From Sensorimotor Categories to Grounded Symbols

Stevan Harnad
Canada Research Chair in Cognitive Sciences
Universite du Quebec a Montreal
&
School of Electronics and Computer Science
University of Southampton

In the beginning was the category. A category is a kind of "thing": objects, events, actions, properties, states. Even individuals are kinds insofar as our brains are concerned, because to recognize an individual, we have to detect that all the instances of that individual that we encounter are instances of that same individual and not another, despite all the variation from instance to instance. So learning that this is a dog and learning that this is Fido are both cases of learning a category. In both cases, there are instances of members and nonmembers: With "dog," there are instances of other members of the same category – other dogs – and instances of nonmembers, such as cats, trees and rocks. With "Fido" there will be other instances of Fido, seen near and far, in different positions, moving or stationary, awake or asleep, as well as instances of things that could be confused with Fido, but are not Fido, such as other dogs of the same breed.

It's important for organisms to get their categories right so that they can do the right thing with the right kind of thing: eat what's edible and not what's toxic, approach friend but not foe, etc. Most of cognition is the acquisition of categories and much of adaptive behavior is doing the right things with the members and the nonmembers of those categories (Harnad 2005).

To be able to categorize correctly, one must be able to distinguish the members from the nonmembers: Zebras have black and white stripes and giraffes are brown with long necks, but telling categories apart is not always that easy. It's always a matter of detecting the features that reliably distinguish the members from the nonmembers but sometimes discovering those features is hard work. For some categories, the feature detector is inborn, as it is with the frog's bug-detector, and hence the hard work of detecting members was done in advance by trial-and-error evolution in the evolution of the species. Most categories, however, have to be learned through trial and error during the lifetime of the organism. The process sounds simple, but success might take long, and require a lot of work: The organism encounters instances of members and nonmembers of the category, tries to do the right thing with them (such as eat them or avoid them), makes mistakes, which are then "corrected" by feedback from the consequences of having done the right or wrong thing with the right or wrong kind of thing (eating toxic things and getting sick, passing up edible things and getting hungry). If all goes well, the organism will eventually learn how to tell them apart and what to do with what. Its brain, which contains powerful feature learning mechnisms, will eventually detect the features that reliably distinguish one category from the other.

Notice that we have not said a word yet about words or language. We tend to think of categories as having names (and most do), but for species other than our own, “doing the right thing with the right kind of thing” does not mean naming it, but doing something more concrete and practical with it, such as eating it, or fleeing from it. Nevertheless, the cognitive lives of many other species consist, as our own do, in acquiring new categories, except that they can only acquire them by direct trial-and-error sensorimotor experience, as just described, guided by the feedback from the consequences of correct and incorrect categorization. Let us call this acquiring categories through “sensorimotor induction.”

Our species has another way of acquiring categories, a better way, one that is freed of the delays and risks of trial and error learning from direct experience. We have shown in a-life simulations that simple virtual creatures in virtual worlds that must learn to do the right thing with the right kind of thing in order to survive and reproduce are able to do this through trial and error experience, with the help of neural nets that are able to learn to detect the features that reliably distinguish one category from another. So far that’s not news. But then we showed that these creatures are out-survived and out-reproduced by creatures that can acquire categories in a much faster and surer way: though “hearsay.” They are “told” which features distinguish the members from the nonmembers, hence they do not have to go through the time-consuming and risky process of learning them by trial and error (Cangelosi & Harnad 2001). But this other way – symbolic instruction – is not autonomous. It can’t be symbolic instruction all the way down, for how would you learn what the words in the intruction stood for if all you ever heard was words?

This is the “symbol grounding problem” (Harnad 1990). The best illustration is a dictionary. Suppose you had to learn what Chinese words stood for, but all you had was a Chinese/Chinese dictionary. If you did not know any Chinese at all, looking up the definitions of words in the Chinese/Chinese dictionary would get you nowhere. It would just take you on a merry-go-round, from one meaningless definition to another. But if you already knew the meaning of some Chinese words, then that might be enough for you to learn the meanings of others via the definitions alone.

This is also how our a-life simulations worked: In order to derive the adaptive advantage of hearsay over experience the creatures had to “ground” some of their categories the old, hard way, via direct experience. But once they had done that, they could assign an arbitrary name to those categories, and the names could then be combined and recombined to “define” new categories. That, we think, was the revolutionary advantage that language conferred on our ancestors: the advantage of symbolic instruction over sensorimotor induction, transmitting categories from those who already have them to those who do not by a means that has evolved in no other species.

Our a-life simulations were just in a toy world, with a few trivial categories, some of them being boolean combinations of other categories. So “edible” might be learned by induction, and “markable” (i.e., location needs to be marked) might also be learned by induction. But then a 3rd, higher-order category that is really just based on the conjunction of the features “edible” and “markable” might define the new composite category “returnable” (i.e., return to this location for eating later) which would not itself have to be learned by trial and error in the way its two component categories were, if the creature were capable of learning through symbolic instruction. But a simple toy world and a simple pairwise conjunction do not yet show that this can scale up to full-blown natural language.

