From Observer-Relativity to Assignment-Dependence

John Preston[1]

Abstract. John Searle produced two arguments against cognitivism in his 1992 book *The Rediscovery of Mind*. I set out the more fundamental of the two, and argue that its terms are problematic. However, I also identify what I take to be an important point underlying this argument, which is that devices are digital computers in virtue of some person or other having effected an assignment or labelling of their internal states. Consequently, although computational devices do literally compute, we shouldn't think of their computational operations as intrinsic to them, or as machine computation as a *natural*-scientific property. A lesson is also drawn here for our understanding of 'natural' computation.

1 INTRODUCTION

A decade after producing his notorious 'Chinese Room' argument (henceforth 'CRA') against 'Strong AI', John Searle returned to reconsider the notion of computation. The conclusions he then drew, published in his 1992 book *The Rediscovery of the Mind* [1], suggested to him that one of the CRA's assumptions, that computers follow syntactic rules, wasn't as obvious or unproblematic as he had originally thought. His withdrawal of this assumption seems to have been his only major change of mind on the issues. He took it to mean not that the CRA no longer goes through at all, but rather that Strong AI and computationalism (which he didn't define, as far as I can see, but which I think he took to be the same as *cognitivism*, 'the view that the brain is a digital computer' (ibid., p.202)) aren't sufficiently well-formed enough to have any truth-values.

2 TURING'S DEFINITION OF COMPUTATION

Finding little agreement among contemporary cognitive scientists on fundamental questions about computation, Searle proposed to return to the original definition of the kind of computation that machines are supposed to perform (ibid., pp.205-6). What he meant by this is simply Alan Turing's specification, in his paper on the *Entscheidungsdproblem* [2], of what we now call Turing machines. That specification talks about the machine's elementary operations, which include the ability to print a '0' or a '1' in each square of its indefinitely long tape. These 0's and 1's, Searle pointed out, are not to be thought of as physical inhabitants of the computer: one wouldn't find them if one opened the machine up. As he puts it,

> [t]o find out if an object is really a digital computer, it turns out that we do not actually have to look for 0's and 1's, etc.; rather we just have to look for something that we could *treat as* or *count as* or *could be used to* function as 0's and 1's ([1], p.206).

Searle went on to explain the irrelevance of any particular hardware, that is, the fact that a Turing machine (indeed any computational device) could be made out of *anything* whose parts or states can perform the right kinds of physical operation.

The key point about Turing's definition, Searle emphasised, is that it defines computation *syntactically*, 'in terms of the assignment of 0s and 1s' (ibid., p.207). Syntax, though, Searle insisted, 'is not the name of a physical feature, like mass or gravity' (ibid., p.209). He then took these facts, that the relevant properties are purely syntactical, and that syntax isn't a matter of physics, to have two consequences which he considered disastrous for computationalism.

The first consequence concerns the phenomenon known as *multiple realizability*, that is, the fact, beloved by computationalists and 'functionalists' in the philosophy of mind, that 'the same function admits of multiple realizations' (ibid., p.207). The supposed ensuing disaster is expressed in what we might call Searle's *trivialisation argument* (ibid., pp.207-9), according to which this multiple realizability implies a *universal* realizability, which trivialises the cognitivist doctrine. This argument has already received plenty of attention in the literature, and I shan't consider it here.

My interest here is in the *other* consequence of Turing's definition of computation, and the argument which Searle drew from it, which I shall call his 'observer-relativity' argument (ibid., pp.209-212). I shall set out this argument, and then proceed to critique it, while also trying to draw out what I think is its most important and successful point.

3 SEARLE'S OBSERVER-RELATIVITY ARGUMENT

According to Searle, the key point that computation is defined syntactically, 'in terms of the assignment of 0s and 1s' has the consequence that, as he puts it, *syntax is not intrinsic to physics* (Searle ibid., p.208, emphasis added). In other words, 'syntax' isn't the name of a physical feature or property. And this is supposed to be because, as he puts it, 'the ascription of syntactical properties is always *relative to* an agent or observer who treats certain physical phenomena as syntactical' (ibid., emphasis added), or alternatively '[s]omething is a symbol only relative to some observer, user, or agent who assigns a symbolic interpretation to it' ([3], p.16, see also [4], pp.209-10). Searle quickly allows that one might be able to tighten up Turing's original definition of computation, probably by imposing some *causal* conditions, in order to block the inference to universal realizability. However, he says, 'these further restrictions on the definition of computation are no help in the present discussion *because the really deep problem is that syntax is essentially an observer-relative notion*' ([1], p.209, emphasis in the original). This is why he considered this argument to be more fundamental than his trivialisation argument (ibid., p.208).

