

Stevan Harnad

## There is Only One Mind/Body Problem

### Abstract

In our century a Frege/Brentano wedge has gradually been driven into the mind/body problem so deeply that it appears to have split it into two: The problem of “qualia” and the problem of “intentionality.” Both problems use similar intuition pumps: For qualia, we imagine a robot that is indistinguishable from us in every objective respect, but it lacks subjective experiences; it is mindless. For intentionality, we again imagine a robot that is indistinguishable from us in every objective respect but its “thoughts” lack “aboutness”; they are meaningless. I will try to show that there is a way to re-unify the mind/body problem by grounding the “language of thought” (symbols) in our perceptual categorization capacity. The model is bottom-up and hybrid symbolic/nonsymbolic.

I don’t know about you, but I find that ONE mind/body problem is more than enough for me. Yet these days the problem seems to have bifurcated, and one of its successors seems to be capable of spawning still more mind/body problems. The original mother of all mind/body problems was simply enough stated: It was a persistent conceptual difficulty we all had with equating mental and physical states — with squaring the felt quality of pain, for example, with any functional or neurophysiological story anyone might tell us. The problem was clearly with the felt quality of mental states — qualia — and so far no philosopher or cognitive scientist has solved the problem of what to make of qualia. Yet it is felt that cognitive science has made some overall progress in explaining mental states — not their felt quality but another of their unique properties, namely, their “intentionality” or aboutness.

Qualia are no longer the unique mark of the mental, for philosophers had noted that mental states have another property that seems to be peculiar to them, namely, that they are ABOUT something: I believe X, I want X, I think X. There is always an object of my mental states; this cognitive connection between mental

states and the world seemed more tractable than the qualia problem, so cognitive science has accordingly focussed on it almost to the exclusion of the other mind/ body problem:

How are intentional states to be accounted for physically? Several candidate answers exist — intentionality could be a property of certain neural states, or it could be a property of certain functional states (independent of the specifics of their neural implementation). The functional states could be “narrow” ones, consisting only of what goes on in the brain of the organism that has the intentional states, or the functional states could be “wide” ones, including in them objects and properties of the outside world. Wide functionalism is clearly no longer a theory of MENTAL states, since mental states should surely fit within the head of the organism that has them. But narrow functionalism and other forms of neurophysicalism are in principle candidates for explaining intentional states as mental states. And they have the advantage that “aboutness” is such an abstract property that it does not excite the kinds of contrary intuitions that qualia invariably do, whenever someone ventures a physical or functional account of them.

The intuition pumps are the same in both cases, but in the case of intentionality they simply fail to dredge up much of anything. In both cases we can readily imagine a candidate that is constructed according to the favorite theory — physical or functional — and our concern is whether the candidate really has the mental property at issue: qualia in the one case and intentionality in the other. In both cases we can admit the possibility that the candidate may lack the mental property in question — qualia or intentionality, respectively — and that, because of the other dimension of the mind/body problem, the other- minds problem, there is no way we can know for sure. But in the case of qualia, the immediacy of our sense of what the candidate would lack (if it lacked it) — because we all know exactly what it would lack, as we all know what it feels like<sup>1</sup> to have qualia — makes us much less credulous about assurances that qualia have actually been captured by whatever properties our theory — whether physical or functional — singles out and recommends to us. In the case of aboutness, on the other hand, we are much more easily persuaded that a theory has successfully captured it, because, apart from qualia, we actually have no idea of what such a candidate would be LACKING if it lacked aboutness!

Some intuitions can help make us more steadfast about aboutness: We can be reminded, as we are by Searle, that the aboutness of mental states cannot be like the aboutness of sentences in a book, because the latter are only about what

<sup>1</sup> I have substituted „what it feels like,” passim, for Nagel’s justly famous „what it’s like” criterion in order to bring out the critical felt-quality that really underlies its application in all instances.

they are interpretable as being about because creatures like us, who have mental states with REAL aboutness, actually do interpret them that way. So it would not do if the candidate system that allegedly had real aboutness were merely like a book, with derivative aboutness. So far so good; but what is REAL, nonderivative, intrinsic aboutness, then, apart from the fact that it only seems to reside in the heads of creatures like ourselves?

