REPLY TO COMMENTARIES ON REMARKS COLUMNS

Murray Sidman

I was surprised and to a considerable extent pleased that Per Holth and Jay Moore, two behavior analysts for whom I have the greatest respect, thought that my six Remarks columns, published back in the late 1970s and early 1980s, were still sufficiently stimulating to be republished now, and that they had something more than just historical interest. I could therefore not resist the request that I prepare a seventh column to go with the earlier ones.

I was especially pleased because it has long been a strong conviction of mine that the significance of any scientific publication lies in the changes it brings about in the behavior of others. If a contribution does not change the behavior of others in one’s own field and elsewhere, then one might just as well never have published it. In this set of commentaries, a group of respected colleagues indicate that they have indeed learned from the Remarks columns—that is to say, that their behavior was changed by them. In addition, however, most of the commentaries suggest in one way or another that the original columns did not have as great an influence as they might have, that some behavior analysts and other potential readers have failed to alter their behavior in directions to which the various Remarks pointed. The reasons this has happened are many, but among the most relevant are my own failures to think several points through completely enough to convince all those whom I wanted to convince. I find many of the concerns expressed in the commentaries quite cogent; they will hopefully bring about changes in my own behavior, both here in my replies and in the future.

I will discuss the commentaries in the order in which I received them. First, however, I want to thank all of the participants for what I consider to be constructive, well thought out contributions. They constitute welcome departures from the more usual self-serving and often vitriolic criticism that I have noticed in many other commentary projects. With such a positive starting point I hope that, together, we might produce a larger effect than any of us would have produced alone.

Bryan Roche. I find the gentle humor of this commentary congenial and helpful. I know Bryan Roche by reputation, of course, but his commentary makes me wish I had come to know him more personally. First of all, I thank him for classifying me as a philosopher of science, even though I have usually considered myself an experimenter, leaving philosophical considerations to others. I had always thought of Tactics (Sidman, 1960/1988) simply as a description of the

AUTHOR’S NOTE: Please address correspondence to the author at:
murraysidman@comcast.net
practice of scientific experimentation. It is, after all, possible to appreciate philosophy without actually being a philosopher, but as I noted in my latest Remarks column (Sidman, 2010b) I have learned that the methodology and the philosophy of science are closely related. Roche, it seems, knew this long before I did. I now recognize that if I had known it earlier, I might have been able to make my contributions strike more deeply and more generally than they did. One does not change the behavior of thoughtful, rational, intellectually aware people without giving them good reasons for change.

Roche’s comments also added significantly to my own thoughts about the inferential nature of stimulus control. He defines the inference of stimulus control behaviorally as “A discrimination of the effectiveness of our own experimental manipulations across time.” This definition, by itself, represents a significant contribution to our understanding of a process—inference—that behaviorists had abandoned to cognitivists for analysis. Per Holth started my own rethinking about the role of inference in behavior analysis by asking me whether or not I would also consider reinforcement and extinction to be inferred processes (Holth, 2010, p. 193). I had to agree with that suggestion because we do indeed require multiple observations to identify and measure both reinforcement and extinction. We can, of course, recognize individual instances of reinforcement and extinction operations—the presentation or the withholding of behavioral consequences—but to identify any change in response probability that such operations produce requires several observations.

Roche takes an additional step by suggesting that “Stimuli and responses are, of course, themselves classes that must be discriminated across multiple observations. In other words, they too must be inferred.” The inferential nature of stimulus and response classes is a thought that had not occurred to me, but on considering it I find my own and Roche’s views beginning to diverge.

Even though stimulus control, reinforcement, and extinction are inferred processes, the identification and measurement by the observer of the particular stimuli and responses that make up a class does not necessarily require repeated observations. I am grateful to Roche for emphasizing that any identification of a particular stimulus or response is itself an instance of stimulus control of the observer’s behavior. He seems to be arguing, however, that the inferential nature of stimulus and response units is a consequence of the inferential nature of stimulus and response classes.

Although stimulus control is inferential, one must still be able to identify individual occurrences of the particular stimuli and responses that are related (see, for example, Iversen, 2010). That is to say, before one can determine whether a stimulus is a member of a particular class, one must already have identified the stimulus. At this point in his commentary, it seemed to me that Roche slipped from one definition of inferred processes to another. His main reason now for considering stimuli and responses as inferred units seems to be that they form classes which themselves require multiple observations to be discriminated, rather than that they require several measurements. The identification of any stimulus or response class surely does require several observations, but the inclusion of any
particular stimulus in a class requires only a single observation or measurement of a controlling stimulus that has already been identified.

Thus, it seems to me that Roche now defines stimuli and responses as inferred units because they form classes that require multiple observations to be identified, not because they require several observations before we can identify them. He goes on to use this expanded basis for classifying stimuli and responses as inferred units as a justification for freeing stimulus-control relations from the “constraints of observable discrete activities,” thereby expanding the operant “spatio-temporally to infinity.” He extends the meaning of “inference” from a statement about the stimulus control of the observer’s behavior to the view that “the operant and its consistent parts have been freed from the constraints of observable discrete activities.” This extension appears to constitute his justification for the treatment of derived relations as operant units.

Can it be that Roche is arguing that because he holds stimuli and responses to be inferential concepts—because they belong to inferred classes—then the operants that related stimuli and responses define need not be directly observable? He seems to be concluding that because stimuli and responses are always members of inferred classes, then the operants that related stimuli and responses define need not be directly observable, thereby helping to justify Relational Frame Theory’s postulation that derived relations are to be included as examples of learned operants (e.g., Hayes, Barnes-Holmes, & Roche, 2001).

