

Leibniz and the Problem of Consciousness

Tim Crane

What would Gottfried Wilhelm Leibniz have said about today's problem of consciousness? Some philosophers claim that Leibniz was one of the first to argue that there is an 'explanatory gap' between our knowledge of matter and our knowledge of consciousness, and that he thought this posed a problem for materialism (see for example Churchland 1995: 191-2; Kriegel 2015: 49; Seager 1991; Searle 1983: 267-8). This is supposed to be the point of the famous passage in the *Monadology* (1714), in which Leibniz argues that perception is 'inexplicable in terms of mechanical reasons':

If we imagine that there is a machine whose structure makes it think, sense, and have perceptions, we could conceive it enlarged, keeping the same proportions, so that we could enter into it, as one enters into a mill. Assuming that, when inspecting its interior, we will only find parts that push one another, and we will never find anything to explain a perception.

(*Monadology* §17)

It seems as if Leibniz is arguing that no matter what we know about the material structure of an object, we will never be able to explain why it is conscious; and this 'Mill Argument', as I will call it, therefore seems like an early version of Joseph Levine's (1983) explanatory gap argument.

Despite its fame, Leibniz's Mill Argument does not seem to have persuaded many philosophers of the falsity of materialism (*pace* Landesman 2011). It's not hard to see why. The argument asks us to imagine being small enough to enter into a machine that is supposed to genuinely perceive. Leibniz then says that all we would find would be parts that 'push one another' and nothing that would explain perception. But the hypothesis he is attacking is precisely that perception is a mechanical phenomenon: on this hypothesis, 'parts that push one another' just is what explains perception. So, whatever the truth of this mechanist hypothesis, Leibniz's argument looks as ques-

tion-begging as it could get. Searle brings this out with the following analogy: if you were shrunk so small that you could walk around the parts of an H₂O molecule, you would never find anything that would explain the wetness of water. How can this show that the wetness of water is not explained by its molecular structure?

Here I do not want to defend Leibniz's argument; but I want to cast doubt on whether it is an early version of the explanatory gap argument about consciousness. There are two simple reasons for this doubt: first, Leibniz's concern is with what he calls 'perception', and this is not the same as consciousness; and second, the immediate target of the argument is not materialism, but the view that perception can be a property of a complex or composite thing, rather than a simple substance. Here I will explain why the argument should be understood in this way, and I will conclude by drawing two lessons about how to read the philosophers of the past, one general and one more specific to the study of consciousness.

First, let's look at the main aspects of Leibniz's mature metaphysical system, and then we can introduce his concept of perception. According to Leibniz, the world is made up of an infinity of simple substances ('monads') which 'express' more or less clearly everything else that happens, or appears to happen, in the world. The world we experience is a world of *phenomena* (or appearances). This is not to say that the world consists of mental entities, such as 'ideas' as Berkeley conceives of them. Rather, the best way to think of phenomena in Leibniz's sense is as the *intentional objects* of their expression by monads (see Adams 1994: 219; Furth 1967: 172; Calabi 2004). If we think of expression as a kind of primitive and ubiquitous representation, then we can say that the world of phenomena is what is represented by monads, not the representing itself.

All monads express phenomena, and they all express them with a coherence or 'harmony' which is established by God. This is where the appearance of causal relationships in the world comes from: phenomena appear to causally interact, and we can give mechanical explanations of these interactions. But causation is not part of the true reality of the world, merely itself another phenomenon which is a consequence of the pre-established harmony. Note that the pre-established harmony is not the kind of 'parallelism' (Taylor 1974: 18), between mental and material sub-

stances sometimes attributed to Leibniz. There are no material substances on Leibniz's view, for reasons I will explain shortly.

Perception is defined in *Monadology* §14 as 'the passing state which involves and represents a multitude in the unity of in the simple substance'. Perception is the representation of the multitude by a unity (i.e. by the monad itself). It is important to recognise that Leibniz does not mean what we mean these days by perception — the apprehension of the external world by the senses — rather, he means something closer to what today we call thought, mental representation or intentionality. And, to complicate matters, he does not mean what Descartes and the Cartesians meant by 'thought' either; for they took thought to be essentially conscious, and Leibniz denies this. He says that perception 'should be distinguished from apperception or consciousness' and that 'this is where the Cartesians have failed badly, since they took no account of the perceptions that we do not apperceive' (*Monadology* §14). In the *Principles of Nature and of Grace* (1714), Leibniz defines apperception as 'consciousness, or the reflective knowledge of this internal state'.

