for it to rain tomorrow in Moscow. It may also be possible for it not to
rain. But it’s not possible for it both to rain and not to rain.

Vasilyev adds the further claim that the second state of affairs is likely given the relevant history. But this makes no difference. It may be possible, given current conditions, for it to rain tomorrow, yet likely to be entirely dry: it still doesn’t follow that it can both rain and be dry. There can be a world where P is true and a world where Q is true. There can even be a world where P is true and Q is likely given the relevant history. There still need not be a world where both P and Q are true.

We can see the point in another way by substituting ‘physical’ for ‘mental’ in the original argument. Again, imagine that you and I now have the same physical properties but different causal histories. Those different histories make it likely that we’ll be physically different. Other things being equal, physically different pasts tend to produce physically different states and events. The reason you and I are now physically different (in real life, I mean, not in my imagined scenario) is at least largely that we were physically different in the past. So (returning to the thought experiment) it’s likely, given our different pasts, that we’ll now be physically different, even though we’re in fact physically identical. But it doesn’t follow that we can be at once physically identical and physically different. This argument has precisely the same form as Vasilyev’s, yet has true premises and a false conclusion.

I may have misunderstood Vasilyev’s intent. Here is an argument superficially similar to the one I’ve presented that is logically valid:

1. It’s possible for two people to be physically identical despite having very different causal histories.
2. It’s necessary that people with very different causal histories will have different memories. Thus,
3. It’s possible for two people to be physically identical yet have different memories.

(Again, the modalities are nomological. Let’s not worry about what counts as ‘very’ different.) The first premise is the same as before. But the second says not just that people with different causal histories are likely to have different memories, but that they will have them. In fact they must, given the laws of nature. Imagine once more that you and I are now physically identical despite differences in our histories that are likely to give us different memories. That our memories are likely to differ does not imply that they will: unlikely things can happen. That was the problem with the first argument. But this time our memories are guaranteed by the laws of nature to differ. So you and I will now be mentally different but physically identical. The argument has the valid form: It’s possible that P; it’s necessary that Q; thus, it’s possible that P and Q. (If P is true in some world and Q is true in every world, there is a world where both are true.) Someone might take the original argument to be logically valid by confusing it with this one.

The trouble with the new argument is that the second premise is not compelling. It’s true that very different causal histories are likely to give people different memories. But it still seems possible, by an extraordinary coincidence, for their memories to be the same, contrary to the second premise. It seems possible for the same reason that people with very different causal histories can be physically identical. Our reasons for accepting the first premise tell against the second.

It may be that the argument Vasilyev had in mind is neither of these. In that case I hope my misunderstandings will help him to state it more clearly. I look forward to his reply.

I am grateful to Evgeny Loginov for very helpful discussions of an earlier version.
on a memory argument against local supervenience

Daniel Stoljar, Australian National University

Vadim Vasilyev has put before us an interesting argument against the local supervenience of phenomenal properties on physical properties based on considerations of memory. The basic idea is neatly summarized in this passage from his paper:

According to the first premise…my brain could have reached its current state by different trajectories. There is thus a real possibility that it and the organism of which it is a part could have had a different causal history—that is, a different past. But, according to the second premise, a different past tends to produce different memories. Therefore, while being physically the same, I could have had different phenomenal mental states. (Vasilyev, forthcoming, p. 4).

As I understand it, we may represent this argument as follows. Its first premise is:

Premise 1: It is possible for two agents who are intrinsic physical duplicates to have different causal histories.

To motivate this premise, consider Vadim’s brain and the organism of which it is a part, namely Vadim himself. Vadim is in some intrinsic physical state; that is, there is some total collection of intrinsic physical properties he has. Presumably, Vadim came to be in that state because of some history, perhaps a history leading all the way back to the big bang. Now consider a duplicate organism, Duplicate Vadim. Duplicate Vadim is physically intrinsically the same as Vadim, and also came to be in that state because of some history. Yet the causal history that produced Duplicate Vadim may well be different from the one that produced Vadim. Hence it is possible that different causal histories produce agents that are intrinsically the same.

The second premise of the argument is:

Premise 2: It is possible for two agents who have different causal histories to have different memories.

