Language and the game of life
(to appear in Behavioral and Brain Sciences, 2005)

Stevan Harnad
Centre de Neuroscience de la Cognition, Université du Québec à Montréal, Montréal, Québec H3C 3P8, Canada and Department of Electronics and Computer Science, University of Southampton, Highfield, Southampton SO17 1BJ

harnad@ecs.soton.ac.uk
http://www.ecs.soton.ac.uk/~harnad/

Abstract: Steels & Belpaeme's simulations contain all the right components, but they are put together wrongly. Color categories are unrepresentative of categories in general and language is not merely naming. Language evolved because it provided a powerful new way to acquire categories (through instruction, rather than just the old way of other species, through trial-and-error experience). It did not evolve so that multiple agents looking at the same objects could let one another know which of the objects they had in mind, co-coinining names for them on the fly.

Contra Wittgenstein (1953), language is not a game. (Maynard-Smith 1982), would no doubt plead nolo contendere.) The game is life, and language evolved (and continues to perform) in life's service – although it has since gained a certain measure of autonomy too.

So are Steels & Belpaeme (S&B) inquiring into the functional role for which language evolved, the supplementary roles for which it has since been coopted, or merely the role something possibly resembling language might play in robotics (another supplement to our lives)?

For if S & B are studying the functional role for which language evolved, that role is almost certainly absent from the experimental conditions that they are simulating. Their computer simulations do not capture the ecological conditions under which, and for which, language began. The tasks and environments set for S & B’s simulated creatures were not those that faced any human or prehuman ancestor, nor would they have led to the evolution of language had they been. On the contrary, the tasks faced by our prelinguistic ancestors (as well as our nonlinguistic contemporaries) are precisely the ones left out of S & B's simulations.

S & B do make two fleeting references to a world in which foragers need to learn to recognize and sort mushrooms by kind – with color possibly serving as one of the features on the basis of which they sort. But a task like learning to sort mushrooms by kind is not what S & B simulate here. They simulate the task of sorting colors, and not by kind, but by a kind of “odd man out” exercise called the
“discrimination game.” In this game, the agent sees a number of different colors (the “context”), of which one (the “topic”) is the one that must be discriminated from the rest. If this is done by two agents, it is called the “guessing game,” with the speaker both discriminating and naming the topic-color, and the hearer having to guess which of the visible context-colors the speaker named. Both agents see all the context-colors.

Now the first thing we must ask is: (i) Were any of our prelinguistic ancestors ever faced with a task anything like this? (ii) And if they had been, would that have led to the evolution of language? (iii) Indeed, is what is going on in S & B's task language at all?

I would like to suggest that the answer to all three questions is no. S & B's is not an ecologically valid task; it is not a canonical problem that our prelinguistic ancestors encountered, for which language evolved as the solution. And even if we trained contemporary animals to do something like it (as some comparative psychologists have done, e.g. Leavens et al. 1996), it would not be a linguistic task — indeed it would hardly even be a categorization task, but more like a joint multiple-choice task requiring some “mind-reading” (Premack & Woodruff 1978; Tomasello 1999) plus some coordination (Fussell & Krauss 1992; Markman & Makin 1998).

On the other hand, there is no doubt that our own ancestors, once language had evolved, did face tasks like this, and that language helped them perform such tasks. But language helps us perform many tasks (even learning to ride a bicycle or to do synchronized swimming) for which language is not necessary, for which it did not evolve, and which are not themselves linguistic tasks. (This is S & B's “chicken/egg” problem, but in a slightly different key.)

Let's now turn to something that is ecologically valid. Our prelinguistic ancestors (and their nonlinguistic contemporaries, as well as our own) did face the problem of categorization and category learning. They did have to know or learn what to do with different kinds of things, in order to survive and reproduce: what to eat or not eat, what to approach or avoid, what kind of thing to do with what kind of thing, and so forth. But categorizing is not the same as discriminating (Harnad 1987). We discriminate things that are present simultaneously, or in close succession; hence, discrimination is a relative judgment, not an absolute one. You don't have to identify what the things are in order to be able to discern whether two things are the same thing or different things, or whether this thing is more like that that thing or that thing. Categorization, in contrast, calls for an absolute judgment of a thing in isolation: “What kind of thing is this?” And the identification need not be a name; it can simply be doing the kind of thing that you need to do with that kind of thing (flee from it, mate with it, or gather and save it for a rainy day).

So categorization tasks have not only ecological validity, but cognitive universality (Harnad 2004). None of our fancier cognitive capacities would be possible if we could not categorize. In particular, if we could not categorize, we could not name. To be able to identify a thing correctly, in isolation, with its name, I need to be able to discriminate it absolutely, not just relatively — that is, not just from the alternatives that happen to be copresent with it at the time (S & B's “context”), but from all other things I encounter, past, present, and (one hopes) future, with which it could be confused. (Categorization is not necessarily exact and infallible. I may be able to name things correctly based on what I have sampled to date, but tomorrow I may encounter an example that I not only cannot
categorize correctly, but that shows that all my categorization to date has been merely approximate too.)

Notice that I said categorize correctly. That is the other element missing from S & B's analyses. For S & B, there are three ways in which things can be categorized: (N) innately (“nativism”), (E) experientially (“empiricism”), and (C) culturally (“culturalism,” although one wonders why S & B consider cultural effects nonempirical!). To be fair, the way S & B put it is that these are the three ways in which categories can come to be shared – but clearly one must have categories before one can share them (the chicken/egg problem again!).

