International Journal of Psychology and Psychological Therapy 2002, Vol. 2, Nº 2, pp. 115-130

Addressing the Limitations of Protocol Analysis in the Study of Complex Human Behavior

Francisco Cabello*[1] and Denis O´Hora**

**Universidad de La Rioja, España, **London Metropolitan University, UK*

ABSTRACT

Many commentators have argued that the protocol analysis method (Ericsson & Simon, 1993) has considerable utility in the study of complex human behavior. In particular, it has recently been suggested that this method allows for detailed analyses of human language and cognition from a behavioral perspective. Despite this utility, however, relatively few behavioral studies have employed this technique. In the current article, we point out certain empirical difficulties that we have encountered in attempting to employ the protocol analysis method to study human language and cognition. We then present a number of solutions to these problems, which we suggest will allow for wider use of the protocol analysis method in behavioral psychology.
Key words: protocol analysis, "silent dog" method, verbal reports, verbal behavior, rule-governed behavior.

RESUMEN

Diversos autores han argumentado que el análisis de protocolos (Ericsson y Simon, 1993) tiene una utilidad considerable en el estudio del comportamiento humano complejo. En particular, recientemente se ha sugerido que el método permite un análisis detallado del lenguaje humano y la cognición desde una perspectiva conductual. Pese a esta utilidad, han sido escasos los estudios que han empleado tal técnica. En este artículo, señalamos algunos de los problemas a nivel empírico que hemos encontrado a la hora de emplear el método de análisis de protocolos para estudiar el lenguaje humano y la cognición. Asimismo, presentamos una serie de soluciones para esos problemas, que creemos permitirán un uso más extendido del análisis de protocolos en la psicología conductual.
Palabras clave: análisis de protocolos, método del "silent dog", informes verbales, conducta verbal, comportamiento gobernado por reglas.

An early research strategy in the experimental analysis of behavior was to study the behavior of non-human animals to better understand the variables that control human behavior (Skinner, 1938, 1953). In recent years, however, it has become increasingly clear that, in certain contexts, human performances diverge significantly from non-human performances (Hayes, 1987; Hayes & Hayes, 1992). For example, language-able humans do not demonstrate the expected performances on particular schedules of reinforcement that are almost ubiquitous in non-human populations (Baron, Kaufman

[1]Reprints may be obtained from the author: Unidad de Ciencias Sociales del Trabajo. Edificio Quintiliano. Universidad de la Rioja, Calle Cigüeña, 60, 26004 Logroño (La Rioja, España). E-mail: francisco.cabelloluque@wanadoo.es.

& Stauber, 1969; Harzem, Lowe & Bagshaw, 1978; Matthews, Shimoff, Catania & Sagvolden, 1977; Shimoff, Catania & Matthews, 1981; Weiner, 1970). Also, verbally competent humans readily demonstrate derived relational responding (e.g., stimulus equivalence; Sidman & Tailby, 1982), but non-humans rarely do so (Dube, McIlvane, Callahan & Stoddard, 1993; Hayes, 1989a). That is, when human subjects are trained that a particular Stimulus A is related in a specific way to a second Stimulus B (e.g., A is greater than B), and that Stimulus B is related in a specific way to a third Stimulus C (e.g., B is greater than C), in the absence of further training, subjects respond in accordance with a number of untrained or derived relations (e.g. A is greater than C, both B and C are smaller than A).

Most commentators (for a book-length review, see Hayes, 1989b) have explained the foregoing differences between humans and non-humans in terms of language capabilities and, in particular, the human ability to describe contingencies and to generate verbal rules. Indeed, Lowe (1979) suggested that verbal behavior introduces a fundamental difference between human and nonhuman behavior, a position that has been termed the "language hypothesis". As a consequence, the empirical literature noticeably shifted from non-human to human participants during the early nineties (Dymond & Critchfield, 2001; Hyten & Reilly, 1992; Navarick, Bernstein & Fantino, 1990). Furthermore, at that time, many experimental procedures were devised to study verbal behavior and the influence of verbal processes in other behavior. Some of them concentrated on the effect of instructions provided by the experimenter on subject performance, whereas others were interested in self-verbalized instructions that participants state about their own responding (Chase & Danforth, 1991; Zettle & Young, 1987).

The current paper is concerned with protocol analysis, a recent method for analyzing verbal behavior and the effect of verbal responding on other behavior using verbal reports from participants. The main feature of this technique is that subjects are instructed to say aloud everything they are thinking about, thus providing a concurrent verbal stream that allows a moment to moment analysis of the relation between what subjects are saying and what they are doing, and of the role of verbal behavior in human responding. In the first part of this article, we briefly outline the self-reports approach to verbal control and the protocol analysis method and review some of the recent behavioral interest in the methodology. We then present some conceptual and methodological considerations of the protocol analysis procedure that might have restricted the number of published behavioral studies that employed these techniques. Finally, we describe a number of recent developments, presenting some of our own work specifically aimed to address the considerations indicated above. It is our hope that these developments will encourage the wider use of protocol analysis in the study of complex human behavior.