The Turing Test (TT) has been proposed as the criterion: Surely the internal states of a candidate that could interact with us as a pen-pal till doomsday, completely indistinguishably from a real pen-pal, would have real aboutness, and not just interpretable-as-if aboutness, like a book. One might even have thought that the TT was the best criterion we could hope for, empirically and intuitively, had Searle (1980) not shown that, at least in one special case, there was a way of intuiting exactly what the candidate would be lacking — and that it would indeed be lacking it: For if the candidate happens to be a computer passing the TT in Chinese, and the claim is the usual functionalist one to the effect that it is only the programme that matters, the details of the implementation being irrelevant, then Searle could memorize the program — it's really just a bunch of meaningless formal symbols and symbol manipulation rules — and then pass the TT till doomsday just as the computer does, and by exactly the same means, going through exactly the same functional states, yet he could truthfully report that he does not understand Chinese in so doing, and that the mental processes with the understanding (aboutness) that his real pen pal is imputing to him are not really there; and hence that they are not there in the computer either.

This successful demonstration would on the face of it seem to count in favor of the independence of the aboutness problem from the qualia problem, and hence it would seem to support, rather than refute, the existence of more than one mind/body problem. But if we look a little more closely, we will discover that the success of Searle's argument is completely parasitic on qualia in just the same way that the derivative, extrinsic intentionality of a book is parasitic on the elusive "real," nonderivative, intrinsic intentionality in our heads. For the intuitive discrimination on which Searle's Argument draws is just the distinction between understanding and not understanding what something is about, in other words, it is a qualitative distinction, exactly the distinction that underlies the difference between what I am saying now and "es amit pedig most mondok" — at least for those of you who do not understand Hungarian.

It is, in other words, the fact that we all know what it FEELS like to mean X, to understand X, to be thinking about X, that allows us to see what the TT-passing computer would be lacking. Hence the absence of "aboutness" is really just the absence of certain kinds of qualia — the kind that accompany our understanding of natural language. I will expand on the implications of this in

a moment — the most important of them is that there is only one mind/body problem after all — but first I want to lay to rest a few potential objections.

The most typical of these objections is: How can Searle KNOW that he doesn't understand Chinese? First of all, no one has even defined what it means to understand Chinese. And second of all, there are plenty of cases when we don't know things we think we do know and vice versa: Why can't Searle's successful TT PERFORMANCE be the criterion for whether or not he understands Chinese? So if that criterion is met, let's just think of it as another one of those "implicit" skills we have without knowing it consciously, and without knowing how we do it.

I think this is an entirely unsatisfactory line of objection. The short version of the refutation would be that this kind of lifelong coherent "speaking in tongues" is about as likely as telepathy and clairvoyance in ordinary human life, so the analogies on which it is based — partial knowledge of a foreign language, the unconscious substrates of skills, implicit learning, knowledge and memory, etc. — are simply irrelevant. Nor is there any need to define "understanding a language" in order to know whether you do or you don't (that was the point of the Hungarian example a moment ago). No, this kind of objection is motivated rather by something akin to the wrong-headedness of Freudianism, in which minds were likewise multiplied beyond necessity, and unconscious minds were posited, very much like our conscious ones in every respect, except that they were nonconscious. Now for the original mind/body problem it is quite obvious why positing such an entity would be gratuitous: If you thought it was hard to equate conscious qualia with physical stuff, imagine the problem with elusive "unconscious" qualia: Who, as a first approximation, would be the subject of those qualia, and is there anything it feels like to be that Freudian alter ego? For if not, then what justification is there for speaking about any of this unfelt stuff as another "mind" at all?

