Other Bodies, Other Minds: A Machine Incarnation of an Old Philosophical Problem

STEVAN HARNAD
Department of Psychology, Princeton University, Princeton, NJ 08544, U.S.A.

Abstract. Any attempt to explain the mind by building machines with minds must confront the other-minds problem: How can we tell whether any body other than our own has a mind when the only way to know is by *being* the other body? In practice we all use some form of Turing Test: If it can *do* everything a body with a mind can do such that we can't tell them apart, we have no basis for doubting it has a mind. But what is "everything" a body with a mind can do? Turing's original "pen-pal" version of the Turing Test (the TT) only tested linguistic capacity, but Searle has shown that a mindless symbol-manipulator could pass the TT undetected. The Total Turing Test (TTT) calls instead for all of our linguistic *and* robotic capacities; immune to Searle's argument, it suggests how to ground a symbol manipulating system in the capacity to pick out the objects its symbols refer to. No Turing Test, however, can guarantee that a body has a mind. Worse, nothing in the explanation of its successful performance requires a model to have a mind at all. Minds are hence very different from the unobservables of physics (e.g., superstrings); and Turing Testing, though essential for machine-modeling the mind, can really only yield an explanation of the body.

Keywords. Artificial intelligence, causality, cognition, computation, explanation, mind/body problem, other-minds problem, robotics, Searle, symbol grounding, Turing Test.

"*Yes, but can it be creative? Can it do what Einstein did?*" So goes the usual objection to "artificial intelligence" – computer models that appear to work the way our mind does [Minsky (1961); Newell (1980); Haugeland (1985)]. Most people have the intuition that these machines may be doing something clever, but not clever enough. There's also doubt that the machines are doing it the "right way." We, after all, are not digital computers: chunks of solid-state hardware and accessories running according to a tape that carries someone else's programmed instructions.

The only trouble is that those who know what kind of "hardware" we are – the anatomists and physiologists, especially those who study the brain – have absolutely no idea how our mind works either. So our intuitive conviction that computers work the wrong way is certainly not based on any knowledge of what the "right" way is. Yet I think those intuitions are largely correct.

As to Einstein – well, it seems a rather high standard to hold computers to, considering that almost every one of us would probably likewise fail resoundingly to meet it. Yet the intuition that a convincing computer model has to be able to do many (perhaps most, perhaps all) of the things a real person can do – that too seems to strike a proper chord.

The Turing Test (TT)

Alan Turing, a distinguished logician and father of a lot of the modern theory

Minds and Machines **1**: 43–54, 1991.

about what computers and computation are, had a similar idea about what ought to be the critical "test" of whether or not a computer is intelligent [Turing (1964)]. He proposed that if you had a person in one room and a computer in another room and you could communicate with both of them (by teletype, as with a pen-pal), talking about anything you liked, for as long as you liked, and you could never tell which one was the human and which one was the machine – you would never even have suspected they weren't both people, in fact – then it would be pretty arbitrary to deny that the candidate was intelligent (or that it understood, or had a mind, which all come to the same thing) simply because you were informed that it was a computer.

The purpose of keeping the computer out of sight in the "Turing Test" was so that your judgment would not be biased by what the candidate looked like – by its "body." But the teletype version of the Turing Test has a big disadvantage entirely independent of the question of bodily appearance. Assuming the intuition is correct (and I think it is) – that if it can do everything we can do indistinguishably from the way we do it, then the way a candidate *looks* is not a good enough reason for denying that it has a mind – then there's still the problem that people can do a lot more than just communicating verbally by teletype. They can recognize and identify and manipulate and describe real objects, events and states of affairs in the world.

So the candidate should be a robot, not just a teletype machine connected to a computer. This new version of the Turing Test (TT) I've dubbed the "Total Turing Test (TTT)": The candidate must be able to do, in the real world of objects and people, *everything* that real people can do, in a way that is indistinguishable (to a person) from the way real people do it.

