

## The Trouble with Babies and the Value of Bathwater: Complexities in the Use of Verbal Reports as Data

Thomas S. Critchfield  
and  
L. Kimberly Epting  
Auburn University

Recent interest among behavior analysts in protocol analysis techniques prompts a consideration of some general measurement issues and some special issues relevant to protocol analysis. The development of behavior-analytic method and theory specific to verbal report research is a good thing, and Ericsson and Simon's (1984) book, *Protocol Analysis*, provides a useful model of integrating psychological theory and the craft of research. But protocol analysis techniques do not provide a magic window to the "world within the skin," and individual researchers should adopt these techniques only after confronting thorny issues such as how to determine the operating characteristics of verbal reports about private events, how to identify public performances to which protocol analysis can be applied productively, and how to maintain theoretical integrity in the empirical search for private events. We also caution against letting enthusiasm (and controversy) regarding protocol analysis distract behavior analysts from the benefits of using verbal report methods to study interesting events that are public in principle but difficult to measure in practice.

Recent commentaries (Austin & Delaney, 1998; Hayes, 1986; Hayes, White, & Bissett, 1998) and experiments (e.g., Wulfert, Dougher, & Greenway, 1991; Wulfert, Greenway, & Dougher, 1994) highlight a growing interest among behavior analysts in protocol analysis. As described in Ericsson and Simon's (1984) landmark book, protocol analysis is a set of strategies and tactics for research in which subjects verbalize thoughts, normally those taking place concurrently with a public performance. For example, Wulfert et al. (1994, Experiment 2) gave the following instructions to subjects working in a stimulus equivalence experiment:

We want you to think aloud when you perform the task on the computer. ... We will record and transcribe what you say because this will help us understand how you learn and how you arrive at the solution of the problems you will see. ... Be sure to describe clearly everything you think while you are solving the problems. (p. 426)

We are indebted to Rudy Vuchinich, Jalie Tucker, two anonymous reviewers, and Elizabeth Flanagan for comments that helped to shape this article.

Address correspondence to T. Critchfield, Department of Psychology, Auburn University, Auburn, Alabama 36849 (E-mail: tcritch@mail.auburn.edu).

Independent raters placed verbal responses into categories (e.g., naming experimental stimuli and describing physical features of stimuli). When intersubject differences appeared on equivalence tests, inspection of verbal response patterns "unambiguously identified aspects of the experimental procedure responsible for these performance differences" (Wulfert et al., 1994, p. 435).

The recent interest in protocol analysis is striking given the historical rarity of verbal report data in behavior-analytic research (Critchfield, Tucker, & Vuchinich, in press; Perone, 1988). Reasons for this omission rarely are specified, but casual observation suggests that (a) some in the field regard all self-report methods as misapplying psychometric theory in service of a dualistic philosophy of science; and (b) historically, most behavior analysts have conducted research on individuals who have little to say (e.g., nonhumans and persons with severe developmental disabilities). Given the latter point, it may be no accident that the growth of human operant research in recent years (Hyten & Reilly, 1992) has been accompanied by increased interest in verbal report methods.

The past century has produced verbal report methods that were neither theoretically nor empirically defensible. Although it is important not to overgeneralize from mistakes of the past, advocating the development and use of verbal report methods is not as simple as admonishing behavior analysts of the past for having thrown out the baby with the bathwater. Many standard arguments for and against verbal report methods obscure the complexity of the measurement issues involved. The purpose of this article is not to prescribe good research practices per se (in fact, it may prove necessary to develop and validate verbal report methods specifically for each research program) but rather to identify some general issues worth considering in the use of verbal report methods, including protocol analysis.

### PROTOCOL ANALYSIS, INTROSPECTION, AND THE MEGAPHONE HYPOTHESIS

Ericsson and Simon (1984) make clear that protocol analysis is different from classical introspection, in which subjects know the goals and hypotheses of the research and thus are susceptible to pronounced audience control. Following in Watson's (1920) tradition, Ericsson and Simon are interested in methods that publicize inner events (such as private verbalization) of experimentally naive individuals working under minimal audience control. In theory, the verbal responses studied in protocol analysis are not descriptions but rather are a direct sampling of the contents of consciousness. Protocol analysis is held to channel thinking into speech much as Pavlov's surgeries redirected canine salivary output from an internal to an external terminus.

