THE MIND-BODY WORLD-KNOT

Jesper Kallestrup

Here Kallestrup presents a succinct introduction to some of the latest thinking about the notorious mind-body problem.

The mind-body problem is about how to find a place for consciousness in a world that is entirely composed of aggregations of physical entities. We can take an entity to be physical just in case it counts as physical by best physical theory, where this is understood very broadly as comprising physics, chemistry, biochemistry, physiology, neurology, etc. This problem has seemed intractable for centuries. Schopenhauer (*The Fourfold Root of the Principle of Sufficient Reason*, (Illinois: Open Court Publishing 1847/1974): §42) famously called it a "world-knot". It consists of two problems: the first is about mental causation and the second is about phenomenal consciousness. We will present it as a Hegelian dialectical set-up: Section I defends the thesis that mental properties are identical to physical properties. Section II defends the anti-thesis that mental properties cannot be identical to physical properties. Finally Section III sketches some syntheses that resolve the conflict between the thesis and the anti-thesis – some ways of unsnarling the knot.

I

We are all familiar with the idea of one property causing another. You strike a bottle with a baseball bat. What caused the bottle to shatter? Not the colour or the price of the bat. These properties are causally irrelevant. If the bat had had a different colour or a different price, the bottle would still have Shattered. What matters is the mass or density of the bat. Its being made of metal rather than paper made the difference.

doi:10.1017/S1477175608000365
Think 21, Vol. 8 (Spring 2009)

Epiphenomenalism is the view that mental properties such as believing it's going to rain and desiring not to get wet never cause behavioural properties such as staying inside. Mental properties are caused by properties of the immediate physical environment. The rain hitting the window through which you are looking causes your belief that it rains outside. But this view has it that mental properties are in turn causally impotent. That's highly implausible by any reckoning. For instance, the causal explanation of my staying inside in terms of believing that it's raining and desiring not to get wet would be false were this view true. Or consider my judgement that I am in pain or I am having a visual experience of something yellow. If such mental properties were causally inefficacious, they would be explanatorily irrelevant to my judgements about them. Here's Jerry Fodor:

> . . .if it isn't literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my scratching, and my believing is causally responsible for my saying. . . if none of that is literally true, then practically everything I believe about anything is false and it's the end of the world.[1]

We better thus think of mental properties as having causal powers to bring about behavioural properties. In fact, mental properties also have causal powers to bring about other mental properties. Think of my pain causing not just withdrawal behaviour but also a desire for pain-relief. Circumstances may in actual fact never be suitable for the manifestation of these powers but that's irrelevant. A sugar cube remains water-soluble even if never immersed in water. Call this:

> *Mental Causation*: Mental properties cause behavioural properties

If epiphenomenalism is false, there's a compelling argument to the effect that mental properties are identical to

https://doi.org/10.1017/S1477175608000365 Published online by Cambridge University Press

physical properties of the brain. By 'identity' we mean numerical identity, just in the same way that Superman is identical to Clark Kent. There's only one man there. But before addressing that argument we need to reflect on two key principles. The first is called:

> *Completeness*: Every physical effect has a sufficient physical cause

What this means is that the physical world is causally closed. If you take any caused physical property and trace its causal ancestry or posterity that will never take you outside the physical domain. You will never need to appeal to mysterious alien properties or spurious spiritual forces. Contemporary physical science sustains *Completeness*; or so we are told. Here's an example. At the end of August 2005 an intense low pressure area was formed over warm ocean waters in South Eastern Bahamas. Water vapour evaporated from the ocean surface, and caused a tropical storm to become the hurricane Katrina. When it hit the American Gulf Coast it caused catastrophic flooding several kilometres inland. Consequently New Orleans is now steadily sinking into the mud of the Mississippi Delta causing economic recession. The second is called:

> *Exclusion*: No effect has more than one sufficient cause

If a property had two distinct sufficient causes, it would be causally overdetermined. And while there may be some rare cases of such overdetermination – think of two assassins independently and simultaneously shooting a convict – it is incredible that all mental causation is like that. For instance, if a property had two distinct sufficient causes, that property would also have two independent and complete causal explanations. But if one cause is capable of fully accounting for that property, there's no explanatory work left for the other putative cause to do.

https://doi.org/10.1017/S1477175608000365 Published online by Cambridge University Press

Imagine a detective telling you the following regarding the circumstances of Mrs X's death: Mr X had a motive, the opportunity was clearly there, the murder weapon has his finger prints all over it, Mr X's DNA was found on Mrs X's body, and no one but Mr X was anywhere near Mrs X when she was fatally attacked. We find little room for the thought that Mr Y could also be the murderer of Mrs X, having a different motive, using a different murder weapon, and so on.

