JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2004, **82,** 225–234 NUMBER 2 (SEPTEMBER)

GENERIC RESPONSE CLASSES AND RELATIONAL FRAME THEORY: RESPONSE TO HAYES AND BARNES-HOLMES

DAVID C. PALMER

SMITH COLLEGE

Hayes and Barnes-Holmes (2004) assert that the concept of a topographically unconstrained response class, the concept that carries the explanatory burden of relational frame theory, appeals to no new principles. Operants are properly defined functionally. I argue that they have stretched the concept of the generic response class beyond its appropriate limits. Skinner conceived of response classes as empirically defined units, mutually interchangeable in quantitative functions. The notion of overarching, generalized operants is an uncritical, analogical extension of this concept. I hold that the conceptual work of relational frame theory is incomplete, that a statement of principle is necessary, even if not new. Finally, I distinguish a supposed commitment to a philosophical ''mediationism'' from a valid inquiry about mediating behavior; that is, behavior with stimulus products that participate in the control of the behavior of primary interest.

Key words: generic units, mediating behavior, relational frame theory, relational frames, response classes, stimulus classes, units of analysis, verbal behavior

In preparing for my review (Palmer, 2004) of *Relational Frame Theory: A Post-Skinnerian Account of Human Language and Cognition* (Hayes, Barnes-Holmes, & Roche, 2001a), I struggled to understand relational frame theory, and I was puzzled by the magnitude of that struggle. Why should I have so much trouble understanding the prose of other behavior analysts? In their vigorous rejoinder to my review, Hayes and Barnes-Holmes (2004) ask the same question and offer the following suggestion: Together with other critical reviewers (Burgos, 2003; Malott, 2003; Tonneau, 2001, 2002), I represent a mechanistic philosophical tradition, incompatible with the authors' contextualistic stance. The term *mechanistic* is used pejoratively, but I don't understand why. I simply want to know how the world works.

Nevertheless, to the extent that the contrast reflects different approaches within behavior analysis, I think it does indeed help explain why I have found it so hard to understand relational frame theory. The theory provides a vocabulary for describing relational behavior and a framework for studying and interpreting it, but I had expected that which Hayes and Barnes-Holmes might call a mechanistic account. That is, I had expected a statement of an empirically derived principle of this form: Under specified conditions, a specified response class will emerge that shows the properties of mutual and combinatorial entailment and specified transformations of stimulus function. Such a statement would presumably be an inductive generalization from a host of studies and would stand as the impetus of a reinvigorated behavior analysis. It would indeed justify the excitement of the authors, for it would lead to new applications and to the interpretation of more complex cases outside the laboratory, and it would enable us to account for much more of the troublesome variance in behavior.

However, such a statement of principle is not on offer. As a result, I argued, relational frame theory is not a theory but a proposal; it is a proposal that the ultimate form of the above statement of principle will take a certain form. Specifically, the reader is told that relational behavior emerges from exposure to a history of exemplars as a kind of generalized response class. But the details of this history, its particular effects, and what counts as an exemplar, are not specified, nor is justification offered for the concept of a response class in which responses are not mutually interchangeable. These are not trivial omissions.

Hayes and Barnes-Holmes address this central criticism by insisting that no such principle is required, that relational frames are simply another operant, albeit an overarching, generalized operant. Given enough ex-

Address correspondence to the author at Department of Psychology, Smith College, Northampton, Massachusetts 01063 (e-mail: dcpalmer@smith.edu).

amples of the shaping of behavior under control of stimulus relations, a subject will emit generalized relational behavior on future occasions without training. The problem, as they see it, is that I, and others of like mind, have failed to grasp the functional nature of the operant. This, I now believe, is the heart of the controversy, and their reply has greatly clarified that point.

I will confine my response to this central point and two closely related topics. Hayes and Barnes-Holmes raise many detailed objections to my review; I believe they can be answered, but I will leave most of that labor as an exercise for the reader. Exchanges of this sort can assume a polemical tone, which I wish to avoid, and I don't want to abuse the privilege of having the last word. However, it may be helpful to clear up some minor misunderstandings—misunderstandings that are probably inevitable in matters as complicated as those under discussion. I did indeed err when I asserted in my review that the authors had contradicted themselves (Palmer, 2004, p. 195). I did not understand that when they used the phrase, "target of a contingency" they meant, "the primary, not secondary, effect of that contingency." It was an error—an understandable one in my view—but it did not affect my conclusion. However, on their side, Hayes and Barnes-Holmes misunderstood my comments about using the term relational frame to refer to the history that produces behavior under the control of stimulus relations. My objection is not to the notion that relational frames entail a certain history, just as operants entail a history of differential reinforcement; I was objecting to the apparent use of the term relational frame to *refer* to that history. Whiskey entails a history of distillation, but we do not use the term whiskey to denote distillation. The problem is not logical, but practical: In a context that is already fraught with difficulty, using words in multiple senses imposes an unnecessary burden on the reader. But that, too, was a misunderstanding. It is now clear that when the authors say that the term relational frame is a process concept, they do not mean that the term is the name of the process, only that it entails the process. Finally, Hayes and Barnes-Holmes are too eager to economize: They wish to dismiss my review, the review of Burgos (2003), and the commentary of Tonneau (2002) with one stroke, but the reviews have little in common. In particular, neither of the other reviewers endorsed what Hayes and Barnes-Holmes call "mediational associative processes," and Tonneau is decidedly critical of such accounts. The reader is advised to weigh those reviews on their own merits. On the other hand, I believe that my review is entirely compatible with that of Malott (2003), although they differ in detail. But these are peripheral matters; I turn now to a discussion of the generic nature of response classes, which I believe is at the center of our disagreement.