Self-reports and The Experimental Analysis of Verbal Control

Despite the controversial status of verbal reports within the experimental analysis of behavior (e.g., Critchfield, Tucker & Vuchinich 1998; Perone, 1988), an important body of empirical evidence obtained through verbal reports has demonstrated that different

types of verbal regulation such as counting, describing or planning, may be consistently related to participants' performances, (Barnes & Keenan, 1989, 1993; Holland, 1958; Leander, Lippman & Meyer, 1968). For example, Leander et al. examined the behavior of 80 subjects on different fixed-interval (FI) schedules, and found that the behavior of subjects on the schedule was mirrored in subject's post-experimental descriptions of the schedule contingencies. If subjects responded at low rates on the FI schedule, then they described the schedule as interval based, but if subjects responded at high rates, they described the schedule as rate based. In a later study, Lowe, Harzem and Hughes (1978) exposed human adults to different FI schedules of reinforcement during several hours, after which participants were given a questionnaire with different questions (e.g. "how did you gain points", "why did you press the panels when you did?"). The answers to these questions were related to the schedule performances, and the results demonstrated that subjects who counted during the intervals showed patterns of responding significantly different than the patterns of subjects who did not count (and whose responding was very similar to that of non-humans).

Most examinations of such verbal control have however examined post experimental verbal reports. This practice is troublesome as it relies on the assumption that verbal behavior after the experiment accurately reflects the behavior during the experiment (for a more detailed discussion of this issue, see Shimoff, 1986).To address these deficiencies, some authors have suggested that concurrent "think aloud" procedures might constitute useful tools for studying verbal behavior. In "think aloud" procedures, subjects are asked to verbalize their thoughts as they perform an experimental task. In this way, self-reports are obtained concurrently with other measures of behavior (i.e. task performance). This practice avoids assumptions about whether *verbal behavior that is recorded after the experiment corresponds to verbal behavior that may have occurred during the experiment*. Furthermore, it also allows the moment-to-moment relation between different behaviors (e.g. verbal reports and button pressing) to be examined.

"Think aloud" Procedures and Protocol Analysis

From the early years of the twentieth century, one approach that has often been employed in the psychological study of thinking has been to instruct subjects to say aloud everything they are thinking about (Watson, 1920). However, for much of this time, "think aloud" procedures constituted a group of heterogeneous techniques under a single label. It was not until the early eighties that Ericsson and Simon aimed to systematize "think aloud" procedures by providing a robust theoretical model for reporting cognitive processes, and a well-defined methodological procedure (Ericsson & Simon, 1980, 1993).

Protocol analysis was designed to gain information about the course and mechanisms of cognitive processes (Crutcher, 1994; Ericsson & Simon, 1980; Payne, 1994). For this purpose, subjects are asked to "think aloud" and to verbalize their thoughts at the same time they are engaged in the experiment as if they were alone, in a process that is not conversational and tries to minimize audience control. What follows

is a typical example of the method, based on the recommendations by Ericsson and Simon (1993):

1. Before starting the experiments, participants receive instructions about the experimental task, and are also given specific instructions about the "thinking aloud" requirement: "In this experiment we are interested in what you think about when you find answers to some questions. In order to do this I am going to ask you to think aloud as you work on the problem given. What I mean by think aloud is that I want you to tell me everything. I would like you to talk aloud constantly, I don't want you to try to plan out what you say or try to explain to me what you are saying. Just act as if you are alone in the room speaking to yourself."
2. After receiving instructions, subjects also receive practice at "thinking aloud" using simple exercises such as mental calculations. These exercises are used to prompt the participants to think aloud and to shape their verbal behavior, as "So that you understand what I mean by think aloud, let me give you an example. Assume I asked you 'How much is 127 plus 35?' Now think aloud so I can hear how you solve this problem."
3. Once that the training is completed, subjects are exposed to the experimental task, while prompted to keep "thinking aloud" if they remain silent for some time. The participants' verbalizations are recorded during the experiment (e.g. using audio tapes).
4. The following step involves the transcription into text of these verbalizations. Once that the content of the tapes is converted to written form, it is divided into several segments (e.g. speech sentences, or trial by trial).
5. These written segments are then assigned to different categories devised by the experimenter, usually by at least two independent raters. The categorizations of the different raters should be check for interrater reliability.
6. Finally, the categorized protocols are ready to be analyzed. In the case of cognitive psychology, the protocols are used to build an information processing model that is later simulated in the computer (Newell and Simon, 1972).

Although the possibility of reporting the content of cognitive processes has been a polemical issue within cognitive psychology (e.g., Nisbett and Wilson, 1977; Russo, Johnson & Stephens, 1989), protocol analysis attracted great interest in the psychological study of thinking and cognition. Subsequently, it generated a great amount of empirical research, and was widely applied to many areas (for a review, see Crutcher, 1994; Ericsson & Simon, 1993; Payne, 1994).

Protocol Analysis in The Experimental Analysis of Behavior

Protocol analysis, as used in the experimental analysis of behavior, is based on the assumption that, although some instances of verbal behavior can occur privately, the boundary between private and public does not introduce a fundamental difference in behavior (Hayes & Brownstein, 1986). Under certain conditions, what participants say overtly may constitute a functional analogue of private verbal behavior that would otherwise facilitate performance, and the verbal reports obtained using the protocol analysis method may be used as a model for the study of self-instructions (Barnes-

Holmes, Hayes & Dymond, 2001; Hayes, White & Bissett, 1998; Hayes, Zettle & Rosenfarb, 1989). The main utility of protocol analysis, however, is to provide an analysis of the effect of specific instances of verbal behavior, and of the functional properties of specific instructions and strategies in the completion of certain tasks.

Hayes (1986) was the first behavior analyst to examine the possible utility of the protocol analysis method. Since then, protocol analysis has piqued the interest of other behavior analysts. Indeed, the advantages of protocol analysis were examined in a special 1998 issue of *The Analysis of Verbal Behavior* that included several theoretical papers (Austin & Delaney, 1998; Critchfield & Epting, 1998; Hayes, White & Bissett, 1998). More recently, a paper by Potter (1999) addressed additional methodological issues, and researchers on Relational Frame Theory (Barnes-Holmes, Hayes & Dymond, 2001; Hayes, Zettle & Rosenfarb, 1989) have suggested the utility of thinking aloud for the study of rule-generation and rule-following.