Although we can reasonably grant that “inferential” implies “unobserved,” an important question is whether inferential implies “unobservable.” Once one has performed a sufficient number of measurements, an original inference becomes either confirmed or disproved. If it is to be at all useful, an inference must eventually prove correct or false. I am sure that Roche believes this just as strongly as I do. Somewhere, though, we have encountered a disconnect. I do not really follow how my own view of stimulus control as an inference leads to a justification for Relational Frame Theory.

In another place, too, I think Roche gives me credit where none is due. I have a problem with his interpretation of my views about the learning process. I have proposed that the discovery of errorless learning demonstrates that theories of learning should stress the behavior not of learners but of teachers (e.g., Sidman, 2010a). I find it difficult to conclude from this, as Roche does, that “behavioral units are not to be bound by the observable activity of our subjects.” I was simply proposing that the study of learning should concentrate on the behavior of teachers rather than of learners. This is a statement not about the location of behavior but simply about the location of determining variables.

It is, of course, always possible that one’s thoughts may possess greater relevance than was originally intended. When that happens, one should be grateful to those who attribute greater significance to one’s work. Perhaps conventional views of the nature of scientific activity continue to constrain my own thinking here. Roche has certainly contributed more than enough to our field to give me pause when I find myself in disagreement with his position. Perhaps a more
sophisticated philosophic orientation would relieve my apprehensions and allow me to reorient my own views in unconventional directions.

William H. Ahearn. This commentary deals more with the significance of my Remarks columns for the motivation and personal conduct of behavior analytic practitioners than with the logical and systematic implications of the columns. I have known Bill Ahearn for many years, but I would never have guessed that he himself ever had to wrestle with the kinds of decisions he outlines in his commentary, decisions not just about how to influence his clients’ behavior but about how to control his own behavior. His admirably noncoercive conduct in his relations with clients, students, and professional and nonprofessional colleagues gave no indication that he was ever seriously tempted to behave differently. I suspect his behavior is attributable more to his own working through the implications of aversive control than to the persuasiveness of my writings. In any event, he does bring up several matters that I believe are worth more extensive development.

For example, it is indeed true that behavior analysis is popularly associated with aversive control. It is a common belief that both its theory and its therapeutic techniques are based on the “carrot and stick” analogy. Even though standard behavior analytic practice strongly discourages the use of the stick, coercive techniques are so widely practiced in most communities that the general public—along with some behavior analysts—finds it impossible to believe that consistent positive reinforcement is even possible. To accept the proposition that positive reinforcement of desirable behavior can replace the punishment of undesirable behavior as a means of behavioral control requires an intellectual about-face that ranks in significance with historically major revolutions in thought. From child-rearing practices to international diplomacy, the replacement of punishment and threats by a search for reinforceable behavior does not rank even as a possibility in standard lists of acceptable options. Positive reinforcement may be recognized as effective in limited circumstances, but it is rarely accepted as a way of life. One can only hope that positive reinforcement will begin to replace coercion at all levels of human interaction before we have blown each other into oblivion.

Although the side effects of punishment—escape, avoidance, anxiety, behavioral rigidity, countercontrol, etc.—have been extensively investigated, the need for research continues. How about the possibility of using mild punishment to stop behavior temporarily and at the same time reinforcing appropriate alternate behavior? How successful would this strategy be? As Ahearn points out, this way of using punishment in emergency situations remains a still-to-be-investigated possibility. A commendable public reaction against coercive control in education is, however, rarely accompanied by public support of effective alternatives to coercion. When one encounters extreme delinquency that seems unstoppable except by punishment, might we instead engineer environmental changes that will generate acceptable behavior? Do we know how to do that? The unthinking public reaction of distaste for coercion threatens to spill over into a ban on further research. Preventing an increased understanding even of harmful natural processes is to nobody’s advantage.
Such fundamental changes in our way of thinking about our own behavior are responsible for the tremendous excitement many behavior analysts feel about the significance of their work. As another example, many of us are immensely stimulated by the proposition that the behavioral science we have been developing can be included among the natural sciences. This revolution in thought has not yet received general acceptance. How many scientists, philosophers of science, or just intellectually curious lay people would classify the study of behavior as a natural science? A general acceptance of this proposition will require a major philosophic and practical reorientation of people’s views of where human behavior comes from. Does our behavior have an internal controller like a self or a mind or is it controlled by directly observable and measurable changes in genetic and environmental variables?

Another example of an intellectual revolution inspired by the science of behavior analysis is the discovery that learning does not require trial and error—one can deliberately generate errorless learning. Again, this conceptual revolution has not yet been generally accepted, even where it is most immediately relevant. Most professional educators, members of governing bodies, and other individuals who are concerned with educational effectiveness will shake their heads in disbelief if they just hear the term “errorless learning.” In spite of its empirically demonstrated reality, errorless learning is a concept whose general acceptance appears to require more than just data. This is because techniques that produce errorless learning shift the study of learning from the behavior of learners to the behavior of teachers. A general acceptance of this shift in emphasis will require a major reorientation in prevailing educational practice and philosophy. It is also relevant to the question of where behavior comes from; learning is produced by the behavior of teachers, not by supposed controllers situated within learners.

