DAVIDSON ON TURING: RATIONALITY MISUNDERSTOOD?

JOHN-MICHAEL KUCZYNSKI
University of California

Abstract

Alan Turing advocated a kind of functionalism: A machine M is a thinker provided that it responds in certain ways to certain inputs. Davidson argues that Turing's functionalism is inconsistent with a certain kind of epistemic externalism, and is therefore false. In Davidson's view, concepts consist of causal liasons of a certain kind between subject and object. Turing's machine doesn't have the right kinds of causal liasons to its environment. Therefore it doesn't have concepts. Therefore it doesn't think. I argue that this reasoning is entirely fallacious. It is true that, in some cases, a causal liason between subject and object is part of one's concept of that object. Consequently, to grasp certain propositions, one must have certain kids of causal ties to one's environment. But this means that we must rethink some old views on what rationality is. It does not mean, pace Davidson, that a precondition for being rational is being causally embedded in one's environment in a certain way. If Turing's machine isn't capable of thinking (I leave it open whether it is or is not), that has nothing to do with its lacking certain kinds of causal connections to the environment. The larger significance of our discussion is this: rationality consists either in one's ability to see the bearing of purely existential propositions on one another or rationality is simply not to be understood as the ability see the bearing that propositions have on one another.

1

In this paper, I'd like to make a narrow point and a broad point. The narrow point is that what Donald Davidson says about Turing machines — specifically, his argument for holding that they don't think — doesn't go through.[1] The broad point is: Given that these

© *Principia* 9 (1–2) 2005, pp. 111–124. Published by NEL — Epistemology and Logic Research Group. Federal University of Santa Catarina (UFSC), Brazil.

reasons of Davidson's don't go through, there is some reason to believe that *one* of the following two claims holds.

(a) Intelligence is *not* to be understood as the ability to see the bearing that propositions have on other propositions.

(b) Intelligence is to be defined as the ability to see the bearing of *purely existential* propositions on other purely existential propositions.

Of course, (a) and (b) cannot both hold.

2

First some background. Putnam showed that one's concepts of objects — one's abilities to have thoughts about objects — often have a causal *component.* At least a *part* of one's concept of water or of Bob consists in there being some kind of causal connection between oneself and those things.[2]

Putnam showed this through thought-experiments like this one. Twin-John and John are molecule for molecule duplicates of each other. So John and Twin-John are exactly alike *leaving aside* facts about the (distal) causes of their conditions. Given this, suppose that John has a visual sensation that is caused by Bob, and that twin-John has a qualitatively identical visual sensation that is caused by twin-Bob. John is perceiving *Bob*, not twin-Bob. And twin-John is perceiving twin-Bob, not Bob. From his perception, John thus 'uploads' a concept of Bob, not of twin-Bob. And from *his* perception, twin-John uploads a concept of twin-Bob, and not of Bob.[3]

The fact that John has a concept of Bob, and not twin-Bob, derives from the fact that John has a causal connection to Bob that he doesn't have to twin-Bob. And the same thing *mutatis mutandis* explains why twin-John has a concept of twin-Bob, and not of Bob.

Thus John has thoughts about Bob *in virtue of* (at least in part) his having a certain causal connection to Bob. Thus John's concept of Bob — his ability to have thoughts about Bob — consists, at least in part, in his having a certain causal connection to Bob. (The same thing is true *mutatis mutandis* of twin-John and twin-Bob.) In this way, Putnam (successfully) shows that a causal connection to an object may be, at the very least, a *part* of one's concept of that object.

In many publications, Jerry Fodor says our concepts of objects *consist* in causal relations of a certain kind between ourselves and objects.[4] As we'll see in a moment, Donald Davidson takes the same view.[5]

I now want to show that this view of Davidson's seriously lead him astray in his analysis of a famous problem posed by Alan Turing.

