A reductive type physicalist account of psychology

Orly Shenker
The Hebrew University of Jerusalem
orly.shenker@mail.huji.ac.il

This is the text of a lecture given at the PSA 2016 meeting. For this reason it contains no references.

Consider a *token* event according to some theory of fundamental physics. For example, think of a *particular* event, at some *particular* place and time, in which (say) the fundamental physical structure (described, say, by quantum field theory) is such that certain atoms exist and interact to form certain molecules, which – in turn – interact to form the cells of a given piece of cork that floats in a give glass of water.

By the name "physicalism" I will call the idea that when this token obtains, it is everything that there is in the universe.

Now focus your attention on an aspect of this particular token, given by some partial description of it. This aspect exists as part of the *particular* token (together, of course, with other aspects of it). Additionally, the set of all the tokens that share this aspect, namely, all the tokens (that are counterfactual at that time and place) that form an equivalence set relative to it, form a type. For instance, two instances of the above floating cork (that may obtain at different moments), that differ only in the details of the arrangement of the field excitations described by quantum field theory, may form such a type.

And if, when a token obtains, it is – as we assume here – *everything* that there is in the universe, then these equivalence sets, these partial description sets, are the *only* sort of types that can exist as part of the physics of the world. I will call this idea "type identity physicalism". Maybe it isn't what other people have in mind when they use the term "type identity physicalism" – and in this case what I present here is a new thesis.

The account of the special sciences types, in this picture, is as follows. Consider two tokens, call them A and B, that share a special sciences kind P. Then, tokens A and B share a physical aspect, and this aspect is *identical* with the special sciences kind P. Let me emphasize: the physical aspect that A and B share doesn't merely give rise to the special science kind P: rather, it is the same as P, it *just is* P. This is what I mean by calling this a type *identity* theory.

When I am saying that this is the physicalist account of the special sciences kinds I am not saying that we know *which* are the shared physical aspects. Of course, we don't know that – the world is complex and science is young. But, what I'm saying is that the kinds couldn't be anything else, because there *isn't* anything else. There are only tokens, and aspect of tokens given by their partial description. This is what I mean by type identity physicalism.

Notice, that in such a type identity theory there are no levels of reality: instead of high level and low level, what we have are different descriptions of the state of the universe. Perhaps some of these descriptions are more coarse-grained and others are more fine grained descriptions of it (up to the level of the precise full description of the token). But there needn't be such a precise ordering relation between the sets of tokens that make up the kinds. And if there is only one level of reality, there is no room, and no need, for discussing inter-level relations. Notions such as realization, grounding, constitution, etc. ought (in this picture) to be understood as convenient ways of talking about relations between aspects of the single token (at each moment) of the universe, or between aspects of tokens at different moments; here causal relations may hold (in some concept of causation). The only relations that obtain, ontologically, are those of identity and causation. Identity means that the kind (or property etc.) that is discussed by the special sciences is nothing but an aspect of the physical fundamental state of the universe. Once the right physical aspect is identified, there is nothing more to say or explain concerning the special science kind: the identification of the physical aspect of the state of the universe, which just is the special sciences kinds, is its full explanation. What remains to be described is the structural and causal relations between the various aspects of the token.

Returning to the floating cork example, in this picture the cork is nothing but an assembly of atoms (or of field excitations) that interact in a certain way, and so are the water and the glass: all are nothing but aspects of the state of the world as described by the fundamental theories of physics. Both their making and their relations are fully described by these theories. In these terms one might say that "atoms float": floating is nothing but a certain interaction between certain atoms. Nothing more.

Here is an example of how this identity theory is applied in contemporary physics. According to each and every physical theory (such as classical mechanics or quantum mechanics) the world at any given time is in a well defined state, that is described, at least in general terms, by that theory, and the theory takes it to be the complete and exhaustive state of the world, without remainder. The instantaneous complete state according to a given theory is called a 'microscopic state' (or 'microstate'). In other contexts the term 'microscopic' sometimes means small, or part of a whole, but in physics the term 'microstate' denotes the complete state of the system.