Given these two principles, we can now mount an argument that mental property M is identical to physical property P. It's a proof by contradiction. First we assume for the sake of argument that M and P are distinct. Then we derive a contradiction. Contradictions are always false. So, given that our argument is valid, we conclude that our assumption is false. A valid argument is one that must have a true conclusion if all the premises are true. Note also that B in the argument is a behavioural property. Consider:

The Exclusion Argument

(1)	M and P are distinct properties	(Assumption)
(2)	M causes B	(Mental Causation)
(3)	B has a sufficient physical cause P	(Closure)
(4)	B isn't caused twice over by M and P	(Exclusion)
(5)	So, M and P are identical properties	

Here's an example. Neuroscientists tell us that pains and nociceptive-specific neuronal activity (NNA) are correlated. Whenever you are in pain your brain has this physical property. Now suppose for the sake of argument that these two properties are distinct. *Mental Causation* tells us that your pain causes your arm to withdraw, say from the burning candle. *Closure* then says that there must be a

https://doi.org/10.1017/S1477175608000365 Published online by Cambridge University Press

physical cause, NNA as it happens, sufficient to bring about that withdrawal behaviour. Remember, we are counting that behaviour as physical. But *Exclusion* then rules out that both pain and NNA could be causes of the movement of your arm. For these two properties are assumed to be distinct, and no effect has more than one sufficient cause.

Note finally that although we have taken M, P, and B to be properties, we could equally well think of them as substances. In that case, we would have an argument that mental substances are identical to physical substances. While a physical substance has at least one physical property, a mental substance is one with only mental properties. Think of ectoplasm or ghost-stuff. Descartes famously advocated substance dualism according to which each of us is a composite of two distinct such substances: a physical body located in space and our thinking mind located outside space. But he was well aware of the problem about mental causation facing this view. In an important exchange of letters with Descartes, Princess Elisabeth of Bohemia first put this objection by asking:

> '. . .how the human soul can determine the movement of the animal spirits in the body so as to perform voluntary acts—being as it is only a conscious substance'.[2]

In response (op. cit.) Descartes remarked that it's simply an empirical fact that mind and body do unite and interact, something that we learn from everyday experience, and he suggested that we have an innate idea that allows us to comprehend how they interact, and together constitute a unity. Not many contemporary philosophers subscribe to substance dualism, but some are convinced that property dualism is true. It's striking that any kind of dualism – any view according to which there are two essentially distinct kinds of things in the world – is targeted by the Exclusion Argument. Jaegwon Kim (*Mind in a Physical World* (1998), p. 38) calls it Descartes' revenge! He shows that it poses a challenge even for those

https://doi.org/10.1017/S1477175608000365 Published online by Cambridge University Press

property dualists who hold that mental properties are in some strong sense determined by physical properties.

II

So much for the thesis. Let's now turn to the anti-thesis, which negates the thesis. In *Meditation VI* Descartes offers a so-called conceivability argument for substance dualism:

> I know that whatever I clearly and distinctly understand can be made by God just as I understand it. [. . .] I have, on the one hand, a clear and distinct idea of myself taken simply as a conscious, not an extended, being; and, on the other hand, a distinct idea of body, taken simply as an extended, not a conscious, being; so it is certain that I am really distinct from my body, and could exist without it.

Note how Descartes relies on two principles:

> *Conceivability*: If something is clearly and distinctly conceivable, then God ensures that it is possibly the case

In other words if what seems to be the case meets a certain condition, namely that of being clear and distinct, then God vouches for its genuine possibility. The idea that suitably constrained conceivability is a reliable guide to possibility is also expressed by Hume (*A Treatise of Human Nature*: I. ii. 2) albeit without invoking divine intervention:

> Whatever the mind clearly conceives, includes the idea of possible existence, or in other words, that nothing we imagine is absolutely impossible.

The second is the logical principle called:

> *Leibniz's Law*: If two things are identical, then they have all the same properties.

https://doi.org/10.1017/S1477175608000365 Published online by Cambridge University Press

Thus if my mind is identical to my body, and my body has the property of being divisible, then so must my mind. Consequently, if my mind doesn't have that property, but my body does, then my mind is distinct from my body.