So Searle's second argument, as I understand it, is that computation is defined in terms of syntax, but syntax is observer-relative, rather than a matter of physics, therefore computation must be observer-relative too. Computation can therefore never suffice for semantics, since whether or not a state has a given semantic 'content' is one of its observer-*independent*

or 'intrinsic' features. So there's no prospect of our ever *discovering* that something (be it an electronic device, a brain, a mind, or anything else) is a machine carrying out computations *independently of someone's having assigned it such a role*.

This new argument concedes less to computationalism than the CRA, since it implies that computationalism doesn't even succeed in being false, but rather is incoherent, having no clear sense ([3], p.15, [4], p.209, [5], p.14, plus [6], p.194). Whereas the CRA, if successful, shows that computation isn't *sufficient* for cognition, this new argument is supposed to show that it can't be *necessary* for cognition, either.

4 ON THE 'OBSERVER-RELATIVE' (AND THE 'OBSERVER-DEPENDENT')

Some of my concerns about this argument relate to the terms in which it is framed. The expression 'observer-relative' seems to be a misnomer, in three respects. First, it's not clear that the relevant category is that of *observers*, or the relevant activity that of *observation*. Precious few properties seem to be brought into existence by *observation*, that is, very few statements about any phenomenon are made true simply by the fact that someone is observing it. The most obvious cases that are of this kind come from human social activity, cases such as Jean-Paul Sartre's waiter, who is diligently "playing the part of a waiter" at least partly because he is being observed ([7]). One can imagine examples in the area of human computation, too: whether a person is computing, and/or what function s/he is computing, *might* depend on whether s/he is being observed (and by whom, in what context, etc.). The domain of machine computation, though, features no such cases. That is, whether an electronic device is computing, and what function it's computing, and how, have nothing to do with whether it's merely being *observed*, strictly speaking.

Second, the term '*relativity'* also seems unhelpful to me, although I find it harder to say why. It seems to me that genuine and paradigm relativities involve a possible difference in certain properties being consequent upon a difference in framework. (The relativity of properties like the mass and velocity of physical bodies to inertial frameworks, as specified in Einstein's theory, is the paradigm I have in mind, of course). But in these cases the relativity isn't a matter of *causal* dependence. In fact, where one thing does depend causally on another, I think the term 'relativity' is inappropriate. I don't think we do or should say that effects are 'relative to' their causes. A person will die if deprived of oxygen for long enough, for example, but this would make us say their death was *due to* a lack of oxygen, not 'relative to' it. Perhaps philosophers (and others) misuse the concept of relativity in ways like this, but I don't think we should follow them. In the cases I think Searle has in mind, though, as we'll see, just such a causal dependence is in question.

I shall reflect this by mainly replacing the term 'relativity', from now on, with another term that Searle later used, *dependence* (see [8], for example). Even when one has done this, though, the phrase 'observer-dependence' still has the potentially misleading connotation that the dependence in question is a dependence upon *a person*. This makes it look as if whether some device is computing might depend on whether someone interprets it as computing, even perhaps whether someone merely *thinks* it's computing. Some of the ways in which Searle expresses himself encourage this: when he says, for example, that the processes going on within a device '*depend on an interpretation from outside*' ([1], p.209, emphasis in the original). This is the third respect in which the phrases 'observer-relativity' *and* 'observer dependence' are misnomers. In my human examples of observer-dependence, whether what someone is doing is Φ may depend on whether the activity is being observed *at all*, or it may depend on *which* observer is doing the observing. But as we shall see soon, the dependence in the case of computation is not of this kind, for the dependence is not on a person but on some historical *act* that a particular person must have performed.