None of these notions—the effectiveness of positive reinforcement, the study of behavior as a natural science, the possibility of errorless learning—are just beliefs. Copious experimental and practical evidence supports them. Why they have not been generally accepted is a problem that behavior analysis itself has failed to confront. The processes that are involved in the public acceptance and rejection of scientific findings and principles need themselves to be investigated. This need, of course, exists much more widely than just in behavior analysis. Witness the public reaction to data-based theories of evolution, global warming, origins of the universe—even of life itself.

*Kathryn Saunders*. In her characteristically well-reasoned, concise, yet thorough commentary, Kate Saunders describes what she used to think of as the rookie stimulus-control error more clearly than I was ever able to do. She also points out that this error continues even among veteran researchers. Perhaps I can add my own personal conviction that the identification of the actual controlling stimulus (what McIlvane & Dube, 2003, have called stimulus control topography coherence) is the most pressing problem in the study of how stimulus control is developed and maintained in both experimental and applied situations. That is to say, how closely does the stimulus control of the subject’s or student’s behavior correspond to the stimulus control of the experimenter’s or teacher’s behavior?
For example, if we are teaching someone to read what to the teacher are words, is the student responding, like the teacher, to whole words or instead, to the number of letters in the words, to just the first or last letters of the words, or to some other feature of the stimuli? In teaching a pigeon to distinguish horizontal from vertical lines, is the pigeon responding, like the experimenter, to the line orientations or rather, to the amount of space between the lines and the top (or bottom or sides) of the display? Very often, too, we find that a student who has learned to respond perfectly accurately to particular stimuli no longer responds accurately when tested in another room, or a student or experimental subject who has learned to discriminate stimuli on a computer screen no longer responds correctly when those same (to the teacher or experimenter) stimuli appear on the table top. Although such a finding is often called a failure of generalization, it really means just that the original controlling stimuli included not just the material being presented but also some other aspect(s) of the original teaching environment.

A different kind of generalization, usually called primary stimulus generalization, is based on a quantitatively specifiable degree of resemblance between the original and the current controlling stimulus. A characteristic primary generalization gradient shows a particular sample of behavior declining in probability as the original controlling stimulus varies (for example, as a light is made less or more bright, a form is made larger or smaller, a sound is tuned to a higher or lower pitch, etc.) It turns out, however, that the independent variable in the famous generalization curve is not the amount of variation between the originally reinforced and the tested stimulus. Rather, the independent variable reflects the probability of control by nonspecified stimuli as the original stimulus varies. If one is alive to the likelihood of such variation, one can often identify the competing sources of control in a particular context and thereby specify the actual controlling stimuli in a mistakenly labeled “generalization” curve (e.g., Sidman, 1969). Unfortunately, the beautiful quantitative regularity of the so-called generalization curve has in many instances led researchers (at one time, including me) to invent a nonexistent process called primary stimulus generalization and to ignore the need to develop techniques for identifying actual controlling stimuli.

Kate Saunders emphasizes a particular kind of instance in which a teacher mistakenly assumes that the stimuli being presented are actually controlling a pupil’s behavior. This occurs at or near the start of a teaching program when the teacher presents the same stimuli on every trial and the pupil comes always to select the “correct” one. A common name for such procedures is single-stimulus training. For example, on every trial the teacher says “dog” and the child receives candy by selecting the picture of a dog and not selecting the picture of a cat. Saunders points out that the child here never actually has to learn the relation between the spoken word and the picture. In fact, the child can ignore the spoken word and always just select the same picture. As Saunders points out quite clearly, the controlling stimuli here are not the same for teacher and pupil.

Saunders, however, charitably credits this procedure with the virtue of at least teaching the child to discriminate the two pictures. I find myself hard heartedly disagreeing with this generous assessment of single-stimulus training. The child
can always select the correct picture while learning nothing about the incorrect one. Or, the child can always reject the incorrect picture without learning anything about the correct one. Kate, of course, knows this and I suspect that she was just trying to avoid additional complexity in her commentary.

Although, as Kate Saunders points out, stimulus control is basic to instructional programming, its analysis is barely touched on in the training of behavior analytic practitioners. One unfortunate consequence of this neglect within our own field is the failure to adopt behavior analytic principles and techniques in our educational system. Most schools of education are unaware of and therefore ignore the relevance of stimulus-control analysis to the technology of teaching. Indeed, technology plays a small or nonexistent role in the training of teachers. Because standard teaching practices are so inefficient, teacher training emphasizes what is to be taught rather than how to teach it. As long as behavior analytic practitioners themselves are kept largely in the dark about the analysis of stimulus control, there is little hope of extending that analysis to educational practice.

William J. McIlvane. I find myself amused, touched, and stimulated by learning of Bill McIlvane’s concern, when he was a graduate student, that I was wasting my time writing in my Remarks columns about things that Bill thought all behavior analysts must surely have known already. One direction in which his revelation stimulated me was to ask myself, “How many of us have the same view of ourselves that other people have of us?” It is quite common for people to claim that their real selves are inside, deep in their feelings and thoughts, and are therefore unknown to anyone else. Although answers to questions about our feelings, about our central nervous system, and about other internal systems can indeed help us understand how our behavior meshes with all the other natural processes that we call life, our feelings and our brains are not independent organisms that direct our behavior from within. Variables in the external environment direct our feelings, the activity of our brains, and our visible behavior. A basic philosophic tenet of Radical Behaviorism is that we are what we do. In my Remarks columns, therefore, Bill McIlvane was seeing the real me. I am touched that he wanted me to do—and thought me capable of doing—better things with my time. I am amused by the implication that he must therefore have wanted me to be someone else. When one’s philosophic position clashes with the customary thinking of one’s culture, it is often difficult to adhere to the philosophy rather than to the mores one has been trained to follow.