So let us let go of the notion of "unconscious aboutness," which is just as arbitrary and superfluous as "unconscious qualia," for all attributions of MENTAL properties to UNconscious processes are, until further notice, parasitic on the mental properties of conscious processes, once again parasitic, in a word, on qualia: Only minds with conscious qualia can be spoken of as having unconscious processes at all. Otherwise, what we are talking about is the kind of NONconscious process that occurs in a rock — in which case it is clearer, I hope, why it should not be called UNconscious.

At the other extreme from the supernumerary minds in Freudian thinking are the superordinate mental states that are the focus of the child and animal "theory of mind" literature. Here too, some theorists think they have a phenomenon that is independent of the qualia problem: How does a child or an animal know that another child or animal knows, or even that it itself knows?

I think it should be even clearer here how this second-order problem — of mental states that have other mental states as their objects — is based completely on the presence of first-order mental states in the first place. From that standpoint, higher-order intentionality and other-minds theories are simply particular mental states, with their own particular qualitative contents: As conscious human beings, we not only know what it feels like to feel pain, but also what it feels like to understand English (and not understand Hungarian), what it feels like to do something deliberately and rationally (versus what it feels like to do something automatically or even inadvertently: “The Devil made me do it!”), as well as what it feels like to know that you have such experiences, and to infer that others do too. Those are all just mental states with different qualitative contents.

Well, having argued at length for the primacy of qualia, suppose the point was accepted; what next then? What is one to do about it? If all the questions about the presence of real mental states as opposed to as-if lookalikes are really just questions about the presence or absence of certain qualia, what are we to make of that, especially since qualia are the toughest conceptual nut to crack, when it comes to the mind/body problem.

First, having established myself as a friend of qualia, let me immediately reveal a much less popular methodological stance I also advocate, one that immediately seems to take back what I’ve been at pains to establish primacy for: I am a methodological epiphenomenalist: Although I believe qualia are real and irreducible, and the sine qua non of mental states, I also believe that they will not enter at all into that Utopian true, complete theory of behavior and brain function that will one day be generated by that branch of reverse bioengineering that is cognitive science. The reason is that qualia play no independent causal role: They merely piggy-back on the real causality. The best we can ever hope to provide is a model (or models) that can pass, not only the TT (which is discredited until further notice) but the TTT, the Total Turing Test, which requires the candidate to exhibit not only a lifetime’s worth of our pen-pal capacities (i.e., our linguistic capacities) Turing-indistinguishably from ourselves, but also a lifetime’s worth of our robotic capacities — all of our sensorimotor interactions with the world, categorizing and manipulating all the objects, events and states of affairs that our words are interpretable as being about. It is on top of the functional capacity to pass the TTT — having ruled out all the degrees of freedom of lesser candidates — that I believe qualia will piggy-back. And if not, then we cannot hope to be any the wiser, because the TTT, unlike the TT, is immune to Searle’s Argument, and hence subject to that extra order of underdetermination

- over and above the ordinary underdetermination of scientific theories by data
- arising from the mind/body problem and its other-minds lemma.

Let me close with a sketch of what I would recommend as a stand-in for qualia in cognitive theory. I think the problem of aboutness is closely related to

what I've called the "symbol grounding problem." The approach to cognitive theory that Searle showed to be impracticable has been variously called "Strong AI," "computationalism" and "symbolic functionalism" (SF). According to SF, mental states are just computational states, in other words, implemented but implementation-independent symbols manipulated according to formal (syntactic) rules. Searle has shown that this cannot be correct. My hypothesis as to WHY it can't be correct is that in an implemented formal symbol system, as in a computer program passing the TT, the meanings of the symbols are "ungrounded." They don't mean anything TO the system; they are merely systematically interpretable as meaningful to us by us. Such a system is like a Chinese-Chinese dictionary, which, though fully interpretable and even a source of new word meanings to a Chinese speaker, cannot provide meanings to a non-speaker of Chinese. For the latter they are just ungrounded squiggles and squoggles with certain systematic formal interrelations (which is what they are in reality).