Bodily appearance clearly matters far less now than it might have in Turing's day: Our intuitive judgment about an otherwise convincing candidate no longer runs much risk of being biased by a robotic exterior in the Star-Wars era, with its loveable tin heroes, already well ahead of the AI game in their performance capacities on the cinematic screen. On the other hand, to settle for anything short of our total performance capacity – say, for only the capacity of AI's "toy" models in their "microworlds" – would be to increase the degrees of freedom of Turing Testing far beyond those of the normal underdetermination of scientific theories by their data [Harnad (1989a)]. It would be like devising laws of motion in physics that applied only to some motions of some objects in some situations. There is an infinity of (wrong) toy "laws" of motion that one could come up with that way, most of which had no hope of "scaling up" to the true laws. The TTT constrains mind-modeling to the normal scientific degrees of freedom, thereby maximizing the likelihood of converging on the true necessary and sufficient physical conditions for having a mind.

Similar arguments can be made against behavioral "modularity". It is unlikely that our chess-playing capacity constitutes an autonomous functional module, independent of our capacity to see, move, manipulate, reason, and perhaps even

to speak. The TT itself is based on the pre-emptive assumption that our linguistic communication capacity is functionally isolable. I suggest instead that the constraints of the TTT itself provide the most likely route to a discovery of any functional modules (if they exist) that may underlie our behavioral capacity [cf. Fodor (1985)].

The Mind/Body Problem and the Other-Minds Problem

It happens that philosophers have been struggling with a related problem at least since Descartes (probably much longer), the mind/body problem [see Nagel (1974, 1986); Searle (1985b)]. How can mental states – those private, subjective experiences such as pain, joy, doubt, determination, and understanding, that we each know *exactly* what it's like to have – be reduced to physical states? How can a pain be the same thing as a physical-chemical condition of the body? Suppose a biologist came up with a complete description of the structure and function of the "pain" system of the brain. How could that capture what it's like to *feel* pain? Wouldn't that physical system's description be equally true of an organism or machine that to all appearances – on the basis of every scientific measurement we could make, both external and internal – behaved identically to the way we do when we're in pain, but that did not feel anything at all, only *acting* exactly as if it did? Where's the *pain* in the biological description?

The mind/body problem does not stop with the problem of how to equate mental states with physical states. There is also the "other minds" problem. How do I know that anyone else but me really has a mind? Couldn't they just be behaving exactly as if they had a mind – exactly as I would under the same conditions – but without experiencing anything at all between their ears while doing so: just mindless physical bodies going through the motions?

There is a very close connection between the philosopher's old other-minds problem and the modern computer scientist's problem of determining whether artificial devices have minds [Harnad (1984)]. Unfortunately, there is no solution in either case – if a solution would require certainty (as in mathematics) or even high probability on the available evidence (as in science). The reason there's no solution is that the mind/body problem is unique. There is in fact *no evidence* for me that anyone else but me has a mind. In the case of any other kind of objective, scientific data (such as laboratory observations and measurements), I can in principle check it out directly for myself. But I can never check whether anyone else but me has a mind. I either accept the dictates of my intuition (which is the equivalent of "If it looks like a duck, walks like a duck, quacks like a duck . . . it's a duck") or I admit that there's no way I can know.

Suppose we adopt another strategy and treat "have a mind" as an unobservable property or state of others, one whose existence we *infer* on the basis of indirect evidence in exactly the same way that physicists infer the existence of their unobservables, such as quarks, gravitons, superstrings or the Big Bang. This still

won't do, unfortunately, because unlike the case of, say, superstrings, without whose existence a particular contemporary theory in physics simply would not work (i.e., it would not succeed in predicting and explaining the available evidence), in the case of the mind all possible empirical evidence (and any theory of brain function that explains it) is just as compatible with the assumption that a candidate is merely behaving exactly as if it had a mind (but doesn't) as with the assumption that a candidate really has a mind. So unless we're prepared to be dualists (i.e., ready to believe that there are two kinds of "stuff" in the world – physical and "mental" – each with causal powers and laws of its own [Alcock (1987)], "having a mind" can't do any independent theoretical work for us [cf. Fodor (1980)] in the way that physics' unobservables do. Hence, consciousness can be affirmed or denied on the basis of precisely the same evidence [Harnad (1982, 1989b)].