To motivate this premise, suppose that Vadim has an episodic memory of some event that happened 10 years ago, for example, stubbing his toe on the doorframe while leaving his apartment. It is plausible he has that memory because of his particular history. Had he had a different history, e.g., one which his toe missed the doorframe by a centimetre or two, he would not have had that particular memory. Likewise, consider someone who, like me, has a quite different history from Vadim. I have different memories from him, and the explanation is that I have a different history. Hence it is possible that two agents with different histories have different memories.

Vasilyev himself expresses this premise in slightly different language. He says that the episodic memories of agents ‘reflect’ their histories and also that different histories ‘tend to produce’ different memories (p. 4). These suggestions go a little beyond Premise 2 as I have stated it here, but not in ways that affect the assessment of the argument. After all, if different causal histories tend to produce different memories and if these memories reflect these different pasts, Premise 2 follows.

The conclusion of the argument is:

Conclusion: It is possible for two agents who are intrinsic physical duplicates to have different memories.

This conclusion is inconsistent with local supervenience. Local supervenience entails that, necessarily, if two agents are intrinsic physical duplicates, they are phenomenal duplicates. This in turn entails that it is not possible for two agents who are intrinsic physical duplicates to have different memories (at least if ‘memory’ here is understood, as Vasilyev does understand it, in phenomenal terms.) But the conclusion of the argument says directly that this is possible.

There are several interesting features of this argument. First, the modalities here — that is, expressions such as ‘possible’ and ‘necessary’ — do not need to be understood,
as they often are in philosophy of mind, as metaphysical or logical modalities, modalities in the widest sense. They can instead be understood as tacitly restricted to situations in which the empirical or contingent laws of nature obtain — ‘real’ possibilities, as Vasilyev says in his paper.

Second, because the modalities here are comparatively weak, the conclusion of the argument is correlative strong. This is because it rules out a version of local supervenience that is very widely held, and indeed is widely held to be non-controversial. The question of whether it is metaphysically possible for intrinsic physical duplicates to differ phenomenally is somewhat controversial; to assess it we need to consider remote possibilities. But if we restrict attention to cases in which the laws hold, it is extremely plausible to think that if agents have the same physical intrinsic properties, they will have the same phenomenal properties. Indeed, the practice of neuroscience and cognitive science, and the parts of such sciences that develop theories of phenomenal consciousness, arguably rest on this assumption.

Third, the premises of the argument taken individually seem extremely compelling, especially given that the modalities here are real rather than metaphysical. Surely it is a real possibility, a possibility consistent with the empirical laws, that duplicate physical agents have different histories. And surely it is a real possibility that agents with different histories have different memories.

Finally, despite the strength of the conclusion of the argument, Vasilyev makes a good case that it is not impossible to believe, and so that what we have here is an argument for a potentially acceptable conclusion, rather than a paradox. For one thing, one might give up local supervenience while accepting global supervenience. Global supervenience entails that any world that is a physical duplicate of the actual world is a phenomenal duplicate of that world; this is different from local supervenience in several ways, not the least of which is its focus on worlds rather than agents. Moreover, rejecting local supervenience does not commit you to epiphenomenalism. Indeed, as Vasilyev suggests, it may even be helpful in responding to positions of that sort.

However, while Vasilyev’s argument has these attractive properties, I doubt it is persuasive. One way to see this is to ask the following question about the two agents mentioned in the second premise: are they intrinsic physical duplicates or not? I think the argument is unsound whichever option you take here.

Clearly there are two possible answers to this question, which means there are two different versions of the premise.

**Premise 2.i**: It is possible for two agents, who (a) are intrinsic physical duplicates and (b) have different causal histories, to have different memories.

**Premise 2.ii**: It is possible for two agents, who (a) are not intrinsic physical duplicates and (b) have different causal histories, to have different memories.

These premises are different from each other. Premise 2.i is not going to be accepted by any proponent of local supervenience, since it directly entails that that thesis is false. That means that you can’t legitimately rely on this premise in mounting an argument against local supervenience. Premise 2.ii by contrast is certainly acceptable in the context. But when you combine it with Premise 1, the conclusion no longer follows. Either way therefore the argument is unpersuasive. If the second premise is understood as Premise 2.i, that premise is false, or at least begs the question against local supervenience. If the second premise is understood as Premise 2.ii, by contrast, the argument is invalid.