Leibniz gives a number of arguments for the view that some perception is non-conscious. In *Monadology* §23, for example, he claims that when we wake from an unconscious 'stupor', we become conscious of our perceptions. But since 'a perception can only come naturally from another perception', we must have had some perceptions before the moment of awakening; but these perceptions are *ex hypothesi* unconscious. So there must be unconscious perceptions: the Cartesians are wrong.

The important underlying principle here, again, is that there are no casual relationships between phenomena and monads, only pre-established harmony. This is why perceptions can only come from other perceptions. Monads can change, but change (e.g. in perceptions) comes about because of an 'internal principle' in each monad, not because of external causes. This activity of this internal principle is what he calls *appetition* (§15), which he elsewhere defines as 'tendencies from one perception to another' (*Principles of Nature and Grace* §2). For present purposes, appetitions can be considered to be simple appetites or drives. Monads therefore can be characterised — and distinguished from one another — in terms of their perceptions and appetitions. In a letter to Burchard de Volder Leibniz wrote that 'there is nothing in the world except simple substances, and,

in them, perception and appetite' (letter to de Volder 30 June 1704). However, although all souls are monads, not all monads are souls (§19) — since souls have memories too, and their perceptions are more distinct than those of other monads. But the distinctness of a perception does not imply that it is conscious.

My first point about the common reading of the Mill Argument, then, is that it is not about consciousness, but about perception in Leibniz's distinctive sense. Perhaps it would be less misleading these days if Leibniz's 'perception' were called thought or intentionality. But this was the word he used (the French *perception*) so we must work with this. The explanatory gap argument, however, is explicitly about the problem that *consciousness* poses for materialism, not intentionality generally (Levine 1983). So the Mill Argument is not the explanatory gap argument.

This brings me to my second point, about materialism. Despite what I have just said, it might be thought that Leibniz's target is still materialism, even if the problem for this doctrine is not consciousness. After all, Leibniz was a well-known anti-materialist: he discussed whether matter can think at great length in, for example, the *New Essays*, his critique of Locke. In the Mill Argument he explicitly says that perception cannot have a mechanical explanation, and in Leibniz's day, all materialist explanations were considered to be mechanical explanations (i.e. explanations in terms of matter in motion). So if the Mill Argument were an argument that perception (or thought, or intentionality) cannot be given a mechanical explanation, this would imply that it cannot be given a materialist explanation either (see Duncan 2012 for this point). Isn't the argument best seen as directed against materialism?

This is an initially persuasive line of thought. But we should pause before accepting it, not least because the term 'materialism' (French: *matérialisme*) does not occur in the *Monadology*. Nor is there any mention of the doctrine under any other name; indeed, matter itself is barely mentioned in this work. If Leibniz's aim was to attack materialism, why didn't he mention the doctrine? This consideration is not absolutely decisive, of course. It may be that Leibniz had some reason for not mentioning the doctrine of materialism even though he is arguing against it. But we would want some reason why this is so, given that Leibniz did not hold back from criticising materialism in other works, and in general he did not hold back in identifying his opponents (Locke, Descartes, Bayle

etc.). Since, as far as I know, there is no such reason, we should take the fact that Leibniz does not mention materialism in the *Monadology* at face value, as an indication that this was not his real concern.

So what was his real concern? Leibniz's worldview is without question very strange, and very remote from many of the metaphysical systems we are familiar with today. However, it is important to acknowledge that its source is not something arbitrary or mysterious, but derives ultimately from the perfectly intelligible conviction (present throughout Leibniz's extant writings) that the fundamental reality of the world must be something unified. This is the truth that he identified in the Aristotelian tradition of theorising about substance, according to which substances are the natural unities, and therefore the only ultimately real beings. Genuine realities must be genuine unities: or as Leibniz himself put it, 'what is not truly *one* being is not truly one *being* either' (letter to Arnauld 30 April 1687).