The variant of the TT proposed here – the TTT – accordingly just amounts to calling a spade a spade. If our only basis for judging that other people have minds is that they behave indistinguishably from ourselves, then there's no ground for withholding the same benefit of the doubt from robots. I differ from Turing in holding that our sensorimotor capacities – our ability to interact bodily with the things in the world in the many nonverbal ways we do – are as important to this test as our linguistic capacities; moreover, there are strong reasons to doubt that a device could pass the teletype version of the Turing Test if it were not also capable of passing the robot version. It is highly unlikely that our linguistic capacities are independent of our robotic capacities. Successfully passing the teletype version of the Turing Test alone may be enough to convince us that the candidate has a mind (just as written correspondence with a never-seen pen-pal would), but full robotic capacities (even if only latent ones, not directly exhibited or tested in the TT) may still be necessary to generate that successful linguistic performance in the first place.

What's the evidence for this? Well, first, as already suggested, this is not really a matter of "evidence" at all, at least not scientific evidence. You either accept Turing's criterion or you don't (and most of us instinctively accept and use it all the time, without even realizing it). But biological evolution[1] does seem to suggest that robotic capacities come before linguistic ones, because there are many species that have robotic capacity without linguistic capacity, but none (so far as I know) that have linguistic capacity without robotic capacity. It is hard to imagine, for example, that a TT candidate could chat with you coherently about the objects in the world till doomsday without ever having encountered any objects directly – on the basis of nothing but "hearsay," so to speak. Some prior direct acquaintance with the world of objects through sensorimotor (TTT) interactions with them would appear to be necessary in order to *ground* the candidate's words in something other than merely more words. This "symbol grounding problem" is analogous to the problem of trying to learn the meanings of words from a dictionary. Some of the words' meanings would have to be

grounded in previous nondictionary experience; otherwise, the dictionary could only send us round and round in endless circles, from one meaningless string of symbols to another. Our linguistic capacity must be similarly grounded in our robotic capacity [Harnad (1990a)].

Searle's Chinese Room

More support for the primacy of robotic capacity comes (inadvertently) from a very provocative "thought experiment" originally proposed by the philosopher John Searle as a challenge to the Artificial Intelligence (AI) research community ten years ago [Searle (1980a)]. His thought experiment, called the "Chinese Room Argument," shook up the entire AI field considerably, and things still have not settled down since. Searle's challenge was also aimed at Turing Testing itself, but, as I will try to show, it turns out that only the teletype version, the TT, is vulnerable to it.

For those who believe that the TT is the decisive test for having a mind, suppose there is a computer that is capable of passing it in Chinese. Whatever Chinese symbols are typed on the teletype, the machine will respond with symbols exactly the way a Chinese pen-pal would. No one can tell (even if they keep testing for a lifetime) that it's a machine and not a person. Now those who are convinced by this would be prepared to conclude that the computer had a mind and hence understood the Chinese symbols. But how could this be, challenges Searle, given that the only thing the computer is really doing is following rules for manipulating symbols on the basis of their shapes? What the machine does can all be summarized by a table of symbols such as "If 'squiggle-squiggle' comes in, send 'squoggle-squoggle' out."

Now, says Searle, suppose I myself took the place of the computer and followed the instruction table for manipulating the symbols. I could do that without ever understanding what "squiggle-squiggle" or "squoggle-squoggle" meant. In particular, if the symbols were Chinese, I would no more understand Chinese if I were doing the symbol-manipulation than I do now! [Searle understands no Chinese.] So if I would not be doing any understanding under those conditions, neither would the computer whose functions I was duplicating. So much for the Turing Test and the minds of machines.

The AI field was left in a great turmoil by this simple argument [e.g., Wilensky (1980); Dennett (1982); McDermott (1982); Carleton (1984); Harvey (1985); Rey (1986); Macqueen (1989); Searle (1980b), (1982a), (1982b), (1985a), (1989), (1990a), (1990b); Dietrich (1990); Harnad (1990)], because its practitioners had become accustomed to confidently claiming that their little toy demos of artificial intelligence – chess-playing programs, question-answering programs, scene-describing programs – were *already* displaying the rudiments of mind, and that just more of the same would eventually be enough to pass the Turing Test and capture all of the mind.