It is therefore surprising to find that a comparatively small number of studies have been conducted within the experimental analysis of human behavior using the method. Of these, the vast majority have examined the stimulus equivalence phenomenon (Rehfeldt & Dixon, 2000; Rehfeldt, Dixon, Hayes & Steele, 1998; Rehfeldt & Hayes, 2000; Wulfert, Dougher & Greenway, 1991; Wulfert, Greenway & Dougher, 1994). Others have focused on areas such as the distinction between different types of verbal behavior (Potter, Huber & Michael, 1997), the role of rule-following in human operant resurgence (Dixon & Hayes, 1998), or the analysis of behavior during work (Austin & Mawhinney, 1999). Critically, however, the number of studies is quite small and, arguably, no systematic attempt to analyze verbal behavior has used this method in a detailed series of experiments.

Considerations in The Use of Protocol Analysis

It seems appropiate, at this point, to analyze the reasons for the lack of a behaviorally oriented empirical research program in protocol analysis. We believe that some considerations should be faced, when employing protocol analysis, that impede the use of the technique. The next section outlines these considerations. For the sake of clarity, a distinction will be made between theoretical considerations (related to the nature of verbal behavior and the validity of the method to gather valid data) and empirical considerations (related to procedural issues and the implementation of the research procedure). However, it is important to note that these two categories are fully interrelated, and neither can be fully understood in isolation.

Theoretical considerations

Apart from the methodological considerations raised above, some other conceptual issues need to be addressed, because they constitute a threat to the validity of the inferences that can be raised from the data obtained by the application of protocol analysis. In this section, we will focus in particular on the issues of validation of verbal

reports and the use of a correlational analysis of behavior.

Regarding validation, several authors have noted that in the use of self-reports, two components are involved: a referent event, and the public act of reporting (Critchfield & Epting, 1998; Critchfield, Tucker & Vuchinich, 1998). In the case of self-reports about private events, the referent is not readily available, and thus a process of validation is neccesary to establish "the degree of correspondence between self-reports and their referents that can be expected under a given set of circumstances" (Critchfield, Tucker & Vuchinich, 1998; p. 445). In other words, it should be determined how sure we are that the self-reports correspond to the self-instructions, a question that is not easy to answer (Hayes, White & Bissett, 1998). Indeed, most authors have relied on conceptual ellaborations about self-reports (e.g., about the nature of verbal reports, about private events, or about verbal behavior itself). This "theoretical validation", however, is limited by the theory on which it is based and is supported by uncertain conceptual assumptions that are open to debate (Critchfield & Epting, 1998; Critchfield, Tucker & Vuchinich, 1998; White, 1988). Subsequently, establishing the correspondence between self-reports and self-instructions on the grounds of assumptions about this correspondence represents a weak point that undermines the validity of protocol analysis.

Finally, protocol analysis is predominantly employed as a correlational methodology for the study of verbal behavior. From Ericsson and Simon's cognitive model, verbal reports may accurately reflect the content of cognitive processes that are regarded as cause of task performance (Ericsson & Simon, 1980, 1993). This explanation, however, is unacceptable from a behavioral view. Within a behavior-analytic perspective, both verbal reports and task performance are instances of behavior (for a review of mentalism and cognitive causality, see Hayes & Brownstein, 1986). Even if verbal reports appear to be consistently related to the behavior of participants, it is not clear whether (a) verbal reports are caused by task performance, (b) task performance is caused by verbal reports, (c) both verbal reports and task performance are caused by a third variable. Therefore, although correlations may interest the behavior analyst, a satisfactory explanation must demonstrate the historical and current contexts that are responsible for the relation among behaviors (Barnes, 1989; Barnes & Keenan, 1989, 1993; Critchfield & Epting, 1998; Hayes, 1986; Hayes & Brownstein, 1986; Hayes, White & Bissett, 1998; Perone, 1988; Shimoff, 1986).

Empirical Considerations

A first problem that researchers interested in protocol analysis must face is the considerable effort that it requires (see Critchfield, Tucker & Vuchinich, 1998). The proper implementation of the procedure requires the recording of participants' verbal reports, which often results in many hours of tapes. The content of these tapes has to be transcribed and divided into fragments, which are assigned into different categories. The categorization of the transcribed protocols requires devising the categories, selecting and training at least two independent raters, and checking agreement between them. The requirements in terms of time and effort, therefore, can be a barrier insofar as those interested in the analysis of verbal behavior might opt for less complex techniques.

Secondly, although Ericsson and Simon standardised certain aspects of the methodology, there is still no common procedure for protocol analysis, with different researchers using procedures that differ in important ways, such as the presence or absence of the experimenter during the experiment, the use of "thinking-aloud" practice exercises, the type of instructions provided, and the measures used to maintain a tight experimental control (see Normand, 2001). Furthermore, the impact of different procedures on the results is typically unanalyzed, and similar inferences are sometimes derived from very different procedures. These are crucial issues, because the relationship between verbal reports and task performance might well be sensitive to differences in procedure (see Austin & Delaney, 1998; Critchfield, Tucker & Vuchinich, 1998).