Obviously, a means of publicizing covert verbal behavior would be a welcome addition to the research arsenal of behavior analysis. Skinner's (e.g., 1953, 1957) radical behaviorism has always placed heavy emphasis on covert verbal processes, and in recent years, behavior analysts increasingly have invoked rules or other verbal mediation to explain public performances (e.g., Horne & Lowe, 1996). Behavior analysts can object to the cognitive theoretical foundations

of protocol analysis but remain intrigued by the notion that well-constructed verbal report methods can provide "pure data" (direct access to thought) in the sense described by Ericsson and Simon (1984). This notion bears further inspection because it can be easily misconstrued.

Direct access does not imply mere amplification of covert verbal behavior, a caricature view that we think of as the "megaphone hypothesis." Taken out of context, Skinner (1957) sometimes can sound like an advocate of this view:

There are, then, important variables which determine whether a response will be overt or covert. But they do not greatly affect its other properties. They do not suggest that there is any important distinction between the two levels or forms. (p. 437)

If there is "no important distinction" between public and private responses, then amplifying private ones for measurement purposes should be a simple matter. But Skinner's comment defends the use of common principles to account for private and public verbal behavior, not a necessary equivalence of the two forms. In fact, the megaphone hypothesis is inconsistent with a functional analysis of verbal behavior.

Skinner (1957) described verbal behavior as implicated in complex functional relations that dictate both its occurrence and its formal properties. In the latter case, verbal responses occur covertly for distinct reasons, including efficiency of emission and a history of punishment. Any technique that seeks to make the verbal behavior public necessarily alters the relevant functional relations. Some approaches, for instance, may be designed to neutralize stimulus control related to punitive factors that suppress verbal behavior (consider hypnosis, or "automatic writing" as discussed by Skinner, 1934, 1957, pp. 388-390). In lay terms, these techniques may be thought to "release" verbal behavior (Skinner, 1957, p. 437) that previously has been driven within the skin. But to speak in such terms is to approach the logical fallacy of assuming that properties of behavior can exist independently of responses (e.g., recall Skinner's own, 1938, ill-fated notion of the reflex reserve). Responses are part of atomic, indivisible functional relations. Altering a component of

these relations (e.g., neutralizing factors that cause verbal behavior to be emitted covertly) does not illuminate previously hidden behavior. It produces new functional relations.

A related point is that covert verbal responses may not take linguistic form. Unlike Watson (1920), Skinner (1957) did not assume that private verbal behavior necessarily consists of silent talking. In fact, one reason for verbal behavior to become private is to circumvent the mechanical constraints of speaking – responding is faster and easier when the cumbersome formal properties of spoken language are abandoned (Skinner, 1957, pp. 435-436). Thus, private verbal responses, when they exist, may be in a form that is incompatible with spoken expression. When covert verbalization is not linguistic in nature, it must be translated in order to become public. That is, new functional relations must emerge (Skinner, 1957, chap. 8) from some unspecified combination of factors that once promoted covert behavior and factors that now require public verbalization.

### MEASUREMENT AND REPRESENTATION

Within the context of protocol analysis, verbal reports are neither descriptions (as in introspection) nor copies (as in the megaphone hypothesis) of private verbalizations. And if thinking out loud is not identical to thinking, then verbal reports are of dubious value to those seeking to investigate Skinner's (1953) "world within the skin," unless verbal reports can be presumed to correspond systematically to private verbal behavior. The required link is often assumed to be one of shared causation. For example, the common basis of private and public verbal behavior is, to Skinner (1957), a constellation of environmental variables, and to Ericsson and Simon (1984), information residing in short-term memory. Whatever the underlying mechanism, protocol analysis is designed to take advantage of a naturally occurring *representational system*, or systematic, structural correspondences between events in two domains (e.g., see Barsalou, 1992, pp. 52-55).

It is important to note that all measurement involves representation (e.g., see

Cohen, Swerdlik, & Smith, 1992; Poling, Methot, & LeSage, 1995). Computer data, mechanical counter readings, and paper data printouts are not identical to responses. They merely correspond to them. In good measurement, this correspondence is dependable and well understood, but is still imperfect and still subject to the influence of a host of variables. When verbal reports serve as measurement, therefore, it makes sense to consider them within the context of general measurement issues.

#### *Measurement Operating Characteristics*

The classic conundrum in verbal report methodology pits the potential contributions of information gleaned from verbal reports against ambiguities regarding the scientific utility of verbal report data. In considering this issue, many behavioral scientists rely on traditional, but relatively uninformative, distinctions among measurement systems (e.g., subjective vs. objective, direct vs. indirect). At best, these distinctions only hint at characteristics that make measurement systems desirable; at worst, they falsely imply that measurement systems have static properties. To avoid distracting dichotomies, Critchfield et al. (in press) recommended thinking of measurement in terms of *operating characteristics*, that is, the functional relations governing the production of any measurement system's data records and thus the properties of the underlying representational system.