I think Bill McIlvane theorizes correctly about why I was concerned with many of the matters I wrote about in the Remarks columns. A theme that can be found in almost everything I have written, in the Remarks and elsewhere, is, as Bill states, “. . .the criteria by which we evaluate whether a given problem has been posed and/or answered adequately.” The key word here is “evaluate.” It is likely that few readers recall that my book, Tactics of Scientific Research, has a subtitle, Evaluating Experimental Data in Psychology (Sidman, 1960/1988). That subtitle was what I thought the book was about and I originally proposed it as the full title. The editor with whom I worked felt that my suggested title was a bit too buttoned down and that she would prefer something that would be more likely to stimulate.
prospective readers. With the help of Larry Stoddard, we came up with Tactics. ... I was willing to accept this suggested title because I realized that the criteria by which we evaluate data require examination of the procedures that were used to obtain the data. Those procedures are the tactics for carrying out research. Still, however, my desire to be honest and clear about the book’s contents made me insist on retaining my original title as a subtitle.

Unfortunately, in evaluating data, researchers seem more and more to be subordinating examination of the investigative procedures to examination of the data’s consonance or dissonance with some theoretical formulation. If this statement strikes Bill McIlvane as “edgy,” he would be correct. I believe that there are good data and there are bad data, but their classification as good or bad has nothing to do with how well they support or refute some theory. Theories must indeed be supported by data, but if bad data are used to support a theory, then as far as I am concerned that is also a bad theory.

B. F. Skinner, for example, was not only a brilliant theoretician but was also a master experimenter. His contributions to science—and more generally, to culture—began in the laboratory. The procedures by means of which he gathered data were innovative, clean, reliably replicable, and generally applicable both within and outside the laboratory. Although some find his theories philosophically distasteful, the data that gave rise to the theories are impeccable. If his radical behaviorism fails to survive, the failure will not come from any faults revealed by the evaluation of his experimental data. Although Skinner’s first lengthy report, known informally as the B of O (Skinner, 1938), taught me how to produce and to evaluate behavioral data, few of today’s students read it. That neglect shows up as the methodology described in behavioral journal publications is coming to resemble that found in nonbehavioral publications. The evaluation of data in their own right is becoming less important than the evaluation of the data’s relevance to particular theories. Although I do not, like some teachers, insist that my students review everything I had to go through, I do believe that reading the B of O will provide even today’s students with a view of science they will rarely find elsewhere.

Finally, Bill McIlvane notes that behavior analysts tend to avoid talking to each other informally. He suggests that both students and teachers would benefit from more opportunities to do so, as in the Remarks columns and the commentaries being presented here. Although I am pleased that Bill values those columns as a useful channel of informal communication, his comments made me recall that informal exchanges of information were at one time much more usual than they are now. Such exchanges, however, were not carried out via generally available publications. How was it done, why has it changed, and how might it be resumed?

During the starting years of behavior analysis—then called operant conditioning—the few of us who were involved at the time felt that we were pioneering meaningful developments in an area that was potentially of great importance. Our innovative fervor, the excitement we shared as we kept producing and observing behavioral phenomena that nobody had ever seen before, and our
REPLY TO COMMENTARIES ON REMARKS COLUMNS

recognition that if what we were doing was to achieve more general acceptance, we had to make sure we were doing it right, all caused us not to compete but to cooperate. We not only cooperated but we formed friendships. We visited back and forth, greeting each other at airports, going out to meals together, confiding in each other about personal experiences, and sharing our hopes, fears, and dreams. The glue that held us together was our data. We were always eager to show each other our latest observations, to describe the methods we had used to make those observations, and to question each other about any details that had not been described. Moreover, we questioned each other mercilessly, always searching for overlooked factors that might have been responsible for the data. We knew that we were each other’s best critics and we accepted each other’s critical analyses as acts of friendship.

There are now so many behavior analytic researchers and practitioners that such intimate and informal communication is no longer possible on a large scale. We communicate via formal instruments—conventions and journals—arranged by professional organizations. Such communication is rarely two-way. Contrary to our basic principle that teaching is most effective when prospective learners participate actively rather than just reading or listening, we usually just hand out information that readers and listeners are supposed to absorb. Commentary projects like the present one are attempts to solve that problem, but they are rare and their format is limited. Poster sessions at conventions are supposed to help counter the usual large-scale distribution of information, but even those sessions are now so crowded and brief that little two-way communication usually takes place. Poster presentations are also touted not as methods for receiving friendly criticism but as opportunities to show off one’s accomplishments.

So then, how to generate more opportunities for behavior analysts to talk less formally to each other? How about invitations to productive researchers and thinkers from elsewhere to visit and make a presentation to local students and faculty? I have found that such visits usually accomplish little more than one-way communication. One delivers a lecture, perhaps joins a few faculty and students at dinner, and then leaves.

On the other hand, I have experienced a few occasions in which my hosts organized my visit in such a way as to produce productive and enjoyable two-way interactions. Although I usually gave a formal lecture—sometimes more than one—the visits consisted largely of scheduled meetings with individual students or professors during which they acquainted me with the rationale for some particular research and showed me relevant data. I was able to respond with questions, evaluative comments, and accounts of my own pertinent experiences. In those interactions speakers were also listeners, and listeners were also speakers. The communication was productive, informal, and friendly, even though it might have included expressions of doubt about some particular aspect of the research or its interpretation.

