

## Symbol Grounding, Turing Testing and Robot Talking

Stevan Harnad<sup>a,b</sup>, Alexandre Blondin-Massé<sup>a</sup>, Bernard St.-Louis<sup>a</sup>,

Guillaume Chicoisne<sup>a</sup>, Yassine Gargouri<sup>a</sup> & Olivier Picard<sup>a</sup>

<sup>a</sup>Institut des sciences cognitives  
Université du Québec à Montréal  
Montréal, Québec, Canada H3C 3P8

<sup>b</sup>Department of Electronics and Computer Science  
University of Southampton  
Highfield, Southampton, United Kingdom SO17 1BJ

**ABSTRACT:** Turing proposed a methodological criterion for Cognitive Science, the Turing Test (TT): If a system can talk with people, indistinguishably from a real person, for a lifetime, then the mechanism underlying its performance capacity is also a candidate explanation for human language capacity. The original TT just targeted language input and output, but language has to be grounded in the robotic sensorimotor capacity to categorize and manipulate the referents of words and sentences if those are to be systematically connected to the outside world other than by rote association. We report some results on (1) the age of acquisition, concreteness, and imageability of the basic grounding vocabulary of dictionaries and (2) the brain correlates of human sensorimotor category acquisition.

Whatever language is, it is not just associations between words and things; nor is it just social agreement on what to name things: Language is not just words; and words are not just the names of things (even if “things” include objects, actions, events, states and traits). Language is *the capacity to say and understand anything and everything that any speaker can say and understand*. And what speakers spend most of their speaking/listening time doing is not naming things, nor agreeing on what to name things, but exchanging *propositions*.

Propositions are sequences of words (sentences) that make assertions<sup>1</sup> with *truth values* (true or false): They *tell* what is and is not the case in the world. If you cannot *tell* it, you can only *show* it. That is the difference between propositions and pantomime. And it is also the radical adaptive advantage of language: the advantage

---

<sup>1</sup> Let us set aside the special case of questions and commands for now: They are just assertions with an interrogative or imperative sign: Is it true that P? Make it true that P!

of acquiring and conveying information by telling rather than showing (Cangelosi et al. 2002; Dror & Harnad 2009).

Propositions consist of both content words and function words. “The cat is not on the mat” contains three content words (cat, mat, on) and four function words (the, is, not, the).<sup>2</sup> With a sufficient vocabulary of content words, plus the requisite function words (according to the grammar of the language), any language can express any proposition at all. And propositions describe and define everything that is (and is not) the case.

Very few content words are proper names of individuals. Most content words name *kinds*, i.e., categories with an infinite number of members: Not only is there a potential infinity of individual cats, but each individual cat has an infinity of sensorimotor “affordances” for the categorizer, depending on the “shape” (including the dynamics) of the distal cat’s contact with the categorizer’s proximal sensorimotor surfaces (Gibson 1966).

So the reason language is not just an association between words and things is that most (content) words name *kinds*, not individual things, and both kinds and things have an infinity of sensorimotor “shapes.” Hence naming is an “association” between words and the *invariants* of the sensorimotor shapes of things and kinds, and those invariants need to be *detected and abstracted* by a means much more demanding than just rote association between names and individual instances. This is also why learning to name is much more demanding than just coming to a social agreement about what to name what.

Before we can name kinds, we need to be able to categorize them, and before we can categorize them, we need to have learned their sensorimotor invariants. Those invariants are *sensorimotor* rather than just sensory, because to categorize is to *do* (or be able to do) something *different* with the members and the nonmembers of the category. *Doing* can in turn be anything from eating or not eating the kinds of things that do and don’t afford safe edibility, to calling the one kind “edible” and the other kind “inedible” (Harnad 2005).

The primary functional burden of language, then, is on sensorimotor category learning and identification. That – and not rote association – is how (content) words are grounded in whatever they refer to (Harnad 2003). Do all words need to be grounded directly through sensorimotor category learning? Clearly not: words can also be grounded indirectly with words alone, through explicit definition, as in a dictionary, or through implicit word usage in context. But the meaning of an

---

<sup>2</sup> The word “on” is borderline between content and function. In general, for content words, you can point to (i.e., show) positive and negative examples: “That is green. That is not green.” You can do that with “on”: “That is on. That is not on.” But you cannot point to examples of what is and is not “the.” (“The” is a definite article, similar in function to a demonstrative like “that,” which merely functions to point out something.) And you cannot point to an “is.” It merely functions to assert, while “not” functions to deny.

unknown word can only be grounded through definition or context if the other words in the definition or context are already grounded.

Alan Turing proposed a methodological criterion for cognitive science: To explain cognitive function, design a system that can do anything a human mind can do: its performance capacity must be totally indistinguishable *from* that of a human, *to* a human. How that system manages to pass the “Turing Test ” (TT) will then be a candidate explanation of how the human brain manages to do it (Harnad 2008).