A measurement system's operating characteristics will encompass complex, situational, dynamic variables. Even the best measurement systems provide useful data only under certain conditions. Good measurement detects some target events and produces data records in response, but all measurement, whether automated or implemented by humans, produces imperfect data records. All measurement leaves some potentially important events undetected (e.g., the peck on the periphery of the response key). All measurement necessarily distorts, in that data records are imperfect reproductions of target events, so that information is both lost and altered during measurement. And all measurement incorporates noise, in the form of data records influenced by

something other than target events (e.g., a microswitch registers a key-peck response regardless of whether the key was depressed by pecking or by accidental contact with a flapping wing).

Thinking of measurement as a representational system with operating characteristics renders lay conceptions of data accuracy meaningless. "Accuracy" can never imply perfect point-to-point correspondence between every feature of the measured environment and the features of data records. At issue, instead, is the extent to which data records promote effective behavior with respect to target events. To a baseball hitter, for example, it matters little whether sensory and perceptual systems capture the true essence of hurtling baseballs, as long as these systems allow the ball to be located (predicted) and hit (controlled). To the behavioral scientist, it is unimportant whether measurement provides a veridical snapshot of target events, as long as data records spawn prediction and control in the form of useful theory, ready integration with the findings of other studies, and valuable applications. To preface a later point, however, it is worth noting that the similarity between measurement and other representational systems breaks down in special cases. Because the baseball hitter receives regular feedback (e.g., "Steerike!") that calibrates his perceptual-motor systems, he can make do with only a tacit understanding of the operating characteristics of these systems. By contrast, target events of interest to many psychologists, including those related to events within the skin (Skinner, 1953), provide less vivid feedback, often requiring a more systematic approach to estimating measurement-system operating characteristics.

The challenge inherent in any measurement system, therefore, is to understand its operating characteristics well enough to use it intelligently. Investigators who use verbal report measurement actually must be concerned with two sets of operating characteristics: those of the measurement system (which affect the quality of inferences that be drawn about target events as they occur in the experimental environment) and those of target behaviors (which affect the generality of data-based conclusions to situations

outside the experiment). In the latter case, one must be concerned that the process of measurement does not substantially alter target behaviors, because verbal reports, and the variables that prompt them, may be implicated in the operating characteristics of target behavior. Ericsson and Simon (1984) amass an impressive array of evidence to demonstrate that, under certain circumstances, verbal report data can be collected without disrupting ongoing public performances. This is a crucial point, perhaps the most convincing of Ericsson and Simon's book.

But what evidence exists that public verbal responses can promote meaningful inferences about covert verbal behavior? Most writers are quieter on this matter (cf. Hayes et al., 1998), except, primarily, to note that the misuse of protocol analysis can result in verbal reports with little direct bearing on the events they are supposed to represent. The underlying theme is an important one, because human research subjects will do almost anything other than remain silent. Survey respondents, for example, may readily express opinions on topics about which further questioning reveals them to know little (Sudman & Bradburn, 1974). And protocol analysis subjects may, in response to experimental queries, be prompted to make guesses, draw inferences, and construct hypotheses that become an unwelcome part of the data set (Ericsson & Simon, 1984; Nisbett & Wilson, 1977; White, 1988). Obviously, it is important to structure the social environment of the study to limit such influences. But this is not the same as demonstrating that verbal reports correspond well to the target events they are supposed to represent.

#### *Validating Verbal Report Data*

To chart a measurement system's operating characteristics is to identify the tolerances under which it can provide useful data. Standard measurement concepts such as reliability and validity can be thought of as identifying strategies for estimating operating characteristics, but various approaches can contribute to this effort. Validation of verbal report methods is any means of establishing that inferences about target events can be adequately supported by verbal report data.

*Empirical validation.* Sometimes it is possible to evaluate directly the correspondence between verbal reports and target events, preferably under circumstances much like those of the study. Most of the behavioral assessment literature on self-monitoring, for example, is devoted to charting the conditions under which self-records reliably track public behaviors like smoking, studying, and eating (R. O. Nelson, 1977). Typically in these studies, experimenters have observed both the verbal reports and the behaviors being reported. When researchers lack easy access to target events, however, such a direct approach is not feasible, and creative alternatives are required. For example, most eating and drug use, although public in principle, take place away from the researcher's view, but reports corresponding to these events can be partially corroborated using physiological assays that track by-products of food or drugs in the body (e.g., Bandini, Schoeller, Cyr, & Dietz, 1990). Another approach is similar to that of assessing interobserver agreement. Alcoholics' verbal reports about naturalistic drinking patterns have been validated through comparison with reports by conspecifics (e.g., spouses) who know the drinkers well and have presumably observed the same bouts of drinking (e.g., Gladsjo, Tucker, Hawkins, & Vuchinich, 1992; Samo, Tucker, & Vuchinich, 1989).