One may bring out the mistake in the argument in a different way by noting that in structure it is analogous to this argument about colour:

**Premise 1**: It is possible for two objects that are precisely the same shade of red to have different shapes.

**Premise 2**: It is possible for two objects that have different shapes to have different colours.

**Conclusion**: It is possible for two objects that are precisely the same shade of red to have different
colours.

The premises of this argument do not entail its conclusion. Suppose I have three objects before me: a red sphere, a red pyramid and a blue pyramid. The red sphere and the red pyramid are (we may suppose) precisely the same shade of red; this illustrates the truth of the first premise. The red sphere and the blue pyramid have different shapes and different colours; this illustrates the truth of the second premise. Nevertheless, there is no case in which two objects are of precisely the same shade of red but are of different colours. Hence the argument is invalid.

Is there a way to modify the colour argument so that it is valid? The obvious move is to interpret the second premise so that the two objects here are understood to be precisely the same shade of red. But in that case, the argument has a false second premise; indeed now the second premise is false for the same reasons that the conclusion is already false. In sum, the possible moves and counter-moves in relation to this argument are analogous to those in relation to Vasilyev’s argument.

I have suggested that his memory argument against local supervenience is unpersuasive, but Vasilyev himself considers a related objection and rejects it. Perhaps what he says there will apply here too? Let me end by looking briefly at this objection, which starts from the so-called identity theory, namely, the theory that phenomenal properties are strictly identical to intrinsic physical properties.

If the identity theory were true, it would immediately follow that the conclusion of Vasilyev’s argument is mistaken. After all, given the identity theory, if two agents had different memories they would have to have different physical properties. Moreover, if the identity theory is true, the second premise of the argument must be understood as entailing that the two agents mentioned in the premise are different physically, since they have different memories. And this in turn means that the argument itself is invalid.

In response to this objection, Vasilyev points out that the argument as he formulates it takes no stand either way on the identity theory. Nothing in either premise entails either that the identity theory is true or that it is false. I agree with him about this, and so I would agree that the objection to the argument founded on the identity theory is no good. But the objection I have offered here is quite different. It takes no stand on the identity theory either. It simply asks whether the two agents mentioned in premise 2 are assumed to be physical duplicates or not. If they are not, then the argument is invalid; but if they are, then the argument assumes the falsity of local supervenience in its second premise. Hence, whichever way you go, the argument is unpersuasive.

can local supervenience be saved?

Colin McGinn

I am happy to respond to Vadim Vasilyev’s interesting, clear, and well-argued paper. The question is whether the nature of memory undermines the claim that memory impressions are supervenient on current local brain conditions. Memories track the past in that they “tend to reflect their causal histories”: what a person remembers is conditioned by his or her past causal interactions with the environment (including the inner environment). But if that is so, they cannot depend completely on the contemporaneous state of the brain, since this state can be caused in multiple ways. Memory depends on past causal history not on current brain state, and the same effect can be brought about in many ways. Of course, the memory may well be supervenient on the current state of the brain and the prior state of the brain, so that no contravention of physicalism is to be entertained; but Vasilyev is concerned only with the thesis of current local supervenience — the dependence of the phenomenal on what is going on in the brain now.

I think the question turns on what it means to say that memories “tend to reflect their causal histories”. True, what we remember is a function of what we have experienced, but is every aspect of the cause reflected in the phenomenal content of the memory? Surely some aspects are not so reflected: I might seem to remember playing tennis yesterday because I actually did play tennis yesterday, or because I played tennis two days ago, or because I dreamt of playing tennis but didn’t play it, or because someone has given me a drug that fools me into thinking I played tennis then. I have the same memory impression in all these situations but the impression is caused in quite different ways. Why is this? It is because in each case my brain is in a similar state despite the variety of causes. The memory doesn’t reflect each of its possible causes: it has a constant phenomenal content. So in this sense memory doesn’t reflect its causal history; it is invariant under changes of causal history, presumably because of a constant contemporaneous brain state. Compare perceptual experience: the very same perceptual impression can be caused in many possible ways, failing to reflect its causal history. For example, an impression of a blue ball can be caused by an actual blue ball, or by a trick of the light, or by a blue photograph of a ball, or by a drug administered so as to create a hallucination of a blue ball. It is true that visual impressions tend to reflect their external cause, but in exceptional circumstances they can fail to do this — presumably because the same internal brain state is caused in different ways. So the defender of local supervenience will insist that the principle of reflection is only valid within a certain class of cases; it is perfectly possible for the principle to break down in abnormal cases. Thus in the cases I have described local supervenience is still being respected: there is no violation of it deriving from the general nature of memory. And the claim then is that in all cases of memory variation we must find a variation in internal local contemporaneous brain state. No counterexample has been produced to this claim once we understand correctly the limited way memory reflects causal history. There is a weak kind of causal dependence, but Vasilyev’s argument requires a very strong (and implausible) kind of causal dependence.