This lack of consensus among researchers is of critical importance when considering the categories used for the analysis of the data. Some researchers have suggested the use of task-independent categories, based on theoretical analyses. For example, Potter, Huber & Michael (1987) used the verbal operants suggested by Skinner to analyze verbal utterances (e.g. tact or intraverbal). Other researchers, in contrast, have recommended categories that are specific to each task, although also depending on conceptual elaborations. For example, Dixon and Hayes (1998) devised categories such as "comments about reinforcement", "reading screen" or "describing accurate contingencies" after an inspection of the data, to account for all the protocols obtained.

The main advantage of task-independent categories is that the same set of categories can be applied across different tasks, so that results are easier to generalize. However, the categorization process is less specific, and the categories may not account well for all of the verbal utterances that participants produce. In contrast, task-dependent categories are more specific and will account for all of the self-reports; but the results are less general and the comparison among studies is difficult. Although some cognitive psychologists have focused on the types of categories used in protocol analysis (e.g. Crutcher, 1994; Crutcher, Ericsson & Wichura, 1994; Ericsson & Simon, 1993), the issue has not been addressed by behavioral researchers. As a consequence, there are no clear guidelines regarding the categories, and whether to opt for task-independent or task-specific categories is not clear.

Finally, another methodological consideration is related to the analysis and representation of the data. Usually, a moment-to-moment analysis of the interaction between task performance and verbal reports is called for when using protocol analysis (Austin & Delaney, 1998; Critchfield & Epting, 1998; Hayes, White & Bissett, 1998). However, such analysis is difficult given the considerable amount of data that a protocol analysis typically generates (imagine having to analyze and graphically represent several hours of sessions on a moment-to-moment basis). Because of these difficulties, a common practice is to present cumulative frequencies listed by response category. This method for reducing complexity has been used with animal data, in which a small sample of data is typically offered out of a massive pool. However, it eliminates much of the variability as well as particular sequences of responding, which are arguably crucial features in analyses of the interactions among different behaviors (e.g., Critchfield, Tucker & Vuchinich, 1998). For example, two subjects that produce equal numbers of reports from specific categories might emit them in very different sequences. Furthermore,

it is difficult to determine the relation between verbal reports and successful task performance by considering only cumulative frequencies. Without a detailed view of the results, a proper analysis of the effects of verbal reports may not be feasible, so presenting frequencies sorted by categories may hinder the analysis (for further information on the analysis of protocol analysis data, see Wallander, 2001).

Addressing These Considerations: Some Proposals

Given the considerations described in the foregoing section, we can venture a number of suggestions as to why so few studies have been conducted using protocol analysis. Specifically, the technique is complex and requires considerable time and effort, there are no clear guidelines on how to implement the method, and the conclusions that can be drawn are limited. However, we believe that it is possible to address these considerations, and more importantly, that it is through empirical research and the accumulation of data that this objective should be achieved, and not through theoretical elaborations. The following section describes three contributions which we hope will facilitate this crucial empirical work: the use of the "silent dog" method, the use of computerized tools for the analysis of the data, and the development of standard procedures for the implementation of protocol analysis.

The "Silent Dog" method

In his review of Ericsson and Simon's work, Hayes (1986) suggested that additional methodological measures should be implemented when using protocol analysis to determine whether the performance of participants is governed by covert self-instructions, and the concurrent talk is functionally equivalent to these self-instructions. These measures were termed the "silent dog" method after one of Sherlock Holmes' famous cases[1]. The method consists of three different experimental controls that are describe below (for further detail, see Barnes-Holmes, Hayes & Dymond, 2001; Hayes, White & Bisset, 1998):

1. *No differences are found due to concurrent think-aloud*: if verbal reports are functionally identical to private self-instructions, then subjects are saying aloud what they are already saying privately. Thus, saying out loud what they are already saying to themselves should not affect task performance. In other words, performance on a task should be similar whether participants are required to think aloud or not.
2. *Disrupting verbal behavior must affect subjects' performances*: a lack of effect on performance does not necessarily mean that verbal reports are equivalent to self-instructions: it could mean that they are irrelevant to the task. To avoid this problem, verbal behavior is interrupted (e.g. by asking subjects to say out loud what they thought about over the last minute). Self-instructions should not be present during such disruption; therefore, if disruption affects task performance, the lack of differences cannot be attributed to the irrelevance of verbal behavior.
3. *Presenting other subjects with the protocols should affect their performances in a consistent and replicable manner*: the protocols should be presented to other participants

about to engage in the experiment, indicating that they should consider the material. This must alter the performance of the new subjects in a consistent and replicable manner, related to the content of the protocols. This control demonstrates that the content of the protocols itself is task-relevant.

In summary, the silent dog method uses three controls to ensure the functional similarity between the verbal reports of subjects required to verbalize their own thoughts, and the self-instructions that are controlling their responding: the comparison between conditions with and without concurrent think-aloud, the disruption of verbal behavior, and the replication of the effects of verbal reports across subjects. If the three requirements are met and the expected pattern of results is obtained, it can be concluded that "a) the behavior is in part governed by concurrently available rules; and that b) the lack of a difference between performance with and without concurrent talk-aloud is explained by the functional similarity of the rules present in the two conditions. That is, the self-rules formulated in the silent condition and the overt verbalizations in the talk-aloud condition are functionally the same" (Barnes-Holmes, Hayes & Dymond, 2001, p. 137). If this pattern is not obtained, then it is as though the "silent dog" has "barked" and we can therefore make no inferences about the status of the recorded verbalizations.

