Metaphysics // Fall 2016

Handout 19

Free will: Searle

The mind-body problem. The traditional mind-body problem was the problem of mental causation. How can mental states cause bodily movements if mind is not material? How can the two substances, mind and matter, causally interact with each other? To state the problem in this way is to presuppose some form of Cartesian dualism with its division of the world into two substances. The problem disappears once we see that mental states are themselves 'realizations' of brain processes.

Once we remove the Cartesian presupposition, we are left with no problem to solve. Searle then asks whether a similar move can be made with regard to the problem of free will: remove the confusion, and the problem vanishes.

The two worlds. We have no problem attributing some form of necessity to the events in the natural world. Given the initial conditions, an earthquake *had* to happen. This is a claim of determinism (how difficult it is, we have already seen). We take a different view of the human action. Given my past history, I did not have to vote for a candidate X. I could have voted X, but I also could have voted Y. It was my choice not caused by the past events.

And even if I refuse to exercise my choice, the refusal itself is only intelligible on the assumption of a free decision taking place. If, that is, there is any action taking place, the intelligibility of that action presupposes an exercise of free will.

Remark 1. The link between intelligibility of action and exercise of free will was made by Kant. See, e.g., *Groundwork*, chapter III.

But the pull of determinism is also strong. This is because in our interactions with the material world we find that world causally determined.

Remark 2. As Searle understands it, determinism entails negation of free will. This view should be compared to the views described in Earman's discussion.

Mind the gap. Our experience of freedom is beyond dispute (i.e. psychological determinism is false). A legitimate question is whether this experience represents a genuine feature of reality, or is a mere illusion. The freedom experience comes down to a gap. We grant that our actions have causal antecedents, that there is a causal chain leading up to a point in our deliberations and decisions. But then we think there is a causal gap between those antecedents and the decision and the beginning of the action itself. This gap is filled by free will in our own representations of our actions. We can ask, however, whether the gap can be filled neurobiologically, i.e. whether no real causal gap exists.

The status of consciousness. Searle holds the view that consciousness is (i) just a higher level feature of the brain, (ii) causally efficacious, (iii) not reducible ontologically to neuronal structures, but (iv) whose causal powers *are* reducible to the causal powers of neurons.

Remark 3. See the referenced chapter in the *Rediscovery of the Mind* for a lucid defence of the claim (iii).

The discussion is not entirely satisfying. First of all, there is an obvious tension between the claims (ii) and (iv). This will become evident later. Second, the wheel-solidity analogy obscures the matter. Solidity is a feature of the wheel, and the wheel's behaviour is explained by the way solidity functions. Yet ultimately, solidity is nothing but a combination of molecules composing the wheel. So this gives us a good sense of talking about solidity as 'nothing above and beyond the wheel molecules'. The same move cannot be made for consciousness, since first-person ontology is not reducible to third-person ontology. Then it is unclear why we cannot say that consciousness is something beyond neuronal structures, and that its causal powers are *sui generis*.

In other words, we may have difficulty rejecting ontological reduction, whilst endorsing causal reduction. One might think that, as soon as you have done the causal reduction of X to Y, you have also provided for a possibility of ontologically reducing X to Y. (See below for more on the analogy.)

Two styles of explanation. We often, indeed nearly always, give reason-based explanations of our actions. But such explanations are not causal, that is, they do not cite causally sufficient conditions. It is not clear exactly why—i.e. the alleged difference between the three explanations in page 51 is somewhat elusive. Perhaps the idea is that explanations by reasons are not meant to convey inevitability. If I say:

> I ate goat cheese because I was hungry, (19-1)

my statement is ambiguous. In one situation (19-1) can be unpacked thus:

> Being presented at a lunch-break with goat cheese, cucumbers, water, and tea, and having little time to spend, I ate goat cheese because I was hungry. (19-2)

My hunger is not a causally sufficient condition: it is implied in (19-2) that a choice was made. The locution 'because' here indicates that the person acted on reasons (e.g., that goat cheese is more nutritious than the other dishes). In a different situation, however, (19-1) can be unpacked thus:

> Having been lost in a desert, having had no food for three days, I found goat cheese splashed on the road; and though I absolutely detest it, this time I ate goat cheese because I was hungry. (19-3)

Here it is implied that no choice was made, that the person was propelled toward eating goat cheese. The person did not act on reasons, and indeed, he did not act in the full sense of the word.