The success I have experienced in such visits makes me wonder if it might be better not just to invite a particular person to visit but to invite that person to bring along all the students in his/her laboratory. Perhaps instead of sending one’s own
students to a big convention, might more effective communication take place if students from elsewhere came and interacted informally? This would, of course, involve a change in the allocation of financial resources from paying the convention travel expenses of our own students to paying the expenses for students from elsewhere to visit us. I suspect that such explicit arrangements for students to talk to each other would justify the shift in expenditures.

I have found another kind of visiting experience to yield not only useful two-way communication but to result in new friendships as well. In those visits, I was scheduled not to interact with individuals but to join and take part in laboratory meetings in various research areas. Even in areas with which I was unacquainted, I was able to teach more by asking questions than by making a presentation. I was able to get students to think about problems they had not recognized, or even just to get them to describe their work in terms understandable to an outsider like myself. In addition, I learned much about areas with which I had been either completely unfamiliar or in which I had not kept up with recent developments. And then, at meals and parties, I was able to share with students experiences, thoughts about behavior analysis in general, and possible roles an individual might play in future developments.

The fields of experimental and applied behavior analysis have grown large. Such informal methods of talking to each other, productive though they may be in limited environments, could not possibly replace the one-way forms of communication that have become traditional in conventions, journals, and in most college and university courses. Might it be possible to establish a leading institution based on instructional techniques that encourage students and teachers to talk to each other? How about setting up an Institute of Behavioral Technology? Students there would not attend lectures or seminars in which they report on assigned topics but would instead gather informally to exchange information, questions, and thoughts about their own and others’ behavioral research and behavioral engineering. Would such an Institute produce a corps of behavior analysts who talked to each other and enjoyed doing so? Would such communication produce a new group of leaders who would take behavior analysis in directions that none of us would have thought of if we continued the limited kinds of communication that are now customary? I do believe such a dream is worth pursuing.

Well, so much for some of the thoughts Bill McIlvane’s commentary generated. I can only hope those thoughts might prove nearly as provocative and productive as Bill’s own way of establishing communication among behavioral researchers in the United States and elsewhere, particularly in Brazil—its own a story that needs telling.

Iver H. Iversen. I know from personal experience that whenever Iver Iversen expresses doubt or confusion about some methodological issue, then a closer examination of that issue results in greater clarity. Here, he expresses his perplexity about a matter that probably few others have considered; his commentary provides a clear exposition of his reasons for finding it difficult to provide a general definition of stimulus control.
The first difficulty is one that Iver understands perhaps better than anyone else. He is well acquainted with the problem of stimulus definition when we want to or have to know which aspect of what he calls a *complex stimulus assembly* is the controlling stimulus for some particular behavior. Because such control cannot be identified in a single observation, I have called stimulus control inferential; until we have made several observations, we cannot specify which aspect of a complex event is a controlling stimulus. For example, when we observe a child matching colors, we do not know until we have run special tests whether the particular property being matched is hue, brightness, or saturation. When a child is matching shapes, we need additional tests if we want to know whether the critical stimulus property for the child is shape, size, color, or details of form like curved and straight lines.

On the other hand, Iversen also recognizes that we often do accept a single observation as a demonstration of stimulus control. Under what circumstances might it be sufficient to identify stimulus control in a single observation? Iversen makes the excellent point that the three-term contingency “has to be an actionable unit that can be identified, counted, and communicated each time it occurs.” For example, in teaching a child a discrimination between green and red, we have to be able to recognize and record each occasion on which the child selects green and not red. At this point, we do not have to ask which aspect of the green stimulus is controlling the child’s behavior. A single observation is all we require in order to say that the child has selected what we call the green stimulus and for us to deliver a reinforcement. When we ask someone to pass the salt and the person does so, we do not have to know whether that person’s response was controlled by the shape of the salt shaker, by the color of its contents, or by the size of the holes in its top; we just say “Thank you” without having to ask several more times.

Do we need separate terms to refer to these two types of instances of stimulus control, one of which requires several observations and the other only one? I am not sure we do. The same basic question applies to both: “Does the same stimulus control both the pupil’s or experimental subject’s and the observer’s responses?” (see McIlvane & Dube, 2003, for a more sophisticated development of this question). Sometimes we are satisfied to answer “Yes” or “No” without specifying which aspect of what Iver calls an *entire stimulus assembly* controls the behavior we have recorded. A single observation tells us all we need to know. We do not have to fine-tune our own discriminations.

At other times, however, we may want to or have to identify the actual controlling stimuli. In a teaching situation, for example, having taught the learner a particular discrimination, the teacher may then want to use that discrimination as the basis for teaching or testing something else. If, however, teacher and learner have not been responding to the same stimulus aspect in the original teaching situation, the additional teaching or testing will probably be a failure. For the learner, the new situation may not involve the same stimuli that the teacher thought, erroneously, had been in control.

We are not, then, dealing with two types of stimulus control over the observed person’s responses but only with two types of stimulus control over the observer’s
responses. I am not even sure we need to classify these as different types of stimulus control. Is control of the observer’s behavior by an entire stimulus assembly different in principle from control by a particular aspect of the assembly? It does not seem necessary to me to have different names for both kinds of instances. We only have to ask whether we need or want to know more precisely whether or not the stimuli controlling one person’s behavior are the same as those controlling somebody else’s behavior.