*Validation by design.* Automated "transducers" of target events, such as computers and electromechanical equipment, rarely are subjected to formal tests of reliability and validity. What substitutes is faith in processes of design and manufacture that rely on firmly established scientific principles. Understanding the principles on which a measurement device is built allows relatively easy prediction of conditions under which the device will provide useful data records (Critchfield et al., in press).

Ericsson and Simon's (1984) *Protocol Analysis* approximates an effort at validation by design. Specifically, Ericsson and Simon draw upon consensus views in cognitive science to argue that information is available to verbal awareness only while in short-term memory. They directly acknowledge that subjects who lack knowledge called for in a research protocol may infer or otherwise

construct an answer, and thus stress the importance of using verbal report methods only to gain access to knowledge at subjects' disposal. Much of the book is devoted to explaining conditions under which information of interest can be expected to reside in short-term memory and be reported in uncorrupted form. But this approach is an imperfect parallel to that used in the manufacture of mechanical measurement devices. Cognitive scientists do not always agree about what events reside in short-term memory, are verbally represented, or are inherently accessible to awareness. And it remains to be seen whether recent data on self-awareness and apparently nonconscious processes (e.g., Fox, 1995; T. O. Nelson, 1992; Roediger, 1997) requires any reappraisal of Ericsson and Simon's (1984) assumptions about the relationship between short-term memory and verbal awareness.

When validation efforts are steeped heavily in psychological theory, we must confront the fact that psychologists are still struggling to understand the scientific principles on which humans, who might serve as measurement tools, are "constructed."

When responses are recorded by specialized apparatus, the events to be observed and explained operate primarily in one domain (behavior), while the transducer of these events operates primarily in others (e.g., physics). Interpreting the output of devices that transduce behavior is relatively straightforward because the scientific and technical domains in which they operate are fairly well understood. By contrast, self-observers ... operate in the same scientific and technical domain as the [events] they are used to measure. ... Attempts to use one response pattern to transduce another represent a bootstrapping operation that is feasible only when the investigator's understanding and control of the "transducing response" ... exceeds that of the [events it is supposed to measure]. (Critchfield et al., in press)

In the end, Ericsson and Simon's (1984) approach may be regarded as *theoretical validation*, which can be only as persuasive as the theory on which it is based. Nevertheless, theoretical coherence seems essential to any useful verbal report methodology (Critchfield et al., in press). To dismiss the cognitive basis of Ericsson and Simon's work is to obscure the fact that it elegantly unites theory and the craft of research in a way that behavior analysts thus far have failed to attempt with respect to verbal reports.

Like *Protocol Analysis*, a thoroughly developed behavior-analytic approach would have to employ a single set of principles both to articulate its theoretical foundations and to prescribe strategies and tactics of good verbal report methods. On the theoretical front, it would have to respect conceptual analyses such as Skinner's (1957) *Verbal Behavior*, be consistent with the empirical literature on stimulus control, and account for critical nuances of a rich and massive literature on remembering. Ideally, it would also be informed by the practical experience of those already studying verbal reports as behavior, for example, in the context of behavioral assessment (Babor, Stephens, & Marlatt, 1987; R. O. Nelson, 1977) or survey methods (Sudman & Bradburn, 1974; Wentland & Smith, 1993).

### WHEN TO USE PROTOCOL ANALYSIS?

Common sense suggests that research programs are most complete when they draw upon multiple methods and designs to examine phenomena of interest at many levels of analysis. For this reason, we advocate verbal report methods, like any other, as part of a multifaceted plan of attack rather than as the sole focus of any research program. In this section, we address a few broad issues worth considering when contemplating the use of protocol analysis.