3

Turing posed the following problem. Let M be some machine, and suppose that M responds to inputs in exactly the way that a thinking being would respond to those inputs.[6] So if you say to M 'should we go back to the gold-standard?' or 'what did you think of the game last night?', M gives you just the kind of answer that a cognitively normal human gives you. And, of course, the same is true for any other question or assertion that you might direct to M. Turing asks: under that circumstance, does M qualify as a thinker? Turing's implicit answer is 'yes'.

Turing is not, I think, making an epistemological point. He is not saying that, since there is as much data to support the view that the machine is thinking as there is to support the view that the person is thinking, there is *good reason* to hold that the machine thinks; he is not saying the machine's behavior is merely *evidence* of thought. (For it is obvious and trivial to say that if M acts just like a person that is *evidence* — though, perhaps, defeasible evidence — that he has the mental capacities of a person.). Turing is saying, it seems, that if the machine behaves indistinguishably from the thinking person, then *ipso facto* the machine is a thinker.

Davidson thinks that Turing is wrong. Now I myself *agree* with Davidson that Turing is wrong; I don't think that M thinks. (More specifically, I don't think that M qualifies as a thinker *merely* in virtue of behaving in the right way in response to certain inputs. M might qualify as a thinker for some other reason.) But my purpose here is not to discuss whether Turing is right or wrong. It is to discuss Davidson's criticisms of Turing. I submit that Davidson has given us no reason to think Turing wrong.

4

Davidson's criticism is this. Let John be some normal human being whose thoughts, and therefore whose sentence-tokenings, are causally 'hooked up' to external objects in the right kind of way. So John has de re thoughts about Bush, Socrates, light and water; and he expresses these when uses sentences containing the terms 'Bush', 'Socrates', etc.

Now let us suppose that we assemble some being that behaves just like John in terms its verbal and behavioral responses to inputs — externally, it responds in the same way as John to the sound 'was Plato more conservative than Aristotle?' and 'is there an even prime?' Let Robo-John be this artificial John-proxy. Of course John actually *means* Plato by 'Plato', actually *means* water by 'water', and so on. This is not (merely) because John is possessed of some mysterious subjectivity or mentality; it is (at least in part) because John is *causally* connected in the right kind of way to Plato and water, and his verbal behavior derives from this connectedness. John's current psychological condition is the fall-out, at least in part, of his being embedded in a certain physical environment in a certain way: and it is this embeddedness which enables him to think about Plato and thus to *mean* Plato by 'Plato'. This, in effect, is what we learn from the Putnamian thought experiment described above.

Now Robo-John was put together in some laboratory; so *his* psychological condition is an artifact; it is definitely *not* the fall-out, even in part, of his being embedded in a certain environment in a certain way. (Of course, one might say: 'well, the scientists who put Robo-John together knew about Plato and water, and so on, and *their* ability to thinks about things was incorporated into Robo-

John.' But we can easily side-step this complication by supposing that Robo-John wasn't created in a laboratory, but rather came together as a result of a sandstorm or some random quantum event.[7]) In any case, his relation to his environment is definitely *not* such as to underwrite his having concepts of Plato or water. So Robo-John *just cannot think about Plato or water or Bush.* He acts just like John; and he says (or rather mouths) the same things. But when he says 'Plato was wise', he isn't *referring to Plato.* He is shooting a blank; he cannot *mean* Plato by 'Plato', for he hasn't got the right causal connection to the man. And, Davidson says, this point applies mutatis mutandis to anything else Robo-John says or does.

Davidson concludes that Robo-John *does not think*: for he hasn't got the causal connections to his environment to underwrite the having of the concepts needed to think. Thus, even though Robo-John is indistinguishable (externally) from John, Robo-John doesn't think, while John does think. And Turing is therefore wrong.