AI had had two other reasons – over and above (1) its successes in getting computers to do intelligent things – for believing that computers could have minds. Despite some abstract limitations, about which only a minority worried [e.g., Lucas (1961); Slezak (1982); Davis (1985); Penrose (1990)], (2) computation was provably very powerful and general [e.g., Kleene (1969)], capable of simulating just about anything ("Turing Equivalence"). If computers could simulate airplanes and furnaces, why not minds too? And finally, (3) the software/hardware distinction – the implementation-independence of the computational level of function – seemed to bring some closure to the mind/body problem [Pylyshyn (1984)]. For if mental states were just computational states then that would explain their peekaboo relation to their physical implementations! What makes them mental is their computational properties, not their physical properties: the very same mental states can be implemented in countless physically different ways.

Now, it is precisely the latter "teleportability" property (3) that Searle's argument exploits: Ordinarily, the other-minds barrier ensures that you cannot know one way or the other whether your guess that another body has a mind is correct. The only way you could know would be by *being* the other body. But this is exactly what Searle manages to do, because of teleportability. He still can't say whether that computer over there understands Chinese. It might, just as surely as a stone might. But he can say for sure that, if it does understand Chinese, it's certainly *not* because of the computational state it's implementing, because Searle is implementing the very same computational state – and he is able, unlike the computer, to *tell* us that he is understanding no Chinese! Searle thereby shows not only that teleportability (3) was much more of a liability than an asset when it came to the other-minds problem, but that Turing Equivalence (2) was no help either. For just as surely as a computer-simulated airplane, be it ever so Turing Equivalent to a real airplane, cannot really fly, and a simulated furnace cannot heat, so a computer-simulated mind cannot think, even if it's Turing Equivalent to a real mind that can.

AI's favored "System Reply" to Searle [e.g., Wilensky (1980); McDermott (1982); Dietrich (1990)] – to the effect that Searle had erred about *where* the mind would be in his Chinese Room: it wouldn't be in his head but in the "system" as a whole – was, I think, a predictable piece of hand-waving, given the rest of the far-fetched claims AI had already gotten so used to making. And Searle successfully lampooned it in the eyes of all but the unshakeable true believers in pointing out that he didn't see anything else on which to hang a mind amongst the blackboard on which the rules were written or the walls holding up the ceiling of the Chinese room. That, after all, was all there was to the "system", aside from Searle himself.

And if that was not enough to convince System Repliers that there was no one else home, Searle went on to remind them that he could memorize all the symbols and symbol manipulation rules. Then the entire "system" would be between his ears. Yet even this reply has been challenged with a "multiple personality"

counterargument [Dyer (1990)], which is in turn easily laid to rest as pure hypothesis-saving at a mounting counterfactual price; cf. Harnad (1990). Why are people willing to resort to sci-fi fantasies to continue believing that their "systems" think? Perhaps it's partly the hermeneutical power of the linguistic version of the Turing Test (TT) itself – the same tendency that made us so ready to believe that chimpanzees could speak once we had projected an English translation onto their symbols [Terrace (1979)]. All the more reason for turning to the stronger version of the TT that is being proposed here, the TTT, to immunize us against the kind of overinterpretation of symbols that can blur even the distinction between simulated airplanes and real ones.

Searle's argument against the TT and symbol manipulation has even been underestimated by thinkers who are no friends of AI. According to Churchland (1990), who is normally a neuromolecular (TTTT) partisan – as is Searle too, for that matter – Searle's failure to understand Chinese when he manipulates the symbols in the Chinese room would not refute the theory that thinking is symbol manipulation any more than his failure to see light when he waves a magnet would refute the theory that light is electromagnetic oscillation. But slow electromagnetic oscillations would indeed be "light"! One would fail to see to see it simply because its frequency was below the retinal threshold for visible light. And, as far as I know, there's no basis for expecting a corresponding quantitative threshold – some sort of "phase transition" from the physical into the mental – in the case of symbol manipulation (whether in terms of speed, capacity, or complexity). Besides, the "visibility" of the visible portion of the spectrum is irrelevant to physics, be it ever so central to mind-modeling. So Searle's argument is correct after all.