Aristotle's view was that the primary cases of substances are organisms. Organisms are natural unities, and they are also complex things, with many working parts. But at least by the time he wrote the *Monadology*, Leibniz had rejected this Aristotelian view. He thought of unity in terms of simplicity, and simplicity in terms of lacking parts. That 'simple' means 'without parts' is stated in §1 of the *Monadology*. He then goes on to say that 'there must be simple substances, because there are composites' (§2) and that where simple things are concerned, there can be no divisibility, no shape, no extension; simple substances cannot come into being 'by composition' and they cannot 'perish naturally' (§§3-5). Simple substances (monads) can only come into being by creation. And nor can they be changed by anything but themselves — as we saw, there can be no causal influences on monads.

And yet although they are simple, monads must differ from one another — otherwise there would be no reason to say there is more than one monad. This is a consequence of Leibniz's principle of the identity of indiscernibles: that objects which share all their properties are identical. He does not argue for this at this stage in the *Monadology*, but assumes it (§9). The fact that there is more than one monad follows from the existence of composites (assumed in §2) because compos-

ites must be made of more than one simple. So monads must differ from one another in their properties.

What properties do monads have? Leibniz's answer, as we saw above, is perception and appetite. His route to this conclusion moves first from the claim that there is real change in created substances (§10) to the claim that there is diversity in these substances themselves (§13), even though they have no parts. This is because in any change, something changes and something remains: hence there must be a plurality in things. The next section (§14) involves a jump: the plurality is *represented* in the simple substances, and this representation is (by definition) perception. But why should we conclude that the plurality is *represented*, and not *instantiated* in these substances? The best answer I can think of on Leibniz's behalf is that if there were actually something changing and something remaining in a monad, this would have to come about through some kind of 'internal motion' which would bring about the 'change among its parts', as in a composite thing (see §7). But monads have no parts. So the only way they can involve change or a plurality (which he also calls a 'variety' or a 'multitude') is by representing it. Leibniz says that we can recognise this multitude in 'the least thought we ourselves apperceive' and that therefore 'all those who recognise that the soul is a simple substance should recognise this multitude in the monad' (§16).

It is natural to object again that a multitude *represented* in the monad is not the same as the multitude actually *being* in the monad. (A tree being represented by the mind is not the same as the tree actually being in the mind.) The best way to understand Leibniz's position is to return to the point outlined earlier — that everything we encounter in the world of experience, including material objects with their apparent changes and causal relations, are phenomena, or intentional objects. So the multitude which we experience must ultimately be a matter how things are represented by the mind: this is what Leibniz calls perception, the representation of a multitude by a unity.

This brings us, finally, to target of the Mill Argument in §17. Up to this point in the *Monadology*, Leibniz has argued that there are simple substances (monads), that they do not causally interact, but they are distinguished from one another by their perceptions and appetitions. Leibniz now says that perception cannot be explained in mechanical terms ('that is, through shapes and

motions'). Having already said that perception is a property of monads, he now wants to rule out that it can be a property of composite things. For if it were explicable in mechanical terms, then it would be a property of a composite, since all machines are composite. If perception is not a property of a composite, then it would follow that it cannot be mechanically explained. But as Paul Lodge and Marc Bobro (1998) have argued, the claim that composites cannot perceive follows from the characterisation of perception as something that only a unity can have, plus the fact that no composite is (*contra* Aristotle) a unity. The point of the mill thought experiment, then, is simply to illustrate that perception cannot be a property of a complex object. Here Leibniz is merely drawing out the consequences of his conception of perception; or as Lodge and Bobro put it, 'Leibniz is simply reminding us of what we already know' (1998: 566). This is why he concludes:

And so, we should seek perception in the simple substance and not in the composite or in the machine. Furthermore, this is all one can find in the simple substance — that is, perceptions and their changes. It is also in this alone that all the *internal actions* of simple substances can consist. (*Monadology* §17)

Before §17, Leibniz took himself to have established that perception is a property of monads, which are the only unities. He then shows in §17 why it follows from this view that it cannot be a property of composites, and therefore has no mechanical explanation. This is the target of the Mill Argument.