5

Again I agree with Davidson that Turing is wrong. But Davidson's argument has little force against Turing. A brain in a vat cannot think about Socrates or about water. But it can still think.[8] When you dream about unicorns and goblins, your dream-images are images of *nothing*; and the narrative of your dream represents nothing real.[9] But in dreaming — in experiencing this play of dream images — you are still thinking. Whatever the line is between thinking and non-thinking beings — between being a rock and being a cognizant entity — you are, in dreaming, well on the right side of it. Obviously a brain in a vat can dream. What it cannot do is have thoughts that are object-involving with respect to Socrates or water. But clearly *something* mental — indeed, something *cognitive* — can still go on in it.

Turing's machine may or may not be thinker: I leave it open whether it is. But if it *isn't* a thinker, that has nothing to do with its lacking the right kinds of causal connections. The fact that there are *certain* things Turing's machine cannot think about — e.g. Socrates — has to do with its not having the right causal connections. But if Turing's machine cannot think *at all*, that is not to be explained by saying that it doesn't have the right causal connections. What led Davidson astray here is his holding that concepts *consist* in causal connections with the external world, rather than merely *involving* them.

6

Of course, the following objection will be made:

> To think is to entertain propositions or, at any rate, is to handle information. Turing's machine — and its cousins: the subject of a Cartesian nightmare, the brain in a vat — just cannot grasp things; it can't grasp the constituents of propositions (water, Plato, redness…). Therefore it can't grasp propositions, and thus cannot think at all.

There are a number of things to say in response to this. First of all, when you dream of griffins or unicorns, it certainly seems that you are grasping a great many *existential* propositions, even though you are not grasping any object-involving ones. Suppose you are dreaming of a non-existent pink elephant. Now given somebody who was in a phenomenologically identical state who *was* causally connected to a pink elephant in a certain way, that person *would* be grasping an object-dependent proposition[10] that you are not: in the cognitive 'place', to so speak, where that person is grasping an object-dependent proposition, you are grasping nothing.[11] But you are still grasping various existential propositions: e.g. *there is a four-legged creature moving about in such and a such a way…*

Now one might *counter*-respond by saying:

> The constituents of these existential propositions are such that your grasping them consists, at least in part, in your standing in certain causal relations to the external world. Consider, for example, the existential proposition: there is a four-legged creature moving about … Surely your concept of

> a leg is an object-involving one, the same being true of concept of a creature and perhaps even of your concept of movement. Indeed, So once you take away all of a causal liaisons, you take away its ability to entertain even existential propositions like *there is a four-legged creature*.

I find this highly dubious. A brain in a vat, it seems to me, has a concept of a leg, of a creature, of movement, as much as anyone else, notwithstanding that the kinds of connections it has to the outer world don't sustain de re awareness of any of its constituents.[12]

Suppose you discovered — let us set aside, as irrelevant, *how* you discover this — that you are a brain in a vat. Your Plato-concept will thus turn out to be empty, and so will your Nathan Salmon-concept. But your leg-concept? Your movement-concept? Would you then think to yourself: "Gosh, I guess my leg-concept$_o$ turned out to be empty after all, just like my Socrates-concept"? No — *leg* is a functionally defined concept. You don't have to be causally connected to an actual leg to grasp the concept of a leg; you just have to know what functional role legs have. And it is very hard to see how your brain-twin could lack such knowledge.

7

What is true of one's concept of Plato or of water seems not to apply to *all* of one's other concepts, but only some of them. Suppose there is a brain in a vat that is just like your brain *modulo* those properties of you (and your brain) that supervene on your *not* being a brain in a vat. Further, suppose that, one day, you put that brain in a body — so it can speak, move, and so forth. That creature will produce sounds like 'that creature is moving rather quickly'. Is that creature really 'shooting blanks' when it says such things? To be sure, where attempts to make *certain* statements are concerned, it will be shooting blanks. If it says 'Socrates drank hemlock' it will not mean by that sentence what you and I mean by it. But a disembodied brain in a vat can have an appreciation of what it is for an object to move from one place to another, of what sentience is, of what a leg does for a sentient creature.

Also that creature would surely have just as much a grip on mathematical and purely rational truths as you. It would be nonsense to say that, in virtue of the fact that you had certain causal liaisons to the external environment, you were a better mathematician than your recently embodied brain-twin.