#### *To What Public Performances Can We Apply Protocol Analysis?*

As noted by Austin and Delaney (1998), protocol analysis may be especially useful for studying well-defined tasks that draw upon a common knowledge base shared by many typically developing individuals. Well-defined tasks have a known endpoint or can be completed successfully using only a finite number of different response sequences. For example, most people, confronted with the problem  $43+18$ , can determine the answer to be 61, and to achieve this answer many people will carry the 1 from the digits to the tens column. Well-defined tasks provide obvious benchmarks against which verbal reports can be compared, and thus offer a means of estimating validity. But basic research programs

in behavior analysis usually exist to examine behavioral processes that currently are ill defined, and thus provide few such benchmarks (consider, for example, the various strategies that can "solve the problem" of a fixed-interval schedule by generating a response at the moment reinforcement becomes available). In such cases, the information sought through an a priori task analysis may include precisely that sought from the research itself.

#### *When Does Thinking Occur?*

The appeal of protocol analysis derives in part from the promise that it can bring thinking under the purview of behavior-analytic research, but what is thinking? And how is it implicated in performances of interest to behavior analysts? Skinner (1957, e.g., chap. 19) addressed the former question in some detail, but not the latter. As an alternate point of departure, recall that the thinking of interest to Ericsson and Simon (1984) is covert processing of verbal information in short-term memory. In a lengthy discussion of this point, Ericsson and Simon question the extent to which verbal representation can be expected in overlearned performances that have become "automated" such that they require little conscious cognitive effort. It is important to note that automated performances arise from extensive practice involving consistent relations between stimuli and responses (Barsalou, 1992), precisely the recipe for most steady-state conditioning procedures. Some cognitive psychologists now view many conditioning phenomena as automated performances that operate largely outside of awareness (Barsalou, 1992), an assumption that seems to be bolstered by studies providing evidence of operant conditioning without subject awareness (e.g., Hefferline & Perera, 1963). It is even possible that relatively autonomous neural substrates underlie self-observation and other behavioral processes. For example, different types of human brain injury differentially impair classical conditioning and verbal explanations of the events involved in the conditioning (Bechara et al., 1995).

In short, there simply may be cases in which thinking, as implicated in protocol

analysis, is not an important component of performances of interest to behavior analysts. Some behavior analysts may object to Ericsson and Simon's (1984) theoretical framework as a basis for this conclusion, but no behavior-analytic alternative has been articulated with sufficient clarity to generate testable predictions. In the absence of clear theoretical guidelines for suitability, research will have to progress in trial-and-error fashion until it becomes clear which research questions are best served by protocol analysis procedures. It remains to be seen, therefore, how much impact methods of protocol analysis can have in basic behavior analysis research.

#### *What Does It Mean to Look for Thinking?*

Unresolved issues about the role of thinking in overt performances highlight the fact that all psychological methods have philosophical underpinnings. Often, researchers of conditioning phenomena have considered reliable behavior-environment relations to be satisfactory explanations of behavior. A prototypical example was Pavlov who, early in his research, was frustrated by attempts to understand the internal psychological states that might contribute to the formation of conditioned reflexes. As rigorous experiments gradually revealed the environmental determinants of conditioned reflexes, Pavlov followed his successes: "I decided finally, in regard to the so-called psychical stimulation, to remain in the role of ... an objective external observer and experimenter, having to do exclusively with external phenomena and their relations" (Pavlov, 1928, pp. 38-39).

Cognitive psychologists have tended to view reliable behavior-environment relations not as explanations but as events to be explained in terms of internal mediational processes (Baars, 1986; Barsalou, 1992). Clearly, cognitive and behavioral approaches draw upon broadly different assumptions (e.g., Hayes, Hayes, & Reese, 1988). Pepper (1942) proposed that all theoretical systems are subsumed under a handful of metatheoretical "world hypotheses" that are mutually exclusive. Thus, according to Pepper, theoretical progress is likely only within the confines of a single world

hypothesis. This may be an overstatement, true only if a theoretical system's premises constrain conceptual revision (see Harzem & Miles, 1978), but Pepper at least reminds us of the pitfalls of "confused eclecticism," or the development of theory without regard to conceptual integrity. Depending on how research questions are cast and how thinking is defined, one wonders what influence protocol analyses might have on the theoretical direction of behavior analysis research programs. Where reliable behavior-environment relations have been demonstrated, for example, introducing protocol analysis implies dissatisfaction with these relations as a thorough account of behavior. A tool analogous to protocol analysis, if one can be conceived for dogs, might have taken Pavlov's research in very different directions, both empirically and theoretically. Although open-mindedness is a virtue, one should step outside one's world hypothesis only cautiously and with a plan (Hayes et al., 1988), and it seems prudent for researchers considering the use of protocol analysis (or any measurement system, for that matter) to consciously evaluate the correspondences between their research methods, research questions, and theoretical frameworks.