The critical feature of the TT is its “*total* indistinguishability” criterion: It is not enough for the system to be able to do only a partial fragment of what the human mind can do – e.g., play chess well enough to fool people into thinking it is a real person playing chess. The performance mechanism that can generate only an arbitrary fragment of our total performance capacity has too many degrees of freedom, because the same performance can be generated in many different ways: it is too underdetermined. The degrees of freedom have to be reduced to the normal degree of underdetermination of a scientific theory, by requiring the system to be able to do *all*, not just *part*, of what a real human cognizer can do.

Of course one does not scale up to a successful generator of full TT capacity overnight: it has to be reached by degrees, starting with “toy” fragments of performance sub-capacity (such as chess-playing). It is also important to try to target coherent, natural, self-contained “modules” of cognitive sub-capacity, rather than arbitrary fragments. Turing proposed a natural module in the form of language capacity itself.

It is not clear whether the reason Turing formulated his TT as a purely linguistic test rather than a sensorimotor robotic test was that he thought language was an autonomous module of cognitive performance capacity, or that he thought any system that could successfully pass the linguistic TT would also have to possess and draw indirectly upon all the rest of our cognitive capacities too, even though the sensorimotor ones are not directly tested by the TT. In any case, the original TT was purely linguistic – words in and words out -- and could be conducted and passed via email alone: the candidate would have to be totally indistinguishable *from* a human pen-pal, *to* real human pen-pals.<sup>3</sup> It is another question entirely, however, whether it is possible to scale up to the full linguistic TT bypassing sensorimotor robotic capacities altogether. Our hypothesis is that it is not possible: word meanings first have to be grounded in sensorimotor categorization capacity.

This is not to underestimate the expressive power of language, which is the expressive power of propositions. Every natural language can express every possible proposition, and propositions can in principle describe anything and everything that is the case, whether empirical or formal. Among other things, the

---

<sup>3</sup> Turing mentions no time limit, but it is evident that the capacity is not just meant to be a short-term trick: The candidate should in principle be testable and totally indistinguishable lifelong, if need be.

expressive power of language includes the expressive power of mathematics and computation, for anything that can be said in formal symbols can be said in any natural language.

It is important to appreciate how much greater is the expressive power of *propositions* than that of *pantomime*. This is the power of *telling* over *showing*. One can show a lot, but one can tell a lot more than one can show (and that includes being able to tell everything that one can show). Moreover, as noted earlier, showing itself has no truth value: Mimicking an object or event may resemble the object or event to various degrees, and it may or may not succeed in calling to mind the object or event for the person to whom one is mimicking them. But the mimicry itself is merely a matter of similarity and associations. It is only if we *subtitle* the pantomime with words – “Ah, you’re trying to tell me that the cat is not on the mat!” – that it can become something that is either true or false.

It is true that any picture (or pantomime) is worth more than 1000 (or even an infinite number of) words; but any verbal description can be lengthened to approximate the content of the picture (or pantomime) as closely as we like. Approximate it in what sense? A string of digital symbols can never *be* the analog object it describes.<sup>4</sup> “The cat is not on the mat” is not the same thing as the state of affairs consisting of the cat not being on the mat. The mediating link between (1) the string of symbols that encodes the proposition and (2) the state of affairs that the proposition describes is a human mind that *understands* the proposition.

The human body (with its brain) is an autonomous sensorimotor system that can produce and understand propositions, its symbols are grounded, and it can pass the TT. But do all of the symbols it understands have to be grounded? Clearly not, because, as noted, we can and do learn and hence ground the meanings of new words through definitions, as long as the other words in the definitions are already grounded. They in turn might have been grounded directly in sensorimotor experience or they too may have been indirectly grounded through definition. *But it cannot be indirect grounding by definition all the way down: Some words have to be grounded directly in sensorimotor experience.* It hence seems natural to ask: how many? And which ones?

To answer this question, we have been analyzing digital dictionaries. We have applied an algorithm that reduces dictionaries to their “grounding kernel” by eliminating all words that can be “reached” through definition alone (Blondin-Massé et al. 2008). Applying our algorithm to three special dictionaries -- Longmans Dictionary of Contemporary English (LDOCE), Cambridge International Dictionary of English (CIDE), and WordNet -- we find we can reduce them to about 10% of their size in this way. It is not yet clear whether the resultant grounding kernel is

---

<sup>4</sup> Even when a string of symbols describes “itself” (“I have 4 words”), the description is not the same thing as the thing it is describing – and especially because it has to be mediated by an interpretation, which is yet another thing; otherwise we just have “squiggle, squiggle, squiggle, squiggle”).

minimal, unique, or universal across dictionaries (in the same and different languages). But we have found, using the MRC Psycholinguistic Database (Wilson 1988) that the words in our grounding kernel are significantly more *concrete* and *imageable*, and that they are *acquired at a significantly earlier age* than the words in the rest of the dictionary (Chicoisne et al. 2008).