The lesson Searle wants to learn here is that reason-based explanations are intelligible on the assumption that a choice-making self is involved, an agent capable of acting on reasons. Such an agent would also be capable of deliberation where different reasons will be weighed against each other. No such assumption is necessary in the case such as (19-3).

The two hypotheses. We know, or in any case assume, that every mental event is realised by the brain. So the beginning of deliberation at t_1 and its end at t_2, immediately before the action, before the contraction of muscles (what Hobbes called the 'will'), both of them are realised in the brain, in two brain states. We also know, or assume, that the two brain states are causally related. But we think that the two extreme ends of deliberations are gappy. The link between them, even if causal, does not contain causally sufficient conditions. This is the idea behind the parallelogram in page 59.

Now at this point we might make the following claim:

Hypothesis 1. The brain states realising the deliberation at t_2 are causally sufficient.

On this view, the experience of freedom has no causal or explanatory role to play. It is an epiphenomenon. We have an experience that we act freely, and while the experience itself is not an illusion (I think), freedom is. Or one might say, the conviction that this experience produces is false through and through.

As one could guess, Searle has to reject Hypothesis 1. Consciousness, as we saw above, *is* causally apt. According to Hypothesis 1, however, it is causally irrelevant, just like the colour of my shirt is when I break a glass (the example in page 67).

But *why* to believe that consciousness is causally apt? I think Searle gives two independent reasons. One is that to think otherwise is simply 'incredible'. Another is that such an important element of our lives cannot be fashioned by evolution to have no causal role.

Another view to take is simply a negation of Hypothesis 1:

Hypothesis 2. The brain states realising the deliberation at t_2 are causally insufficient.

We now say that the transition in deliberation from t_1 to t_2 can only be explained by the operations of the conscious self, i.e. an agent able to act on reasons. Here we have a challenge of explaining the relation between consciousness and brain states. We have already conceded, with Searle, that consciousness is ontologically irreducible to brain states, but that its causal powers *are* reducible to the causal powers of the brain.

Searle repeatedly appeals to the wheel analogy and then, abruptly, to the diagram in page 65. Things are not going well, when the argument has to rely on analogies and diagrams to make a crucial point. Indeed, I do not understand the diagram. What about the analogy? We are told: The wheel is rolling down the hill. It is composed of molecules (rubber and iron molecules, say). Solidity is a

condition of the system of those molecules—that is, it is a condition of the wheel. And the motion of each molecule is causally explained by the solidity of the whole wheel, i.e. of the collection of other molecules.

So the properties of the system have two characteristics: they are not properties of individual parts (molecules are not solid), and they causally influence the behaviour of individual parts. I think this is a strange way of describing the situation. Of course we *can* say that solidity is causally efficacious. Yet, granted the ontological reduction in place, this is a shortcut for saying that other molecules act on an individual molecule causing it to move the way it does.

Consider yet another analogy: sports teams. We occasionally say that a team demonstrated 'team spirit', that the win was due to team spirit, or that individual players were affected by team spirit. Yet, unless we think in mystical terms, this talk is an abbreviation of (or a metaphor for?) the more complicated talk about the attitudes of individual players. It is not clear why the two cases, of teams and of wheels, should differ in principle.

Another complaint may be that 'conditions' cannot be causal agents. We say:

> Abe shouted because he was angry. (19-4)

Presumably Searle would want to say that anger is a condition Abe was in. Individual elements of Abe's body were affected by his anger. Or so it seems. But the causal agent was still Abe, not *Abe's condition*. Again, we occasionally talk of emotions (conditions of the body) causing things to happen. And once again, it is simply unclear how seriously we should take this kind of talk.

RATIONAL INDETERMINISM. Aligned with the Hypothesis 2 is the following problem. If the gaps between the deliberations at t_1 and t_2 exist, are they deterministic? Each deliberation state is realised by a brain state. Thus if there is determinism at the level of brain states, there must be determinism at the level of deliberation states. And vice versa. But, on the other hand, the whole point is presumably to say that there is no deliberation determinism. Thus we must account for indeterminism at the level of brain states.

We might resort to quantum indeterminism, arguing that brain states manifest quantum behaviour. This is a lazy move, and anyway, it is of no help. For the quantum indeterminism of a system entails that the system has random behaviour. Yet being random is not at all the same as being rationally indeterministic. That is, we do not want to say that decisions happen randomly.

The discussion is thus ultimately inconclusive, since there are good reasons to reject both hypotheses. Hypothesis 1 is simple, but incredible. Hypothesis 2, while not incredible, is fundamentally confusing.

YSB