But suppose that we wish to say something about the world, which – at least prima facie – seems not to be well captured by its physical description. For example suppose that we want to talk about biology, or about psychology. And suppose that we want our description to be compatible with physics. Given what we just said about the microstate being complete, the only thing which we can do which will be informative and nonetheless not repetitive is say less, that is give a partial description of the microstate, focus on an aspect of it.

Indeed, the special science of thermodynamics is understood in physics in this way. The complete mechanical microstate of a system is the complete list of all the positions and velocities of its particles, but the temperature of an ideal gas in equilibrium is identified with the average kinetic energy of the particles of the gas – which is an aspect of the complete microstate, given by a very partial description of it. In the jargon of physics, an aspect of the microstate is called a macrovariable. So the macrovariable of partial kinetic energy is temperature: it is not merely associated with temperature, it doesn't give rise to temperature, it doesn't ground temperature, and – most importantly – it doesn't realize temperature: it just is temperature.

This is one sort of special sciences kinds, and this is the type identity physicalist account of it. But there is another sort of special sciences kinds that we need to look into. Suppose that our tokens A and B do not share any relevant physical aspect, but they are nevertheless said to be two instances of the high level kind P. Of course, mathematically we can always find some aspect that they share, but suppose that none of these aspects is identified with the high level property P that they share. In this case we shall say that A and B are physically heterogeneous relative to P, and for short I will just say that they are physically heterogeneous.

According to type identity physicalism there is only one possible account of this case. There is another system, that interacts with the system of interest, and following this interaction it reaches the same state P (or a state with the same aspect P) in both cases. Notice that this is not a case of measurement in the sense that – by assumption – A and B do not share any aspect that can be associated with the state P of the device. In these cases if we look only at the system of interest, for example only at the tokens A and B, we may mistakenly conclude that the special sciences kind P is multiply realized by the physical kinds of A and B. But this is a mistake. It is a mistake because P is identical to a shared physical aspect, only this time it is not of A and B themselves but of another system, that interacted with them. So what seems, on first sight, to be a case of multiple realization, is actually a case of type identity.

Notice that if we do not have such a device, then in order to explain why A and B are P but C isn't we need to postulate some brute fact – and this is indeed a prevalent move in the literature. It is a bad move (and I will later say a little about why I think it is bad). But suppose that you want to make this "brute fact" move anyway. Then, the thing to point out is that this brute fact is not in the physics, which is – by our assumption – entirely in the tokens, and so it must be a non-physical fact. This non-physical fact makes token A a case of P and token B a case of P, and token C a case of non-P. It is, in this sense, in each and every token, and so this is a case of token dualism.

The conclusion is that if the appearance of multiple realization is explained by shared physical aspect of another system, then it isn't genuine multiple realization. And if it is genuine, since the realizing tokens are physically heterogeneous, then it is a form of

dualism (and I don't care if it is property dualism or substance dualism). And so, the so-called non reductive physicalism is an incoherent idea. There are only two coherent options: reductive type identity physicalism, and non-reductive token dualism.

Let me emphasise that I don't pretend to know which of them is true. Maybe type identity physicalism is true, maybe dualism is true. This is a question of fact. But what I would like to stress is the need to be clear about our ontological commitments: If type identity physicalism is not true, then the alternative is dualism (of some sort). So-called non-reductive physicalism is not an option, since it is incoherent. I will now focus on the type-identity option, just in order to examine some of its implications.

And I would now like to point out an interesting consequence of the understanding gained so far. And the consequence is that the special science of psychology is very different from the other special sciences like biology. And the difference has nothing to do with the hard problem or explanatory gap or things like that: the unique nature of psychology comes out even if we accept that each and every token is completely physical, and it is a result of the idea of type identity physicalism as presented so far.