How could one ground the meaning of the symbols in such a system in something other than just the interpretation of an outside mind? The TTT suggests a way: The TT-scale symbolic capacity must be grounded in the TTT-scale robotic capacity. The "aboutness" of the system's words and thoughts is no longer just parasitic on their systematic interrelations and interpretability: It is also grounded in the system's capacity to categorize and manipulate that analog world of objects, events and states of affairs that the symbols are interpretable as being about. The TTT-system is doubly constrained — symbolically and robotically — whereas the TT-system is constrained only symbolically, and hence ungrounded.

Grounding is by its nature a bottom-up affair. Hence my approach emphasizes sensorimotor categorization and manipulation. The idea is that the system learns to identify concrete perceptual categories from sampling positive and negative instances together with corrective feedback from the consequences of miscategorization. Once a perceptual category can be reliably identified, it can be assigned a name. Such a name is a grounded symbol. It can then be combined with other grounded symbols into symbol strings that are interpretable as statements about category membership, with new, higher-order symbols inheriting the grounding. For example, if "horse" and "stripes" are grounded, "zebra" is grounded by the proposition: "A zebra is a striped horse." And so on, all the way up to goodness, truth and beauty.

There are prominent objections to this kind of hybrid bottom-upism, and I have tried to answer them elsewhere. The short answer is that this kind of approach cannot be evaluated from the armchair and it has never really been tried. Moreover, ungrounded "top-down" alternatives to this approach are hanging from a symbolic skyhook. Let me close, however, by returning to qualia. It turns out that certain perceptual categories display a surprising qualitative effect that has been called "categorical perception" (CP). Equal sized physical variations in

the stimulus are in some special cases not perceived as equal-sized: They are perceived as smaller if they are differences within a category and bigger if they are differences between categories. CP has been interpreted as an instance of that interaction between language and perception that also goes by the name of the “Whorf-Sapir” hypothesis. How we name and describe the world influences what it looks like to us.

Most of the examples of CP involve innate categories such as colors and phonemes. In recent experiments with Jan Andrews and Ken Livingston at Vassar we have confirmed that CP can also arise purely as a consequence of learning to sort and label a set of objects in a specified way. In comparisons between subjects that have and have not been trained to categorize a set of stimuli in a particular way, it was found that the space of perceived pairwise between-stimulus similarities is “warped” by the categorization training, in such a way as to compress within- category distances and expand between-category, cross-boundary distances.

In an attempt to investigate what functional role this warping of qualitative similarity space might play in categorization, in some recent neural net simulations we have shown that CP arises as a natural functional component of category learning in both supervised and unsupervised nets, and that it plays a specific functional role in how these nets actually manage to partition the input space correctly into categories.

Neural nets, being especially well suited for pattern learning, are natural candidates for the mechanism that learns to categorize sensory inputs so that they can be assigned the right identifying label, in other words, for connecting symbols to the sensory projections of the objects to which they refer. The limited evidence we have so far suggests that the way nets categorize their inputs is by modifying the distances between their internal representations to position them so that they can be readily separated by the category boundaries. If this were the mechanism of CP, it would mean that the subjective counterpart of the quantitative difference in the relative positions of the internal representations is a qualitative difference in the way we perceive the objects they represent.

A priori, there seems to be no reason this should be the case: Why should learning to categorize things in one way rather than another result in any change in the way they look? If the neural net CP effect turns out to be reliable and general, one possible explanation is that it is merely an epiphenomenon — a side-effect of the way our perceptual system succeeds in finding the features that allow it to categorize things correctly, on the basis of feedback from miscategorization. The TTT task the environment actually imposes on us — successful categorization — is the real constraint. Changes in perceived qualitative similarity are merely a side-effect.