Certainly, the use of the "silent dog" controls increases the complexity of the experimental procedures, and requires more subjects than might otherwise be required. However, we believe that it specifically addresses the theoretical considerations raised about the validity of the verbal reports obtained, and, critically, suggests an experimental analysis of self-instructions. Regarding the validity of the data, in the "silent dog" method the correspondence between referent and reports is an issue of demonstrating a pattern of empirical results (no differences due to the "thinking aloud" requirement, alterations in performance caused by the disruption of verbal behavior, and the replication of the effect of verbal reports using new subjects). This kind of validation has been termed "empirical validation" and is to be preferred whenever possible, as it does not rely on either underlying assumptions or theoretical elaborations (Critchfield & Epting, 1998; Critchfield, Tucker & Vuchinich, 1998).

Furthermore, in what probably constitutes the most important strength of the "silent dog" method, the procedure permits an experimental analysis of self-instructions and not merely a correlational one. Although a correlational strategy is also used at a first stage, and relations between verbal reports and other behaviors are obtained, these self-reports are used to affect the performance of new subjects. Self-reports are now the independent variable and are used to produce definite patterns of responding; therefore, verbal behavior is brought under explicit and replicable experimental control (Taylor & O'Reilly, 1997).

Analysis of the data

Regarding categorization, we have developed a computerized program to assist in the process[2]. The program, written in Visual Basic, is based on previous work by Crutcher, Ericsson & Wichura (1994), and is composed of three modules:

1. In the first module, protocols are added to a database by either typing with the keyboard or importing a word-processor file. Additional information about each protocol can be included (for example the arrangement of stimuli presented in that specific protocol).
2. The second module is used to assign the protocols to the different categories devised (see Figure 1). To perform this process, the computer randomly selects one of the protocols in the database and presents it on screen (together with any additional information and the different coding categories), and the rater assigns the protocol to one of the categories by clicking in the corresponding button.
3. Finally, the third module presents the data from the categorization process to the experimenter, calculates some basic descriptive statistics, and (because the program allows for multiple coders to work in the same computer), calculates agreement between different coders.

The categorization process is eased and reduces fatigue because raters are focused on individual protocols one at a time and can work at their own pace. Furthermore, validity is increased because information can be included to help raters and because the random selection of protocols eliminates the problem of raters assigning categories

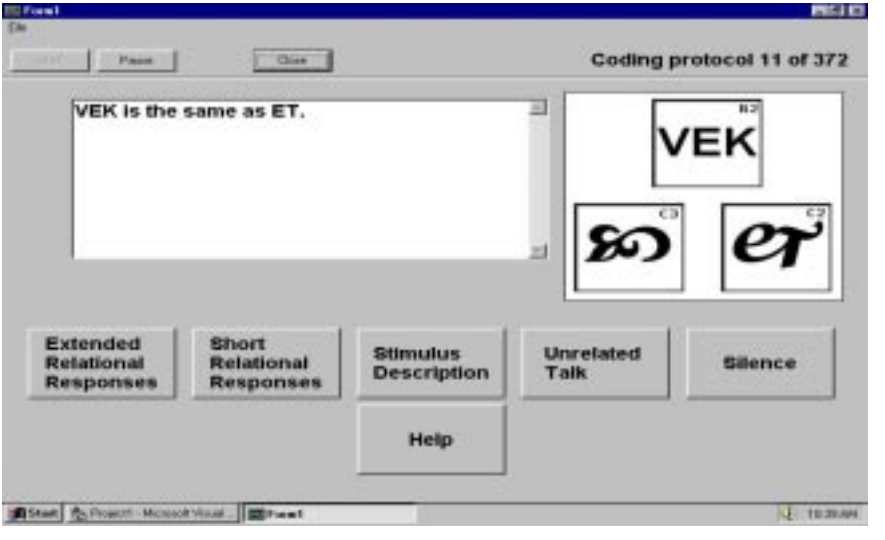

Figure 1. Capture of the program to assist raters in categorizing the verbal protocols of experimental subjects. Protocols are presented at the top of the screen, additional information (in this case, the arrangement of stimuli) is displayed in the right side, and the different categories are presented at the bottom of the screen.

based on previous selections, rather than on the protocol itself. The semi-automated encoding process reduces greatly the effort of categorizing protocols and increases the consistency of the process, as other authors have demonstrated in the cognitive field (Bashkar & Simon, 1977; Crutcher, 1994; Crutcher, Ericsson & Wichura, 1994; Fisher, 1988; Sanderson, James & Seidler, 1989).

Computer-based procedure

Regarding the lack of a sound methodological procedure for implementing protocol analysis within the experimental analysis of behavior, we are conducting some empirical work focused on the development of such a procedure. The procedure we have used is based in some recommendations about the collection of verbal reports (Critchfield & Epting, 1998; Critchfield, Tucker & Vuchinich, 1998; Gomez & Luciano, 2000; Zettle & Young, 1987).

In this procedure, the experimenter provides a general introduction about the task to the participants, and then leaves the experimental room. Subjects are seated in front of a computer equipped with a table-top microphone and are provided with detailed instructions about the task and about the "think aloud" procedure through screen messages. Next, they are exposed to some "think aloud" practice exercises (e.g., using a simple identity matching task), and they are prompted to think aloud through a written reminder that appears in the computer screen: "REMEMBER TO THINK ALOUD". After completing practice exercises, participants are exposed to the experimental task, while are reminded to keep "thinking aloud" through a message in the computer screen: "REMEMBER TO THINK ALOUD" that is presented in regular intervals. Finally, disrupting activities are also presented during the task using on-screen messages.