And the problem is this. This latter type-identity account of the appearance of multiple realization, by brining in another device, is open for special sciences like thermodynamics and biology and economics, but not for psychology. And the reason is, that in psychology the tokens are physical states of the observer, and so bringing in the extra device means bringing in another observer, which is a case of the homunculus fallacy. Since the homunculus brings with it an infinite regress, it is unhelpful, and so we reject this option.

But then the only explanation for why A and B are P but C isn't is some "brute fact". And since by assumption all the physics is in the tokens, this additional fact is not physical.

Suppose now that we encounter another token D for the first time, and see that it is of the psychological kind P. For example, suppose that the world, including my brain, is in the physical token state D, and that I'm in pain. My mental state is of the kind P.

We have already seen that the fact that makes D a case of P is not in the physics of D; and since P is a mental types, there is no homunculus outside the physical tokens which unifies them under the type P. So in this particular token case, the only thing that makes me of kind P is the brute non physical fact. Either I interact with this non-physical fact; or I am that non-physical fact. In both cases, in this particular case, the non-physical fact obtains in addition to the physical token D. There are two sorts of facts in the world: physical ones and non-physical ones. And so this is a case of token dualism.

Let me emphasize that this dualism is not only about kinds, and is not only about sets, but is a fact about each and every individual token. Whether it is property dualism or substance dualism doesn't matter.

Tom Polger and Larry Shapiro, in their recent book "The Multiple Realization Book", argue that whether or not multiple realization holds in our world, is a question of fact. I agree with them. But if multiple realization holds in our world, then, dualism holds as well. And this if-then conclusion is not empirical, but analytical.

From what I said so far, the conclusion is that only two options are coherent: reductive type identity physicalism, and non reductive token dualism. Non-reductive physicalism is not coherent, and therefore it is not an option.

I would now like to expand on two attempts to salvage non reductive physicalism, that appear in contemporary literature (and that Meir mentioned briefly). One is functionalism, in all its versions, and the other uses the relation of supervenience. I start with functionalism.

The main idea of functionalism is that the tokens A, B, C, and D are characterized by their functional roles, such that A and B realize the same functional role and C realizes a different functional role. Can this explain why A and B are P but C isn't?

Now, I come from physics, and in physics there are no functions. When I look at a token of a function, all I see is a sequence of physical states. And when I look at tokens A and B as having functional roles, what I see are two states that partake in

two sequences of physical states. The claim is, of course, that the two sequences of states realize the same function. But my question is this: what fact makes it the case that these two physical sequences, in which A and B partake, are of the same functional kind?

You will notice easily, that this question reiterates the question we posed before, only this time it is about sets of sequences, instead of just about sets of states.

If the fact, that makes the two sequences of A and B be of the same functional kind, is in the physics of these sequences, then these sequences share some physical aspect, and we have a case of type identity physicalism. But this isn't what we want: we want functionalism to salvage non-reductive physicalism.

But to have non-reductive physicalism we need the sequences A and B to be physically heterogeneous. We need sequences A and B to not share any physical fact that will render them of the same functional kind. But if they are heterogeneous, then the fact that makes them of the same functional kind must be a non physical fact. And therefore functionalism is a form of token dualism.

And so, from the analytic point of view, functionalism is in exactly the same situation as when we thought of A, B, C and D as non-functional mental states. We remain with the two options: either reductive type identity physicalism, or non-reductive token dualism.

But I'm sure you have noticed that so far I ignored the concept that many take to be the essence of non reductive physicalism, namely, supervenience. So what about supervenience? Can supervenience salvage non-reductive physicalism?

Many contemporary thinkers take supervenience to be sufficient for physicalism, and it is for this reason that they take the so-called "non reductive physicalism" to be a form of physicalism and not – as I just suggested – a form of dualism. The idea is, that if supervenience obtains, then the physical tokens fix the mental kinds, or ground them, or gives rise to them, or something like that. What does this mean?