I am not saying the argument is good or bad, only that it is not an argument against materialism. It may have the consequence that materialism is false — as noted above, this would follow if materialism requires that perception be explained, and that all explanation is mechanical explanation. And it is also true that Leibniz himself rejected materialism. But this does not mean that the aim of the Mill Argument is to refute materialism.

Why does this matter? I think there are two general lessons one can learn here. The first is methodological: when interpreting a philosopher, it is important not just to understand which beliefs they hold, which propositions they think are true. It is equally important to understand what is cen-

tral to their view, and what is peripheral. This is particularly important when reading philosophers who are remote from us, temporally and culturally. Which philosophical questions are important changes with time, and we should not assume just because a philosopher seems to be talking about a question that concerns us, that this is what they are really doing. In the case of Leibniz, some of the aspects of his metaphysics are familiar to us (anti-materialism), but others are very remote from things we might think today (substances as simples). It is unlikely, therefore, that someone with the worldview described here would be asking exactly the same questions as the materialist or anti-materialist philosophers of today.

This brings me to my final point. Leibniz was certainly interested in consciousness, and his distinction between perception and apperception is very important in his rejection of Cartesianism. But was he interested in the problem of consciousness as we find it today? If the problem of consciousness is — as is often claimed by contemporary philosophers — the problem of how consciousness can exist in a fundamentally material or physical world, then it seems that the problem simply cannot arise for Leibniz. The fundamental entities in the world for Leibniz are monads whose properties are perception and appetite. Monads are not material objects, and perception and appetite are not material properties: the world is not fundamentally material. If an idealist is someone who thinks that matter is a mere appearance (as Leibniz himself does), and that the fundamental beings are substances with something like intentional and motivational states, then it is easy to see why Leibniz is often classified as an idealist (see Adams 1994). But even if he is not straightforwardly an idealist (see Garber 2009), Leibniz's view that monads are the fundamental beings is certainly incompatible with a materialist view of the world. Does this mean that there is no problem of consciousness for a philosopher like Leibniz? Surely not; so it cannot be that the only thing worth calling the 'problem of consciousness' is the question of how it fits into a fundamentally material world.

REFERENCES

All quotations from Leibniz are from *G.W. Leibniz: Philosophical Essays*, edited and translated by Roger Ariew and Daniel Garber (Indianapolis: Hackett 1989).

Adams, Robert M. (1994) *Leibniz: Determinist, Theist, Idealist* (Oxford: Oxford University Press)

Calabi, Clotilde (2004) 'Are Monads Intentional Systems?' in M.Carrara et al. (eds.) *Individuals, Minds and Bodies: Themes from Leibniz* (Stuttgart: Franz Steiner Verlag) 159-72

Churchland, Paul (1995) *The Engine of Reason, the Seat of the Soul* (Cambridge, MA: MIT Press)

Duncan, Stewart (2012) 'Leibniz's Mill Arguments against Materialism' *Philosophical Quarterly* 62: 250-72

Garber, Daniel (2009) *Leibniz: Body, Substance, Monad* (Oxford: Oxford University Press)

Kriegel, Uriah (2015) *The Varieties of Consciousness* (New York: Oxford University Press)

Landesman, Charles (2011) *Leibniz's Mill: a Challenge to Materialism* (South Bend, IN: University of Notre Dame Press)

Levine, Joseph (1983) 'Materialism and Qualia: The Explanatory Gap' *Pacific Philosophical Quarterly* 64: 354–361

Lodge, Paul and Bobro, Marc (1998) 'Stepping back inside Leibniz's Mill' *Monist* 81: 554–73

Seager, William (1991) 'The Worm in the Cheese: Leibniz, Consciousness, and Matter' *Studia Leibnitiana* 23: 79–91

Searle, John (1983) *Intentionality* (Cambridge: Cambridge University Press)

Taylor, Richard (1974) *Metaphysics* (Englewood Cliffs NJ: Prentice Hall)