THE GENERIC NATURE OF THE OPERANT

Relational frame theory "treats relational responding as a generalized operant, and thus appeals to a history of multiple-exemplar training. Specific types of relational responding, termed relational frames, are defined in terms of the three properties of mutual and combinatorial entailment, and the transformation of functions" (Hayes, Barnes-Holmes, & Roche, 2001b, p. 141). The various relational responses embraced by this definition are conceived of as a generalized operant: "From this perspective, therefore, responding to B given A and to A given B may be considered a single response unit controlled by a relevant contextual cue (or cues) by virtue of its previous correlation with differential reinforcement. In effect, the RFT approach invokes a purely functional concept of an operant, and the term 'overarching operant class' . . . is used to emphasize this fact" (p. 146). As I understand these statements, all of the behavior under the control of a web of stimulus relations is viewed as a single response class. Because a relational frame is a class of behavior showing not only mutual and combinatorial entailment but also transformation of stimulus function, it typically embraces a topographically heterogeneous set of responses. It is on this point that the explanatory power of relational frame theory rests, for the puzzle at hand is how to explain responses that are derived or transformed, that is, that have not been taught. If they are all members of a single class, then this emergence is not a puzzle: "A single specified relation between two sets of relata might give rise to myriad derived relations in an instant. Entire sets of relations can change in an instant'' (Hayes, Fox, et al., 2001, p. 40). Although this claim seems plausible in the abstract, it can be startling in the particular: If A has appeared in a frame of opposition to both B and C, a subject is likely to report, for example, that ''C is the opposite of A,'' but that ''C is the same as B.'' If the subject has been trained to press a lever with maximum force in the presence of A, he can be expected, without training, to press it with minimal force in the presence of C. According to relational frame theory, all of these responses are members of the same response class and change in strength together. One assumes that there are no limits to the kinds of responses that can be glued together into a single response class by relational contingencies; flying an airplane into the World Trade Center under control of verbal instructions to do so is presumably relational behavior and therefore a member of the same response class as responses such as reading the words or pointing to pictures.

This conceptualization is defended on the grounds that the operant is defined functionally, not topographically: ''The concept of a response class with an infinite range of topographies is a defining property of operant behavior, and has been from the very beginning (e.g., see Skinner, 1938, p. 33–41)'' (Hayes et al., 2001b, p. 147). They cite imitation, attending, identity matching, exclusion, novelty, randomness, and several other examples from the behavioral literature as precedents.

In my view, this stretches the notion of a topographically heterogeneous class beyond its intended limits. To evaluate this claim, let us consider the origins of the argument that the appropriate unit of analysis in the science of behavior is a generic term, embracing responses with a variety of topographies. Skinner first discussed the topic in *The Generic Nature of the Concepts of Stimulus and Response* (1935/1999a), and he repeated much of his discussion verbatim in *The Behavior of Organisms* (1938). We will find that Skinner's analysis by no means supports the interpretation of the operant adopted by the proponents of relational frame theory. His analysis does not specifically preclude adopting their interpretation on other grounds, but he provides no precedent.

When first faced with the problem of identifying units of environment and behavior, Skinner (1935/1999a) recognized that the experimenter cannot simply define appropriate units in advance; doing so might yield perfectly objective terms, but such terms might not be related to controlling variables in orderly ways. (Much of modern linguistics has been following a false scent because of the a priori assumption that the sentence, formally defined, is the appropriate unit of analysis in language; see, for example, Palmer, 1986/2000a, 2000b; Palmer & Donahoe, 1992.) Rather, the units should be determined empirically, using the orderliness of one's data as a criterion. When we do so we will find that, as we gradually restrict our definitions, appropriate units emerge at a level short of complete specificity. That is, to get smooth curves in our functional relations, we need not specify every detail of the target response:

> Suppose that we are studying the behavior of such an organism as a rat in pressing a lever. The number of distinguishable acts on the part of the rat which will give the required movement of the lever is indefinite and very large. Except for certain rare cases they constitute a class, which is sufficiently well-defined by the phrase ''pressing the lever.'' Now it may be shown that under various circumstances the rate of responding is significant—that is to say, it maintains itself or changes in lawful ways. But the responses which contribute to this total number-per-unit-time are not identical. They are selected at random from the whole class—that is, by circumstances which are independent of the conditions determining the rate. Not only, therefore, are the members of the class all equally elicitable by the stimulation arising from the lever, they are *quantitatively mutually replaceable.* The uniformity of the change in rate excludes any supposition that we are dealing with a group of separate reflexes and forces the conclusion that ''pressing the lever'' behaves experimentally as a unitary thing. (Skinner, 1935/1999a, p. 508)

Note the italicized phrase: the different topographies that the response can take, when defined at a certain level of specificity, are mutually replaceable in the quantitative measures of behavior. For example, a run of relatively forceful responses, or a run of responses with the left paw, would not affect the smoothness of the curve representing the rate of behavior in a particular schedule of reinforcement. Moreover, Skinner (1938) notes that if one tries to restrict the definition of the environment or of behavior too narrowly—for example, by defining the lever press as only those responses made with the left paw, with a certain latency and force—the orderliness of one's data actually decreases:

> If only such responses as had been made in a very special way were counted (that is, if the response had been restricted through further specification), the smoothness of the resulting curves would have been *decreased*. The curves would have been destroyed through the elimination of many responses that contributed to them. The set of properties that define 'pressing a lever' is thus uniquely determined; specifying either fewer or more would destroy the consistency of the experimental result. (pp. 37–38)

Skinner (1938) was cautious about the status of responses of widely discrepant topographies:

> It is true that the non-defining properties are often not wholly negligible and that the members of classes are consequently not exactly mutually replaceable. On the side of the response, the data will not show this in most cases because of the present lack of precision. But it is certain that there are outlying members of a class which have not a full substitutive power; that is to say, there are flexions and pressings that are so unusual because of other properties that they do not fully *count as such.* It ought to be supposed that lesser differences would be significant in a more sensitive test. If we should examine a large number of responses leading to the movement of the lever, most of them would be relatively quite similar, but there would be smaller groups set off by distinguishing properties and a few quite anomalous responses. It is because of the high frequency of occurrence of the similar ones that they are typical of the response 'pressing the lever,' but it is also because of this frequency that any lack of effectiveness of atypical responses is not at present sufficiently strongly felt to be noted (p. 38)

It should be apparent from these passages that Skinner by no means endorsed ''the concept of a response class with an infinite range of topographies'' (Hayes et al., 2001b). Skinner's response classes were dominated by responses of similar topography; outlying responses were permitted into the class only if they did not seriously disrupt his curves. As members of a response class, they might not ''fully count as such,'' but ''the data will not show this in most cases because of the present lack of precision.'' The size of the appropriate response class emerges from an experimental analysis; it does not arise from an a priori philosophical commitment, nor need it coincide with an experimenter's criterion for delivering reinforcement. Moreover, Skinner's criterion for determining class membership was that the various members of the response class be mutually replaceable in quantitative functions. This criterion is entirely inappropriate for the putative relational operants of relational frame theory. Pointing to ''same'' is not mutually replaceable with pointing to ''opposite,'' or with pressing a lever forcefully or lightly. The responses in relational frames are often topographically disparate if not entirely unrelated, and it is self-evident that they do not have ''full substitutive power.''

I am not arguing that relational responses cannot be shown to hang together as response classes according to some other criteria. But they do not fall neatly into Skinner's scheme. Mutual substitutability in quantitative functions will have to be abandoned as a criterion, and new criteria will have to be advanced and defended. What are the implications of having two independent procedures for determining response class membership? Must we throw out Skinner's concept of the operant along with his interpretation of verbal behavior? Hayes, Barnes-Holmes, and Roche note that ''a large number of studies . . . have explicitly examined the possibility that arbitrarily applicable relational responding can be thought of as a functional response class. So far as we are aware, every examination of this issue to date has been supportive of a functional conception'' (2001b, p. 149). But I must assume that they were not using Skinner's criteria for defining a response class; what did they use, and why? It is not enough that a putative response class be shown to be sensitive to contingencies of reinforcement; Skinner's point is that there are many possible definitions of a response class of varying generality, all of which might show some sensitivity to controlling variables, but they differ in the orderliness of the relevant quantitative relations. The functionally defined operant that lies at the heart of relational frame theory may be an illusion. At present it is too poorly worked out to carry any explanatory burden.