The main feature of this procedure is that computers and other electronical devices eliminate the variability that might be caused by the presence of the experimenter during the experiment: instructions, practice exercises and reminders to keep thinking aloud are presented by the computer through on-screen messages. By serving as an audience, the experimenter can affect the verbal behavior of the participants, as argued and demonstrated by many authors (Barrett, Deitz, Gaydos & Quinn, 1987; Williams, 1985). Furthermore, social interaction is kept to a minimum, an issue that has been specifically recommended when studying verbally-governed behavior (Austin & Delaney, 1998; Gomez & Luciano, 2000; Zettle & Young, 1987).

In developing this procedure, we have included verbal disruptions (see the section about the use of the "silent dog" controls above). While they should disrupt verbal behavior, these disruptions should not be incompatible with the task (e.g. in the case of pressing a button to obtain points, reciting letters of the alphabet would disrupt verbal behavior but is not incompatible with button pressing, whereas writing on a paper would not only disrupt verbal behavior but would also be clearly incompatible with the ongoing task). If the disruption is incompatible with the ongoing task, effects on performance could not be attributed exclusively to the disruption of verbal behavior. Our studies have employed a variety of activities for verbal disruptions. In one ongoing study, many different activities were presented to participants: counting forward or

backwards, adding three numbers, multiplying two numbers, forming words from syllables, reciting words backwards, and repeating a word many times. Some of them produced almost no disruption, whereas others were incompatible with task performance. In later experiments, we used counting forward and reciting letters of the alphabet as disrupting activities. These two activities showed similar levels of difficulty and disrupted self-talk to the same extent, thus achieving the two requirements indicated above.

Conclusion

The main objective of the current paper was to address some of the limitations of protocol analysis as a method for the experimental analysis of verbal control. Protocol analysis, as a tool for the study of private verbally-governed behavior (and more specifically, for assesing the function of specific instances of verbal behavior) produced much interest in the late nineties, in the form of conceptual papers (Austin & Delaney, 1998; Barnes-Holmes, Hayes & Dymond, 2001; Critchfield & Epting, 1998; Hayes, 1986; Hayes, White & Bissett, 1998; Hayes, Zettle & Rosenfarb, 1989; Potter, 1999) as well as empirical studies (Austin & Mawhinney, 1999; Dixon & Hayes, 1998; Potter, Huber & Michael, 1997; Rehfeldt & Dixon, 2000; Rehfeldt, Dixon, Hayes & Steele, 1998; Rehfeldt & Hayes, 2000; Wulfert, Dougher & Greenway, 1991; Wulfert, Greenway & Dougher, 1994). This interest, however, has decreased and, currently, there is no agreed systematic empirical approach.

We have considered the lack of clear methodological guidelines for applying protocol analysis as an important barrier to its extension to the analysis of verbal behavior. Although Ericsson and Simon suggested a standardized procedure, their suggestions were in some ways inconsistent with the objectives of a behavioral approach and so their techniques had to be adapted. In this adaptation, different researchers used different procedures without clearly analyzing their impact on their results (see Normand, 2001). Other issues affecting the validity of verbal reports as data for an experimental analysis have also been considered, such as the relation between referents and verbal reports, and the need for a thoroughgoing experimental analysis.

To address these issues, we have made some recommendations based on ongoing research. Specifically, the use of the “silent dog” method (Hayes, 1986; Hayes, White & Bissett, 1998) is aimed at empirically demonstrating verbal-governance of task performance, and functional correspondences between self-reports and private self-instructions. The main feature of the “silent dog” method is that it relies heavily on empirical demonstrations and not on conceptual assumptions; the validity of self-report data is established through the demonstratation of patterns of empirical results, rather than through theoretical ellaborations (Hayes, White & Bissett, 1998). Indeed, the method provides an experimental analysis of self-instructions, in contrast to the correlational analyses usually performed when studying verbal behavior (Barnes, 1989). Although some researchers have been sensitive to these issues (Rehfedlt & Dixon, 2000; Taylor & O’Reilly, 1997), our research constitutes the first attempt to fully implement the different methodological measures described in the “silent dog” method to the area of protocol analysis. Furthermore, and as part of these studies, we have targeted the

development of a robust methodological procedure based on the recommendations of different commentators. As described above, a computerized system avoids the presence of the experimenter during the experiment, minimizes audience control, and assists in the analysis of the data.

A lot of work remains to be done, however, and the current article constitutes a preliminary attempt to facilitate a systematic experimental analysis of concurrent verbal behavior. As Navarick, Bernstein & Fantino (1990) stated, advances in the experimental analysis of human behavior depends largely on innovations and refinements in methodology. We hope that the current paper may inspire other researchers to conduct further studies in the area. Protocol analysis is a complex technique, its implementation increases the complexity and resources needed, and it is not the only tool available to behavior analysis for the analysis of by self-generated verbal behavior (see Barnes-Holmes, Hayes & Dymond, 2001). Indeed, it should not be applied to every behavioral problem, and a careful application of protocol analysis relies on recognizing its benefits as well as its limitations (Austin & Delaney, 1998; Critchfield & Epting, 1998). However, we have argued that the technique bears promise in the analysis of private verbal behavior, and we believe that protocol analysis is a tool that deserves a place in the experimental analysis of human complex behavior.

Notes

[1] In the Silver Blaze adventure (Doyle, 1892), a murderer is discovered because the guarding dog didn't bark, indicating that the criminal was its owner.

[2] The program can be obtained from the first author. For additional information about the use of Visual Basic in the experimental analysis of human behavior, see Cabello, Barnes-Holmes, Stewart & O'Hora (under submission).

References

Austin, J. & Delaney, P.F. (1998). Protocol analysis as a tool for behavior analysis. *The Analysis of Verbal Behavior, 15,* 41-56.