The same thing can be said of belief, pain, and other mental states. They all tend to reflect their causes in that different causes tend to produce different effects, but it is entirely possible to describe cases in which the same mental state can be produced by quite different causes. The way is then open for the local supervenience theorist to claim that in all cases there must be a common core of identical local brain states upon which the mental state supervenes, even though we can say that mental effects generally tend to reflect their causes. The problem for Vasilyev’s argument is that the reflection principle is not true in a version strong enough to deliver his conclusion: it is not true that every aspect of the cause must be reflected in the phenomenal content of the mental state. What the memory impression truly reflects is the current state of the brain not its contingent causal history; so long as you produce that brain state you will get the same memory impression, despite differences in how that may be done. For the local supervenience theorist memory
phenomenology does not reflect causal history, except in the commonsense sense that memories tend to track what actually happened—but not in the sense that allows differences of memories under identity of local brain states.

It is another question whether and why such a supervenience claim is true. A full-blooded Cartesian dualist will dispute it, holding that memory supervenes on the current state of one’s immaterial self-substance; and a behaviorist might argue that behavior is the correct supervenience base not the internal state of the brain. I think most people gravitate towards local cerebral supervenience so as to avoid dualism of this type (and possibly also behaviorism detached from brain state), as well as the empirical evidence that changing the brain is necessary and sufficient for changing a subject’s state of consciousness. Certainly we can agree that it is not a transparent intelligible necessity that consciousness is supervenient on current brain state—hence the feeling (endorsed by Kripke and others) that we can imagine the same brain state in the absence of its usual conscious correlate. I think there are real issues about the validity of the local supervenience claim, but I don’t see that it is brought into question merely by the fact that memories tend to reflect causal history. It all depends on how that principle is interpreted; and it needs a very strong—and clearly erroneous—interpretation if it is to deliver the result that Vasilyev intends. I certainly find it very hard to believe that a memory of a blue ball could be based on the same contemporaneous neural state as a memory of a red brick, or a memory of a tennis game have the same neural basis as a memory of a dinner party. We surely couldn’t have a case in which twin brains house completely different sets of memories—as it might be, memories of a life in Moscow and memories of a life in London. I find it credible that some aspects of memory content might be subject to externalist thought experiments (Twin Earth etc.), but I don’t find it credible that the phenomenology of memory might float free of internal brain states—any more than perceptual phenomenology might similarly float free. So I think Professor Vasilyev has made a valiant effort to refute local supervenience, but I remain unconvinced by his argument.
memories and brains: reply to vasilyev

Keith Frankish, University of Sheffield

Vadim Vasilyev argues for the radical conclusion that an organism’s mental states do not supervene even naturally on its physical states. It’s an interesting argument, and though I don’t believe it works, it highlights something important about the mind.

The argument has two premises. The first is that it is really (naturally, nomologically) possible for qualitatively identical events to have qualitatively different causes, and, more specifically, for two different organisms to reach the same total brain state via different causal paths. The second premise is that an organism’s episodic memory tends to reflect the organism’s past and, thus, the causal path to its present state. Combined, these claims suggest that it is really possible for two physically identical organisms to have different memories, from which it would follow that at least some mental states do not supervene on the physical state of the organism that possesses them. (Vasilyev allows that they may still supervene on global physical conditions.)