It is tempting to interpret such changes as more than a side-effect, as occurring in the service of the task itself, which is to get the sorting done right. For what

would be the best way to learn to sort a large number of highly interconfusable objects in a specified way? To come somehow to see the ones that belong in the same category as looking unmistakably more like one another, and the ones that belong in different categories as looking unmistakably more different.

Here we are in danger of stepping into the homuncular hermeneutic circle: We have an artificial neural net model that demonstrably accomplishes successful sorting by quantitatively altering distances in internal representational space. We see the functional role that those quantitative internal changes perform for accomplishing the task. We can even suppose that something similar goes on in our heads when we learn to categorize. But why should things LOOK different as a result? Why should the quantitative functional differences translate into qualitative phenomenal ones?

Isn't it enough that the unconscious function performed by my cerebral net computes the right sorting? If there is nothing for the artificial neural net to FEEL when its internal representations change, why should it be otherwise in my own case? Yet it is. The circularity, however, is in yielding to the temptation to ascribe an independent functional role to feeling rather than merely detecting those differences. That's where the homuncular hermeneutics get the better of us; for, looked at objectively, the fact (and it is indeed a fact) that those internal, quantitative, functional differences do have qualitative phenomenal counterparts is not a functional fact; it is not based on the performance of an independent causal role. In other words, the fact that equidistant pairs of wave-lengths within the yellow range look more alike than equidistant pairs of hues that cross the yellow/green boundary, is a fact, but it is an epiphenomenal fact rather than a functional one.

Indeed, not only are such qualitative differences in appearance epiphenomenal, but by exactly the same token, the fact that things should have any qualitative appearance at all is epiphenomenal, since only its detection, discrimination, identification, and manipulation matter functionally — not only to the TTT, which is blind to anything subjective, but to the Blind Watchmaker who shaped us, likewise blind to our subjective lives and hewing only to our adaptive structure and function.

The only ones NOT blind to TTT-indistinguishable differences are ourselves — and only in our own singular cases. By the same token, even if we managed to scale up to a successful TTT-scale theory of all of our robotic capacities, we would still be left with the one mind/body problem, to which we could only try in vain to turn a blind eye.