The Total Turing Test (TTT) and Sensorimotor Grounding

But was Searle too glib? And is our knee-jerk skepticism about machines – the kind that impels us to make unreasonable demands, such as Einsteinian performance – really justified? I will close by showing that the variant of the Turing Test I have proposed – the TTT, calling for both linguistic *and* robotic capacity – is not only immune to Searle's Chinese Room Argument but, as already suggested, turns out to be no less (nor more) exacting a test of having a mind than the means we already use with one another in our everyday, practical solutions to the "other minds" problem. It's just a more rigorous and systematic version of the same test. And there is no stronger test, short of *being* the candidate.[2]

Suppose that the critical question we focused on in our TTT candidate's performance was not whether it understood symbols, but whether it could *see*. The questions, "Is it really intelligent?", "Does it really understand?", "Can it really see?", are all just variants of the question, "Does it really have a mind?"[3] It is important to bear in mind in what follows that all these questions are embedded in the TTT, which, among other things, calls for both seeing *and* understanding in the same candidate. Nor is the particular sense-modality singled

out here (vision) critical. It might have been any other modality. But before critics hasten to remind me that blind people have minds too, let me remind them that it's more advisable methodologically to capture the normal case first, before trying to model its pathologies (which is not at odds with Claude Bernard's suggestion that an organ's pathologies might yield *clues* to its normal function). And although vision can be subtracted from a body leaving the mind intact, the converse does not follow: An independent visual model is not possible, one that can do nothing but "see." (Nor is it likely that *all* sensory capacities could be subtracted and still leave a body with a mind.)

So in the TTT variant of Searle's thought experiment there would again be two possibilities, just as there were in the Chinese Room. In the original TT case, the machine could either really be understanding Chinese or it could be merely going through the motions, manipulating symbols *as if* it understood them. Searle's argument worked because Searle himself could do everything the machine did – he could *be* the whole system – and yet still be obviously failing to understand.

In the TTT case of seeing, the two possibilities would again be whether the machine really saw objects or simply *acted exactly as if* it did. But now try to run Searle's argument through. Searle's burden is that he must perform all the internal activities of the machine – he must *be* the system – but without displaying the critical mental function in question (here, seeing; in the old test, understanding). Now machines that behave as if they see must have sensors – devices that transduce patterns of light on their surfaces and turn that energy into some other form (perhaps other forms of energy, perhaps symbols). So Searle seems to have two choices. Either he gets only the *output* of those sensors (say, symbols), in which case he is *not* doing everything that the candidate device is doing internally (and so no wonder he is not seeing -- here the "System Reply" would be perfectly correct); or he looks directly at the objects that project onto the device's sensors (i.e., he is *being* the device's sensors) – but then he would in fact be seeing!

What this simple counterexample points out is that symbol-manipulation is not all there is to mental function and that the linguistic version of the Turing Test just isn't strong enough, because linguistic communication could in principle (though perhaps not in practice) be no more than mindless symbol manipulation. The robotic upgrade of the TT to the TTT is hence much more realistic. It requires the candidate to interact with the world (including ourselves) in a way that is indistinguishable from how people do it in *every* respect, both linguistic and nonlinguistic.

That mere sensory transduction can foil Searle's argument should alert us to the possibility that sensorimotor function may not be trivial – not merely a matter of adding some simple peripheral modules (like the light-sensors that open the doors of a bank when you approach them) to a stand-alone symbol manipulator that does the real mental work. Rather, to pass the Total Turing Test, symbolic

function may have to be grounded "bottom up" in *nonsymbolic* sensorimotor function in an integrated, non-modular fashion not yet contemplated by current computer modelers. For example, in Harnad (1987), one possible bottom-up symbol-grounding approach is described in which the elementary symbols are the names of perceptual categories picked out by sensory feature detectors from direct experience with objects. Nonsymbolic structures, such as analog sensory projections and their invariants (and the means for learning them), would play an essential role in grounding the symbols in such a system, and the effects of the grounding would be felt throughout the system.

The Mind/Body Problem Revisited

What more can we ask, over and above the TTT? In fact, what more *do* we ask if ever we are challenged to say why we believe that other people have minds? All we can offer by way of justification for the assumption that other bodies have minds is that we can't tell them apart from someone with a mind, like our own. And the candidate certainly doesn't have to be an Einstein for that.