I think it is also worth pointing out that if it is to be at all useful, an inference about stimulus control must eventually prove correct or false. As I said in my reply to Roche (this volume), once one has performed a sufficient number of measurements, an original inference becomes either confirmed or disproved. Once we have made enough observations to identify the controlling stimulus, its definition is no longer inferential. From then on, we need only a single observation to identify an instance of stimulus control even by a particular aspect of a complex stimulus.

This reasoning, of course, does not apply if there is any suspicion that the controlling stimulus may change. If it does, a single observation is likely to be deceptive. To deal with such a situation it would be necessary to institute procedures either for preventing change or for detecting any changes that might still occur, or for both. Such procedures are not peculiar to behavior analysis. They must form part of the behavioral repertoire of any scientist, basic or applied. Iver Iversen does not need any advice from me about the prevention and evaluation of stimulus control variations during the course of an experiment or an application. His own work may be taken as a model.

Javier Virués-Ortega. Virués-Ortega expresses valid concerns about experimental and clinical methodology that are crucial for other kinds of scientists’ and the general public’s appreciation of behavior analytic data and conclusions. In particular, he questions behavior analysts’ “inflexible adherence” to what have become standard experimental and treatment designs. I cannot discuss several of the nonstandard design options he offers because I am not familiar with the relevant literature—the field has grown too large for anyone to know it all—but his general concerns are certainly matters that need consideration.

I join Virués-Ortega in his unease about what he sees as a narrowing of research methodology in spite of a broadening of the problem areas into which behavior analysis has moved. When I wrote Tactics I was worried that this kind of methodological restriction might take place. In the Preface I wrote:

. . . I must caution the student not to expect a set of rules of experimental procedure, to be memorized in classic textbook fashion. . . . Even those who find their activities most accurately described here would feel uncomfortably restricted if they had to proceed solely as I have outlined. Neither the practice of experimentation nor the evaluation of its products can be bounded by any specific rules—a qualification that lends a certain note of irony to any book on experimental methodology. (Sidman, 1960/1988, pp. vi-vii)
Virués-Ortega points out that we might generate a more general acceptance of behavior analysis by explicitly explaining in our reports how our investigative techniques are particularly appropriate to the problems being investigated. He seems to be suggesting here that our journals, for example, require us to write not just for our fellow behavior analysts but also for scientists in other areas, and that in our presentations we not only describe our investigative techniques but also explain how those techniques, although seemingly inconsistent with conventional research standards, do in fact meet those standards. For example, how do the features of multiple baseline designs allow us—and everyone—to evaluate data from individual subjects more precisely than a statistical study that compares an experimental to a control group? Virués-Ortega also points out that such explicit justifications for an experimental design would require the behavior-analyst experimenter or practitioner to understand his/her own techniques and not just use them because they are described in a textbook. A greater understanding would expose the weaknesses of the techniques and perhaps generate new methods that are more generally applicable to problems that behavior analysis has not yet attempted to solve.

What particular areas of research and practice does Virués-Ortega feel have suffered from an unquestioning dependence on standard behavior analytic techniques? He cites verbal behavior and complex human behavior in typically developed individuals as problem areas in which our standard approaches—like individual subject methodology and an emphasis on the frequency of behavior—have so far restricted rather than promoted generally acceptable data. I do, of course, agree that experimental methodology has to be adjusted to the particular problem being investigated but I am a bit concerned that Virués-Ortega may be confusing investigative methodology with broader scientific and philosophical considerations.

For example, he is worried that our emphasis on behavioral frequency is keeping us from studying low frequency or highly variable behavior in the natural environment. I do not believe, however, that our concern with behavioral frequency stems originally from methodological considerations. It is just the other way around; measures of behavioral frequency are the outcome of our conviction that the probability of behavior is its most important characteristic. What we most want to know about any behavior is the likelihood that it will occur. Coming from this angle, then, we would not attack the analysis and treatment of low-frequency behavior by looking at some other datum than frequency, but rather by adapting our frequency measurements to the particular analytic problems posed by low frequencies.

How might that be done? One approach to this problem has been the use of interval rather than continuous recording. Although discontinuous measurement procedures are sometimes revealing, they also possess a number of serious deficiencies (for a general review, see Johnston & Pennypacker, 2009, pp. 120-124). Touchette, MacDonald, and Langer (1985), however, reported a potentially fruitful method for using interval recording to help identify the controlling stimuli for particular behavior. Dividing the day into, say, half-hour segments, they then
made scatter diagrams of the occurrences of the behavior of interest throughout successive days. Although a scatter plot based on interval recording may provide an inaccurate specification of response frequencies, it can indicate a pattern of high and low frequencies that helps to identify the setting events for behavior of highly varying frequencies.

More generally, the investigation and therapy of infrequent behavior may profit from a more deliberate application of what we already know about the stimulus control of behavior. As far as I am aware, most applied behavior analysts deal almost exclusively with reinforcement variables, with little attention being paid to matters of stimulus control. It seems to me that much could be accomplished by placing even undesired behavior under stimulus control so that it could be made reliably available for therapeutic measures.

Infrequent or variable behavior, instead of requiring a redefinition of behavior analysis into something other than a concern with behavioral frequency, might successfully be studied and changed by the application of stimulus control techniques.

Finally, behavioral dimensions like latency, force, and topography may indeed be important, but even here frequencies are still basic. Under various conditions, we want to know how often and under what conditions particular latencies, forces, or topographies occur. Whatever behavioral dimensions interest us, we want to know their frequencies.