The first premise is probably true. It is certainly the case that two physical systems could arrive at the same configuration by different routes. (Think, for example, of two Rubik’s Cubes that have been solved in different ways.) It is, however, enormously unlikely that two brains would converge like this. A brain is an immensely complex system, composed of billions of functional components, whose total state evolves in hugely complex and context-dependent ways. Even if two brains started in exactly the same initial state, they would rapidly diverge as they encountered even slightly different stimuli, and the chances of their ever again converging on the same overall state would be infinitesimally small. Still, for the sake of argument, let us grant this premise.

This second premise is true: memory ‘tends to reflect’ the past (though we might question Vasilyev’s claim that it does so by producing images of it). However, this is not enough to support the conclusion. For it is compatible with memories being realized in physical states of the brain. Indeed, there is every reason to think that this is the case and that our memories tend to reflect the past because our brains tend to encode information about it. We know all too well that brain disease and damage can produce memory loss, and there is a large and rapidly growing body of neuroscientific work devoted to identifying the neural mechanisms involved in memory storage. So even if two brains did miraculously converge on the same overall physical state, there is no reason to think that they would retain different memories. All the evidence indicates that they would not, and that physical convergence would involve a parallel convergence in memory.

A little thought experiment may make the point more vivid. Imagine an artificial memory device which really does produce images of past events. It works like this. A tiny camera mounted next to your eye continually records snapshots of the scene in front of you and stores them on a high-capacity memory drive implanted in your skull. The device has a rapid search facility, which is speech-activated. If you say a date and a time, the device’s controller searches the drive for snapshots with that timestamp and projects them directly onto your retina from a tiny projector on the back of the camera. Such a device is feasible, and not far beyond the limits of current technology.

Now it is, I take it, uncontroversial that the visual memories stored in this way supervene on the device’s physical state. They are encoded in patterns of 1s and 0s on the drive, and we could in principle identify the precise location where each image was stored. Moreover, the kind of supervenience involved is not merely natural but logical; basic physical laws entail that a device with these physical properties stores images of this character.

Yet the premises of Vasilyev’s argument apply to this artificial memory just as much as to natural memory. It is possible (though extremely unlikely) that two artificial memory devices might arrive at the same physical configuration by different routes. And each device tends to reflect past events in the life of its user and thus the causal history that brought it to its current state. So, the case for thinking that the device’s memories are not
naturally supervenient on its physical state is every bit as strong as that for thinking that our memories are not. Indeed, if a person had grown up using such a device without ever learning how it worked, they might employ the same argument to infer that the visual images it stores do not supervene on physical states. Yet that is false. If the premises do not support the conclusion in this case, then they do not in the case of natural memory either.

So, I think Vasilyev’s argument fails. Still, it does tell us something important about the mind. It tells us that we do not conceive of memory as a physical process. Memories and other mental states do not present themselves as physical. From the first-person perspective, we are oblivious to their physical nature and find it conceivable that they might vary independently of the physical state of our brains.

Indeed, it may be that introspection positively obscures the physical nature of our mental states. Grant for the moment that our mental states are complex brain states. Then our awareness of our mental states must depend on evolved neural mechanisms of introspection, which monitor other brain systems. Such mechanisms will have been designed by natural selection for practical purposes — to provide information that is useful for social coordination and self-control. We don’t need to have accurate, detailed representations of the brain states involved, or even to know that they are brain states.

We just need a way of recognizing what kind of states they are and what significance they have, so that we can respond appropriately. (Or, more accurately, higher-level control systems in our brains need this. We are not something distinct from our brain systems.) As Daniel Dennett points out, we should expect introspection to present us with a radically simplified version of events — a user illusion, like the graphical interface on a computer, which creatively distorts internal reality in a way that facilities self-manipulation [Dennett 1991]. (Again, the user here is just another set of brain systems; the illusion is the brain’s interface on itself.) It is this user illusion, Dennett suggests, that creates our sense of the distinctness of mind and brain; the introspective perspective is so different from the third-person one that we cannot accept that it is a perspective on the same reality. (It might even be the case that evolution has actively shaped our brain processes to create the impression that our minds are not physical entities. Nicholas Humphrey has defended this view at length, arguing that it would be adaptive to have a belief in one’s own metaphysical specialness [Humphrey 2011].)

This is only a suggestion, of course, but it points to a different approach to the mind-body problem. As Vasilyev’s paper beautifully illustrates, we have strong intuitions about the distinctness of the mental and physical realms, which can be employed to construct an elegant argument for the failure of supervenience. It may be more profitable, however, to take a more detached attitude to our intuitions and focus on understanding how and why they arise.