#### *Rules, Protocol Analysis, and Variability*

It is clear from Ericsson and Simon (1984) that protocol analysis was developed largely as a means of exploring the moment-to-moment correspondences between internal events and public performances. Theoretical arguments about the nature or causal importance of inner events notwithstanding, behavior analysts, as students of intrasubject variability and advocates of a natural science approach (Sidman, 1960), should find favor with Ericsson and Simon's emphasis on individual-subject events occurring in real time. It is interesting to note, therefore, a tendency in behavior-analytic research toward describing only aggregate verbal report data. In some studies, verbal responses are coded, and cumulative frequencies are listed by response category (e.g., Wulfert et al., 1994). This approach retains a broad description of variability, and may be adequate for many purposes, but it obscures momentary

variability, including possible covariance of verbal responses with other aspects of performance (Critchfield et al., in press).

Some analyses, in an attempt to identify overarching rules that may guide public performances, obscure all intrasubject variability. For example, Horne and Lowe (1993) collected postexperimental verbal responses from subjects working on concurrent variable-interval schedules. They then determined the most common type of comment made by each subject, and from this made inferences about the performance rule that guided each subject's performance. In promoting static attributions about individuals, this strategy has more in common with trait theories than with the usual behavior-analytic approach, and Ericsson and Simon (1984) probably would recommend that it be undertaken with caution. Although Ericsson and Simon, as cognitive scientists, assume that people form rules and acquire general knowledge, they continue to stress the importance of dynamic processes:

Knowledge and rules reside in [long-term memory]. In order to be heeded and to influence cognitive processes, they need to be accessed by appropriate retrieval cues. Even though we have been able to identify them in a protocol on one problem, we cannot be certain that they will be accessed and used on other problems. (p. 311)

Accordingly, Ericsson and Simon attempt to distinguish verbal responses indicating broad rules from those indicating fleeting hypotheses. "Generalized entities," like rules, they conclude, "are directly verbalized only under extraordinary circumstances" (p. 310). In cases in which they advocate aggregating individual responses, Ericsson and Simon (1994) search for sequences of thoughts under apparently common control – what behavior analysts might call a response chain – which can be viewed as a single behavioral unit and therefore subjected to real-time analysis like any other behavioral unit. This stands in stark contrast to the practice of pooling numerous verbal responses, irrespective of time and situation, and categorizing individuals according to their most common class of verbalization.

Rule governance has commanded substantial interest in behavior analysis at least since Skinner (e.g., 1969) distinguished it from contingency-controlled behavior. But

aside from acknowledging that rules can be inferred from behavior patterns that are impervious to situational exigencies, behavior analysts have reached little consensus on the nature of rules and analogous forms of mediation, or the means by which they may influence behavior (e.g., Buskist & DeGrandpre, 1989; Cerutti, 1989; Horne & Lowe, 1996). Protocol analysis may prove to be useful in the study of rule governance from a behavior-analytic perspective, but further theoretical development would be helpful in guiding the creation of appropriate verbal response categories and the choice of a proper level of analysis for verbal report data. In the absence of this assistance, if protocol analysis data are to be useful in the sense of behavioral assessment, it likely will be in their capacity to measure events in real time.

#### OTHER OPPORTUNITIES IN VERBAL REPORT MEASUREMENT

Verbal report procedures, properly devised and executed, may well help to bring thinking under the empirical scrutiny of the experimental analysis of behavior, and in the process, one may suppose, help to demonstrate the utility of behavior analysis to psychologists who assume that private events have no place in radical behavioral epistemology. In our haste to pursue thinking as an empirical subject matter, however, we should not lose sight of the fact that verbal report measurement can be useful to behavior analysts in other ways. For example, to increase their impact in psychology and society, behavior analysts might also look for ways to demonstrate empirically that functional relations revealed so precisely and reliably in the operant laboratory apply to the everyday affairs of humans without developmental disabilities. Obvious constraints on this enterprise include difficulties in measuring and manipulating events in uncontrolled settings. A small number of descriptive studies, however, show clear operant influences in readily observed, naturally occurring behavior (Mace, Lalli, Shea, & Nevin, 1992; Moerk, 1990). Verbal report data might permit the extension of this approach to behavior that is public in principle



but impractical to observe. Tucker and Vuchinich, for example, have drawn upon research establishing the operating characteristics of alcoholics' reports about drinking episodes over extended time frames (e.g., Gladsjo et al., 1992; Samo et al., 1989) to develop methods for examining environmental precursors of drinking, abstinence, and help seeking within a framework informed by behavioral economics and behavioral choice theory (e.g., Tucker, Vuchinich, & Pukish, 1995; Vuchinich & Tucker, 1996). Compared to research on intrinsically private phenomena, research of this sort lends itself to a broader range of verbal report validation strategies, and can be more readily informed by existing research that also focuses on verbal reports of public events (Critchfield et al., in press).

