

Distributed Cognition: Cognizing, Autonomy and the Turing Test

Stevan Harnad & Itiel Dror
University of Southampton
Southampton SO17 1BJ
United Kingdom

ABSTRACT: Some of the papers in this special issue distribute cognition between what is going on inside individual cognizers' heads and their outside worlds; others distribute cognition among different individual cognizers. Turing's criterion for cognition was individual, autonomous input/output capacity. It is not clear that distributed cognition could pass the Turing Test.

Cognition is what cognizers do: What are cognizers? They are autonomous systems with certain I/O (input/output) performance capacities; and cognition is the internal process that generates that I/O capacity.

What is I/O capacity? It is what higher (and possibly lower) vertebrates are able to do in their worlds: In the case of our own species, it is the capacity to pass the Turing Test. Cognition is what organisms' brains do to enable them to do all the things that organisms can do.

And, until further notice, cognition takes place entirely within the brains of cognizers: It starts with the I in the I/O and ends with the O, skin and in, stretching from the proximal projection of distal objects, events and states onto the cognizer's sensory surfaces to the proximal projections of motor patterns onto the cognizer's effector surfaces, but no further (this is sometimes called "narrow cognition," or "internalism"). The causes and effects stretch more distally, but not the cognition; cognition begins and ends at the cognizer's sensor and effector surfaces. (Fighting words, in our ecumenical era of "distributed cognition"!)

So if, to a first approximation, cognizing is whatever goes on inside the brains of organisms to give rise to their performance capacity, then cognition is the functional substrate -- the generator -- of their know-how. And their performance is that of individual, autonomous systems. It's a solo performance, not like a symphony orchestra, which is in fact many performers. A violinist is a performer; a string quartet is not: it's four performers, performing together. The violinist knows how to read music; the quartet does not. An organist knows how to play 12 notes at once (and to sing and beat rhythm with any remaining appendages at the same time); a musician may be able to play, successively, a string, wind, keyboard and percussion instrument (possibly even a few of them at once), but no musician can play an orchestra -- and the conductor merely leads it, as a shepherd leads his flock; he does not play it. When a conductor is auditioned, it is to test how he directs an ensemble of players, not how he plays the oboe.

There exists an audition for cognition too: The Turing Test (TT). The candidate must be able to exhibit the performance capacities of a normal human being, indistinguishably from a human being, to a human being. But just as it would be regarded as cheating on an exam if the "candidate" consisted of a student plus all his friends and hirelings, sitting the examination jointly, there is a loop-hole in the original version of the TT, which was meant to be conducted by email: If the candidate has performance capacity indistinguishable from that of a real human pen-pal, then it successfully passes the TT. But the purpose of the TT was so that we could design a machine that could pass it, in order that we should at last understand what cognition is, what generates the know-how of an autonomous system like us. A whole team of real human pen-pals could of course also pass the test, pretending to be just one pen-pal, but that would not teach us anything at all about cognition; it would simply be cheating, and the outcome would be trivial and uninformative (except perhaps concerning the social psychology of composite play-acting).

How to close this loop-hole that would otherwise trivialize the TT? Don't restrict it to email: Require that the candidate should be a robot that we can see is just one individual autonomous system like ourselves. That way we not only eliminate the possibility of collective play-acting, but we can also test the candidate's full sensorimotor I/O capacity to confirm that it is indeed completely indistinguishable from our own. After all, despite the remarkable and undeniable expressive power and universality of natural language, human cognizers are capable of a lot more I/O than just email in and email out.

But what is and isn't a single autonomous system? Certainly a robot controlled remotely by a team of technicians would be neither an autonomous system nor a TT candidate in any interesting sense. An organism is a single autonomous system, but so are its cells or even its organs, in the right support media. So an autonomous system may itself be composed of autonomous systems as parts. And so far we are talking about biological autonomy. The logic of autonomy probably breaks down with engineering and physics, where a plane, a pendulum and a positron are all "autonomous systems" in some sense. So autonomy is really only interesting in the cognitive/behavioral domain -- which brings us back to Turing-scale I/O capacity.¹

Let us agree, without further hair-splitting, that insofar as the TT is concerned, any sort of *conscious* multi-person collaboration in generating the robot's I/O capacity would be cheating, or begging the question that the TT is meant to answer: How does the brain generate its I/O capacity? It doesn't generate it by recruiting the brains of a team of collaborators. Is it logically or empirically impossible that the TT could be passed by a bona-fide superordinate cognizer emerging from the interactions of many individual cognizers in the same way that a superordinate slime mold emerges from many individual amoebae? It is not logically impossible, but let us admit that nothing even faintly like it is anywhere in sight empirically, outside of science fiction. What do we see, then, that aspires to be called "distributed cognition"?