Is Descartes' argument cogent? Take the Superman story. Superman is Clark Kent although Lois Lane doesn't know (prior to their marriage) that Superman is identical to Clark Kent. Suppose she argues like Descartes: I have a clear and distinct idea of Superman as a flying hero, and a clear and distinct idea of Clark Kent as a non-flying non-hero. Whatever I can conceive clearly and distinctly, God can so create. So, Superman isn't identical to Clark Kent!

Something has gone wrong. But Leibniz's Law is impeccable. Or rather once we ensure that the relevant properties are genuinely individuating properties, applying Leibniz's Law cannot lead us astray. Suppose Lois Lane reasons as follows: I know that Superman fights for the American way. I don't know that Clark Kent fights for the American way. So, Superman and Clark Kent are distinct. What's wrong with this reasoning is that what Lois Lane knows or fails to know about Superman aren't properties that individuate Superman. They are rather properties that Lois Lane has. Contrast with the following: Superman fights for the American way, but Clark Kent doesn't, so Superman is distinct from Clark Kent. Now there's no fault with the reasoning, but the second premise is false. Clark Kent does fight for the American way.

But how can we trust *Conceivability*? Despite God's alleged omnipotence even He cannot separate what is identical to itself. If Superman is identical to Clark Kent, then Superman is necessarily identical to Clark Kent. There's just one man in the fable and he couldn't possibly be someone other than whom he is. Call this:

> *Identity*: If a and b are identical, then a and b are necessarily identical

But it also seems obvious that we can have a clear and distinct idea of Superman as a flying hero, and a clear and

https://doi.org/10.1017/S1477175608000365 Published online by Cambridge University Press

distinct idea of Clark Kent as a non-flying non-hero. So, it looks like something is amiss with *Conceivability.* Descartes himself invoked God to ensure that his Evil Demon wasn't deceiving him about its reliability. But the existence of God was supposed to follow in part from this principle, and so God cannot then in turn safeguard its reliability. This is the Cartesian circle: you can know that whatever is clear and distinct is true only if you first know that a non-deceiving God exists, but you can know that a non-deceiving God exists only if you first know that whatever is clear and distinct is true. Compare with: I can only get a scholarship if the University has already accepted me, but the University can only accept me if I already have a scholarship.

But the fact that *Conceivability* is flawed doesn't mean there's no other tight connection between what is conceivable and what is genuinely possible. In fact there better be such a connection – otherwise we would be cognitively screened off from the realm of such possibilities. Here's an example. Pigs don't actually fly but that's not necessarily so. We can imagine circumstances where both the gravitational force and the physical constitution of pigs are sufficiently different. There's nothing incoherent in that thought. After all Newton's law of gravitation already tells us that as an object moves away from the surface of the Earth, the gravitational force decreases. But then we would want to say that given the right changes in the laws of physics and matters of particular fact, pigs might have flown. That's a genuine way things might have been. Note also that we need to idealise on conceivability. Someone might think they can conceive of married bachelors. They can do so only if they aren't paying sufficient attention or don't fully master the concept of a bachelor. On this background, consider the following revision:

> *Conceivability**: If something is conceivable on ideal reflection, then it is possibly the case

Let's turn to psychophysical identities. There's a very strong case for the claim that we can ideally conceive of, say, pain

https://doi.org/10.1017/S1477175608000365 Published online by Cambridge University Press

without NNA and of NNA without pain. We need only envisage a zombie: a being that is physically just like us from the skin in but nevertheless lacks consciousness. There's nothing it is like to be a zombie. It's all dark inside. We can even imagine the zombie being a functional duplicate of us: she behaves in every way just like we do when we are in pain and so on. This sets our zombie aside from Hollywood 'zombies'. The bioengineered replicants in the science fiction film Blade Runner from 1982, for instance, are mere proximate duplicates of adult Humans, e.g. they lack certain emotional responses when subject to a Voight-Kampff test. If they were perfect functional duplicates, viewers would never be able to tell the difference! We can call this:

> *Zombie*: Psychophysical identities are conceivably false

Using Descartes-style reasoning we can now mount an argument to the effect that no mental property M is identical to a physical property P. Again, it's a proof by contradiction. First we assume for the sake of argument that M and P are identical. Then we derive a contradiction. So, given that our argument is valid, we conclude that our assumption is false. Consider:

The Conceivability Argument

(6)	M is P	(Assumption)
(7)	If M is P, then M is necessarily P	(Identity)
(8)	It's conceivable that M isn't P	(Zombie)
(9)	So, it's possible that M isn't P	(Conceivability)
(10)	So, M isn't P	

Let's spell out the step from (9) to (10). If it's possible that M isn't P, then it's not necessary that M is P. But (7) says that M is P only if M is necessarily P. So, if it's not necessary that M is P, then M isn't P. Take our test case. Suppose

https://doi.org/10.1017/S1477175608000365 Published online by Cambridge University Press

pain is NNA. If pain is NNA, then pain is necessarily NNA. There's only one property and it is necessarily self-identical. But you can conceive even on ideal reflection of NNA without pain. So, it's possible that NNA isn't pain. But then it's not necessary that pain is NNA. And so it's not the case that pain is NNA. This completes our anti-thesis.