Notice that my argument applies equally to a claim that responses emitted in experiments showing generalized imitation, novelty, attention, and so on, are all members of a single response class. In my opinion, there is nothing to be gained by making such a claim; the phenomena are so complex and so poorly understood that it is premature to do so. As with relational frame theory, one remains faced with the problem of advancing and defending new criteria for determining one's units of analysis. That is not to say that such responses are not sensitive to reinforcement contingencies. They have been shown to be. But as I have remarked, that does not, by itself, indicate that the response units have been appropriately identified. If we are to follow Skinner's example, deciding on the appropriate definition of the response class calls for a systematic evaluation of competing response definitions, with gradually shifting criteria in search of optimal criteria of both environment and behavior. That analytical task so far exceeds our ability to control relevant variables that we may need to be satisfied with quite tentative understanding of such complex phenomena. No explanatory burden can be borne by assumptions that such heterogeneous responses are all members of the same response class.

Hayes, Fox, et al. (2001) remark that, "Even a large unit of behavior with widely varying topographies, such as writing a novel or driving to the beach, might be usefully analyzed as an operant" (p. 22). I have long advocated the interpretation of complex phenomena in terms of principles that have emerged from basic research when the phenomena cannot be experimentally analyzed (e.g., Donahoe & Palmer, 1989, 1994; Palmer, 1991, 2003), but all such interpretations run the risk of being too facile, and I believe that is the case here. It may seem plausible that one is more likely to write a second novel if one's first has met with critical acclaim, but how do we account for the first instance in a way that does not also account for the second? What are we to make of half-finished novels, or of the prolific writer who never gets published, or of a novelist like Harper Lee, who wrote no second novel after the extraordinary success of her first? Our interpretation does not account for the variance in behavior. In my opinion, the effect of consequences on such heterogeneous activities is best considered to be only an analog of reinforcement (cf. Malott & Suarez, 2004). I cannot claim that interpreting such large and heterogeneous behavioral events as operants is indefensible, but to do so is to abandon Skinner's concept of empirically defined response classes. Moreover, when behavior analysts make such claims, they issue a license to others to do the same. Apparently believing that one can define units of behavior as one pleases, Chomsky (1971) argued that the notion of probability of response is meaningless in the domain of verbal behavior:

> What does it mean to say that some sentence of English that I have never heard or produced belongs to my "repertoire," but not any sentence of Chinese (so that the former has a higher "probability")? (p. 20)

Chomsky's error is in assuming that a sentence one has never heard or used is an appropriate unit of behavior. He made the same error by asking how the act of suicide could ever occur, because it could never have been reinforced. He was unaware that Skinner (1953/1999b) had addressed that problem many years before:

> Another common objection is that if we identify probability of response with frequency of occurrence, we cannot legitimately apply the notion to an event which is never repeated. A man may marry only once. He may engage in a business deal only once. He may commit suicide only once. Is behavior of this sort beyond the scope of such an analysis? The answer here concerns the definition of the unit to be predicted. Complex activities are not always "responses" in the sense of repeated or repeatable events. They are composed of responses, however, which are repeatable and capable of being studied in terms of frequency. The problem is again not peculiar to the field of behavior. Was it possible to assign a given probability to the explosion of the first atomic bomb? The probabilities of many of the component events were soundly based upon data in the form of frequencies. But the explosion

> of the bomb as a whole was a unique event in the history of the world. Though the probability of its occurrence could not be stated in terms of the frequency of a unit at that level, it could still be evaluated. The problem of predicting that a man will commit suicide is of the same nature. (p. 107)

In this nontechnical context, Skinner did not describe how one decides on an appropriate unit of analysis, but he made it clear that they are not arbitrary. Writing a novel or driving to the beach seem to me to be unprofitable units of analysis.

Stimulus Classes

With Donahoe and Burgos, I have argued that the appropriate unit of analysis in our field is an environment-behavior relation (Donahoe & Palmer, 1994; see Donahoe, Palmer, & Burgos, 1997, for an extended discussion with commentaries.) One cannot specify response classes except in relation to controlling variables. For example, the mand *Fire!* is a different operant from the tact *Fire!* It is particularly in this sense that the topography of a response is an inadequate defining criterion of a response class. The claim that all of the responses in a relational frame are members of the same response class implies that all of the correlated stimuli are members of the same stimulus class. In stimulus equivalence procedures, such a claim might be uncontroversial, but relational frame theory embraces all stimulus relations, in some examples of which the implication is less plausible. Recall that ‘‘From [relational frame theory's] perspective, therefore, responding to B given A and to A given B may be considered a single response unit'' (Hayes et al., 2001b, p. 146). Extrapolating to a more complicated example, this suggests that if stimuli A, B, and C are in a frame of opposition to stimuli D, E, and F, the utterances ‘‘A is the same as B'' and ‘‘C is the opposite of D'' are also a single response unit, and it suggests that all six stimuli are members of the same stimulus class. This formulation treats statements under control of different events as members of the same response class and merges stimuli that evoke different activities into the same stimulus class; in neither case are the events mutually interchangeable. This is a departure from traditional conceptions of response classes and stimulus classes, and it ignores functional differences among different environment-behavior units.

