

*EXPERIMENTAL ANALYSIS OF NAMING BEHAVIOR
CANNOT EXPLAIN NAMING CAPACITY*

STEVAN HARNAD

UNIVERSITY OF SOUTHAMPTON

The experimental analysis of naming behavior can tell us exactly the kinds of things Horne and Lowe report here: (a) the conditions under which people and animals succeed or fail in naming things and (b) the conditions under which bidirectional associations are formed between inputs (objects, pictures of objects, seen or heard names of objects) and outputs (spoken names of objects, multimodal operations on objects). The stimulus equivalence that Horne and Lowe single out is really just the reflexive, symmetric, and transitive properties of pairwise associations among the above. This is real and of some interest, but it unfortunately casts very little light on symbolization and language in general and naming capacity in particular. The associative equivalence between name and object is trivial in relation to the real question, which is: How do we (or any system that can do it) manage to connect names to things correctly (Harnad, 1987, 1990, 1992)? The experimental analysis of naming behavior begs this question entirely, simply taking it for granted that the connection is somehow successfully accomplished.

Note that I am not talking about the equally trivial mechanism of pairwise association between particular inputs (I) or between particular inputs and outputs (O). There is no mystery there. And unfortunately, the kinds of arbitrary association between specific pictures and written and spoken words examined in Horne and Lowe's Figure 1 and the arbitrary unique "class members" examined in Figures 15, 16, and 17 simply take us further away from what makes language special and send us off instead on the more general and less informative path of the experimental analysis of associative directionality.

What makes linguistic naming (as opposed

to pairwise I-I or I-O associations) special is that it requires categorization (Harnad, 1987): the abstraction of an invariance shared by the variable sensory projections of the members of the category designated by the name and not shared by the likewise variable sensory projections of members of other categories with which they might be confused. This is not the mere association of a specific X with a specific Y (whether X and Y be input or output, picture or name). And categorization does involve an equivalence, but it is unfortunately not Horne and Lowe's associative equivalence; rather, it is the input equivalence among all sensory projections of the members of the category named by the name. To realize that there is nothing in the analysis of associative directionality that can explain this kind of input equivalence, we must first stop thinking about tasks like pairing pictures of cars and spoken and written names of cars in the laboratory, because those are overlearned categories we already have. Even less relevant are the arbitrary classes in the Saunders and Green (1992) or the Lowe and Beasty (1987) studies cited by Horne and Lowe; those are not categories but associative clusters.

Think instead of a nontrivial naming task, one that the subject has not already overlearned before entering the experiment (which would beg the question): Think of cancer cell identification or chicken sexing (Andrews, Livingston, Harnad, & Fischer, 1994¹; Biederman & Shiffrar, 1987). What good does it do me if you tell me that the way I learned which kinds of cells were and weren't cancerous, or which newborn chicks were male and female, was by "associating the name with the stimulus"! This might be con-

Address correspondence to the author at the Department of Psychology, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom (E-mail: harnad@soton.ac.uk).

¹ Andrews, J., Livingston, K., Harnad, S., & Fischer, U. (1994, June). *Are concepts grounded in categorical perception? Some relevant empirical results*. Paper presented at the annual meeting of the Society for Philosophy and Psychology, Memphis, TN.

ceivable if there were only a small number of stimuli that I was shown over and over again until I memorized the pairwise connections. But that's not what happens with cancer cell identification or chicken sexing: I learn to recognize what *kind* of stimuli fall under which name (i.e., I learn the invariance underlying the variability and interconfusability). Such learning is not easy; it takes time. I cannot say how I am successfully doing it, once I can (so the question cannot be answered by introspection either). But what is certain is that there are some invariant properties that inputs from the members of the category designated by the name share, and that reliably distinguish them from members of other categories with which they could be confused, and that I have somehow managed to learn them. The naming problem requires an answer to the question of how I (or any system that can do it) manage to do that.

It will not do to reply that most of our named categories are not as difficult as cancer cell identification or chicken sexing. First, cancer cell identification and chicken sexing are not difficult for the successfully trained expert; they are only difficult for us, who have not learned how to do it. Well, the child is in that position initially, with respect to all the objects, events, and states of affairs in the world. The fact that children eventually succeed in naming the ones they do is thanks to the same capacity that allows adults to learn to do chicken sexing. Anything in between is simply begging the question: taking for granted the vast but still-unexplained naming capacity of the child (and the adult), or focusing on trivial associations between arbitrary stimuli and "names."

Renaming the capacity as the exercise of listener behavior or echoic behavior does not explain anything either. If the child already knows what category of things names name before ever uttering them, then the critical connection was made through hearing names and interacting with objects in the presence of naming and pointing by others, but it still does not explain *how*. Given that the connection is made, all the associative equivalences can come to be useful (of course I think of objects when I name them, and think of their names when I see or operate on or think about objects), but what can I thank for this capacity in the first place?