Now, if you are still not entirely comfortable with equating having a mind with having the capacity to produce Turing-indistinguishable bodily performance, then welcome to the mind-body problem! By way of consolation, consider the equivalent problem for physicists. Instead of being mind-modelers, they are world-modelers. They try to discover the laws that explain how the world of inanimate material bodies works. But what answer can they give if someone says, "Ah, yes, your laws seems to fit the world, but how do we know that that's the way the world really works? Your theoretical model is only Turing-indistinguishable from the real world. It only behaves exactly *as if* it were the real world. Maybe the real world hews to different laws." The physicist has no answer to this, nor does he need one, because a Turing-indistinguishable world – the ultimate as-if model – is all that physics can ever offer – and all it aspires to offer.

If there are rival candidates, physics always prefers the model that captures the *most* about how the world behaves. This too has a parallel in the "Totality" requirement of the Total Turing Test (TTT). As a model accounts for more and more of the data, the number of alternative lookalike ways that this can successfully be accomplished should be shrinking. Add to this the requirement that a model should be no more complicated or extravagant than necessary, and the number of possible lookalikes shrinks still more. Finally, although Turing Testing can probably go a long way without worrying about anatomy and physiology, eventually brain performance must join bodily performance as part of our Total behavioral capacity, giving the ultimate model its last round of fine tuning as the "TTTT". In the end, our bodily "behavior" must include all the observable data on what every relevant part of our body does, even its neurons and molecules. [My own guess, however, is that the TTTT will be unnecessary –

that all the substantive problems of mind-modeling will already have been solved once we know how to make a candidate that can pass the TTT. On the other hand, neural (TTTT) data might conceivably give us some clues about how to pass the TTT.]

But why is the "as-if" objection so much more troubling in the case of mind-modeling than in the case of world-modeling? Because in physics, what a "Total" model might be missing is a matter that, by definition, can have no bearing on us whatsoever. We have no source of "direct" information about the real world by way of an alternative to physical data and the models that explain them. The difference between being and not being "really" right (as opposed to only "as-if" right) is a difference that can make no palpable difference to us. But in the case of the mind we do have an alternative source of direct information: our own private experience, our own individual minds.[4] So only the candidate itself can know (or not know, as the case may be) whether we're really right in thinking that it has a mind. But that is a potentially palpable difference, at least for one body. And we *all* know perfectly well exactly what it is that the model would be missing if it were just going through the motions, without any subjective experience: There would be nobody home.

The idea that subjective experience does not and cannot figure directly in a model of the mind – not even the way a quark does – but must always be taken on faith, is apparently too much for us to swallow. In fact, it's about as easy to accept as the idea that we don't really have free will, but only feel and behave *as if* we did [see Harnad (1982, 1989b); Libet (1985)]. Yet this too is one of the dictates of mind-modeling, for all modeling is cause/effect modeling, and (except if we are to be dualists or spiritists) there is only room for one kind of cause in this world, and hence in models of it: bodily cause [cf. Alcock (1987)].

Some of us are indeed ready to become dualists because of the unpalatability of this picture, ready to believe that the world of the mind is distinct from and partly (or wholly) independent of that of the body – a world to be understood and explained on its own unique terms [e.g., Popper and Eccles (1977)]. Others take the opposite course and hold either that only the world of bodies is real while the rest is illusion and misconception [e.g., Churchland (1988); Dennett (1988)], or that minds and bodies are really one and the same thing [Armstrong (1981)]. Still others – and I count myself among these – are persuaded by the reality and distinctness of both bodies and minds, and conclude that, whatever we might have expected, objective models – be they ever so "Total" – are doomed to be incomplete, even when they have explained all the objective facts, because there will always remain one fact unaccounted for, the fact of subjectivity itself [Nagel (1974, 1986)]. It thus looks as if no scientific answer can be expected to the question of how or why we differ from mindless bodies that simply behave exactly as if they had minds. The good news is that this difference makes no objective difference. The bad news is that it continues to make a subjective difference, one that cannot be explained away.