### CONCLUSIONS

At a general level, we applaud the current interest in protocol analysis because such methods sensitize us to three general points. First, some questions about behavior do lead us to wonder what is happening outside the sphere of direct observation, and verbal report protocols can extend our scope of observation by allowing subjects to talk about events to which we have no other access. Second, verbal report methods merit the same level of scrutiny as other methods. Whatever its other features, Ericsson and Simon's (1984) *Protocol Analysis* is theoretically coherent, meticulously researched, and clearly articulated. By contrast, most forays by behavior analysts into verbal report measurement have not been impressive in their rigor (Critchfield et al., in press). Third, by applying general principles of measurement, in conjunction with a traditional respect for methodological rigor, behavior analysts who carefully develop verbal report methods very well may be able regard their data as a trustworthy form of behavioral assessment (e.g., Babor et al., 1987; Critchfield et al., in press; R. O. Nelson, 1977).

A thoroughgoing science utilizes all of the resources available to it, and behavior analysts overlook a valuable tool when they reject verbal report measurement out of hand. But developing a uniquely behavior-analytic approach to verbal report methods

is no small agenda. Some issues surrounding verbal report data are demystified (although not necessarily simplified) by considering verbal reports within a framework of measurement operating characteristics. Progress in charting the operating characteristics of verbal reports currently is restricted by limited empirical guidance and by the lack of a behavior-analytic theoretical approach to verbal report measurement that can match, in detail and level of conceptual integration, Ericsson and Simon's (1984) account based on information processing theory. Finally, with the opportunities inherent in verbal report measurement come difficult questions about the compatibility of protocol analysis with behavior-analytic research and theory. As long as these questions are approached rigorously and thoughtfully, however, regardless of the conclusions, behavior analysis is bound to be better for having accepted the challenge.

### REFERENCES

- Austin, J., & Delaney, P. F. (1998). Protocol analysis as a tool for behavior analysis. *The Analysis of Verbal Behavior, 15*, 41-56.
- Baars, B. J. (1986). *The cognitive revolution in psychology*. New York: Guilford.
- Babor, T. F., Stephens, R. S., & Marlatt, G. A. (1987). Verbal report methods in clinical research on alcoholism: Response bias and its minimization. *Journal of Studies on Alcohol, 45*, 410-424.
- Bandini, L. G., Schoeller, D. A., Cyr, H. N., & Dietz, W. H. (1990). Validity of reported energy intake in obese and nonobese adolescents. *American Journal of Clinical Nutrition, 52*, 421-425.
- Barsalou, L. W. (1992). *Cognitive psychology: An overview for cognitive scientists*. Hillsdale, NJ: Erlbaum.
- Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science, 269*, 1115-1118.
- Buskist, W. F., & DeGrandpre, R. J. (1989). The myth of rule-governed behavior. *Experimental Analysis of Human Behavior Bulletin, 7*, 4-6.
- Cerutti, D. T. (1989). Discrimination theory of rule-governed behavior. *Journal of the Experimental Analysis of Behavior, 51*, 259-276.
- Cohen, R. J., Swerdlik, M. E., & Smith, D. K. (1992). *Psychological testing and assessment: An introduction to tests and measurement*. Mountain View, CA: Mayfield.
- Critchfield, T. S., Tucker, J. A., & Vuchinich, R. E. (in press). Self-report methods. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior*. New York: Plenum.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Fox, E. (1995). Negative priming from ignored distractors in visual selection: A review. *Psychonomic Bulletin & Review, 2*, 145-173.

