DEVELOPING A THEORY OF DERIVED STIMULUS RELATIONS

STEVEN C. HAYES

UNIVERSITY OF NEVADA

What do we want from a behavior-analytic theory of derived stimulus relations? Indeed, what do we want from behavior-analytic theories of any kind?

Behavior-Analytic Theorizing

It is ironic that some behavior analysts are afraid of theory, because few fields are more committed to an essentially theoretical purpose. Behavior analysts seek the prediction and control of interactions between organisms and their environments. As an analytic strategy, behavior analysts start with careful observations of relatively simple and controlled examples of organism-environment interactions, and identify how these interactions are influenced by manipulable events. Over time, generally applicable ways of speaking about these interactions and the variables that influenced them are abstracted; that is, we develop *behavioral principles.* These principles are only "true" if they help us to predict and control with high precision and with broad scope. Behavior analysts are extremely interested in such abstractions, and teach courses, write books, and develop applied programs based on them.

But behavioral principles were never an end in themselves. They were meant as analytic tools with which to evaluate complex (especially complex human) behavior. When we do an individual functional analysis, we use these tools with a specific case. Over time, however, we abstract generally applicable ways of speaking about a domain of behavioral interactions in terms of sets of interrelated behavioral principles. This is a *behavioral theory.* The relation between behavioral principles and behavioral theories is very much like that between behavioral observations and behavioral principles. In each case we are moving from the specific to the general case—we are seeking more scope, while retaining precision. This is the sense in which a concern for behavioral principles is an essentially theoretical purpose because behavioral theories are the highest expression of the same analytic strategy. And in both cases they are "true" only if they are useful in organizing our contact with events.

Behavioral theories are very distinct from the kinds of theories that litter mainstream psychology. They are not hypothetical. They are not mediational. They are not tested solely by predictive verification. They are meant to illuminate phenomena and not the other way around.

Some behavior analysts are like makers of fine hammers and drills who have forgotten what the hammers and drills were meant to do in the first place. We polish and shine our tools in endless strings of studies done oh so carefully. Instead of building analytic structures with our tools, we put them in well-lit glass cases so that they can be admired. Meanwhile, where are the generally applicable empirical analyses of important classes of behavioral phenomena? Where are the empirically based behavioral theories of emotion, sexuality, metaphor, suicide, confidence, friendship, consciousness, meaning, religiosity, and so on?

With that as background, I can only applaud what Horne and Lowe are doing here. They are developing a behavioral theory. Serious efforts of that kind are good for the field. I believe that relational frame theory (RFT) is another example of the kind of analytic-abstractive theories that are needed within behavior analysis. All of this will be positive, as long as the focus stays on the phenomena and not who is right, who said it first, how experimental minutia will be used to "prove" one theory over another, and so on.

Preparation of this article was supported in part by a grant from the National Institute on Drug Abuse, Grant DA08634.

Requests for reprints should be addressed to Steven C. Hayes, Department of Psychology, University of Nevada, Reno, Nevada 89557-0062.

Comparing Relational Frame Theory and Naming

Horne and Lowe have me at a huge disadvantage—I have only a few words allocated, most readers will not have read these RFT pieces, and their criticisms often misstate my position (particularly when dealing with issues I have mentioned only in passing). For example, I do not hold to the view of meaning they quote—I stand by these quotes, but only in conjunction with a technical analysis that would have to change what Horne and Lowe take them to mean. I can only state that none of their criticisms of RFT seem difficult to me, and give an example or two (see Hayes & Wilson, in press, for a more detailed analysis of other recent criticisms of RFT).

Horne and Lowe chide me for some ambiguity about the histories necessary for the formation of relational frames. The core of relational frame theory is this simple question: Can we think of deriving stimulus relations as an operant? I have done some speculation on the histories needed to think of relating that way, primarily to orient readers to the issue. Unlike Horne and Lowe's account, most of these imagined histories are quite simple (for a reason that I will discuss shortly). Of course, what the histories actually are is an empirical matter, and that is where I think we need to put our greatest effort. It is not by accident that Horne and Lowe often cite my studies or those of my students when they are in need of empirical support for their theory. Our perspectives are similar, and we have not been standing still.

It seems to me that naming is one example of a frame of coordination. Horne and Lowe deny that directly, but they do so because of their very great emphasis on the behavior required to link hearing with saying. They take this to be a pivotal difference between our positions, but I do not. In fact, none of the data in support of their position are incompatible with RFT. The reverse is not true.