Origins of Response Classes

Together with Donahoe (Donahoe & Palmer, 1994; Palmer, 1997, 1998a), I have speculated that the generic nature of the operant arises, at least in part, from the fact that even the simplest responses observed in the freely moving organism are mediated by populations of many thousands of muscle fibers innervated by as many neurons, typically firing at an appreciable resting rate. What seems to be a unitary act emerges from a boiling anthill of activity. It is unlikely that any two such acts are ever identical. Different, but overlapping, subpopulations of neurons and fibers mediate successive responses and are putatively differentially affected by reinforcement. Even when an effector is reliably actuated, variability in behavior from one occasion to the next is inevitable. This scheme accounts for a certain amount of variability in response form, but it does not explain why, for example, a rat might operate a lever both by nibbling it and pressing it with its paws. The responses are so divergent in form that we can assume different populations of neurons and muscle fibers. There are no a priori grounds for predicting generalization from one form to the other. But a prolonged reinforcement contingency might capture a variety of topographies of different origins, so long as the contingency has been satisfied in each case. That is, nibbling the lever appears more or less interchangeably with pressing the lever in the terminal performance only because it has frequently been reinforced. Thus an operant might embrace responses of conspicuously different topographies, but only if they all share a history of reinforcement in the same context. (Analogously, heterogeneous stimuli can become members of a common stimulus class for the same reason; cf. Vaughan, 1988.) From this perspective, an operant does not include every imaginable topography that will close a microswitch (i.e., every form that is functionally equivalent); it includes only those that share a history of reinforcement in the same context or in similar contexts. The concept of functionally defined response classes is not infinitely elastic.

It is an easy matter to square Skinner's concept of generic response classes with such interpretations, but I cannot see that the concept of the relational operant is compatible with them. This is not evidence that such a concept is incorrect, but it helps explain why I, and perhaps other behavior analysts with a ''mechanistic'' bent, regard relational frame theory as incomplete.

A STATEMENT OF PRINCIPLE IS NEEDED

Even if I accepted the claim of Hayes and Barnes-Holmes that relational behavior is simply another operant—which I clearly do not—I would still insist that a statement of principle is required to explain relational operants. I am puzzled by their arguments in this context. They assert that a relational operant is simply another generalized, overarching operant, appealing to no new principles, but that a new principle is implied by such relational operants. The example that forces them to entertain the necessity of a new principle, as I understand it, is one of transfer of stimulus function in the absence of relevant training. But transfer, or transformation, of stimulus function is one of the defining features of relational frames and should be entailed by whatever produces relational operants in the first place. Nevertheless, I will concede whatever point they are making for purposes of getting on with the discussion. But I still insist on asking them to provide a statement of principle of the form I sketched in my introduction. If it is not ''new and mysterious,'' then the task should be an easy one.

Hayes and Barnes-Holmes snort at examples drawn from *Through the Looking Glass* or from everyday anecdotes, and they point to the number of painstaking studies and the wealth of objective data upon which their arguments rest. But the volume of work generated by a theory is not, by itself, evidence of its merits. The history of psychology is littered with research programs that have led nowhere. Clarity is needed: Where do relational frames come from (cf. Galizio, 2004)? What are the critical features of the history of multiple exemplar training said to lead to relational frames, and what are the specific effects of such a history? If we do not have such a statement of principle, we cannot apply relational frame theory unambiguously to the everyday examples I cited, and any theory of behavior should be applicable to such examples. Hayes and Barnes-Holmes note that relational frame theory would predict Alice's puzzlement when faced with a blizzard of elementary sums. But I didn't think so; that's why I advanced it as a counterexample. Without a statement of principle, the predictions of the theory are in the eye of the beholder.

One suspects that, having defined verbal behavior in terms of relational frames, the authors interpret complex phenomena by simply predicting what they as verbal organisms would do. The reader, a verbal organism, acquiesces. This is perhaps unavoidable, but it should be possible to advance an independent interpretation rooted in a principle of relational frame theory. The covariance of verbal phenomena and relational behavior is not persuasive by itself. Perhaps verbal behavior underlies our ability to respond to relations among stimuli, a point also made by Sidman (1994) and Spradlin (2003), among others.