Austin, J. & Mawhinney, T.C. (1999). Using concurrent verbal reports to examine data analyst verbal behavior. *Journal of Organizational Behavior Management, 18,* 61-81.

Barnes, D. (1989). Behavior-behavior analysis, human schedule performance, and radical behaviorism. *The Psychological Record, 39,* 339-350.

Barnes-Holmes, D., Hayes, S.C. & Dymond, S. (2001). Self and self-directed rules. In S.C. Hayes, D. Barnes-Holmes & B. Roche (Eds.). *Relational frame theory: A post-skinnerian account of human language and cognition*. New York: Plenum Press.

Barnes, D. & Keenan, M. (1989). Instructed human fixed-interval performance: the effects of the experimental settings. *The Psychological Record, 39,* 351-364.

Barnes, D. & Keenan, M. (1993). Concurrent activities and instructed human fixed-interval performance. *Journal of the Experimental Analysis of Behavior, 59,* 501-520.

Baron, A., Kaufman, A. & Stauber, K.A. (1969). Effects of instructions and reinforcement feedback on human operant behavior maintained by fixed-interval reinforcement. *Journal of the Experimental Analysis of Behavior, 12*, 701-712.

Barrett, D.M., Deitz, S.M., Gaydos, G.A. & Quinn, P.C. (1987). The effects of programmed contingencies and social conditions on response stereotipy with human subjects. *The Psychological Record,*

37, 489-505.

Bhaskar, R. & Simon, H.A. (1977). Problem solving in semantically rich domains: An example from engineering thermodynamics. *Cognitive science, 1*, 193-215.

Chase, P.N. & Danforth, J.S. (1991). The role of rules in concept learning. In L.J. Hayes & P.N. Chase (Eds.). *Dialogues on verbal behavior*. Reno, NV: Context Press.

Critchfield, T.S. & Epting, L.K. (1998). The trouble with babies and the value of bathwater: complexities in the use of verbal reports as data. *The Analysis of Verbal Behavior, 15*, 65-74.

Critchfield, T.S., Tucker, J.A. & Vuchinich, R.E. (1998). Self-reports methods. In K. Lattal & M. Perone (Eds.). *Handbook of research methods in human operant behavior*. New York: Plenum Press.

Crutcher, R.J. (1994). Telling what we know: the use of verbal report methodologies in psychological research. *Psychological Science, 5*, 241-245

Crutcher, R.J., Ericsson, K.A. & Wichura, C.A. (1994). Improving the encoding of verbal reports by using MPAS: A computer-aided encoding system. *Behavior research Methods, Instruments & Computers, 26, 167-171.*

Dixon, M.R. & Hayes, L.J. (1998). Effects of differing instructional histories on the resurgence of rule-following. *The Psychological Record, 48*, 275-292.

Doyle, A.C. (1892). The memoirs of Sherlock Holmes. Reprinted in *Sherlock Holmes: Complete novels and stories* (vol. 1). New-York: Bantam Books.

Dube, W.V., McIlvane, W.J., Callahan, T.D. & Stoddard, L.W. (1993). The search for stimulus equivalence in nonverbal organisms. *The Psychological Record, 43*, 761-778.

Dymond, S. & Critchfield, T.S. (2001). Neither dark age nor renaissance: Research and authorship trends in the experimental analysis of human behavior (1980-1999). *The Behavior Analyst, 24*, 241-253.

Ericsson, K.A. & Simon, H.A. (1980). Verbal reports as data. *Psychological Review, 87*, 215-251.

Ericsson, K.A. and Simon, H.A. (1993). *Protocol analysis: Verbal reports as data (rev. ed.)*. Cambridge, MA: MIT Press.

Fisher, C. (1988). Advancing the study of programming with computer-aided protocol analysis. In G. Olson, E. Soloway & S. Sheppard (Eds.). *Empirical studies of programmers: 1987 workshop*. Norwood, NJ: Ablex.

Gómez, I. & Luciano, M.C. (2000). Autocontrol a través de reglas que alteran la función [Self-control through altering function rules]. *Psicothema, 12*, 418-425.

Harzem, P., Lowe, C.F. & Bagshaw, M. (1978). Verbal control in human operant behavior. *The Psychological Record, 28*, 405-423.

Hayes, S.C. (1986). The case of the silent dog: Verbal reports and the análisis of rules. A review of K. Anders Ericsson and Herbert A. Simon, Protocol analysis: Verbal reports as data. *Journal of the Experimental Análisis of Behavior, 45*, 351-363.

Hayes, S.C. (1987). Upward and downward continuity: It's time to change our strategic assumptions. *Behavior Analysis, 22*, 3-6.

Hayes, S.C. (1989a). Nonhumans have not yet shown stimulus equivalence. *Journal of the Experimental Análisis of Behavior, 51*, 385-392.

Hayes, S.C. (1989b). *Rule-governed behavior: Cognition, contingencies and instructional control*. Reno, NV: Context Press.

Hayes, S.C. & Brownstein, A.J. (1986). Mentalism, behavior-behavior relations and a behavior analytic view of the purposes of science. *The Behavior Analyst, 9*, 175-190.

Hayes, S.C. & Hayes, L.J. (1992). Verbal relations and the evolution of behavior análisis. *American Psychologist, 47*,1383-1395.

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.

Hayes, S.C., Zettle, R.D. & Rosenfarb, I. (1989). Rule following. In S.C. Hayes (Ed.). *Rule-governed behavior: Cognition, contingencies, and instructional control*. Reno, NV: Context Press.