Taking these points together, one arrives at the conclusion that Searle's 'observer-relativity' really boils down to *person-dependence*, that is *dependence on some activity of people*. However, having got this far, the contrast Searle needs, between those properties that are person-dependent and those that are *intrinsic*, is imperilled.

5 ON THE 'INTRINSIC'

To begin with, consider Searle's idea that syntax isn't 'intrinsic to physics'. First, this idea needs clarifying, since one can quite well imagine someone thinking: look, the syntax of any given group of symbols is simply their *form* or *shape*, and surely that *is* 'intrinsic to physics', in any meaningful sense! If something is v-shaped, like a flying flock of birds, for example (see [11]), then its being so doesn't depend on whether anyone takes it to be so, whether anyone is observing it, has observed it, etc. Of course the *vocabulary* for describing it thus wasn't available until humans and their alphabets came on the scene, but even so those flocks of birds had that shape entirely independently of such factors. It doesn't help much to counter this thought when we recall that by the fact that the syntactical properties of certain physical phenomena are 'not intrinsic to physics' Searle means that they are 'always relative to an agent or observer who treats [them] as syntactical' (Searle ibid., p.208). This is not clearly the case, as the flying birds example suggests: waterfowl really did fly south in v-formations years before humans were around to observe them doing so.

Second, even if syntax isn't 'intrinsic to physics', this in no way means that it's person-dependent (let alone 'observer-relative'). The dichotomy that Searle assumes which would provide backing for the inference from the one to the other is no dichotomy at all.

To see this note that, as Jeff Coulter and Wes Sharrock already complained some time ago ([6]), Searle gives no clear reason to restrict our conception of which properties are intrinsic to those which are intrinsic *to physics*. 'Intrinsic' presumably means something like 'consequent on the very nature of'. Intrinsic properties would normally be contrasted with *extrinsic* properties, those which although they *are* properties of some object or event or process or other, are not such by virtue of the nature of the event, process or object involved. (Such properties are, or are close to, those which philosophers have thought of as contingent or accidental). But *all sorts of properties can be intrinsic*, not just physical ones. To think otherwise one has to have swallowed the idea that everything objective must reduce to physics. But that's an article of faith (and a rather strange one for *Searle* to have swallowed), not a result that science can be thought to have established.

To see this, take the biological domain. Searle's view would have the consequence that unless a biological property was reducible to purely physical properties, it couldn't be intrinsic. That can't be right. Functional properties (such as certain evolutionary properties) aren't, as far as we know, reducible to purely physical ones, yet they should still clearly count as intrinsic to the organisms and organs in question. Your heart is a pump, your kidneys clean your urine, etc. The fact that these properties of such organs aren't purely physical (that is: the kinds of properties that *physics* specifies) shouldn't be taken to mean that they aren't intrinsic, let alone that they're 'observer-relative'. Whether my heart pumps my blood, for example, has nothing to do with whether or not any observer or observers treat it as doing so, and it doesn't depend (in any *such* way) on the activity of any person, thankfully!

The same goes even when we move from biology to the domain of *artefacts*, from which Searle sometimes takes such supposed examples of 'observer-relativity' as bathtubs and chairs ([1], p.211). Whether some construction of wood and metal is a chair doesn't depend on whether anyone observes it to be, or even whether anyone uses it as, a chair. (Plenty of chairs never get used, I suspect). At the *margins* of such categories, as it were, there is, admittedly, *some* kind of user-dependence. We might say that whether an old tree-trunk is a table, or whether a stone that's been found on a beach is a paperweight (at some later given time) depends on whether anyone is using it as a table or as a paperweight, for example. But *these* are examples of naturally-occurring objects that have been turned into artefacts, not artefacts in the core sense.