In my review, I remarked that appropriate relational behavior seems to emerge from a variety of conditions, including ostensive learning, naming, and even simple coincidence (for example, hearing the name of a piece of music while listening to a passage of that music). Hayes and Barnes-Holmes are quick to point out that relational frame theorists have branched out far beyond the limits of the matching-to-sample procedure and have employed many of these less structured procedures. But if relational behavior emerges from unstructured procedures, the task of deriving a statement of principle becomes more formidable, not less so. What are we to make of the child who learns the term *mongoose* simply from hearing it in the presence of the animal? What is the behavior? What is the contingency? Are we to conclude that stimulus contiguity is, by itself, sufficient to induce a relevant response class? (See Tonneau & González, 2004, for some experimental analyses of the effects of stimulus-stimulus pairings on transfer of function.) If so, why do we not learn names when our model is speaking Xhosa, or Bengali, or some other unfamiliar tongue? The listener is not a passive vessel. Listening is behaving, and it appears to play a role in the acquisition of verbal behavior. (A subject who hears, but does not ''pay attention,'' learns little.) But what is that behavior, and what is its role? (See Donahoe & Palmer, 1994, chapter 9, for a discussion of some suggestive experimental work at the physiological level that bears on interpretations at the behavioral level.) Hayes and Barnes-Holmes do not address these questions except by dismissing mediationism as a mechanistic and worn-out creed. But human behavior is remarkably complex, and nothing is gained by denying its complexity.

MEDIATIONISM

I have pointed out that although the procedures used to study the acquisition of relational frames are complex episodes of behavior, only a few features of these episodes are recorded by experimenters. In typical procedures, the behavior of the subject on each trial is represented only by a single keystroke (or an equivalent punctate response). This is an appropriate strategy, if that is the only relevant behavior in a trial. But is it? Lowenkron (1988) showed that his subjects (mentally retarded children) were unable to demonstrate delayed identity matching until joint control had been established as a discriminative event. (Joint control occurs when two antecedents evoke responses of the same topography, as when a visual stimulus, for example, "9," coincides with an auditory stimulus, for example, the spoken word "nine." For discussion, see Lowenkron, 1991, 1998, 2004.) Specifically, the children could not perform a symbolic matching to sample task until they were taught to make, and hold, distinctive hand signs to the stimuli, as a kind of overt naming and rehearsal of the stimuli. The verbally competent adults and adolescents in the typical relational frame studies have had long histories discriminating joint control and have no need to resort to overt hand signs to name and rehearse stimuli. It is natural to call the hand signs of Lowenkron's subjects "mediating," but they are simply a part of the behavioral episode and have no less important a status than any other behavior. They are mediating only in the sense that the behavior of interest to the experimenter depends partly on the independent variable and partly on the stimulus products of this behavior of the subject. That is, the hand sign served as one of the controlling variables for subsequent behavior. If behavior is influenced by the stimulus products of other behavior, a complete account of performance will include that fact. This is not obeisance to a philosophical "mediationism" or "S-R associationism." It is a policy of describing behavior and its controlling variables as completely as possible. The exiguous details reported by those who study relational frames by no means describe the entire behavioral episode. The question is, are the omitted details relevant? Lowenkron showed, at least in his preparations, that they are. It is unclear how relational frame theory can accommodate his data.

Unfortunately, joint control merits only a few lines in Hayes et al. (2001a), and Lowenkron's account is dismissed as unparsimonious (Hayes et al., 2001b, p. 150). In my opinion, Lowenkron's account is both sound and elegant. I see no way of explaining complex human behavior without invoking joint control, and I think it plays an important role in many examples, if not all examples, of relational behavior. A complete account will illuminate that role, not deny it.

Hayes and Barnes-Holmes say that they too are interested in the web of overt and covert behaviors that comprise complex tasks, but they see relational frames as a prerequisite for such behavior. Relational frames, in their view, are functionally defined operants and therefore "are explanatory terms in behavior analysis: they do not depend on other hypothesized mediating processes" (Hayes & Barnes-Holmes, 2004, p. 220). This is fair enough, but the status of relational frames is not settled. If they are indeed elementary operants, as claimed, the unrecorded details of their procedures are irrelevant. But because of the conceptual difficulties with such a claim, outlined above, I am considering the contrary case. If relational behavior is derivative and heterogeneous, we are faced with the task of accounting, not simply for problem solving and other complex behavior, but for relational behavior itself. I suggest that a more fine-grained account of the behavior of subjects in relational tasks, including a consideration of joint control, will be necessary. Hayes and Barnes-Holmes marvel at the enduring appeal of "covert associationistic mediational analyses" (p. 220), but the pejorative tone is inappropriate. If understanding complex behavior requires an account of all of its components, then we must rise to the task insofar as possible. If covert behavior plays a role in some web of contingencies, then acknowledging that role is not a retreat to some primitive philosophy, but is a step forward.