I am sure that Virués-Ortega understands all this. I know he is not suggesting that behavior analysis abandon its current methodologies, or that it give up our standard treatment of intra-group differences as problems for analysis, or that it drop its concern with the probability of behavior. Much of my reply to his commentary stems from a fear that his strongly expressed desire to see new treatment and therapeutic techniques applied to new problem areas will leave the impression with some readers that he would like to abandon our basic philosophical foundations. I commend his impatience with behavior analysis’s lack of acceptance by other sciences and the general public, and I encourage him to continue to stimulate appropriate changes in the behavior of behavior analysts.

Steven C. Hayes. Almost invariably, I find Dr. Hayes’ scientific contributions worth considering carefully. Although I do not find all elements of his theorizing to be congenial to my own views, I can pay him no greater compliment than to concede that he nearly always requires me to evaluate my own convictions about science and behavior analysis carefully. Such self-evaluation has led me to believe that he has succeeded in advancing behavior analysis into fields that post-Skinnerian developments had left untouched.

For example, in his commentary on some of my Remarks columns, he notes my sensitivity to the need for behavior analysis to take account of scientists’ behavior, particularly their own. My concern was provoked by observations that failures to understand our own behavior sometimes keep us from being able to exert a beneficial influence on the behavior of our experimental subjects and our clients. A different path seems to have led Dr. Hayes to the same conclusion. I am struck by his observation that “Skinner takes the radical step of considering
scientific concepts and data in terms of the contingencies controlling actions of
scientists when they observe and analyze. In so doing, artificial prohibitions
against introspection are overthrown and the door to the analysis of private
events is forever opened inside behavioral thinking.”

It is true that Skinner’s behaviorism is called radical because, unlike earlier
versions of behaviorism, it permits the consideration of private processes and
events like thinking and feeling. This extension of the philosophy of behaviorism,
however, has only rarely been matched by corresponding extensions into research
and application. Few data have been gathered to support either the importance or
the utility of Skinner’s enlargement of behaviorism’s philosophical scope. This, I
think, is where Dr. Hayes and his colleagues have contributed mightily. They have
enlarged the field of behavior analysis to attack phenomena of language in both
thought and action and have even provided both philosophical and practical
justification for successful “talk therapy” (e.g., Hayes, Strosahl, & Wilson, 2003).
Overall, they have made huge strides in addressing problems that many previously
considered out of bounds for behavior analysis and the exclusive province of
cognitive psychology.

I am pleased, too, that some of my own work on equivalence relations (e.g.,
Sidman, 2000) provides a basis for demonstrating that even introspective
phenomena—including scientific thought processes—are amenable to a behavioral
analysis. As scientists, we think—like everyone—in words, pictures, sounds,
colors, and so on; by drawing and reading maps, we plan and make trips through
regions we have never seen; using diagrams, we design new buildings and other
physical structures; by writing and viewing numerical symbols, we deal with
quantities of money and of other physical objects; without ever seeing or feeling a
stock certificate, we buy and sell stocks; and so on. All of this thinking, planning,
designing, and communication involves the establishment or the previous
existence of equivalence relations between what we think and the things we think
about. We have to view even our own thoughts as private activities concerned with
otherwise observable objects and processes. By showing how equivalence relations
are established and how contextual factors determine whether classes with
common elements merge or just intersect, we show how introspections are
formulated and how they determine scientists’ observations and analyses of the
external environment. The study of equivalence relations opens a pathway to the
inclusion of the scientist’s own behavior in his/her interpretation of environmental
processes.

Dr. Hayes decries our tendency to attach greater value to the social reinforcers
provided by our intellectual community than to our interactions with our subject
material. He strongly recommends, “We need to begin to treat ourselves and our
own analytic practices as part of the behavioral systems we study.” I applaud the
goal he sets for us here but a cautionary note is needed. I have stated before, only
partly in jest, that one of the first obligations of a behavior analyst is to obey the
laws of behavior. But which laws are to predominate? After all, we do live within a
social community that sets up powerful reinforcement contingencies. However,
only a few of them are directly related to the products of our intellectual
endeavors. How does one come to ignore those immediate social and tangible reinforcers in favor of the delayed and often intangible reinforcers that our analytic practices generate (Sidman, 2007)? We are really asking very much of ourselves. Like our experimental subjects and clients, we too are creatures of our reinforcement histories. It should not come as a surprise to behavior analysts who observe themselves that they sometimes pay more attention to reinforcers that advance personal and socially approved goals than to reinforcers that advance their science.

Telling people what they should do is not the same as getting them to do it. We must admit, however, that the creation of new reinforcers has not been a major concern either of experimenters or of clinicians. It is an area that needs research. Perhaps such research would be more likely to lead to greater influence by our subject material than by any social consequences of our professional activities. Instead of relying on exhortation, a little constructive humility might create a context that is more likely to get others—and ourselves—to change their reinforcement preferences.

Erik Arntzen. I enjoyed Erik Arntzen’s story about how and why he originally became interested in equivalence relations. These days, in taking up equivalence relations with our students, we rarely convey the feelings of excitement that the basic equivalence phenomena create in many of us when we first read about them or when we actually see them happen. More generally, when we present experimental findings in any area to our students—or to whomever our listeners or readers might be—we usually fail to convey the emotional stimulation that goes along with intellectual enlightenment (Sidman, 2007). Instead of starting with basic research questions, if we began by describing the kinds of reinforcers that got us interested in the first place, I think we would be more likely to create students’ and even the public’s interest in finding out more.