Where do the S & B agents' color categories come from? They seem to think that categories come from the “statistical structure” of the things in the world, such as how much things resemble one another physically, how frequently they occur and cooccur, and how this is reflected in their effects on our sensorimotor transducers. This is the gist of S & B's factor E, empiricism. Where the statistical structure has been picked up by evolution (another empirical process) rather than experience, this is factor N, nativism. But then what are we to make of factor C, culturalism? I think that what S & B really have in mind here is what others have called “constructionism.” With factors N and E, categories are derived from the structure of the world; with factor C they are somehow “constructed” by cultural practices and conventions. It is in this light that S & B introduce the “Whorf Hypothesis” (Whorf 1956), according to which our view of reality depends on our language and culture. But the Whorf Hypothesis fell on especially hard times with color categories, and S & B unfortunately inherit those hardships in using colors as their mainstay.

There are many ways in which color categories are unrepresentative of categories in general. First, they are of low dimensionality (mainly electromagnetic wave frequency, but also intensity and saturation). Second, they have a known and heavy innate component. We are born with sensory equipment that prepares us to sort (and name) colors the way we do with incomparably higher probability than the way we sort the categories named by most of the other nouns and adjectives in our (respective) dictionaries. Nor are most of the categories named by the words in our dictionaries variants on prototypes in a continuum, as colors are.

Yes, there are variations in color vision, color experience, and color naming that can modulate color categories a little; but let's admit it: not much! Moreover, color categories are hardly decomposable. With the possible exception of chromatographers, most of us cannot replace a color's name with a description – unlike with most other categories, where descriptions work so well that we usually don't even bother to lexicalize the category with a category-name and dictionary-entry at all. Even “the color of the sea” is only a one-step description, parasitic on the fact that you know the sea's color. Compare that with all the different descriptions that you could substitute for “chair.”

Why does describability matter? Because it gets much closer to what language really is, and what it is really for (Cangelosi & Harnad 2001). Language is not just a category taxonomy. We use words (category names) in combination to describe other categories, and to define other words, which makes it possible to acquire categories via instruction rather than merely the old, prelinguistic way, via direct experience or imitation. S & B think naming's main use is to tell you which object I have in mind, out of many we are both looking at now. (It seems that good old pointing would have been enough to
solve that problem, if that had really been what language was about and for."

But not only are color categories unrepresentative of categories in general, and the joint discrimination game unrepresentative of what language evolved and is used for, but categories do not derive merely or primarily from the passive correlational structure of objects (whether picked up via species evolution or via individual experience). It is not the object/object or input/input correlations that matter, but the effects of what we do with objects: the input/output correlations, and especially the corrective feedback arising from their consequences. What S & B's model misses, focusing as it does on discrimination and guessing games instead of the game of life, is that categories are acquired through feedback from miscategorization. We have this in a realistic mushroom foraging paradigm, but not in a hypothetical discrimination/guessing game (except if we gerrymander the game so that successful discriminating/guessing becomes the name of the game by fiat, and then that is fed back in the form of error-correcting consequences).

Yet all the right elements do seem to be there in S & B's simulations. They are simply not put together in a realistic and instructive way. The task of mind-reading in context seems premature. Every categorization in fact has two contexts. First, there is its context of acquisition, in which the category is first learned (whether evolutionarily via N or experientially via E) by trial-and-error, with corrective feedback provided by the consequences of miscategorization. The acquisition context is the series of examples of category members and nonmembers that is sampled during the learning (the “training set” in machine learning terms). Until language evolves, categories can only be learned and marked on the basis of an instrumental “category-name” (approaching, avoiding, manipulating, eating, mating). With language, there is the new option of marking the category with an arbitrary name, picked by (cultural) convention.

When a category has already been learned instrumentally, adding an arbitrary name is a relatively trivial further step (and nonlinguistic animals can do it too). But then comes the second sense of “context”: the context of application (for an already acquired category) in which the learned arbitrary category-names are used for other purposes. S & B's paradigm is, in fact, just one example of the context of application (telling you which of the colors that we are both looking at I happen to have in mind), but not a very representative or instructive one. Far more informative (literally!) is a task in which it is descriptions that resolve the uncertainty, and the alternatives are not even present. This is not discrimination but instruction/explanation. But for that you first need real language, and not just a taxonomy of arbitrary names (Harnad 2000).

What follows from this is that a “language game” in which words and categories are jointly coined and coordinated “on the fly,” as in S & B's color-naming simulations, is not a realistic model for anything that biological agents ever do or did. There is still scope for Whorfian effects, but those will come from the fact that both our respective experiential “training samples” (for all categories) and our corrective feedback (for categories about which our culture and language have a say in what's what, and hence also a hand in dictating the consequences of miscategorizing) have degrees of freedom that are not entirely fixed either by our inheritance or by the structure of the external world.

Categories are underdetermined, hence so are the features we use to pick them out. In machine learning theory, this is called the “credit/blame” assignment problem (“which of the many features
available is responsible for my successful or unsuccessful categorization?”), which is in turn a symptom of the “frame problem” (how to anticipate all potential future contingencies from a finite training sample?) and, ultimately, the “symbol-grounding problem” (how to connect a category-name with all the things in that category, past, present, and future?) Underdetermination leaves plenty of room for Whorfian differences between agents.

References