CONCLUSION

Hayes and Barnes-Holmes are quite correct when they point out that no one has offered a complete alternative account of the phenomena embraced by relational frame theory. I have long acknowledged not only the importance of transfer of stimulus function, but also the difficulty of explaining it. Skinner described it as ''conditioning the behavior of the listener'' (1957, pp. 357–367) and noted that the effect does not appear in the naive speaker or listener. ''It is the end result of a long process of verbal conditioning'' (p. 360). But he offered no speculations about that process. Schlinger and Blakely have discussed the problem at length (Blakely & Schlinger, 1987; Schlinger, 1990; Schlinger & Blakely, 1987) and noted that it is central to rule-governed behavior, but they do not offer a full account of its origins. My own speculations on the topic resort to covert conditioning, but they have no empirical status. Sidman (2000) has proposed a principle that, if empirically confirmed, would account for transfer of function in equivalence paradigms, but it will not account for transformation of function in other types of relations. I accept, then, that there is no full alternative account of the full range of phenomena of interest. But I have not been entirely silent. I have offered tentative interpretations of memory (Palmer, 1991), cognition (2003), syntax (1998b), and together with Donahoe, a wide variety of cognitive phenomena (Donahoe & Palmer, 1994). Like Skinner's *Verbal Behavior,* the accounts are mostly interpretive, not experimental. If I am correct that understanding complex behavior requires a consideration of fine-grained behavior, much of which, owing to the limits of technology and to ethical constraints on our ability to control our subjects and their histories, is typically unobserved, then such interpretive exercises are appropriate. They permit us to offer tentative explanations for phenomena that would otherwise remain opaque (Palmer, 2003; Palmer et al., 2004). Some of our interpretations have appealed to physiological data and to neural network modeling. We have been encouraged by the ease with which both sources of support can be integrated with molecular behavioral processes. I accept that this may seem mechanistic, but in this context, surely that is a virtue. Nevertheless, I freely admit that not only are such interpretive exercises incomplete, they are no more than tentative proposals. If our field can overcome the technical and ethical hurdles of working with people and go beyond such interpretations, it should certainly do so.

On their side, Hayes and Barnes-Holmes and their colleagues have the considerable advantage of having a vigorous experimental program under way. In time it may sweep away all objections and make debate superfluous. But data are not enough. The conceptual foundations of the theory need to be clarified and squared with the traditional concepts of stimulus and response classes. A statement of principle is needed, a statement explicit enough that it can be applied to the kinds of examples I cited in my review. To me, they seemed to be cogent counterexamples, but it should not be a matter of opinion; the predictions of the theory should be clear.

Although this exchange of viewpoints may seem to have had an adversarial tone, none is intended. I yield to no one in my eagerness for a behavior analytic account of complex behavior. If Hayes and Barnes-Holmes and their colleagues are correct, then they have made an enormous stride forward in what is a common endeavor; if they are wrong, then they will have sharpened the search and challenged the field to do better.

REFERENCES

Blakely, E., & Schlinger, H. (1987). Function-altering contingency-specifying stimuli. *The Behavior Analyst, 10,* 183–187.

Burgos, J. E. (2003). Laudable goals, interesting experiments, unintelligible theorizing: A critical review of Steven C. Hayes, Dermot Barnes-Holmes, and Bryan Roche's (Eds.) *Relational Frame Theory* (New York: Kluwer Academic/Plenum, 2001). *Behavior and Philosophy, 31,* 19–45.

Chomsky, N. (1971). The case against B. F. Skinner. *The New York Review of Books, 17,* 18–24.

Donahoe, J. W., & Palmer, D. C. (1989). The interpretation of complex human behavior: Some reactions to

Parallel Distributed Processing. Journal of the Experimental Analysis of Behavior, 51, 399–416.

Donahoe, J. W., & Palmer, D. C. (1994). *Learning and complex behavior.* Boston: Allyn & Bacon.

Donahoe, J. W., Palmer, D. C., & Burgos, J. (1997). The S-R issue: Its status in behavior analysis and in Donahoe and Palmer's *Learning and Complex Behavior. Journal of the Experimental Analysis of Behavior, 67,* 193–273. (includes commentaries and reply)

Galizio, M. (2004). Relational frames: Where do they come from? A comment on Barnes-Holmes and Hayes (2003). *The Behavior Analyst, 27,* 107–112.

Hayes, S. C., & Barnes-Holmes, D. (2004). Relational operants: Processes and implications: A response to Palmer's review of *Relational Frame Theory. Journal of the Experimental Analysis of Behavior, 82,* 213–224.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001a). *Relational frame theory: A post-Skinnerian account of human language and cognition.* New York: Kluwer Academic/Plenum.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001b). Relational frame theory: A précis. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.). Relational frame theory: A post-Skinnerian account of human language and cognition (pp. 141–154). New York: Kluwer Academic/Plenum.

Hayes, S. C., Fox, E., Gifford, E. V., Wilson, K., Barnes-Holmes, D., & Healy, O. (2001). Derived relational responding as learned behavior. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame theory: A post-Skinnerian account of human language and cognition* (pp. 21–49). New York: Kluwer Academic/Plenum.