8

There is no denying that the behavior of the recently embodied brain-twin is *replete* with mentality. And it is hard to deny that some of this mentality is cognitive and involves intelligence. Surely your brain-twin is not *stupid*; he is obviously not in the same category as something inanimate. Surely it is as intelligent as you, notwithstanding that it doesn't have a grasp of Plato or Socrates.

But suppose that, despite all this, someone taking Davidson's line digs in his heels and insists that one cannot grasp any propositions without having certain kinds of causal liaisons to one's environment — without being (or having been) embedded in a certain physical environment in a certain way. (The idea would be, perhaps, that such causal liaisons so totally infect our cognitive lives that *nothing* propositional bearing would be left over if one were stripped of these liaisons.)

I would suggest that, if that is really the case — and maybe it is — then we must define intelligence *not* in terms of seeing relations among *propositions*, but in some other way. For it seems to me quite obvious that a brain in a vat and the victim of a Cartesian nightmare can have all the *intelligence* in the world; Einstein's disembodied brain-twin is surely not *stupid*, even though he will shoot a lot of conceptual blanks. In other word, where the real Einstein — who *is* appropriately embedded in a physical and social environment — *will* be entertaining various object-dependent thoughts, Einstein's brain-twin will (at least according to Evans) be entertaining thought-*like* things that have gaps in them, due to its lacking the right causal connections to the outside world. But in virtue of having these pseudo-thoughts, or "empty" thoughts, Einstein's brain-twin still obviously has *intelligence.* Einstein's brain-twin is not stupid: it is obviously not in the same category as a rock or a cactus.

So *if* one needs to be embedded in a physical environment in a certain way to grasp any propositions at all, *then* what that means is that intelligence, and cognitive capacity, are not to be defined in terms of proposition-manipulating ability. It *doesn't* mean that a thing, like Turing's machine, which is not appropriately embedded fails to think.

9

It is, of course, strange to say that intelligence consists in something other than the ability to the bearing that propositions have on other propositions But such a view is not without a foundation; it is, I think, the distillation of epistemic externalism.

Kripke (1977) made a very compelling case that one can assent to a proposition *and* to its negation without being irrational. Pierre believes the proposition expressed by 'London is not lovely' and also that expressed by 'London est jolie'. And this is not because Pierre lacks acumen. It isn't that he has failed to deduce some proposition from some other proposition that he accepts. Pierre is not irrational; he isn't even guilty of logical laziness — of not making the right inferences from what he already knows. (If we *stipulated* that Pierre was logically omniscient, he would *still* be in the predicament that Kripke describes.) One lesson that might be drawn from this is: rationality (intelligence) is not to be defined in terms of what one does with the propositions one grasps. This has a certain similarity to the idea that intelligence (rationality) is not be defined in terms of proposition-manipulating ability.

Actually it seems that this last point — intelligence is *not* to be defined as the ability to manipulate *propositions* — is a result towards which externalism has been trending all along.

In any case, it seems plain that a disembodied brain in a vat can have plenty of *intelligence*, even though it has no de re awareness of anything external, and is therefore incapable of grasping many propositions that you and I can grasp. So *if* Turing's machine lacks intelligence, that is not because it doesn't have right kind of causal connection to the outside world; it would be for some totally different reason. Its lack of causal connections deprives it of certain *concepts* but not of intelligence. So Davidson' has by no means proven that Turing's machine cannot think; for given only what Davidson says, it is still an open question whether Turing's machine is intelligent. Thus Davidson's attack on Turing is misguided.