## References

- Hamad, S., Steklis, H. D. & Lancaster, J. B. (eds.) (1976) *Origins and Evolution of Language and Speech*. Annals of the New York Academy of Sciences 280.
- Hamad, S. (1982) Neoconstructivism: A unifying theme for the cognitive sciences. In: *Language, mind and brain* (T. Simon & R. Scholes, eds., Hillsdale NJ: Erlbaum), 1-11. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad82.neoconst.html>
- Harnad, S. (1982) Consciousness: An afterthought. *Cognition and Brain Theory* 5: 29-47. [http://www.cogsci.soton.ac.uk/~hamad/Papers/Harnad\\_hamad82.consciousness.html](http://www.cogsci.soton.ac.uk/~hamad/Papers/Harnad_hamad82.consciousness.html)
- Harnad, S. (1984) Verifying machines' minds. (Review of J. T. Culbertson, *Consciousness: Natural and artificial*, NY: Libra 1982.) *Contemporary Psychology* 29: 389-391.
- Harnad, S. (1984) What are the scope and limits of radical behaviorist theory? *Behavioral and Brain Sciences* 7: 720-721.
- Harnad, S. (ed.) (1987) *Categorical Perception: The Groundwork of Cognition*. New York: Cambridge University Press.
- Harnad, S. (1987) The induction and representation of categories. In: Harnad 1987. [http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad\\_harnad87.categorization.html](http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad_harnad87.categorization.html)
- Harnad, S. (1989) Minds, Machines and Searle. *Journal of Theoretical and Experimental Artificial Intelligence* 1: 5-25. [http://www.cogsci.soton.ac.uk/~hamad/Papers/Hamad\\_harnad89.searle.html](http://www.cogsci.soton.ac.uk/~hamad/Papers/Hamad_harnad89.searle.html)
- Harnad, S. (1990) The Symbol Grounding Problem. *Physica D* 42: 335-346. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad90.sgproblem.html>
- Harnad, S. (1990) Against Computational Hermeneutics. (Invited commentary on Eric Dietrich's Computationalism) *Social Epistemology* 4: 167-172. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad90.dietrich.crit.html>
- Harnad, S. (1990) Lost in the hermeneutic hall of mirrors. Invited Commentary on: Michael Dyer: *Minds, Machines, Searle and Harnad*. *Journal of Experimental and Theoretical Artificial Intelligence* 2: 321 - 327. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad90.dyer.crit.html>
- Harnad, S. (1990) Symbols and Nets: Cooperation vs. Competition. Review of: S. Pinker and J. Mehler (Eds.) (1988) *Connections and Symbols* *Connection Science* 2: 257- 260.
- Harnad, S. (1991) Other bodies, Other minds: A machine incarnation of an old philosophical problem. *Minds and Machines* 1: 43-54. [http://www.cogsci.soton.ac.uk/~harnad/Papers/Hamad\\_harnad91.otherminds.html](http://www.cogsci.soton.ac.uk/~harnad/Papers/Hamad_harnad91.otherminds.html)
- Harnad, S., Hanson, S.J. & Lubin, J. (1991) Categorical Perception and the Evolution of Supervised Learning in Neural Nets. Presented at Symposium on Symbol Grounding: Problems and Practice, Stanford University, March 1991 In: *Proceedings of the AAAI Spring Symposium on Machine Learning of Natural Language and Ontology* (DW Powers & L Reeker, Eds.) Document D91-09, Deutsches Forschungszentrum fur Kuenstliche Intelligenz GmbH Kaiserslautern FRG, pp. 65-74. [http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad\\_hamad91.cpnets.html](http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad_hamad91.cpnets.html)

- Harnad, S. (1992) Connecting Object to Symbol in Modeling Cognition. In: A. Clark and R. Lutz (Eds) *Connectionism in Context* Springer Verlag, pp 75 - 90. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad92.symbol.object.html>
- Harnad, S. (1992) The Turing Test Is Not A Trick: Turing Indistinguishability Is A Scientific Criterion. *SIGART Bulletin* 3(4) (October) 9-10. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad92.turing.html>
- Hayes, P, Harnad, S., Perlis, D. & Block, N. (1992) Virtual Symposium on Virtual Mind. *Minds and Machines* 2: 217-238. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad92.virtualmind.html>
- Harnad, S. (1993) Grounding Symbols in the Analog World with Neural Nets. *Think* 2(1) 12-78 (Special issue on "Connectionism versus Symbolism," D.M.W. Powers & P.A. Flach, eds.). [Also reprinted in French translation as: "L'Ancre des Symboles dans le Monde Analogique a l'aide de Reseaux Neuronaux: un Modele Hybride." In: Rialle V. et Payette D. (Eds) *La Modelisation*. LEKTON, Vol IV, No 2.] <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad93.symb.anal.net.html> <http://cwis.kub.nl/~fdl/research/ti/docs/think/2-1/index.stm>
- Hamad, S. (1993) Artificial Life: Synthetic Versus Virtual. *Artificial Life III. Proceedings, Santa Fe Institute Studies in the Sciences of Complexity. Volume XVI*. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad93.artlife.html>
- Harnad, S. (1993) Symbol Grounding is an Empirical Problem: Neural Nets are Just a Candidate Component. *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society*. NJ: Erlbaum <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad93.cogsci.html>
- Harnad, S. (1993) Problems, Problems: The Frame Problem as a Symptom of the Symbol Grounding Problem. *PSYCOLOQUY* 4(34) frame-problem. 11 <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad93.frameproblem.html> <http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?4.34>
- Harnad, S. (1993) Exorcizing the Ghost of Mental Imagery. Commentary on: JI Glasgow: „The Imagery Debate Revisited.” *Computational Intelligence* 9(4) 309-333. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad93.imagery.html>
- Harnad S. (1993) Discussion (passim) In: Bock, G.R. & Marsh, J. (Eds.) *Experimental and Theoretical Studies of Consciousness*. CIBA Foundation Symposium 174. Chichester: Wiley
- Harnad, S. (1994) Levels of Functional Equivalence in Reverse Bioengineering: The Darwinian Turing Test for Artificial Life. *Artificial Life* 1(3): 293-301. Reprinted in: C.G. Langton (Ed.). *Artificial Life: An Overview*. MIT Press 1995. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad94.artlife2.html>
- Harnad S, (1994) The Convergence Argument in Mind-Modelling: Scaling Up from Toyland to the Total Turing Test. *Cognoscenti* 1:
- Harnad, S. (1994) Computation Is Just Interpretable Symbol Manipulation: Cognition Isn't. Special Issue on "What Is Computation" *Minds and Machines* 4:379-390 [Also appears in French translation in "Penser f Esprit: Des Sciences de la Cognition a une Philosophie Cognitive," V. Rialle & D. Fiset, Eds. Presses Universite de Grenoble. 1996] <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/hamad95.computation.cognition.html>