Notes

[1] Note that evolution is as blind to Turing-indistinguishable differences as we are. So it's a complete waste of time to search for the "adaptive value" of having a mind [cf. Humphrey (1984); the same error was made in Harnad (1982)]. The reason for this is that TTT (behavioral) and TTTT (neuromolecular) differences exhaust all the possible differences that can make an empirical difference – which of course includes all possible "selective advantages."

[2] There is of course also the TTTT, the Total Total Turing Test, calling for Turing indistinguishability right down to the neurons and molecules, which is as much as an empiricist could ever hope for. My own guess, though, is that the TTTT would be methodologically supererogatory. The TTT already narrows the degrees of freedom sufficiently to converge on a model that is within the bounds of the normal range of underdetermination of scientific theory by empirical data.

[3] Philosophers have tried to partition the mind/body problem into two semi-independent subproblems: the problem of (1) consciousness (or what it is like to be in a mental state, which is what has been the main focus of this paper) and the problem of (2) "intentionality," which is whatever it is that makes mental states be "about" things in the world. But note that if there weren't something it was like to be in a mental state that is about something in the world (i.e., (1)), then the difference between "real" and "as-if" intentionality (2) would vanish completely. So, (2) seems to be completely parasitic on (1). Hence, until further notice, there is only one mind-body problem, and all questions about mental predicates – about real vs. as-if feeling, seeing, desiring, believing, knowing, understanding, etc. – address the same problem: Is there really somebody home, and if so, in what does that state consist?

[4] And to make things still more complicated, even the objective data of physics come wrapped inextricably in the subjective data of direct experience. Scientific data are rightly called "observations."

References

Alcock, J. E. (1987), 'Parapsychology: Science of the Anomalous or Search for the Soul?', *Behavioral and Brain Sciences* **10**, pp. 553–643.

Armstrong, D. M. (1981), *The Nature of Mind*, NY: Cornell University Press.

Churchland, P. A. (1988), *Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind*, Cambridge, MA: MIT Press.

Churchland, P. A. (1990), 'Could a Machine Think?', *Scientific American* **262**, pp. 32–37.

Carleton, L. (1984), 'Programs, Language Understanding and Searle', *Synthese* **59**, pp. 219–230.

Davis, M. (1958), *Computability and Unsolvability*, Manchester: McGraw-Hill.

Davis, M. (1965), *The Undecidable*, New York, NY: Raven.

Dennett, D. C. (1988), 'Precis of: *The Intentional Stance*', *Behavioral and Brain Sciences* **11**, pp. 495–546.

Dennett, D. C. (1982), 'The Myth of the Computer: An Exchange', *New York Review of Books* **XXIX** (11), p. 56.

Dietrich, E. (1990), 'Computationalism', *Social Epistemology* **4**, pp. 135–154.

Dyer, M. (1990), 'Intentionality and Computationalism: Minds, Machines, Searle, and Harnad', *Journal of Experimental and Theoretical Artificial Intelligence* **2**(4), pp. 303–319.

Fodor, J. A. (1980), Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology', *Behavioral & Brain Sciences* **3**, pp. 63–109.

Fodor, J. A. (1985), 'Précis of "The Modularity of Mind"', *Behavioral and Brain Sciences* **8**, pp 1–42.

Harnad, S. (1982), 'Consciousness: An afterthought', *Cognition and Brain Theory* **5**, pp. 29–47.

Harnad, S. (1984), 'Verifying Machines' Minds', *Contemporary Psychology* **29**, pp. 389–391.

Harnad, S. (1987) 'Category Induction and Representation', in S. Harnad, ed., *Categorical Perception: The Groundwork of Cognition*, New York, NY: Cambridge University Press.

Harnad, S. (1989a), 'Minds, Machines and Searle', *Journal of Experimental and Theoretical Artificial Intelligence* **1**, pp. 5–25.

