Leibniz's mill argument doesn't favour dualism over physicalism

We now turn to our second argument for dualism, a very popular one:

> Physical things are incapable of thought and sensation. But human persons are capable of thought and sensation. Therefore, human persons are not physical things.

But why should we believe that physical things are incapable of thought and sensation? I am willing to grant that if we try seriously and in detail to imagine a physical thing having thoughts and sensations, we can find this notion—the notion of a physical thing having thoughts and sensations—very puzzling. There is a famous passage in Leibniz's *Monadology* that very clearly brings out the puzzling aspects of this notion:

> Furthermore, we must admit that *perception*, and whatever depends on it, *cannot be explained on mechanical principles*, i.e. by shapes and movements. If we pretend that there is a machine whose structure makes it think, sense and have perception, then we can conceive it enlarged, but keeping to the same proportions, so that we might go inside it as into a mill. Suppose that we do: then if we inspect the interior we shall find there nothing but parts which push one another, and never anything which could explain a perception. Thus, perception must be sought in simple substance, not in what is composite or in machines.6

To take a more modern example, suppose someone were to claim to have programmed a computer so that it could think (in a sense that implies conscious experience and self-awareness) or to have constructed a thinking robot. If the computer or robot were enlarged so that people could walk about inside it, a party of tourists being led through the vast machine would see nothing but physical things interacting physically. And this would be no illusion. It's not as if the thought and conscious

experience were hidden away in some part of the machine off limits to visitors.

But then where *are* the thoughts and the experience? Where *could* they be? How could the mere physical interaction of bits of metal and plastic and silicon "add up to" thoughts and experience? It is important to realize that this point has nothing to do with the specific kinds of physical material a computer or robot would be likely to be made of. The point has to do only with the fact that the materials are *physical*. The point would be unchanged if we imagined a party of tourists being conducted through *ourselves* (or our bodies), as in Isaac Asimov's interesting science-fiction novel *Fantastic Voyage* (or the unspeakably silly movie of the same title). If we could be greatly reduced in size and go inside a functioning human brain and have a look round, we should see no thoughts or experience, not even if we saw everything there was to see. If God looks inside a human brain, even He sees nothing but unthinking physical things like neurons and Nissl granules and amino-acid molecules and electrons in continuous mutual physical interaction. Where, then, are the thoughts? Where are the sudden feelings of elation or despair? Where are the sensations of heat and pain and pressure and color? The answer is, obviously, that they are elsewhere. And that "elsewhere" must be a place that is receptive to the presence of such things, a place where they *could* exist. They must exist in a non-physical thing. (If we like, we can say that they must exist in a non-physical thing that is *mental*: a mind or a soul. But unless we can say something useful about what we mean by 'mental thing' or 'mind' or 'soul', to say this would be to say no more than that they must exist in a non-physical thing.)

Various physicalists—who must of course believe that physical things are capable of thought and sensation—will reply to this argument in various ways. What follows is my own reply. Some physicalists would reject some parts of it.

Let us begin with the question, *Where* are the thoughts and sensations? The answer is that since these things are changes in the cerebral cortex, they are all around you (you who have in imagination been reduced in size and are physically inside someone's brain). It does not follow from

this that you *see* them, since they may involve the whole cerebral cortex or the whole brain or widely scattered parts of the brain: it may be that you cannot see them for the same reason you cannot see the event called 'the election' on election day. But let us suppose for the sake of argument that these events are sufficiently localized that you can see them. (Or some aspects of them: a human being cannot see every aspect of any event. You can see the street lamps come on in your neighborhood, but you cannot see the flow of electrons that is an indispensable component of this event.) Of course these events do not *look* to you like mental events, but then what would you expect a mental event to look like? ("Well, something like the way mental changes in *myself* look to *me*, as when I experience a sharp pain in my left shoulder or a thrill of fear or an intellectual insight." But that's what it's like to experience *having* or *being the subject of* a mental change. That's what a mental change *in you* "looks like" to you. What would you expect mental changes in someone *else* to look like to you?) And anyway, a change may be of a certain type without its being evident that it is of that type. Suppose a computer has been programmed to compute the orbit of a certain satellite. Suppose the computer were greatly enlarged and that you went inside it, "as into a mill." You would not see any orbital computations going on—or at least you would not see anything that "looked like" orbital computations. (What would you expect orbital computations to look like?) The Leibnizian thought-experiment, therefore, should cause the physicalist no unease. Things inside the brain look just the way they would look if physicalism were correct.

Many physicalists would think that this was a sufficient reply to the charge that the notion of a physical thing that thinks is mysterious. I cannot agree with them. I do not deny that everything said in the preceding paragraph is correct, as far as it goes. Nevertheless, it seems to me that the notion of a physical thing that thinks is a mysterious notion, and that Leibniz's thought-experiment brings out this mystery very effectively. We must remember, however, that our present question is not whether the physicalist is faced with a mystery; our question is whether dualism is to be preferred to physicalism. If thinking is a mystery for the physicalist, this fact will be relevant to our question only if it can be shown that the dualist is not confronted with the same mystery or some corresponding mystery.