So we have taken it a step further, with computer analyses of digital dictionaries: We used an algorithm to systematically reduce a dictionary to a “grounding kernel” by eliminating every word that could be reached by definition from a combination of other words until we reached a kernel (it turned out to be about 10% of the dictionary) for which there was nothing left but the merry-go-round: If you did not already know what those kernel words meant, then the dictionary could not tell you. But if you did already know that 10% somehow, then you could reach all the remaining 90% via definition alone (Blondin-Massé et al. 2008).

But where did the meanings of the grounding kernel come from, if not via definition? We hypothesized that these words were more likely to have been grounded by sensorimotor induction. We tested this by using the MRC psycholinguistic database to compare the words in the grounding kernel with the words in the rest of the dictionary in terms of their degree of concreteness/abstractness and their age of acquisition, and the words in the grounding kernel turned out to be significantly more concrete (i.e., closer to the sensorimotor) and learned at an earlier age (Chicoisne et al. 2008).

We have also done psychophysiological studies comparing category acquisition by sensorimotor induction versus symbolic instruction and found a component of the event-related brain potential that emerges during trial and error learning only if the subject is successful in learning the category; it is absent in those who fail to learn. But then when the unsuccessful learners are told in words what feature distinguishes the members of the category from the nonmembers, the component that accompanied successful induction learning appears in their brain activity as well (St-Louis et al 2008) as they categorize successfully using the verbal instruction.

Clearly our species and others had and have the capacity to acquire categories by induction. Our closest cousins, the apes and simians also have body structure and motor capacities similar to our own. We are all potentially equipped, for example, to both observe and (thanks to our mirror neurons) imitate and even mime “doing the right thing with the right kind of thing.” None of this is yet linguistic, but it could certainly be useful. Now suppose that our species evolves a propensity toward this sort of nonlinguistic gestural communication, because of the adaptive benefits it confers in transmitting certain sensorimotor skills and perhaps even some help it provides in the learning (by induction, not instruction) of some categories (Harnad 2007). Only two things are missing for a transition to language (in the gestural mode): (1) category-names and (2) truth-value-bearing propositions.

Prior to language a "name" is merely an arbitrary response associated with a category. If I first mime "eating" and everyone recognizes that that gesture is associated with eating, then it is no longer necessary that the gesture should resemble eating in order to evoke that association. The "iconicity" of the gesture, the resemblance that first made the associative link, becomes irrelevant, and the gesture can gradually become arbitrary and conventional, as long as everyone keeps making the association. But rote association is definitely not the same thing as linguistic reference. Perhaps pointing and making purposive gestures to evoke an association comes closer to reference, but not linguistic reference, for words are not just category-names. Words can also be combined and recombined to form propositions that define new categories -- and, most important, the proposition can be true or false. A name, "X," cannot be true or false. Only a proposition – "This is an X" – can be true or false, and neither pointing nor naming is yet making a proposition; nor is purposive miming.

I do not, in point of fact, have a compelling hypothesis about what induced the transition from purposive pantomime to propositions, though I can spin a plausible Just-So Story just as well as anyone else (Harnad 2000). I would rather close by noting that once you can produce and understand a proposition at all, you can produce and understand any and every proposition (Katz 1976). Propositions are all statements about category membership. (The foregoing sentence is as good an example as any!) So once you have made the transition from purposive pantomime to truth-valued propositions you have the full power of language to define, describe and explain any category at all: the full power of symbolic instruction, just as long as it is grounded, like the kernel of our dictionaries, in sensorimotor induction. In the beginning was the category; with propositions came the word.

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. & Harnad, S. (2001) The Adaptive Advantage of Symbolic Theft Over Sensorimotor Toil:Grounding Language in Perceptual Categories. Evolution of Communication 4(1) 117-142? http://cogprints.org/2036/

Harnad, S. (2007) From Knowing How To Knowing That: Acquiring Categories By Word of Mouth. Kaziemierz Naturalized Epistemology Workshop (KNEW), Kaziemierz, Poland, 2 September 2007. http://eprints.ecs.soton.ac.uk/14517/

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

Harnad, S. (2000) From Sensorimotor Praxis and Pantomine to Symbolic Representations The Evolution of Language. Proceedings of 3rd International Conference. Paris 3-6 April 2000: Pp 118-125. http://cogprints.org/1619/

Katz, J.J. (1976) The Effability Hypothesis. In: Harnad et al. 1976, 445 - 455. In: Harnad, 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.

St-Louis, B., Corbeil, M., Achim, A. and 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. http://eprints.ecs.soton.ac.uk/16620/

Chicoisne, G., Blondin-Massé, A., Picard, O. and 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. http://eprints.ecs.soton.ac.uk/16618/

Steklis, HD and Harnad, S (1976) From hand to mouth: Some critical stages in the evolution of language, In: Origins and Evolution of Language and Speech (Harnad, S, Steklis , HD & Lancaster, JB., Eds.), 445-455. Annals of the New York Academy of Sciences 280. http://cogprints.org/0866/