Horne and Lowe's focus on listening and echoing comes from their overwhelming emphasis on oral naming. Their analysis relies too strongly on examples that the reader understands because of participation in a verbal community. Let me give a less common example. Imagine a mute child who looks at a dog and then hears her mother saying "look, a dog." Now the child hears "where's the dog?" and the child looks for a dog. In Horne and Lowe's account we must suppose that the mute child somehow had to learn gestural or other forms of echoing first in order to show this performance. I do not see echoing as necessary at all, although it may be helpful.

Horne and Lowe follow Sidman's criticism of RFT: "a linguistically naive organism's abstraction of commonalities from a set of exemplars that share no physical features requires more of an explanation than just a history of experience with the exemplars" (1994, p. 557). Here we have a nice focal point for thinking and research. Is it enough of a behavioral explanation in such cases merely to point to the histories that give rise to these operants? I think it is (and thus my imaginary histories are indeed simple), but Horne and Lowe, like Sidman (1994), make the odd claim that even if equivalence occurs in the manner RFT specifies, it provides "little by way of explanation" (Horne & Lowe, p. 232). What is missing? A hypothetical construct? A mediating variable? What "explains" *any* operant other than the history that produces it?

What then *is* the role of the kinds of behavioral processes (e.g., echoing) described by Horne and Lowe? I believe that the role is to enable the individual to engage in his or her relational repertoire, and to support its further development by making this repertoire more effective. Let me provide some evidence.

We have shown that a 16-month-old infant can show symmetry in matching-to-sample procedures (Lipkens, Hayes, & Hayes, 1993). The training was "see object"–"hear word" (actually "say word," but often it was unsuccessful in training) and the testing was "hear word"–"touch object." Note two things. First, the test requires touching, whereas the training required only seeing. But babies learn very early to touch what they see. We do not need to suppose that the see-touch performance (is this "touch echoing"?) underlies the symmetrical performance. It is presumably established elsewhere and in this context is merely an aspect of the symmetrical performance. Second, the child did not reliably echo the words, despite efforts to produce this, and still showed symmetry.

At that early age, however, the child did *not* show symmetry when the training was "hear word"–"touch object" and the testing was "see object"–"say word." In our analysis we suggested that the child may have needed additional echoic training. But note that the child could clearly derive symmetrical stimulus relations (in the other task). This must mean that the lack of verbal echoic behavior merely limited the applicability of that repertoire to a specific overall performance.

Later in that same study we showed that even when this performance was established (hear-touch leading to see-say), if the child touched a novel object in the presence of a novel name (exclusion), he did not then say the name when shown the object. That performance emerged several months later. In other words, relational abilities developed over time (in this case derived symmetry based on derived exclusion), even when echoing was well established. That suggests to me that the details of a specific performance (e.g., seeing can later lead to touching, hearing can lead to saying, and so on) allow the individual to *engage in* his or her relational repertoire, and thus may support its further development by making this repertoire more useful, but that they are nevertheless *distinguishable* from that repertoire. If Horne and Lowe would expand their examples to include writing, signing, blind children, deaf children, mute children, and so on, this would be more evident. If my argument is correct, RFT can easily incorporate naming as a special case.

The three predictions made by Horne and Lowe as tests of their theory have all been suggested in writing by RFT as well, although for different reasons than those supposed by Horne and Lowe. RFT suggests many additional findings, however, and to the extent that they have been tested, these findings have been repeatedly confirmed (see Hayes, 1994, for a recent review). The Steele and Hayes (1991) study, for example, is obvious from the point of view of RFT. If subjects are given nonarbitrary relational pretraining that brings different relational responses under contextual control, those cues can then be used to alter the relational performance that occurs in an arbitrary matching-to-sample procedure. Arbitrary matching-to-sample procedures thus can produce derived relations of an infinite variety—not just equivalence—given the proper contexts that establish or occasion particular relational responses. The Steele and Hayes study has been repeatedly replicated and extended (e.g., Dymond & Barnes, 1995; Lipkens, 1992). Yet competing theories, including Sidman's and Horne and Lowe's, are simply silent on this whole range of relational behavior that is involved in the actual use of language. As a challenge to Horne and Lowe, I would like to have them explain the Steele and Hayes data using their approach.

Horne and Lowe's theory is somewhat similar to RFT (although I would like to see them actually say "naming is an operant," and I am suspicious about why they do not), and thus it will be difficult to devise "tests" between them (cf. Barnes, 1994). I think it is crucial in this context to keep our eye on the prize. It would be a very bad thing if the development of behavioral theories leads to traditional hypothesis-testing research. The goal is not to test theories. The goal is to predict and control behavior with precision and scope. The great prize is an actual empirical analysis of reasoning, metaphor, rule governance, understanding, and the like. Any tests should point to real differences with phenomena of fundamental importance or the tests should not be done.