And, I believe, the dualist is. For it is thinking itself that is the source of the mystery of a thinking physical thing. The notion of a non-physical thing that thinks is, I would argue, equally mysterious. How any sort of thing could think is a mystery. It is just that it is a bit easier to see that thinking is a mystery when we suppose that the thing that does the thinking is physical, for we can form mental images of the operations of a physical thing, and we can see that the physical interactions represented in these images—the only interactions that *can* be represented in these images—have no connection with thought or sensation, or none we are able to imagine, conceive, or articulate. The only reason we do not readily find the notion of a non-physical thing that thinks equally mysterious is that we have no clear procedure for forming mental images of non-physical things. Still, we are not wholly without resources for constructing mental images of non-physical things. (No doubt most of us associate some sort of mental image with the doctrine of dualistic interactionism: perhaps a human body with a vague "something" inside or above its head.) Let us see what we can do.

Leibniz, in the passage we have quoted, contends that a thinking thing must be a simple, a thing without parts. Well, let us represent, in our thought, a simple non-physical thing by a dot and a composite non-physical thing by a bunch of dots, perhaps a bunch that is in constant internal motion like a swarm of bees. Might a *composite* non-physical thing "think, sense, and have perception"? It is hard to see how. Consider our proposed mental picture of a composite non-physical thing. If the simples that make up a composite non-physical thing do not think individually, where is the thinking in our picture? How can a bunch of things that do not individually think or sense or have perception add up to something that does think or sense or have perception? How could their causal interaction produce such properties? Note that these questions are exactly parallel to the questions Leibniz's thought-experiment raises about thought and composite *physical* things. The only real difference between the two cases is that a mental image of a composite physical thing will have reasonably "sharp" constituents drawn from our experience of actual physical things—images of gears and wheels, say—, whereas (an attempt at) a mental image of a composite non-physical thing will be vague and arbitrary (arbitrary because non-physical things necessarily lack visual characteristics; we

chose dots because dots come as close to having no characteristics as anything we can picture).

Leibniz would no doubt agree that these reflections show that a composite non-physical thing cannot think. After all, his position is that a thinking thing has to be a simple.7 But let us look at our proposed mental picture of a (non-physical) simple. It is just a dot. How can we cause it to change in our imagination in such a way that this change will represent its having a series of thoughts and sensations? Change of position (relative to other imagined dots) will be of no help, because that is a relational change, and thought and sensation are supposed to be intrinsic features of thinking, sensing things. Even a dot must have a shape, but when we use dots to represent non-physical simples we do our best not to attend to their shapes, for insofar as we think of a dot as having a shape, we think of it as being composed of smaller regions and thus as composite.

We might think of the dot as changing color, I suppose. Let's try that. Imagine a dot continuously changing its color in some very complex way. Are you imagining something thinking or having sensation? Where are the thought and the sensation in the picture your imagination has created? My point in asking these unanswerable rhetorical questions is not to suggest that a non-physical simple cannot think. (Although I believe that human persons are physical things made of smaller physical things, I believe that God is a non-physical simple, so I should hardly want to suggest that a non-physical simple cannot think.) My point is that nothing could possibly count as a mental image of a thinking thing. Or at least, nothing could count as a mental image that *shows* or *displays* a thing as thinking (except by convention, as, for example, "thought-balloons" in comic strips do, or via the familiar outward and visible signs of human thought, like those displayed by Rodin's *The Thinker*). And, I am suggesting, we need to keep this fact in mind when we consider Leibniz's thought-experiment. It is only the difficulty of conducting a similar thought-experiment for non-physical things that keeps us from seeing that his thought-experiment does not favor dualism over physicalism. Consider this analogy. We are amazed to see a human figure hurtling through the sky like Superman. "It's a woman!" someone shouts. "Why a woman?" we ask. "Well, it's either a man or a woman,

and it's impossible for a man to fly." This argument is valid, and there are certainly good reasons for thinking that it's impossible for a man to fly. But there are equally good reasons (the same ones) for thinking that it's impossible for a woman to fly. Therefore, the argument gives us no reason to prefer the hypothesis that the human figure we saw in the sky was a woman to the hypothesis that it was a man. And this is exactly parallel to what one should say in response to Leibniz's thought-experiment: Since we are unable to imagine a non-physical thing in a way that displays it as thinking, the fact that we are unable to imagine a physical thing in a way that displays it as thinking does not give us a reason to prefer the hypothesis that we human thinkers are non-physical things to the hypothesis that we are physical things.

These points about mental images can be generalized so as to apply to any type of representation. Mental images are representations of how things are or might be, but there are representations of many other kinds, such as schematic diagrams on paper, three-dimensional cardboard models, computer models, and scientific theories. In general, to attempt to explain how an underlying reality generates some phenomenon is to construct a representation of the working of that underlying reality, a representation that in some sense "shows how" the underlying reality generates the phenomenon. (The best scientists seem to be able to "translate" their verbally and mathematically formulated representations of the workings of things into images, which they are able to manipulate mentally in fruitful ways.) Essentially the same considerations as those that show that we are unable to form a mental image that displays the generation of thought and sensation by the workings of some underlying reality (whether the underlying reality involves one thing or many, and whether the things it involves are physical or non-physical) show that we are unable to form *any* sort of representation that displays the generation of thought and sensation by the workings of an underlying reality. Thought and sensation are therefore a mystery—although not necessarily an insoluble one. But since the mystery, soluble or insoluble, is entirely independent of whether the elements in the representation are supposed to represent physical or non-physical things, the mystery of thought and sensation does not favor dualism over physicalism.