- Gladsjo, J. A., Tucker, J. A., Hawkins, J. L., & Vuchinich, R. E. (1992). Adequacy of recall of drinking patterns and event outcomes associated with natural recovery from alcohol problems. *Addictive Behaviors*, *17*, 347-358.
- Harzem, P., & Miles, T. R. (1978). *Conceptual issues in operant psychology*. Chichester, England: Wiley.
- Hayes, S. C. (1986). The case of the silent dog – Verbal reports and the analysis of rules: A review of Ericsson and Simon's *Protocol Analysis: Verbal Reports as Data*. *Journal of the Experimental Analysis of Behavior*, *45*, 351-363.
- Hayes, S. C., Hayes, L. J., & Reese, H. W. (1988). Finding the philosophical core: A review of Stephen C. Pepper's *World Hypotheses: A Study in Evidence*. *Journal of the Experimental Analysis of Behavior*, *50*, 97-111.
- Hayes, S. C., White, D., & Bissett, R. T. (1998). Protocol analysis and the "silent dog" method of analyzing the impact of self-generated rules. *The Analysis of Verbal Behavior*, *15*, 57-63.
- Hefferline, R. F., & Perera, T. B. (1963). Proprioceptive discrimination of a covert operant without its observation by the subject. *Science*, *139*, 834-835.
- Horne, P. J., & Lowe, C. F. (1993). Determinants of human performance on concurrent schedules. *Journal of the Experimental Analysis of Behavior*, *59*, 29-60.
- Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior*, *65*, 185-241.
- Hyten, C., & Reilly, M. P. (1992). The renaissance of the experimental analysis of human behavior. *The Behavior Analyst*, *15*, 109-114.
- Mace, F. C., Lalli, J. S., Shea, M. C., & Nevin, J. A. (1992). Behavioral momentum in college basketball. *Journal of Applied Behavior Analysis*, *25*, 657-663.
- Moerk, E. L. (1990). Three-term contingency patterns in mother-child verbal interactions during first-language interaction. *Journal of the Experimental Analysis of Behavior*, *54*, 293-305.
- Nelson, R. O. (1977). Assessment and therapeutic functions of self-monitoring. In M. Hersen, R. M. Eisler, & P. M. Miller (Eds.), *Progress in behavior modification* (Vol. 5, pp. 263-308). New York: Academic.
- Nelson, T. O. (1992). *Metacognition: Core readings*. Boston: Allyn and Bacon.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231-259.
- Pavlov, I. P. (1928). *Lectures on conditioned reflexes* (W. H. Gant, Trans.). New York: Liveright.
- Pepper, S. C. (1942). *World hypotheses: A study in evidence*. Berkeley, CA: University of California Press.
- Perone, M. (1988). Laboratory lore and research practices in the experimental analysis of human behavior: Use and abuse of subjects' verbal reports. *The Behavior Analyst*, *11*, 71-75.
- Poling, A., Methot, L. L., & LeSage, M. G. (1995). *Fundamentals of behavior analytic research*. New York: Plenum.
- Roediger, H. L. (Ed.). (1997). Implicit learning: A symposium. *Psychonomic Bulletin & Review*, *4*, 1-78.
- Samo, J. A., Tucker, J. A., & Vuchinich, R. E. (1989). Agreement between self-monitoring, recall, and collateral observation measures of alcohol consumption in older adults. *Behavioral Assessment*, *11*, 391-409.
- Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.
- Skinner, B. F. (1934). Has Gertrude Stein a secret? *Atlantic Monthly*, *153*, 50-57.
- Skinner, B. F. (1938). *The behavior of organisms*. Englewood Cliffs, NJ: Prentice Hall.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Free Press.
- Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1969). *Contingencies of reinforcement: A theoretical analysis*. Englewood Cliffs, NJ: Prentice Hall.
- Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys: A review and synthesis*. Chicago: Aldine.
- Tucker, J. A., Vuchinich, R. E., & Pukish, M. M. (1995). Molar environmental contexts surrounding recovery from alcohol problems by treated and untreated problem drinkers. *Experimental and Clinical Psychopharmacology*, *3*, 195-204.
- Vuchinich, R. E., & Tucker, J. A. (1996). Alcohol relapse, life events, and behavioral theories of choice: A prospective analysis. *Experimental and Clinical Psychopharmacology*, *4*, 19-28.
- Watson, J. B. (1920). Is thinking merely the action of language mechanisms? *British Journal of Psychology*, *11*, 87-104.
- Wentland, E. J., & Smith, K. W. (1993). *Survey responses: An evaluation of their validity*. San Diego: Academic.
- White, P. A. (1988). Knowing more than we can tell: "Introspective access" and causal report accuracy 10 years later. *British Journal of Psychology*, *79*, 13-45.
- Wulfert, E., Dougher, M. J., & Greenway, D. E. (1991). Protocol analysis of the correspondence of verbal behavior and equivalence class formation. *Journal of the Experimental Analysis of Behavior*, *56*, 489-504.
- Wulfert, E., Greenway, D. E., & Dougher, M. (1994). Third-order equivalence classes. *The Psychological Record*, *44*, 411-439.