First, we see the distal objects of the proximal (skin-and-in) cognition that we have already admitted. Is there a "wider cognition," consisting of what is going on in a cognizer's brain, plus the external objects that project their shadows onto the cognizer's sensory surfaces, that are manipulated by the cognizer's effector actions, and that are the ("intentional") objects of the cognizer's thoughts? It seems that if we admit this wider spectrum of objects into cognition, we lose both the autonomy of the I/O system that is being T-tested and the TT itself: How can the I's and O's and their sources and sinks be part of the self-same autonomous system that is being T-tested for its I/O capacity? That would be like making the house (or the world) a part of the furnace whose temperature-controlling capacities we are trying to test, or making its simulated environment a part of the simulated airplane whose flight capacities we are trying to test. Or making an exam's answers part of the exam: No test left to speak of, if we subsume its I's and O's into the test itself! (Yet this seems to be what Sutton, Kirsh, Schwartz and Zhang in their respective position papers [in Harnad & Dror 2006] mean by "distributed cognition")

The second candidate for "distributed cognition" seems to be social, with the interactions among multiple cognizers being called "distributed cognition." (The Cangelosi, Steels, Goldstone, and possibly the Poirier position papers [in Harnad & Dror 2006] seem to be thinking along these lines). We have agreed that a team of cognizers -- collaborating to pretend to be a single pen-pal in the email TT, or remotely controlling a robot in the robotic TT -- would be cheating and uninformative. Is there any sense in which multiple cognizers can be said to be doing unitary rather than multiple (additive, or even multiplicative) cognizing? This would seem to call for a superordinate TT test -- but testing for what I/O capacity? and compared to what? Individual people

¹ Even here, with distributed cognitive processing possible within an individual cognizer according to the paradigms of neural networks, parallel distributed processing, and connectionism, the nature and boundaries of a system and its parts may be fuzzy or ambiguous (Dror & Dascal, 1997). The overall behavior of an individual (arising from distributed processing or from centralized processing), however, is clear, with I/O capacity taken to reflect cognitive processing and capacity (although this view has its challengers, such as Searle 1980).

are indisputable, paradigmatic cognizers: They are cognizing, if anything is. Their individual I/O capacities are accordingly the gold-standard against which the TT candidate is judged. What would be the corresponding standard for a superordinate TT for distributed cognition? What would the distributed cognizer need to be able to do? And compared to what? And if human I/O capacity, TT and autonomy are set aside, what is left to nonarbitrarily call “cognizing,” whether local or distributed? (These questions are merely raised here: They are not necessarily decisive.)

For the remaining position papers – Glenberg’s (in Harnad & Dror 2006) and Harnad’s (2005), as well as the five empirical review papers (in Harnad & Dror 2006), “distributed cognition” is really just synonymous with “information processing” and indifferent as to whether it is indeed cognizing at all, or autonomous. Does it matter? Could we just scrap “cognition” and speak about “functioning,” whether local or distributed? The reader is left to judge. What is indisputable is that the practical potential of both local and distributed function, involving both humans and machines in the online age, is enormous and just beginning to be exploited.

References:

Dror, I. E. & Dascal, M. (1997). Can Wittgenstein help free the mind from rules? The philosophical foundations of connectionism. In D. Johnson & C. Erneling (Eds.), *The Future of the Cognitive Revolution*, (pp. 293-305). Oxford University Press.

Harnad, S. (2005). Distributed processes, distributed cognizers, and collaborative cognition. In: Dror, Itiel E. (ed.), *Cognitive Technologies and the Pragmatics of Cognition: Special issue of Pragmatics & Cognition 13:3* (2005). pp. (pp. 501–514) <http://eprints.ecs.soton.ac.uk/12073/>

Harnad, S. and Dror, I. E. (eds) (2006). *Distributed Cognition: Special issue of Pragmatics & Cognition 14:3* (2006).

Searle, J. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences* 3: 417-424.
<http://www.bbsonline.org/documents/a/00/00/04/84/index.html>

Turing, A. M. (1950) Computing Machinery and Intelligence. *Mind* 49:433-460. <http://cogprints.org/499/>

About the authors:

Stevan Harnad, born in Hungary, did his undergraduate work at McGill and his doctorate at Princeton and is currently Canada Research Chair in Cognitive Science at University of Quebec/Montreal and adjunct Professor at Southampton University, UK. His research is on categorisation, communication and cognition. Founder and Editor of *Behavioral and Brain Sciences*, *Psychology* and *CogPrints Archive*. He is Past President of the Society for Philosophy and Psychology, Corresponding Member of the Hungarian Academy of Science, and author and contributor to over 150 publications: <http://www.ecs.soton.ac.uk/~harnad/>

Itiel Dror is a Senior Lecturer in cognitive sciences at the University of Southampton, UK. He holds a number of graduate degrees, including a Ph.D. in cognitive psychology from Harvard (USA). He specializes in the fields of human cognition & behaviour, training & skill acquisition, technology & cognition, expertise, and biometric identification. Dr. Dror has worked in universities around the world and conducted research and consultancy to numerous organizations, including the UK Passport Services, the USA Air Force, the Japanese Advance Science Project, the European Aerospace Research & Development Agency, the Israeli Aerospace Industry, the BBC, and many commercial companies. He has published widely in both basic science and applied domains. For more information, see: <http://www.ecs.soton.ac.uk/~id/>