10

I should qualify something said earlier. We've seen that Davidson's attack on Turing doesn't go through. We've also seen that this fact gives *some* credence to a conception of intelligence (or rationality) that is different from the traditional one. (The traditional one is: intelligence consists in some kind of *proposition*-manipulating ability.) But the spuriousness of Davidson's attack by no means *proves* that intelligence is not to be thought of the traditional way. As I suggested earlier, whenever one is grasping some *object-dependent* proposition, one is also, it seems, grasping many existential propositions. When I see Toonces the cat chasing Fido the dog, I am aware of some object-dependent fact, namely *Toonces is chasing Fido*. But necessarily involved in this grasp is a grasp of various existential propositions. I don't *just* see Toonces; perception, and representational mental activity generally, is predicational; I must see Toonces *as* having certain properties (as having fur of a certain color, as being four-legged, as being a certain distance from me, and so on). So my visual awareness of Toonces seems to mediated by propositions (or, in any case, by information) that is existential (*there is a four-legged creature with thus and such properties … running after a four-legged creature with such and such properties*). Now notice that, where the proposition *Toonces is chasing Fido* is *object-involving* with respect to a certain cat and a certain dog, the just mentioned existential propositions are not thus object-involving: *there is a creature with pointy ears … chasing a creature with floppy ears …* is not object-involving, at least not with respect to either Toonces or Fido. Basically, 'existentializing' a proposition tends to get *rid* of the objects in that proposition; by replacing *Toonces* with a variable, and then prefixing the resulting open proposition with an existential quantifier, one turns a proposition that is object-involving with respect to a certain thing into one that is not object-involving with respect to that thing. It thus seems that if an object-involving proposition could be *completely* 'existentialized' (i.e. if each spatiotemporal object implicated in that proposition could be subject to the procedure we just described), the result would, of course, be a completely object-independent proposition. And a *completely* object-independent proposition is one the grasping of which *does* supervenes on one's brain-states considered apart form their causal origins (or, if you prefer, if on one's mental states narrowly individuated). To grasp the proposition

(*) *Socrates was bald*,

it is not enough that my brain be in a certain state; my brain-state must have certain causal origins. But that is entirely because (*) is *object*-involving; it has Socrates himself as a constituent. Obviously given a proposition that didn't have Socrates as a constituent, one could grasp that proposition without having a causal connection to Socrates specifically. By analogous reasoning, given a proposition that doesn't spatio-temporal object O as a constituent, one can grasp that proposition without having a causal connection to O specifically. So if a proposition is *completely* object-independent, then one can grasp it, in principle, without having a causal connection to *anything*: one's grasping it would supervene on one's brain-states (or mental states) considered *apart* from their causal pedigrees. So *purely* existential propositions can be grasped by a brain in a vat. (Whether *there are* such things as purely existential propositions is another question. I think there are. Presumably, the principles of mathematics and logic don't have spatiotemporal individuals or kinds for their constituents — though Putnam, I believe, has challenged this orthodoxy.) Given someone who believes that Turing's machine is intelligent (i.e. thinks and does so competently) and who *also* believes that intelligence is to be analyzed as proposition-manipulating ability, that person's best bet , perhaps, would be to say that Turing's machine operates on purely existential propositions (propositions that do not have spatio-temporal objects or kinds for constituents).

References

Davidson, D. 2004. "Turing's Test." *Problems of Rationality.* Oxford: Clarendon Press.

Kripke, S. 1979. "A Puzzle about Belief." *In* Nathan Salmon and Scott Soames (eds.), *Propositional Attitudes.* Oxford: Oxford University Press.

McGinn, C. 1988. *Mental Content.* Oxford: Basil Blackwell.

Putnam, H. 1975. "The Meaning of 'Meaning'." *Collected Papers*, vol. III. Cambridge, Mass.: Harvard University Press.

Keywords
Davidson, externalism, internalism, Turing's Test.