The implication of this finding is not so simple as that we merely (i) ground concrete sensorimotor categories first, when we are young, and then (ii) acquire the rest of our categories and their names by definition alone, through propositions composed of those already-grounded category names. For, although the grounding kernel contains a significantly higher proportion of concrete words than the rest of the dictionary, it contains many abstract words as well. Moreover, once we partial out the effects of age of acquisition (which is also correlated with concreteness) the correlation changes sign, and the remaining words in the grounding kernel turn out to be significantly more abstract than the rest of the dictionary.

The very notion of abstractness/concreteness needs to be examined more closely, however. The MRC database is derived from human judgments about degree of abstractness: But are not adjectives, which designate properties, already more abstract than nouns, which may designate concrete, palpable objects? Are human subjects necessarily right if they rate “weight” as being more abstract than “heavy”? Perhaps (sensorimotor) imageability is a more reliable criterion than abstractness, since, after all, *all* categories (kinds) are abstractions – and from a sensorimotor point of view, even individuals are abstractions. Category learning through invariance detection and abstraction is hence itself a process of abstraction. Moreover, even imageability might be more reliably measurable by physiological means rather than by introspective judgment.

For this reason we are also analyzing the electrophysiological correlates (Event-Related Potentials, ERPs) of the acquisition of (visual) categories (by adults) (St-Louis et al 2008). Except in infants before they comprehend words, words play a role even in the direct visual learning of new categories, for a linguistic species like ourselves: A new visual category can be learned by visual trial and error, in which the learner’s brain must abstract the visual invariant that reliably distinguishes the members from the nonmembers. Or, if the invariant is describable in words that the learner already understands, the abstraction rule can be *told* to the learner, verbally, much the way a new word can be defined by a dictionary.

Our subjects had to learn a visual texture category. The invariant that distinguished the members from the nonmembers was based on the relative proportion of tiny vertical and horizontal U-shaped micro-elements distributed randomly in a matrix to make up the texture. We could then either *tell* the subjects which ratios of which micro-elements determined the categories, or we could let them figure it out for themselves, by trial and error, with feedback signaling to them when they had categorized correctly or incorrectly.

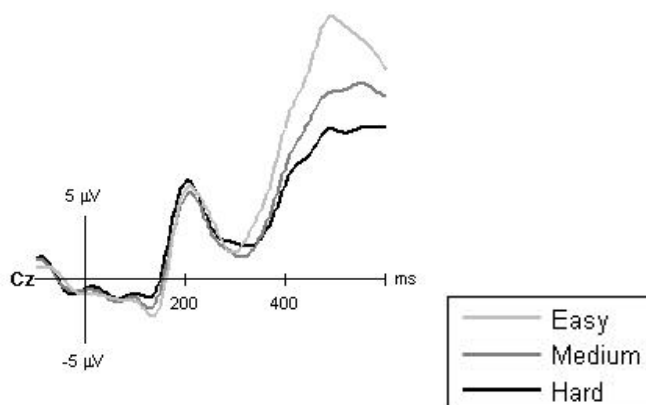
Not all of the trial-and-error subjects succeeded in learning the categorization. But all those who did learn it could verbalize the rule that described the invariant as soon as their categorizing became successful. A late positive component of the accompanying ERPs was found to be reliably correlated with successfully learning the category: it was present in the learners and absent in the nonlearners.

What about those subjects to whom the invariant was told in words in advance? They could of course categorize successfully as of the very first trial (though slowly at first). This is the power of telling over showing, And the late positivity was already present in their ERPs too, though it grew with practice. So, was the positivity a correlate of the visual detection of the invariant? or of the verbal application of the rule? or of confidence that they had mastered the category? More experimental conditions were tested, varying the difficulty of the categorization:

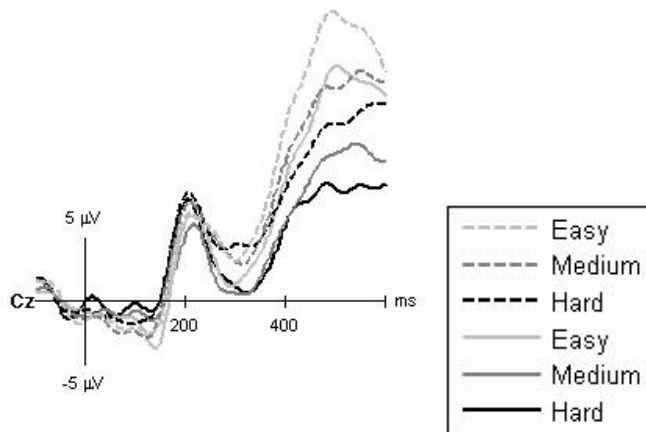
Subjects were all told the categorization rule in advance and given corrective feedback during the training. Stimuli (presented for 200 ms) were again matrices of Horizontal (H) and Vertical (V) micro-elements forming two texture categories according to whether the proportion of H or V was higher. In each category, this difference in proportion was made either high, medium or low (hence, respectively, Easy, Medium and Hard to detect). ERPs were recorded during the performance. Correct and incorrect responses were treated separately, but the two stimulus categories were averaged together.