6 ASSIGNMENT-DEPENDENCE, AND ENDICOTT'S CRITIQUE

So far I've expressed reservations about Searle's argument. But I have to admit that I still think he's onto something important. I think that the key idea underlying his argument, and which emerges at some places in his presentation, is *setting-* or *assignment-dependence*. That is, the crucial thing about machine computation is that there's some ground level of computational operations (machine code, as it were), at which humans must have forged or stipulated an association or correspondence between states of the device (electronic states, usually) and the most basic computational states (see, for example, [9], p.365). So, in the case of electronic digital computers, those designing or making the device must have consistently assigned each binary digit to two different kinds of electronic state (high and low voltages, for example). Without *some* such setting or assignment, we would have no conception of what such devices are doing, and thus no way of understanding their activity or recognising it or using it as computation. This is what's important, for Searle, about Turing's conception of computation. And he would be quite right, I think, to say that no matter what *further* constraints one might put on the notion of computation ([1], p.209), none of them will erase this fundamental dependence of machine-computation on a certain kind of assignment having taken place.

In the most detailed critique of Searle's work on these issues to date, though, Ronald P. Endicott has taken issue with him on this matter. Endicott seems to agree with Searle that in order to determine whether something is a digital computer we have to look inside it and find 'something that we could *treat as* or *count as* or *could be used to* function as 0's and 1's' ([1], p.206, emphasis in the original). He goes on to point out, though, that 'some treatments may be *correct*, and this depends entirely on the attitude and objects concerned' ([10], p.103, emphasis added). It's correct to treat participants at this symposium as adults, for example, rather than children.

This is true, but it doesn't really touch Searle's underlying point, which I take it is that what's going on is an *assignment* or labelling, and that the assignment in question is in a certain respect *arbitrary*. Searle later explicitly denies that by 'observer-dependent' he means 'arbitrary'. But his focus there is the idea that 'you cannot use just any piece of circuitry as an and-gate or an or-gate' ([8], p.68), rather than on this kind of assignment. And the arbitrariness I have in mind means only that there's a range of options between which we must choose, although it doesn't matter *which* one we choose. It *doesn't* mean that there are *no* constraints on our choice, i.e., that the range of options is *unlimited*.

When we look at the workings of the kinds of physical parts or systems that might be, or might become, computational devices we need ultimately to find some states (among their electronic or other processes) which have been or can be assigned to 0's and other, orthogonal states which have been or can be assigned to 1's. So, for example, we might associate '0' with there being *no* voltage in the relevant part of the device, and '1' with there being such a voltage (or a potential difference of 5 millivolts, or whatever). But *we could equally well make any other consistent assignment*: absolutely nothing hangs on *which* such assignment we make, and thus any given assignment *is* truly arbitrary.

The question then becomes: does arbitrariness of *this* kind threaten the idea that whether some device is computing, or what function it is computing, is 'intrinsic'? I think that Searle would be right to believe that it does, and that he's right that there's no genuine analogue to this in the natural sciences. Nowhere in physics, chemistry or biology do scientists think that whether the objects they study have certain system-properties depends on whether someone has made a prior setting or assignment of any such kind to the microstates of those objects. This, I think is, the best way of conveying his ideas that, ironically, far from being too 'mechanical' to be what mental activity consists in, computation *isn't machine-like enough* ([8], p.57), that 'the natural sciences study features that are observer-independent' (ibid., p.62), and that '[i]n an observer-independent sense, the only things going on in the machine [are] very rapid state transitions in electronic circuits' (ibid., p.65).

Of course, given that as a matter of historical fact someone either *has* or *hasn't* made such an assignment in the case of any given object, there's a perfectly objective answer to the question 'Is this object a computer?'. And, given historical facts of that same kind, there can be an objective answer to the question 'What function(s) is it computing?'. That's why I baulk at the expression 'observer-relative', and I think Searle is quite wrong to suggest that whether an object is computing or what it is computing might depend on what anybody now consciously thinks about it (ibid., p.62). That makes it sound as if whether a device is computing, and what it's computing, might depend merely on what someone or other *thinks*. The dependence I have in mind is obviously not of that kind.

So I think Searle is right: computation isn't an *intrinsic* property of machines at all. It's not, as he would put it, that syntax isn't intrinsic *to physics*, but rather that some setting or

other must have been effected, some assignment or other must have been made, if we're correctly to think of any given object as a computer.