REFERENCES


Daniel Dennett, Tufts University

Reading Vasilyev’s essay has been an enlightening exercise for me. It helped me recover something that has been bothering me. I knew from the outset that I disagreed immediately with his whole approach, which draws on Kim and others, and I had the sense that I had sufficiently supported my view somewhere in print, but where was the clearest expression? And then it hit me: it was in my 1978 essay “Where am I?” (in Brainstorms and reprinted in The Mind’s I and elsewhere).

The key sentence in Vasilyev’s essay that provoked my recollection was ‘By the phenomenal aspect of a mental state I understand that aspect that seems to be accessed [my emphasis] exclusively from the first-person perspective.’ My story is a close kin to Vasilyev’s twins thought experiments, and what it shows is that the ‘first-person perspective’ is not as secure a starting point as many have thought.

Recall that in my science fictional — but nomologically possible — tale of having my brain removed and put in a vat, and supplemented with a computer model that duplicated its control powers so that it was a synchronous ‘spare brain’, I could not tell which physical system was responsible for the control of my body at any moment. The switch had no labels. This was a way of drawing attention to the underprivileged access I have to the causes of my mental activities. It was also a way of forestalling Chalmers’ ‘hard problem’ and a host of other artifactual difficulties that have plagued philosophers for decades. My story was, of course, a dramatization of the implications of ‘functionalism’ and, while we’re labeling things, ‘illusionism’ and I don’t think anybody has ever shown my story to be incoherent or self-contradictory. For over forty years I have been bemused by the generations of philosophers who have trudged around inconclusively in the swamp of qualia and phenomenal properties. I invite philosophers to re-read my story and tell us where I went wrong. I think that details in Vasilyev’s discussion help to show that functionalism, not dualism, is the path to follow.
responses

Vadim V. Vasilyev, Lomonosov Moscow State University

In my responses to the objections, I will briefly summarize the positions of my critics and comment on them.

Eric Olson’s objection to my argument is that if it is taken as it was presented in my paper it is invalid, while if we try to modify it to overcome the issue with its validity it becomes unsound. The argument’s first premise states that it is possible for identical events to have qualitatively different causes; its second premise states that individuals’ phenomenal memories do reflect their pasts, i.e. their causal histories; the conclusion drawn from combining and clearing up these two premises is that physically identical individuals can have phenomenally different memories. Olson thinks that the argument would be valid if the second premise stated the necessity of individuals’ memories reflecting their causal histories. But it is evident that there is no such necessity. Our memories may very well be false. In its original form the argument also fails, according to Olson, because even if we grant that each of the two premises is possible individually, nothing guarantees that they are compossible. Olson illustrates this with some examples: for instance, it is possible for it to rain tomorrow in Moscow, and it is also possible for it not to rain tomorrow in Moscow. But it does not follow from this that is possible for it both to rain and not to rain tomorrow in Moscow.

I did consider this kind of objection in my paper. I admitted there that the argument might fail if granting both premises at once would lead to a contradiction, as in Olson’s example. And I tried to show that we have no reason to worry that it would. The general idea is the following one: if a state of affairs A is possible and a state of affairs B is possible, then the only case in which there is no possible world in which they coexist is the case in which A excludes B. If we can rule out the possibility that there is an inconsistency in A and B existing simultaneously, or at least show that such an inconsistency is highly unlikely, then the original argument will be valid and will allow us to draw a conclusion that it is possible (or that it is likely that it is possible) for A and B to coexist. And I do precisely that.

Daniel Stoljar makes several interesting observations regarding my argument that can help one to see it from different perspectives. His main objection is that its second premise, which he states as ‘it is possible for two agents who have different causal histories to have different memories’ (this wording seems to me perfectly acceptable) falls apart once we ask whether the two agents are physically identical. Two answers might be given to this question which will give us two new statements of the premise: (1) ‘It is possible for two agents, who (a) are intrinsic physical duplicates and (b) have different causal histories, to have different memories’ and (2) ‘It is possible for two agents, who (a) are not intrinsic physical duplicates and (b) have different causal histories, to have different memories’. The second version is useless for the argument at hand and the first version will of course be rejected by any proponent of local supervenience of the phenomenal on the physical, as it directly contradicts their position.