The easier the categorization, the bigger the ERP positivity (Figure 1) and the faster the performance. The positivity was also bigger at the end of the task than at the beginning, and so was a negative peak around 300 ms (Figure 2). Both components were bigger for correct categorizations than for incorrect ones (Figure 3).

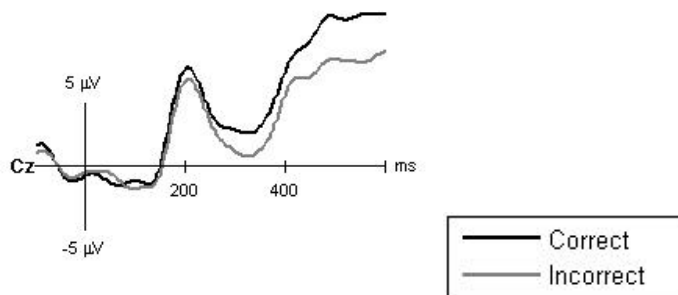
Based on how prior findings have been interpreted in the ERP literature, these results suggest the following: (1) The positivity reflects the subject's confidence in categorizing the stimuli, which depends on difficulty and improves (and accelerates) with practice. (2) The negative peak reflects improved detection of the category invariants with practice, with reduced efficiency on incorrect trials.



**Figure 1.** ERPs for each category difficulty level.



**Figure 2.** ERPs for the first (continuous line) and last (dashed line) 100 trials at each categorization difficulty level.



**Figure 3.** ERPs for correct and incorrect responses

What are the implications of these findings for getting robots to talk? Most of the hard work is likely to be in designing the robots to be able to learn sensorimotor categories, not in getting them to acquire words. But once they have acquired a kernel of grounded, named sensorimotor categories, the robots should be able to acquire further new categories through propositions *describing* the new categories in words. If and when the performance capacity of such a robot model could be

successfully scaled up from the toy level to the TT level, it would be a candidate explanation for how the mind works, just as Turing suggested it would be.

## References

Blondin-Massé, A, Chicoisne, G, Gargouri, Y, Harnad, S, Picard, O, & Marcotte, O (2008). How Is Meaning Grounded in Dictionary Definitions? In TextGraphs-3 Workshop - 22nd International Conference on Computational Linguistics. <http://www.archipel.uqam.ca/657/>

Cangelosi, A., Greco, A. & Harnad, S. (2002) Symbol Grounding and the Symbolic Theft Hypothesis. In: Cangelosi, A. & Parisi, D. (Eds.) Simulating the Evolution of Language. London, Springer. <http://cogprints.org/2132/>

Chicoisne, G., Blondin-Massé, A., Picard, O. & Harnad, S. (2008) Grounding Abstract Word Definitions In Prior Concrete Experience. In: Sixth Annual Conference on the Mental Lexicon, University of Alberta, Banff Alberta, 7-10 October 2008. (In Press) <http://eprints.ecs.soton.ac.uk/16618/>

Dror, I. & Harnad, S. (2009) Offloading Cognition onto Cognitive Technology. In Dror, I. & Harnad, S. (Eds) (2009): Cognition Distributed: How Cognitive Technology Extends Our Minds. John Benjamin. <http://eprints.ecs.soton.ac.uk/16609/>

Gibson, JJ. (1966) The Senses Considered as Perceptual Systems. Greenwood Publishing Group.

Harnad, S. (2003) Symbol-Grounding Problem. Encyclopedia of Cognitive Science. Nature Publishing Group. Macmillan. <http://eprints.ecs.soton.ac.uk/7720/>

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. (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/>

Harnad, S. & 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 (in press), Washington DC. <http://eprints.ecs.soton.ac.uk/14430/>

St-Louis, B., Corbeil, M., Achim, A. & Harnad, S. (2008) Acquiring the Mental Lexicon Through Sensorimotor Category Learning. In: Sixth Annual Conference on the Mental Lexicon, University of Alberta, Banff Alberta, 7-10 October 2008. (In Press) <http://eprints.ecs.soton.ac.uk/16620/>

Wilson, M.D. (1988) The MRC Psycholinguistic Database: Machine Readable Dictionary. *Behavioral Research Methods, Instruments and Computers*, 20(1), 6-11.