III

We have so far canvassed two valid and *prima facie* plausible arguments. While the Exclusion Argument concludes that mental properties are identical with physical properties, the conclusion of the Conceivability Argument is just the opposite. Mental properties obviously cannot both be identical with and distinct from physical properties. So, given that our reasoning is faultless, at least one of the underlying principles has got to be flawed. Which one(s)? There has recently been much discussion about the plausibility of:

> *Completeness*: Every physical effect has a sufficient physical cause

and

> *Exclusion*: No effect has more than one sufficient cause

Some philosophers, e.g. McLaughlin[3], have argued that denying *Completeness* needn't entail commitment to any kind of mysterious alien properties or spurious spiritual forces. On this view, a mental property is best seen as an emergent property: a genuinely novel kind of property of a whole consisting of parts of an old kind that emerges, not because something from the outside is added, but when those parts are put together in the right kind of way. Crucially, the causal powers of an emergent property are

https://doi.org/10.1017/S1477175608000365 Published online by Cambridge University Press

irreducible to the causal powers of the lower-level properties on which it, in some sense, depends.

According to emergentism, our world is a layered world: there is a hierarchy of distinct yet connected levels starting from the physical level. Specific to each level, there are distinct kinds of substances wholly composed of kinds from lower-levels all the way down to elementary material particles. Each kind has specific properties in virtue of a characteristic organizational complexity, and some of these properties will have emergent causal powers. What is more, there are special emergent laws, neither reducible to, nor derivable from, lower-level laws, which attribute these causal powers to the types of properties in question.

The problem with this view is how upward determination from the physical to the mental can be combined with downward causation from the mental to the physical? My headache causes a desire for pain-relief. Presumably both my headache and my desire are determined by distinct physical states of my brain – call them P1 and P2 respectively. How can my headache act directly on P2 without there being some other neuro-physiological causal influences? It would seem more natural that P2 is sufficiently caused by P1. It's true that if I desire to relieve my headache, I have to act on P1. There is no direct way – via telepathy, telekinetics, or what have you – that I can ease my mental condition without intervening in my brain processes. That's why I take an aspirin. I know that in the right circumstances being an ingested aspirin is causally sufficient for relieving my discomfort. So my desire causes my behaviour in conjunction with my belief that aspirin normally has this effect. But what acts on P1 are properties of the aspirin rather than my desire for pain-relief and my belief that aspirin will do the trick.

Other philosophers have aired misgivings about *Exclusion*, which says that no effects can have two distinct sufficient causes. Or rather they have typically argued that a slightly different principle has counterexamples:

https://doi.org/10.1017/S1477175608000365 Published online by Cambridge University Press

> *Exclusion**: If a property F is causally sufficient for another property G, then no distinct property F* is causally relevant to G

Take Yablo's pigeon Sophie[4] who is trained to peck at red cards to the exclusion of cards of other colours such as blue or yellow. A red card is produced and Sophie pecks it. The question is: what caused Sophie to peck? The first answer that comes to mind is: the redness of the card. But of course no card is red without being red in a particular way, say, scarlet or crimson or some other shade of red. In this case, being scarlet is the specific way in which the card is red. Moreover, being scarlet is causally sufficient for Sophie's pecking. But this does not mean that the redness of the card is causally irrelevant to the pecking – as *Exclusion** would have it. The reason is that if Sophie had not been presented with a scarlet card, but with a crimson card, she would still have pecked. Sophie will peck as long as she is shown a card that is some shade of red. And that's why being red is a causally relevant property despite the fact that whatever shade of red is shown will be causally sufficient.

This may all be true, but it doesn't affect our argument. The Exclusion Argument rests on *Exclusion*, but not on *Exclusion**, and Sophie's pecking is no counterexample to *Exclusion*.