Lowenkron, B. (1988). Generalization of delayed identity matching in retarded children. *Journal of the Experimental Analysis of Behavior, 50,* 163–172.

Lowenkron, B. (1991). Joint control and the generation of selection-based verbal behavior. *The Analysis of Verbal Behavior, 9,* 121–126.

Lowenkron, B. (1998). Some logical functions of joint control. *Journal of the Experimental Analysis of Behavior, 69,* 327–354.

Lowenkron, B. (2004). Meaning: A verbal behavior account. *The Analysis of Verbal Behavior, 20,* 77–97.

Malott, R. W. (2003). Behavior analysis and linguistic productivity. *Analysis of Verbal Behavior, 19,* 11–18.

Malott, R. W., & Suarez, E. A. T. (2004). *Principles of behavior.* Englewood Cliffs, NJ: Prentice Hall.

Palmer, D. C. (1991). A behavioral interpretation of memory. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior* (pp. 261–279). Reno, NV: Context Press.

Palmer, D. C. (1997). Selectionist constraints on neural networks. In J. W. Donahoe & V. Packard Dorsel (Eds.), *Neural network models of cognition: Biobehavioral foundations* (pp. 263–292). Netherlands: Elsevier Science Press.

Palmer, D. C. (1998a). On Skinner's rejection of S-R psychology. *The Behavior Analyst, 21,* 93–96.

Palmer, D. C. (1998b). The speaker as listener: The interpretation of structural regularities in verbal behavior. *The Analysis of Verbal Behavior, 15,* 3–16.

Palmer, D. C. (2000a). Chomsky's nativism: A critical review. *The Analysis of Verbal Behavior, 17,* 39–50. (Original work published 1986)

Palmer, D. C. (2000b). Chomsky's nativism reconsidered. *The Analysis of Verbal Behavior, 17,* 51–56.

Palmer, D. C. (2003). Cognition. In K. A. Lattal & P. N. Chase (Eds.), *Behavior theory and philosophy* (pp. 167–185). New York: Kluwer Academic/Plenum.

Palmer, D. C. (2004). Data in search of a principle: A review of S. C. Hayes, D. Barnes-Holmes, and B. Roche (Eds.), *Relational Frame Theory: A Post-Skinnerian Account of Human Language and Cognition. Journal of the Experimental Analysis of Behavior, 81,* 189–204.

Palmer, D. C., & Donahoe, J. W. (1992). Essentialism and selectionism in cognitive science and behavior analysis. *American Psychologist, 47,* 1344–1358.

Palmer, D. C., Eshleman, J., Brandon, P., Layng, T. V. J., McDonough, C., Michael, J., et al. (2004). Dialogue on private events. *The Analysis of Verbal Behavior, 20,* 111–128.

Schlinger, H. (1990). A reply to behavior analysts writing about rules and rule-governed behavior. *The Analysis of Verbal Behavior, 8,* 77–82.

Schlinger, H., & Blakely, E. (1987). Function-altering effects of contingency-specifying stimuli. *The Behavior Analyst, 10,* 41–45.

Sidman, M. (1994). *Equivalence relations and behavior: A research story.* Boston: Authors' Cooperative.

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 74,* 127–146.

Skinner, B. F. (1938). *Behavior of organisms.* Englewood Cliffs, NJ: Prentice Hall.

Skinner, B. F. (1957). *Verbal behavior.* New York: Appleton-Century-Crofts.

Skinner, B. F. (1999a). The generic nature of the concepts of stimulus and response. In B. F. Skinner, *Cumulative record: Definitive edition* (pp. 504–524). Acton, MA: Copley Publishing Group. (Original work published 1935)

Skinner, B. F. (1999b). The analysis of behavior. In B. F. Skinner, *Cumulative record: Definitive edition* (pp. 101–107). Acton, MA: Copley Publishing Group. (Original work published 1953)

Spradlin, J. E. (2003). Alternative theories of the origin of derived stimulus relations. *The Analysis of Verbal Behavior, 19,* 3–6.

Tonneau, F. (2001). Equivalence relations: A critical analysis. *European Journal of Behavior Analysis, 2,* 1–128. (includes commentary)

Tonneau, F. (2002). Who can understand relational frame theory? A reply to Barnes-Holmes and Hayes. *European Journal of Behavior Analysis, 3,* 95–102.

Tonneau, F., & González, C. (2004). Function transfer in human operant experiments: The role of stimulus pairings. *Journal of the Experimental Analysis of Behavior, 81,* 239–255.

Vaughan, W., Jr. (1988). Formation of stimulus sets in pigeons. *Journal of the Experimental Psychology: Animal Behavior Processes, 14,* 36–42.