Harnad, S. (1989b), 'Editorial Commentary on Libet (1985)', *Behavioral and Brain Sciences* **12**, p. 183.
Harnad, S. (1990a), 'The Symbol Grounding Problem', *Physica D* **42**, pp. 335–346.
Harnad, S. (1990b), 'Commentary on Dietrich's (1990) "Computationalism" ', *Social Epistemology* **4**, pp. 167–172.
Harnad, S. (1990c), 'Lost in the Hermeneutic Hall of Mirrors', *Journal of Experimental and Theoretical Artificial Intelligence* **2**(4), pp. 321–327.
Harvey, R. J. (1985), 'On the Nature of Programs, Simulations and Organisms', *Behavioral and Brain Sciences* **8**, pp. 741–2.
Haugeland, J. (1985), *Artificial Intelligence: The Very Idea*, Cambridge, MA: MIT/Bradford Press.
Humphrey, N. (1984), *Consciousness Regained: Chapters in the Development of Mind*, Oxford, UK: Oxford University Press.
Kleene, S. C. (1969), *Formalized Recursive Functionals and Formalized Realizability*, Providence, RI: American Mathematical Society.
Libet, B. (1985), 'Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action', *Behavioral and Brain Sciences* **8**, pp. 529–566.
Lucas, J. (1961), 'Minds, Machines and Gödel', *Philosophy* **36**, pp. 112–117.
MacQueen, N. D. (1989), 'Not a Trivial Consequence', *Behavioral and Brain Sciences* **13**, pp. 163–4.
McDermott, D. (1982), 'Minds, Brains, Programs and Persons', *Behavioral and Brain Sciences* **5**, pp. 339–341.
Minsky, M. (1961), 'Steps Towards Artificial Intelligence', *Proceedings of the Institute of Radio Engineers* **49**, pp. 8–30.
Nagel, T. (1974), 'What Is It Like to Be a Bat?', *Philosophical Review* **83**, pp. 435–451.
Nagel, T. (1986), *The View from Nowhere*, New York NY: Oxford University Press.
Newell, A. (1980), 'Physical Symbol Systems', *Cognitive Science* **4**, pp. 135–83.
Penrose, R. (1990), 'Precis of: *The Emperor's New Mind*', *Behavioral and Brain Sciences* **13**, pp. 643–706.
Popper, K. R. and Eccles, J. C. (1977), *The Self and Its Brain*, Heidelberg, FRG: Springer, 1977.
Pylyshyn, Z. W. (1984), *Computation and Cognition*, Cambridge, MA: Bradford Books.
Rey, G. (1986), 'What's Really Going on in Searle's "Chinese Room"?', *Philosophical Studies* **50**, pp. 169–185.
Searle, J. R. (1980a), 'Minds, Brains and Programs', *Behavioral and Brain Sciences* **3**, pp. 417–424.
Searle, J. R. (1980b), 'Intrinsic Intentionality', *Behavioral and Brain Sciences* **3**, pp. 450–457.
Searle, J. R. (1982a), 'The Chinese Room Revisited', *Behavioral and Brain Sciences* **5**, pp. 345–348.
Searle, J. R. (1982b), 'The Myth of the Computer: An Exchange', *New York Review of Books* **XXIX**(11), pp. 56–57.
Searle, J. R. (1985a), 'Pattern, Symbols and Understanding', *Behavioral and Brain Sciences* **8**, pp. 742–743.
Searle, J. R. (1985b), *Minds, Brains and Science*, Cambridge, MA: Harvard University Press.
Searle, J. R. (1989), 'The Causal Powers of the Brain', *Behavioral and Brain Sciences* **13**, p. 164.
Searle, J. R. (1990a), 'Is the Brain's Mind a Computer Program?', *Scientific American* **262**, pp. 26–31.
Searle, J. R. (1990b), 'Consciousness, Explanatory Inversion and Cognitive Science', *Behavioral and Brain Sciences* **13**, pp. 585–642.
Slezak, P. (1982), 'Gödel's Theorem and the Mind', *British Journal for the Philosophy of Science* **33**, pp. 41–52.
Terrace, H. (1979), *Nim*. New York, NY: Random House.
Turing, A. M. (1964), 'Computing Machinery and Intelligence', in A. Anderson, ed., *Minds and Machines*, Englewood Cliffs, NJ: Prentice Hall.
Wilensky, R. (1980), 'Computers, Cognition and Philosophy', *Behavioral and Brain Sciences* **3**, pp. 449–450.