Hyten, C. & Reilly, M.P. (1992). The renaissance of the experimental analysis of human behavior. *The Behavior Analyst, 15*, 109-114.

Holland, J.G. (1958). Counting by humans on fixed-ratio schedules of reinforcement. *Journal of the Experimental Analysis of Behavior, 1*, 179-181.

Leander, J.D., Lippman, L.G. & Meyer, M.M. (1968). Fixed interval performance as related to instructions and subjects verbalizations of the reinforcement contingency. *The Psychological Record, 18*, 469-474.

Lowe, C.F. (1979). Determinants of human operant behavior. In M.D. Zeiler & P. Harzem (Eds.). *Advances in the analysis of behavior: Vol. 1. Reinforcement and the organization of behavior*. Chichester: Wiley.

Skinner, B.F. (1938). *The behavior of organisms*. New York: Appelton-Century-Crofts.

Lowe, C.F., Harzem, P. & Hughes, S. (1978). Determinants of operant behavior in humans: some differences from animals. *Quarterly Journal of Experimental Psychology, 30*, 373-386.

Luciano, M.C. (1993). La conducta verbal a la luz de recientes investigaciones: su papel sobre otras conductas verbales y no verbales. [Verbal behavior according to recent research: Its role on verbal and nonverbal behavior.] *Psicothema, 5*, 351-374.

Matthews, B.A., Shimoff, E., Catania, A.C. & Sagvolden, T. (1977). Uninstructed human responding: sensitivity to ratio and interval contingencies. *Journal of the Experimental Analysis of Behavior, 27*, 453-467.

Navarick, D.J., Bernstein, D.J. & Fantino, E. (1990). The experimental analysis of human behavior. *Journal of the Experimental Analysis of Behavior, 54*, 159-162.

Newell, A. y Simon, H.A. (1972). *Human problem solving*. Englewood Cliffs, Prentice-Hall

Nisbett, R.E. y Wilson, T.D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological review, 84*, 231-259

Normand, M. (2001). Collecting concurrent verbal reports: Some considerations and methodological refinements. In Miguel, C. (Chair). *The study of verbal behavior and private events: Some methodological considerations*. Symposium conducted at the meeting of the Association for Behavior Analysis, New Orleans.

Payne, J.W. (1994). Thinking aloud: insights into information processing. *Psychological Science, 5*, 241-256

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.

Potter, B. (1999). Some additional considerations of protocol analyses. *The Analysis of Verbal Behavior, 16*, 57-61

Potter, B., Huber, S. & Michael, J. (1997). The role of mediating verbal behavior in selection-based responding. *The Analysis of Verbal Behavior, 14*, 41-56.

Rehfeldt, R.A. & Dixon, M.R. (2000). Investigating the relation between self-talk and emergent stimulus relations. *Experimental Analysis of Human Behavior Bulletin, 18*, 28-29.

Rehfeldt, R.A., Dixon, M.R., Hayes, L.J. & Steele, A. (1998). Stimulus equivalence and the blocking effect. *The Psychological Record, 48*, 647-664.

Rehfeldt, R.A. & Hayes, L.J. (2000). The long-term retention of generalized equivalence classes. *The Psychological Record, 50*, 405-428.

Russo, J.E., Johnson, E.J. & Stephens, D. (1989). The validity of verbal protocols. *Memory & Cognition, 17*, 759-769.

Sanderson, P.M., James, J.M. & Seidler, K.S. (1989). SHAPA: An interactive software environment for protocol analysis. *Ergonomics, 32*, 1271-1302.

Shimoff, E. (1986). Post-session verbal reports and the experimental analysis of behavior. *The Analysis of Verbal Behavior, 4*, 19-22.

Shimoff, E., Catania, A.C. & Matthews, B.A. (1981). Uninstructed human responding: sensitivity of low rate performances to schedule contingencies. *Journal of the Experimental Analysis of Behavior, 36*, 207-220.

Skinner, B.F. (1938). *The behavior of organisms*. New York: Appelton-Century-Crofts.

Skinner, B.F. (1953). *Science and human behavior*. New York: The Free Press.

Sidman, M. & Tailby, W. (1982). Conditional discrimination versus matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37*, 5-22.

Taylor, I. & O´Reilly, M.F. (1997). Toward a functional analysis of private verbal self-regulation. *Journal of Applied Behavior Analysis, 30*, 43-58.

Wallander, R. (2001). The classification and graphical representation of concurrent verbal reports: Methodological considerations in the use of protocol analysis. In Miguel, C. (Chair). *The study of verbal behavior and private events: Some methodological considerations*. Symposium conducted at the meeting of the Association for Behavior Analysis, New Orleans.

Watson, J.B. (1920). Is thinking merely the action of language mechanisms? *British Journal of Psychology, 11*, 87-104

Weiner, H. (1970). Instructional control of human operant responding during extinction following fixed-ratio conditioning. *Journal of the Experimental Analysis of Behavior, 13*, 391-394.

White, P.A. (1988). Knowing more about what we can tell: ''Introspectice access'' and causal report accuracy 10 years later. *British Journal of Psychology, 79*, 13-45.

Williams, R.A. (1985). Schedule performance and verbal behavior under solitary and social training conditions. *Dissertation Abstracts International-B, 46*, 687.

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.J. (1994). Third-order equivalence classes. *The Psychological Record, 44*, 411-439.

Zettle, R.D. & Young, M.K. (1987). Rule-following and human operant responding: conceptual and methodological considerations. *The Analysis of Verbal Behavior, 5*, 33-39.

Received September 30, 2002
Final acceptance December 13, 2002