Let me distinguish between two senses in which a physical token, or even a physical kind, can fix the mental kind. One sense is formal. If I tell you, that A and B are P but C isn't, then from A you can deduce P, by modus ponens. In ontological terms, if it is a fact in the world, that A and B are P but C isn't, then the fact that A, fixes the fact that P; the world satisfies the modus ponens relation. But of course the thesis of supervenience is a metaphysical one, about the world. So noticing the modus ponens doesn't explain much. What I want to know is, what sort of fact makes it the case that these relations hold in our world. In virtue of what A and B are P but C isn't? And here, I am not satisfied by the answer that "God knows" what sort of fact it is. The "god knows" answer is an act of despair and a hindrance to research, and possibly even a fallacy of deriving impossibility from ignorance. So I want to press on: What sort of fact makes it the case that A and B are P but C isn't? Why, say, isn't it the case that A and C are P but B isn't?

In type identity physicalism the answer is clear: it is a shared physical aspect. But what happens in non-reductive physicalism?

Consider token D, which we have never encountered before. To borrow from David Lewis, D can be the state of affairs with respect to a normal person, or a Mad person, or a Martian, or a Mad Martian. And suppose that we want to predict whether this token D is going to be a case of P or not.

And suppose that in order to do that we are given the following information. First, we are given the complete physics of all the tokens, and, if you want, of the entire universe and its history. If you will, we become Laplacean demons. Since we are interested in a case of multiple realization, this complete physical picture reveals to us – by assumption – that the tokens A, B, C and D are heterogeneous: they don't share any relevant physical aspect. This assumption is essential: if the tokens are not heterogeneous we don't have multiple realization, but type identity. So it is crucial that we assume this heterogeneity.

In addition to the full physical information and the heterogeneity, we are told that A and B are P but C isn't. And finally, we are told that the mental supervenes on the physical.

However, all of this information will not help us one bit to determine whether D is P or not P. Since the tokens are heterogeneous, nothing in the physics of the universe determines their partition to sets, and the facts about the sets to which A, B and C belong is absolutely immaterial for the sets to which D belongs. And the formal relation of supervenience between these sets is, ipso facto, totally unhelpful.

So what sort of fact does fix D's type? Clearly, the additional fact is a non-physical "brute fact". When the particular token D obtains we interact with this non-physical brute fact, or we are that fact, and this is how our mental kind is fixed. If the extra brute fact is of the same kind that obtains when A and B obtain, then D is P; and if it the brute fact in D is different from the brute fact in A and B, then D isn't P.

What we have here is a form of token dualism. So we have here multiple realization together with supervenience, and nevertheless this is a case of dualism. And so supervenience is not sufficient for physicalism.

Of course, supervenience is necessary for physicalism in the uninteresting sense that type identity physicalism satisfies supervenience. But if you postulate physical tokens, and add multiple realization, then the tokens cannot fix the types, and you need to add some additional brute fact that will do that. But – and this is important – if a brute non-physical fact fixes the types, then this brute fact could also fix the types in such a way that supervenience would not obtain.

Indeed, the very fact that the brute non-physical fact is constrained to partition the tokens into types, in such a way, that supervenience will obtain, is a mystery in its own right, that is certainly not explained by the physics of the entire universe and its history. But since we are outside the realm of physics anyway, miracles are an option.

And so, even superveniece can't make a case of multiple realization physical, and therefore it can't salvage non reductive physicalism. There are only two coherent options: reductive type identity physicalism, or non reductive token dualism.

Let me end by returning to the theory of type identity physicalism, as I proposed to understand it. I proposed a type identity theory according to which mental types are identical to shared physical aspects of the physical tokens, given by partial descriptions of these tokens. By way of summing up let me illustrate the strength of this sort of identity theory.

Davidson, in his Mental Events, gave up supervenience of the mental on the physical in order to account for the alleged anomaly of the mental. Kim pointed out to Davidson that this is a problematic move, and Davidson made the slight concession by endorsing weak supervenience. I don't know if indeed the mental is anomalous. This is a question of fact. But what I want to show is that Davidson didn't have to give supervenience. He could attain the anomaly of the mental within type identity physicalism, properly understood.