John-Michael Kuczynski
Department of Philosophy
University of California, Santa Barbara
5631 South Hall
Santa Barbara, CA 93106
USA

Resumo

Alan Turing defendia uma espécie de funcionalismo: uma máquina M *pensa desde que responda de certas maneiras a certos* inputs. *Davidson argumenta que o funcionalismo de Turing é inconsistente com certa espécie de externalismo epistêmico, e é, portanto, falso. Na concepção de Davidson, os conceitos consistem em ligações causais de certa espécie en-*

tre sujeito e objeto. A máquina de Turing não tem as espécies corretas de ligações causais com seu ambiente. Portanto, não dispõe de conceitos. Portanto, não pensa. Argumento que esse raciocínio é inteiramente falacioso. É verdade que, em alguns casos, uma ligação causal entre sujeito e objeto é parte do conceito que se tem desse objeto. Conseqüentemente, para alguém apreender certas proposições é preciso que tenha certas espécies de ligações causais com seu ambiente. Mas isso significa que precisamos repensar algumas velhas concepções a respeito do que é a racionalidade. Não significa, pace *Davidson, que uma precondição para ser racional seja estar causalmente imerso de certa maneira em seu próprio ambiente. Se a máquina de Turing não é capaz de pensar (deixo em aberto se é ou não), isso não tem nada a ver com deixar de ter certas conexões com o ambiente. A importância maior de nossa discussão é esta: a racionalidade consiste ou na capacidade que se tem de perceber a relação que proposições puramente existenciais têm entre si, ou a racionalidade simplesmente não deve ser entendida como a capacidade que se tem de perceber a relação que proposições têm entre si.*

Palavras-chave

Davidson, externalismo, internalismo, teste de Turing.

Notes

[1] See Davidson 2004.

[2] See Putnam 1975. The contents of the next few paragraphs are a very watered down version of a famous thought-experiment that Putnam gives in that publication.

[3] Putnam talks more about our concepts of natural kinds than about our concepts of individuals. But in the present context, what is important is that one's concept of spatiotemporal individuals *or* kinds has a causal component; and this point is due, I believe, to Putnam. In the present context, the distinction between our concepts of individuals and our concepts of kinds is not particularly important.

[4] Where the objects are abstract, the causal connection is between ourselves and *instances* of those objects. So my concept of the number two consists in a certain kind of causal connection between brain-states of mine and instances of the number — e.g. pairs of shoes or pairs of hands.

[5] Though Davidson may not go so as to state explicitly that this is his view.

[6] Of course, when I say that M responds to inputs 'in exactly the way' that a thinking being responds to those inputs, I do not mean that M necessarily has the same *thoughts*; I mean that his behavioral output — the sounds and movements he produces — are those of a thinking being. For if I stipulated that M's *cognitive* reactions were just like those of a normal human being, then Turing's question — does M think? — would be *trivial*; for it would be *built into* the way we described the situation that M thinks.

[7] This, in effect, is the tack Davidson takes in another publication.

[8] In this paper, when I talk about 'brains in vats', I mean brains that are just like the brains of *living*, cognitively normal people, and are subject to the some proximal, but not the same distal, stimulations. So the brains in vats I will be talking about are just like my brain and your brain — except that they are in vats. Of course, in reality, there are brains in vats. But they are dead. They are not what I mean by 'brains in vats.'

[9] Of course, your dream might symbolically correspond to something real: the unicorn might be your mother, the griffin might be your father. I am talking about manifest, not latent, dream-content. Also, one can obviously dream about real things (e.g. one's father). But one *needn't* dream about such things. I am discussing dreams that are wholly about unreals like goblins.

[10] When I talk about 'object-dependent propositions', I mean propositions whose existence is dependent on the existence of *spatiotemporal* individuals and kinds, e.g. *Socrates is bald*, *water quenches thirst.* I *don't* mean propositions whose existence depends on that of *platonic* individuals and kinds (e.g. the number two). So *2+2=4* is not 'object-dependent', as I am using this expression, even though (perhaps) its existence is dependent on that of certain objects, namely *two* and *four* and so on. My usage of the term 'object-dependent' is quite conventional.

[11] In any case, this is that Evans (1984) persuasively argues for.

[12] Colin McGinn (1988: 50–1) ably defends a very similar point. McGinn talks about artifact concepts (e.g. *table*). But what he says in that connection applies, it seems to me, to functionally defined concepts like *leg*.