I agree that proponents of local supervenience would not accept this version of the premise. But this is not the version that is offered in my paper. This version is the conclusion to which I argue. And proponents of local supervenience can accept it as a conclusion — after which they of course would cease to be its proponents. The premise does not specify whether the agents are physically identical or not. The conclusion does.

Colin McGinn also thinks that the argument’s main problem has to do with its second premise, the claim that our memories reflect our causal histories. He does not dispute that some past events are in fact reflected in our memories. But couldn’t we induce false memories as well? And if we could, wouldn’t we need to bring the brain into precisely the same state in which it would be while having the genuine memories in order to do that? It seems that we have to answer both these questions in the affirmative. But then local supervenience of the phenomenal on the physical is left untouched by the argument.

I of course agree that phenomenal images that we call memories can reflect real events of the past as well as
not reflect them, being dependent on something else. It is in fact implied by the first premise of my argument, which allows for a variety of causes of qualitatively identical events. But for the argument to work it is enough to assume that memories can (in the sense of real, nomological possibility) reflect real events of our pasts, our causal histories. If a proponent of local supervenience would say that in order to generate images that are qualitatively identical to the ones actually reflecting past events, we need to perfectly reproduce the brain state of the individual whose memories are in question, he would simply assume that which he has to prove and against which I argue in my paper.

Keith Frankish offers several criticisms of my argument and believes it to be on the whole wrong although perhaps useful due to the fact that it shows rather well the force of our erroneous intuitions, which prompt us to distinguish the realms of the mental and the physical. His first remark, which he himself does not regard as crucial though, is that it is utterly unlikely for two brains to come into identical physical states by different causal paths. Looking closer at this remark we may notice that for Frankish this unlikelihood is due to a very low likelihood of two brains coming to be into identical states at all. I of course completely agree with this. But my argument does not require us to assume that it is highly likely. What it requires us to assume is that if two brains would come to be in identical physical states, it would be really possible for them to reach those states by different causal paths.

Frankish’s main objection to my argument reminds me of an objection that was previously made elsewhere by Dmitry Volkov. Frankish claims that my argument is applicable to electronic systems with artificial memory, in which case it would obviously be wrong. Indeed, we can assume that (1) two physical systems come into identical states by different causal paths. And we can also assume that (2) their artificial electronic memories reflect their causal history. But the conclusion that qualitatively identical physical systems may have different artificial electronic memories is false because these electronic memories are parts of these systems and the qualitative identity of these systems implies qualitative identity of all of their parts, including the memories.

I agree that the argument does not work if applied to purely physical systems. And I point out that fact in my paper myself, when I remark that the argument would lose its force if phenomenal states were identical with physical states of the brain. But if they are not or at least if we do not know that they are, it looks like the argument should work.

Daniel Dennett finds my scenarios very similar to scenarios of his own famous philosophical essay ‘Where am I?’. I fail to see the parallels, but I can say that I have always thought that the potential of this essay and its thought experiments is far from being exhausted. Separating the brain from where the sensory receptors are located is a very fruitful idea. One can develop it for example in the following way. Imagine that we connected to one of our eyes, ears, etc. one system of remote sensory receptors and a body to control and to the other eye, ear, etc. — another one. In this case we would feel ourselves to be in two different places and two different bodies simultaneously and could even run into ourselves in the street. What’s interesting is that in this sort of case we would have access from the first-person perspective to experiences of objects that are separate from ourselves, because each body can be considered as a location of ourselves and, consequently, from a point of view of another body, as separated from ourselves, at least in certain situations. This thought experiment may be a way to question a rather common claim that phenomenal states are essentially private. And this in turn may strengthen the position of proponents of the identity of the mental and the physical, which as we know is inconsistent with my argument. Although I do not think that it strengthens it to such a degree as to outweigh the reasons that we have not to accept the claim about such an identity.

In conclusion I would like to thank everyone who took part in this discussion. All the objections and the remarks were extremely interesting and helpful to me. I would also like to thank chief editor of Date Palm Compote Evgeny Loginov and his colleagues for the idea of this discussion and their work to carry it through. I hope that readers will find our debate interesting.