There's a lot more to say about both *Completeness* and *Exclusion*, but let me instead focus on the anti-thesis, and in particular on:

> *Conceivability**: If something is conceivable on ideal reflection, then it is possibly the case

which to my mind is the least plausible of all the principles. If we can successfully reject this principle, then we can block the inference from (8) to (9) in the Conceivability Argument. In order to reject this principle we need an account of why certain statements are impossible even

https://doi.org/10.1017/S1477175608000365 Published online by Cambridge University Press

though they are conceivable, and so in some sense seem possible. Think of the statement that pain isn't NNA. What's called for is some way of explaining away an appearance of possibility as a mere appearance.

At this point some philosophers, e.g. Tye[5], avail themselves of so-called phenomenal concepts. Phenomenal concepts are essentially distinct from physical concepts. The concept of NNA and the concept of pain refer to the same state, but while one can possess the former physical concept without having had an experience of pain, one can possess the latter phenomenal concept only if one has had such an experience. Phenomenal concepts pick out the characteristic phenomenal character associated with the relevant experience. We acquire the phenomenal concept of pain when we undergo an experience of pain, attend to its phenomenal character via introspection, and then form a conception of what it's like to have that experience. Alternatively, some prefer to build experience-dependence into the justification conditions of phenomenal concepts rather than their possession conditions. This means that we are justified in applying such concepts only if we are either undergoing a phenomenal experience of the right kind, or alternatively is recreating it in imagination. Either way, the main point is that there is no corresponding experience-dependence built into any conditions of physical concepts. One can thus justifiably apply a physical concept without having or imaginatively recreating any particular experience.

On this background consider now our psychophysical identity:

11: Pain is NNA

where 'Pain' is assumed to express the phenomenal concept of pain, and 'NNA' expresses the physical concept of NNA. Now we can explain why *11* gives rise to an impression of possible falsity without actually being possibly false. To entertain *11* means to deploy a phenomenal concept, hence it involves a version of the

https://doi.org/10.1017/S1477175608000365 Published online by Cambridge University Press

experience. But it also means to exercise a physical concept, which involves no such experience. Given that the phenomenal concept includes something experiential, which the physical concept leaves out, entertaining *11* makes us wonder why this particular mental state is identical to that particular physical state, rather than some other physical state, or maybe no physical state at all. That is, *11* gives rise to an impression of possible falsity. But this impression is misleading. For the physical concept may fail to activate the experience and yet refer to it. Just as the concept of pain refers to the phenomenal character associated with experiences of pain, so does the concept of NNA. The same concept cannot have two distinct referents, but two distinct concepts can have the same referent. Think of the concept of Superman and the concept of Clark Kent as both referring to the same man, namely Superman.

Note finally certain limitations on this strategy of explaining away appearances of possibility. It seems to Lois Lane that Superman might not have been identical to Clark Kent. What then is she conceiving if not that that man might not have been self-identical? In this case there are no phenomenal concepts in play. One promising answer is to say that Lois Lane identifies the same man in two distinct ways, and those ways might have identified two distinct men. To wit, she conceives of circumstances in which the invulnerable superhero that flies around is distinct from the shy reporter that works for the Daily Planet. Lois Lane confuses this genuine possibility for the merely seeming possibility that Superman isn't Clark Kent. In this case a seeming possibility is explained away by another genuine possibility, but in *11* the phenomenal concept strategy allows us to explain away a seeming possibility without having to invoke some other genuine possibility.

Jesper Kallestrup is Lecturer and Senior Tutor in Philosophy at the University of Edinburgh.

https://doi.org/10.1017/S1477175608000365 Published online by Cambridge University Press

Notes

[1] J. Fodor, 'Making Mind Matter More', in his *A Theory of Content and Other Essays*, (Cambridge, Mass.: MIT Press, 1990), pp. 137-159.

[2] E. Anscombe, E. and P. Geach, (trans.), *Descartes' Philosophical Writings*, (London: Thomas Nelson and Sons Ltd., 1954). pp. 274.

[3] B. McLaughlin, 'The Rise and Fall of British Emergentism', in A. Beckerman, H. Flohr, and J. Kim, (eds.), *Emergentism and Reduction*, (New York and Berlin: De Gruyter, 1992).

[4] S. Yablo, 'Mental Causation', *The Philosophical Review*, Vol. 101, 1992. pp. 245-280.

[5] M. Tye, *Consciousness, Color, and Content* (Cambridge, Mass.: MIT press, 2000).

https://doi.org/10.1017/S1477175608000365 Published online by Cambridge University Press