Why do I emphasize what the experimental analysis of behavior takes for granted (and characteristically relegates to another specialty, such as innate perceptual mechanisms or brain function in general)? Because whatever turns out to be the true substrate of naming capacity will in turn cast light on both the substrate of language and its unique symbolic/propositional power (Steklis & Harnad, 1976), which is decidedly not just the power of associative equivalence between name and object! Yes, perceptual processes are involved in categorization, and some of them may well be innate, but most of them are not, because in naming we are dealing mainly with *learned* perceptual categories: How many of the entries in a dictionary do you think we were born with specific built-in invariance detectors for? No, most of our lexicon was purchased by perceptual learning (analogous to the learning in cancer cell identification and chicken sexing) rather than by inborn Darwinian detectors. If inborn detectors *were* responsible for most of our naming capacity, then the question of the origin and nature of language would become the question of the origin and nature of those detectors (Harnad, 1976), but it is not. Most of our named categories are learned rather than inborn, that is, they are not acquired by phylogenetic "theft" but by ontogenetic "honest toil." The honest toil consists of sampling instances of members and nonmembers and laboriously learning (attention Skinnerians), *from the consequences of miscategorization*, which are which (Skinner, 1984b).

But behavior analysts cannot take too much heart from the familiar "shaping by consequences" phenomenon at work here (Cattania & Harnad, 1988), because the fact that successful categorization performance can be shaped by trial-and-error learning with feedback is not an explanation; it is what *calls* for explanation: What internal properties must a system have in order to be capable of learning the categories we can learn? To see that the behavior-analytic approach is nonexplanatory in this regard, consider how much better off roboticists (Harnad, 1994) would be if, knowing the categorization skills people were capable of but clueless as to how to build a system that could do that, they were informed by behavior analysts that it is accomplished on the basis of feedback from consequences! And that there is a relation of

“equivalence” between the category names and the objects they designate!

Nor does the unfulfilled explanatory agenda stop there, for there is another mystery about language (over and above the mystery of how we manage to name things), and it too is a form of theft. If “Darwinian theft” is the source of the categories we are born already able to detect and “honest toil” is new category learning through trial and error with feedback, then once we have a repertoire of category names earned by honest toil, language gives us the unique further symbolic capacity to acquire new category names from strings of prior names alone: To use an example I have used many times before (Harnad, 1990), if you have learned, by honest toil, to call horses “horses” when you see them, reliably distinguishing them from members of other categories with which they might be confused, and you have learned, likewise by honest toil, to call “stripes” stripes, then, even though you have never encountered one, you are in a position to correctly name your first zebra upon merely being told that a “zebra” is a “horse” with “stripes.” That’s symbolic/propositional

theft; it can spare you an awful lot of honest toil; and it is the true power of language. Nor is it explained by (or equivalent to) associative equivalence. It will be explained by a successful explanation of what internal structures and processes give us the capacity to learn to categorize and name classes of inputs by detecting the invariance in their sensory projections (Harnad, 1992; Harnad, Hanson, & Lubin, 1991, 1995), and then how strings of names in the form of propositions about category membership can give us the capacity to name new members of categories we have not encountered before.

An explanation like this is impossible from just the experimental analysis of behavior. One must also hypothesize and then analyze the internal structures and processes that generate the capacity to exhibit the behavior (Harnad, 1982, 1984). It is only from such research that we will come to understand the origins of naming and symbolic behavior (Harnad, in press). An internal representational analysis of the honest-toil versus symbolic-theft distinction might even cast some explanatory light on Skinner’s named but unexplicated distinction between contingency-based and rule-governed behavior (Skinner, 1984a).

*NAMING AS A TECHNICAL TERM: SACRIFICING
BEHAVIOR ANALYSIS AT THE ALTAR OF POPULARITY?*

DERMOT BARNES

UNIVERSITY COLLEGE CORK

Horne and Lowe present a gold mine of information and deserve our admiration for presenting their material so clearly. Neverthe-

I would like to thank Bryan Roche for discussing with me many of the main issues addressed in this commentary.

Correspondence should be addressed to Dermot Barnes, Department of Applied Psychology, University College Cork, Cork, Ireland (E-mail: IN%“stay8028@iruccvax.ucc.ie”).

less, I have two reservations about their work. First, Horne and Lowe propose that using the term *naming* would “foster productive interaction with scientists from other traditions” (p. 186). I, too, favor greater cooperation with nonbehavioral researchers (e.g., Barnes & Hampson, 1993, in press; Barnes & Holmes, 1991; Cullinan, Barnes, Hampson, & Lyddy, 1993), and would happily adopt naming as a nontechnical term to foster such co-