The proof of the possibility of anomalous mental kinds in a type-physicalist framework requires some technicalities. I will not go into them, and just outline the idea, using some jargon of physics. Consider Figure 1. Here is the state space of a system that is prepared in some macrovariable, that is, in a microstate that belongs to the equivalence set of those that share this macrovariable. The grey rectangles at the bottom left hand side are the sets of microstates that share the prepared macrovariable. The other grey rectangles are the Poincare sections of the bundle of trajectories of the time evolution of that initial set of microstates, at several points of time. If we focus our attention on certain macrovariables, and partition the space into sets of microstates that share these macrovariables, we can have three cases. In the first case, there is a nice harmony between the evolution of the system and the partition to macrovariables, so that the evolution of the system, when described in terms of these macrovariables, appears deterministic. In the second case, the partition is finer, into smaller sets, and at the level of the macrovariables the evolution appears to satisfy a probabilistic law. In the third case there is no harmony at all between the evolution and the partition, and in the long run we cannot see any regularity at all in the evolution of the macrovariables. This case appears to be anomalous. The mental macrovariables could be of this kind. This is how even the anomaly of the mental can be give a reductive type identity physicalist account. This is anomalous physicalism.

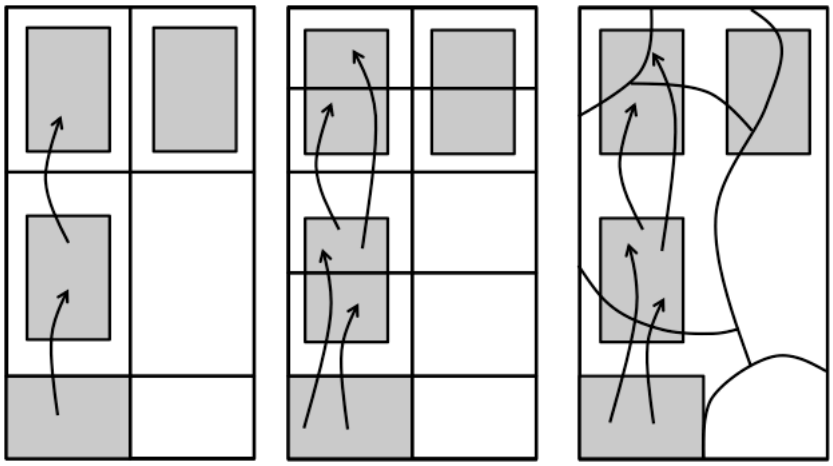

Figure 1

The account of the anomaly of the mental shows that in the type identity theory presented here there is an important sense in which the special sciences are autonomous. They are autonomous in that they discover which are the macrovariables, or the aspects of the physical tokens, that are significant and interesting and that exhibit regularity. Had we been Laplacean Demons, had we known the complete microstate of the universe, we could derive the special sciences kinds and laws. But the complexity of the universe is such that this is not possible, and it is reasonable to guess that it will remain so, and so we can only access these macrovariables via their appearance as high level properties. At the same time we must realize that at bottom they are nothing but aspects of the fundamental physical structure of the world. The special sciences are, fundamentally, branches of physics. This is the only coherent physicalist approach: any attempt to salvage the autonomy of the special sciences by opting for non-reductivism is either token dualism or simple incoherence.

Let me conclude. Whether or not multiple realization holds in our universe, and in particular with respect to mental types, is a question of fact. But if there is genuine psycho-physical multiple realization, either with or without supervenience, then there is genuine psycho-physical dualism: and this conclusion is analytical and not empirical. Many have hoped to escape this conclusion by way of the so-called non reductive physicalism, using ideas such as functionalism and supervenience. What I have shown today is that this is not an option. The only coherent options are non-reductive token dualism, or reductive type identity physicalism.