Harnad, S. (2010) Symbol Grounding Problem: Turing-Scale Solution Needed. IEEE AMD Newsletter 7(1)

Symbol Grounding Problem: Turing-Scale Solution Needed

Stevan Harnad

University du Québec à Montréal, Canada
University of Southampton, UK

Toys. The symbol grounding problem is the problem of causally connecting symbols inside an autonomous system to their referents in the external world without the mediation of an external interpreter. The only way to avoid triviality, however, is to ensure that the symbols in question, their referents in the world, and the dynamic capacities of the autonomous system interacting with the world are nontrivial. Otherwise a toy robot, with exactly two symbols – go/stop – is "grounded" in a world where it goes until it bumps into something and stops.

Turing. From the very outset, the symbol grounding problem – which was inspired and motivated by Searle's Chinese Room Argument – was based on the Turing Test, and hence on a system with *full, human-scale linguistic capacity*. So it is the words of a full-blown natural language (not all of them, but the ones that cannot be grounded by definition in the others) that need to be connected to their referents in the world. Have we solved that problem? Certainly not. Nor do we have a robot with Turing-scale capacities, either symbolic or sensorimotor (with the former grounded – embodied -- in the latter). Designing or reverse-engineering an autonomous system with this Turing-scale robotic and linguistic capacity -- and thereby causally explaining it -- is the ultimate goal of cognitive science. (Grounding, however, is not the same as meaning; for that we would also have to give a causal explanation of consciousness, i.e., feeling, and that, unlike passing the Turing Test, is not just hard but hopeless.)

Totality. Grounded robots with the sensorimotor and learning capacities of subhuman animals might serve as waystations, but the gist of the Turing methodology is to avoid being fooled by arbitrary fragments of performance capacity. Human language provides a natural totality. (There are no partial languages, in which you can say this, but not that.) We are also extremely good at "mind-reading" human sensorimotor performance capacity for tell-tale signs of mindlessness; it is not clear how good we are with animals (apart perhaps from the movements and facial expressions of the higher mammals).

Terms. There are certain terms (or concepts) I have not found especially useful. It seems to me that real objects -- plus (internal or external) (1) analogs or iconic copies of objects (similar in shape) and (2) arbitrary-shaped symbols in a formal symbol system (such as "x" and "=", or the words in a language, apart from their iconic properties), systematically interpretable as referring to objects -- are entities enough. Peirce's "icon/index/symbol" triad seems one too many. Perhaps an index is just a symbol in a toy symbol system. In a formal symbol system the links between symbols are syntactic whereas the links between internal symbols and the external objects that they are about are sensorimotor (hence somewhat iconic). And inasmuch as symbols inside a Turing-scale robot are linked to object categories rather than to unique (one-time, one-place) individuals, all categories are abstract (being based on the extraction of sensorimotor invariants), including, of course, the category "symbol." The rest is just a matter of degree-of-abstraction. Even icons are abstract, inasmuch as they are neither identical nor co-extensive with the objects they resemble. There are also two sorts of productivity or generativity: syntactic and semantic. The former is just formal; the latter is natural language's power to express any and every truth-valued proposition.

Talk. Yes, language is fundamentally social in that it would never have bothered to evolve if we had been solitary monads (even monads born mature: no development, just cumulative learning capacity). But the nonsocial environment gives enough corrective feedback for us to learn categories. Agreeing on what to call them is trivial. What is not trivial is treating symbol strings as truth-valued propositions that describe or define categories.

Harnad, S. (2003) Can a Machine Be Conscious? How? *Journal of Consciousness Studies* 10(4-5): 69-75. http://eprints.ecs.soton.ac.uk/7718/

Harnad, S. (2005) *To Cognize is to Categorize: Cognition is Categorization*, in Lefebvre, C. and Cohen, H., Eds. *Handbook of Categorization*. Elsevier. http://eprints.ecs.soton.ac.uk/11725/

Harnad, S. and Scherzer, P. (2007) First, Scale Up to the Robotic Turing Test, Then Worry About Feeling. In *Proceedings of Proceedings of 2007 Fall Symposium on AI and Consciousness.* Washington DC. http://eprints.ecs.soton.ac.uk/14430/

Harnad, S. (2008) The Annotation Game: On Turing (1950) on Computing, Machinery and Intelligence. In: Epstein, Robert & Peters, Grace (Eds.) *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Springer http://eprints.ecs.soton.ac.uk/7741/

Picard, O., Blondin-Masse, A., Harnad, S., Marcotte, O., Chicoisne, G. and Gargouri, Y. (2009) Hierarchies in Dictionary Definition Space. In: 23rd Annual Conference on Neural Information Processing Systems: Workshop on Analyzing Networks and Learning With Graphs, 11-12 December 2009, Vancouver BC (Canada). http://eprints.ecs.soton.ac.uk/18267/