- Harnad, S. (1995) Does the Mind Piggy-Back on Robotic and Symbolic Capacity? In: H. Morowitz (ed.) "The Mind, the Brain, and Complex Adaptive Systems." Santa Fe Institute Studies in the Sciences of Complexity. Volume XXII. P. 204-220. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad95.mind.robot.html>
- Harnad, S. (1995) Grounding Symbolic Capacity in Robotic Capacity. In: Steels, L. and R. Brooks (eds.) *The Artificial Life Route to Artificial Intelligence: Building Embodied Situated Agents*. New Haven: Lawrence Erlbaum. Pp. 277-286. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad95.robot.html>
- Harnad, S. Hanson, S.J. & Lubin, J. (1995) Learned Categorical Perception in Neural Nets: Implications for Symbol Grounding. In: V. Honavar & L. Uhr (eds) *Symbol Processors and Connectionist Network Models in Artificial Intelligence and Cognitive Modelling: Steps Toward Principled Integration*. Academic Press, pp. 191-206. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad95.cpnets.html>
- Harnad, S. (1995) Why and How We Are Not Zombies. *Journal of Consciousness Studies* 1: 164-167.  
<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad95.zombies.html>
- Harnad, S. (1995) The Warp Factor. *Guardian* (OnLine: OffLine) Thursday February 23 1995. P. 11. <http://cogsci.soton.ac.uk/~harnad/whorf.html>
- Harnad, S. (1995) Grounding symbols in sensorimotor categories with neural networks. In: IEE Colloquium "Grounding Representations: Integration of Sensory Information in Natural Language Processing, Artificial Intelligence and Neural Networks" (Digest No. 1995/103). <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad95.iee.html>
- Harnad, S. (1995) What Thoughts Are Made Of. *Nature* 378: 455-456. Book Review of: Churchland, P.M. (1995) *The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain* (MIT Press) and Greenfield, S.A. (1995) *Journey to the Centers of the Mind*. (Freeman) *Nature* (1995) <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad95.churchland.bookrev.html>
- Harnad, S. (1996) The Origin of Words: A Psychophysical Hypothesis In Velichkovsky B & Rumbaugh, D. (Eds.) "Communicating Meaning: Evolution and Development of Language. NJ: Erlbaum: pp 27-44. <http://www.cogsci.soton.ac.uk/~hamad/Papers/Harnad/harnad96.word.origin.html>
- Hamad, S. (1996) Experimental Analysis of Naming Behavior Cannot Explain Naming Capacity. *Journal of the Experimental Analysis of Behavior* 65: 262-264. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad96.naming.html>
- Harnad, S. (1996) What to Do About Feelings? [Published as "Conscious Ecumenism" Review of PSYCHE: An Interdisciplinary Journal of Research on Consciousness] *Times Higher Education Supplement*. June 7 1996, P. 29. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad96.feelings.html>
- Cangelosi, A., Greco, A. & Harnad, S. (1996) Categorical perception effects in connectionist models. *Proceedings of the Annual Conference of the Italian Psychological Association. Experimental Psychology Section. Capri, 30 September 1996.*
- Greco, A., Cangelosi, A. & Harnad, S. (1997) A connectionist model of categorical perception and symbol grounding. *Proceedings of the 15th Annual Workshop of the European Society for the Study of Cognitive Systems. Freiburg (D). January 1997: 7.*