In reading his commentary, I am a bit surprised to find Erik Arntzen asking me questions about equivalence relations. He knows as much as I do about equivalence and about how to pursue its as-yet unanswered questions. Perhaps he is asking me because he knows how much problems that require creative experimental designs for their solution fascinate me. I am sure, however, that he knows perfectly well how to find out whether or not a critical variable in testing for equivalence is the relative number of times one requires the stimuli to be discriminated simultaneously vs. consecutively. One way to find out, of course, would be to keep what he calls the training structure constant while varying the relative number of times each stimulus is involved in simultaneous vs. successive discriminations. Perhaps this could be accomplished more cleanly by means of pretraining in which the stimuli were used in simple discriminations—both successive and simultaneous—before becoming involved in conditional discriminations and equivalence testing.

For me, working on equivalence relations has revealed many design problems that have more general relevance than just to a particular experimental question. For example, instead of asking whether a particular variable is responsible for
REPLY TO COMMENTARIES ON REMARKS COLUMNS

differences in the number of test trials it takes for equivalence relations to emerge, suppose we look at gradual emergence itself.

It seems clear to me that if equivalence relations emerge only over many trials, even trials for which the experimenter schedules no reinforcement, then learning must be taking place during the test trials. We have an instance of gradual or delayed learning. As I have proposed elsewhere (Sidman, 2010), gradual learning is an artifact; it characterizes not the learning process but the teaching process. If we first teach pupils or experimental subjects everything they have to know in order to perform a task, then learning takes place immediately, that is to say, without errors. From this more general point of view, we must then ask what we have failed to teach before we test for the performance in which we are interested—in this instance, the emergence of equivalence relations.

One of the things our pupils may have needed to learn before being tested for equivalence relations is to discriminate the stimuli from each other both when they are presented simultaneously and when they are presented successively. The critical features of the training procedures that Arntzen describes do not lie in their “structures” but rather in how they expose pupils to these two kinds of discriminations. Other possible factors, too, may help prevent pupils from showing equivalence relations immediately, when first tested. For example, in teaching the baseline conditional discriminations, we may have failed to teach the pupils to discriminate both sample and comparison stimuli no matter where they were located and when they were presented relative to each other. If they had to learn these stimulus features during testing, gradual learning curves might have resulted.

Then, too, all stimuli belong to many classes. When encountering new sample-comparison combinations during unreinforced test trials, pupils may start by classifying certain stimuli together but then discover in succeeding trials that their original classification was not always possible. For example, they might start by classifying sample and comparison stimuli together on the basis of shared colors or shared forms such as curved or straight lines. They may require many trials to discover that the only way sample and comparison stimuli could always be classified together was on the basis of acquired equivalence.

Even though the teacher or experimenter delivers no reinforcement on test trials, no learning would have taken place without some kind of reinforcement. What might it have been? I would guess that the pupil must have learned at least two things in addition to the equivalence relations that are being tested. First of all, the pupil must have had enough experience with the teaching or experimental procedures to have learned that there is always a correct response. Second, the pupil must also have learned that the basis for a correct response does not change from trial to trial. What is relevant here is that pupils benefit from a history of consistent experience with the experimental or teaching procedures. A pupil who has not had sufficient experience with the procedures is going either to demonstrate equivalence relations slowly or to fail to show equivalence at all.

And so, we see that more general questions may arise out of a narrowly defined observation. In the present instance, what looks like a limited problem caused by differences in experimental structures is actually a more general
problem about what goes on during learning. Erik Arntzen’s question is more complex than it seemed at first. Its answer is not only relevant to the understanding of equivalence relations but appears also to suggest the possibility that learning may take place in the absence of teacher- or experimenter-administered reinforcement if we simply make available to the pupil a single consistent basis among many inconsistent bases for relating samples and comparisons.

Experimental demonstrations of conditional-discrimination learning on the basis solely of consistency of sample-comparison relations would provide much more generally applicable information about behavior than would answers to the more limited question of why the location of the nodes (in the samples or in the comparison stimuli) in networked conditional discriminations helps to determine the rapidity of the emergence of equivalence relations. Indeed, design features may prove of greater interest than the data that the particular designs produce.

Finally, Arntzen is concerned about experimental data that we cannot replicate by observing the behavior of individual subjects—as when an original behavioral change proves irreversible. Such irreversibility, which happens often in studies of learning and of stimulus control, would, of course, prevent replication of the behavioral change with the same subject and seems to provide ammunition for critics of individual-subject research. Are group designs the answer?

I think the solution to this problem lies in what one means by group designs. For example, group design usually means the comparison of averaged data from experimental and control groups. This kind of group comparison does, of course, have its value but only when one is not interested in any individual’s data. Another kind of group design that does not require the averaging of data across subjects is direct replication across several individuals. One could still average the data if there turned out to be very little variability from subject to subject, but despite our best efforts, precise quantitative inter- and intrasubject consistency is sometimes elusive. It is the practice of behavior analysis to try to determine the causes of inter- and intrasubject variability, not to hide variability by averaging the data.

We have here a major philosophical distinction between behavior analysis and psychology. Behavior analysis is philosophically devoted to the study of individual behavior but psychology (with the exception of psychophysics and, occasionally, physiological psychology), although it states its findings as applying to individuals, looks only at averages. This philosophical difference shows up in the technology of investigation, that is to say, in the design of experiments and in the evaluation of data. Attempts to solve the problem of data irreversibility, then, by using averaged group data would represent the abandonment of the philosophical commitment of behavior analysis to the understanding and treatment of the individual.
References