(Perhaps there's an analogy here with currency: for any given piece of metal, paper or plastic, there's an objective answer to the question 'is this currency?'. But that answer depends on whether a particular accredited person (in a nation's central bank, for example) has historically assigned that status to such pieces of metal, paper or plastic. Likewise, one can quite well imagine, although it seems sociologically unlikely, that individual pieces of paper with a particular shape, size and design might have been assigned completely *different* monetary values in the banking-systems of two different states. They might be, at the same time, one dollar and five pounds, for example).

7 'NATURAL' COMPUTATION

And here there may be a lesson to be drawn for the subject of 'natural computation'. Just as whether a wood and metal construction is a chair depends primarily on whether it was *designed* to be a chair, whether a device is a computer doesn't depend on whether or how it's being used, but it *does* depend on whether it has been *designed* to be used as a computer (where the design constitutively involves an assignment of the kind I've talked about). This is why, if in the future we come to count purely natural systems as performing computational operations, as some computer scientists now urge, we will thereby have (once again) *extended* our concept of computation. Neither the original concept of human computation, nor the concept of computation we now apply to devices of our own construction, apply as they stand to *natural* systems. There would be nothing *wrong* with thus extending the concept, of course, and there may be good reason for doing so. All I'm suggesting is that we ought not to think or pretend that we've *discovered* that what's going on within natural systems is activity of the same kind as that which is going on within our machines. Rather, we use important similarities between what our computational devices do and what natural systems can be thought of as doing (notably: similarities in their *function*), in order to motivate extending the concept of computation to cover the latter.

8 CONCLUSION

While some of the trappings of Searle's argument against cognitivism are problematic, then, I still find that there's an important point underlying his critique, which is that devices are digital computers in virtue of some person or other having effected an assignment or labelling of their internal states. Since we can't pretend that this is a matter of *discovery*, we can't think of their computational operations as *intrinsic* to them. They're literally computers, all right, but their being so isn't a *natural*-scientific fact about them.

Whether or not I'm following Searle in what I've argued, I'm not sure. On the one hand, he has in recent works backed away from the term 'observer-relative' ([8], for example). On the other hand, and as far as I can see, those works don't clearly identify the dependence on an arbitrary assignment of the kind I have in mind as the reason why computation isn't an 'intrinsic' property of computational devices.

REFERENCES

[1] J.R.Searle, *The Rediscovery of the Mind*, (Cambridge, MA: MIT Press, 1992).

[2] A.M.Turing, 'On Computable Numbers, with an Application to the Entscheidungsproblem', *Proceedings of the London Mathematical Society*, Series 2, vol.42, 1936-7, pp.230-65.

[3] J.R.Searle, 'The Problem of Consciousness', *Social Research*, vol.60, 1993, pp.3-16.

[4] J.R.Searle, 'Ontology is the Question', in P.Baumgartner & S.Payr (eds.), *Speaking Minds: Interviews with Twenty Eminent Cognitive Scientists*, (Princeton: Princeton University Press, 1995, pp.202-213.

[5] J.R.Searle, *The Mystery of Consciousness*, (London: Granta, 1997).

[6] J.Coulter & W.Sharrock, 'The Hinterland of the Chinese Room', in J.M.Preston & J.M.Bishop (eds.), *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*, (Oxford: Oxford University Press, 2002), pp.181-200.

[7] J-P.Sartre, *Being and Nothingness: An Essay on Phenomenological Ontology*, (London: Methuen, 1957, French original 1943).

[8] J.R.Searle, 'Twenty-One Years in the Chinese Room', in J.M.Preston & J.M.Bishop (eds.), *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*, (Oxford: Oxford University Press, 2002), pp.51-69.

[9] J.M.Bishop, 'Dancing with Pixies: Strong Artificial Intelligence and Panpsychism', in J.M.Preston & J.M.Bishop (eds.), *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*, (Oxford: Oxford University Press, 2002), pp.360-378.

[10] R.P.Endicott, 'Searle, Syntax and Observer-Relativity', *Canadian Journal of Philosophy*, vol.26, 1996, pp.101-122.

[11] Y.J.Erden, 'The "Simple-Minded" Metaphor: Why the Brain is *not* a Computer, via a defence of Searle', Proceedings of the AISB Symposium, 2013.