- Harnad, S. (1997) "Lively Flights of Fancy." Book review of M. Boden (ed.) "The Philosophy of Artificial Life." Blackwell. Times Higher Education Supplement. January 31, 1997 <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad97.thes.alife.html>
- Pevtzow, R. & Harnad, S. (1997) Warping Similarity Space in Category Learning by Human Subjects: The Role of Task Difficulty. In: Ramscar, M., Hahn, U., Cambouropoulos, E. & Pain, H. (Eds.) Proceedings of SimCat 1997: Interdisciplinary Workshop on Similarity and Categorization. Department of Artificial Intelligence, Edinburgh University: 189 - 195. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/pevtzow97.textures.html>
- Tijsseling, A. & Harnad, S. (1997) Warping Similarity Space in Category Learning by Backprop Nets. In: Ramscar, M., Hahn, U., Cambouropoulos, E. & Pain, H. (Eds.) Proceedings of SimCat 1997: Interdisciplinary Workshop on Similarity and Categorization. Department of Artificial Intelligence, Edinburgh University: 263- -269. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/tijsseling97.cpnets.html>
- Andrews, J., Livingston, K. & Harnad, S. (1998). Categorical Perception Effects Induced by Category Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/>
- Harnad, S. (1998) Turing Indistinguishability and the Blind Watchmaker. In: Mulhauser, G. (ed.) "Evolving Consciousness" Amsterdam: John Benjamins (in press) <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad98.turing.evol.html>
- Harnad, S. (1998) Beyond Object Constancy. Institution of Electrical Engineers (IEE) Seminar on "Self-Learning Robots II: Bio-Robotics" (Digest 98/248: 2/1-2/3) <http://www.cogsci.soton.ac.uk/~harnad/Papers/Hamad/harnad98.iee.embodiment.html>
- Greco, A., Cangelosi, A., & Harnad, S. (1998) A Connectionist Model for Categorical Perception and Symbol Grounding.
- Harnad, S. (1998) Hardships of Cognitive Science. Review of J. Shear (Ed.) Explaining Consciousness (MIT/Bradford 1997) *Trends in Cognitive Sciences* (in press)
- Csato, L., Kovacs, G, Harnad, S. Pevtzow, R & Lorincz, A. (submitted) Category Learning, Categorisation Difficulty and Categorical Perception: Computational Modules and behavioural Evidence. *Connection Science*.
- Greco, A., Cangelosi, A., & Harnad, S. (in prep.) A Connectionist Model for Categorical Perception and Symbol Grounding.
- Cangelosi, A & Harnad, S. (in prep) On the Virtues of Theft Over Honest Toil: Grounding Language and Thought in Sensorimotor Categories: Grounding Language and thought in Senosimotor Categories. <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad96.language.theft.html>
- Harnad, S. (in prep) From Praxis to Pantomime to Propositions: Communicative Continuum or Cognitive Hurdles? (Proceedings of the Language Origins Society)
- Hamad, S. (in preparation) *Icon, Category, Symbol: Essays on the Foundations and Fringes of Cognition*. Cambridge University Press.