JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 1990, 53, 321-344 NUMBER 3 (MAY)

VERBAL SELF-REPORTS OF DELAYED MATCHING TO SAMPLE BY HUMANS

THOMAS S. CRITCHFIELD AND MICHAEL PERONE

WEST VIRGINIA UNIVERSITY

Undergraduates participated in two experiments to develop methods for the experimental analysis of self-reports about behavior. The target behavior was the choice response in a delayed-matching-to-sample task in which monetary reinforcement was contingent upon both speed and accuracy of the choice. In Experiment 1, the temporal portion of the contingency was manipulated within each session, and the presence and absence of feedback about reinforcement was manipulated across sessions. As the time limits became stricter, target response speeds increased, but accuracy and reinforcement rates decreased. When feedback was withheld, further reductions in speed and reinforcement occurred, but only at the strictest time limit. Thus, the procedures were successful in producing systematic variation in the speed, accuracy, and reinforcement of the target behavior. Experiment 2 was designed to assess the influence of these characteristics on self-reports. In self-report conditions, each target response was followed by a computer-generated query: "Did you earn points?" The subject reported by pressing "Yes" or "No" buttons, with the sole consequence of advancing the session. In some cases, feedback about reinforcement of the target response followed the reports; in other cases it was withheld. Self-reports were less accurate when the target responses occurred under greater time pressure. When feedback was withheld, the speed of the target response influenced reports, in that the probability of a "Yes" report increased directly with the speed of accurate target responses. In addition, imposing the self-report procedure disrupted target performance by reducing response speeds at the strictest time limit. These results allow investigation of issues in both behavioral and cognitive psychology. More important, the overall order in the data suggests promise for the experimental analysis of self-reports by human subjects.

Key words: self-reports, verbal behavior, stimulus control, temporal contingencies, feedback, delayed matching to sample, response speed, button press, humans

---

Among the earliest laboratory investigations in psychology were introspective studies, in which human subjects observed and described their own private events. Perhaps not coincidentally, one of the earliest debates in psychology as a formal discipline concerned the theoretical status and methodological utility of self-observation and corresponding verbal self-reports (Boring, 1953; Danziger, 1980; Lyons, 1986). Although psychology has changed substantially since the heyday of introspection, verbal self-reports still play a significant role in modern theoretical discussions (e.g., Ericsson & Simon, 1984; Giorgi, 1975; Nisbett & Wilson, 1977; Perone, Galizio, & Baron, 1988) and continue to contribute to psychology's data base.

Probably the most common modern use of self-reports is as instrumentation. Numerous procedures rely on the reports to provide information about overt and covert responses not directly observed by experimenters. These procedures include surveys, many clinical and personality assessment strategies, concurrent verbalization (or "think-aloud") procedures, and various other techniques employed under the rubric of *protocol analysis* (e.g., see Ericsson & Simon, 1984). Postexperimental interviews, when used to shed light on a subject's performance during an experiment, also fall into this category (e.g., Case, Fantino, & Wixted, 1985; see also Lowe, 1979).

A second function of self-reports is as intervention. Clinical psychologists have made use of the fact that when clients report and

This report is based on part of the first author's doctoral dissertation at West Virginia University. The research was supported by grants from the University's Office of Academic Affairs and Research and Psychology Alumni Fund. We thank Stephen Menich and Alan Baron for access to their computer programs, Alan Baron for helpful discussions about the general research topic, and Karla Doepke for screening subjects and commenting on drafts of the manuscript. Correspondence may be sent to the first author at the Behavioral Pharmacology Research Unit, D-5-West, John Hopkins/Key Medical Center, 4940 Eastern Avenue, Baltimore, Maryland 21224, or to the second author at the Department of Psychology, West Virginia University, P.O. Box 6040, Morgantown, West Virginia 26506-6040.

maintain records of clinically important behavior, socially desirable changes often follow. Self-monitoring procedures derived from this discovery have been used extensively with a broad range of clients and behavior problems (e.g., see Nelson, 1977).

The two categories described above, instrumentation and intervention, encompass large numbers of investigations and applications. By contrast, psychology has paid less attention to the third role of self-reports, that of dependent variable. Relatively few studies have examined events that may influence an individual's reports of actions and personal characteristics under conditions in which the content of the reports can be corroborated. Even fewer have used laboratory conditions to examine the reports as behavior under potential environmental control.

Within the operant arena, several studies have investigated clinically relevant self-reports in natural settings (e.g., O'Farrell, Cutter, Bayog, Dentch, & Fortgang, 1984), and descriptions of response–consequence relations in the laboratory (e.g., Catania, Matthews, & Shimoff, 1982; Wasserman & Neunaber, 1986). More common, and more relevant for present purposes, are studies of self-reports by nonhumans (hereafter referred to as "animals"; e.g., Kramer, 1982; Maki, Moe, & Bierley, 1977; Reynolds, 1966; Ziriax & Silberberg, 1978).

The typical procedure requires a pigeon to make a *reporting response,* implicitly or explicitly defined, under discriminative control of another response, which we will refer to as the *target response* (drug-discrimination studies employ analogous procedures using drug states as the target event; see Lubinski & Thompson, 1987). A case in point is the work of Shimp, who arranged contingencies to produce a bimodal distribution in the interresponse times (IRTs) of pigeons (Shimp, 1981, 1982, 1983). The IRTs served as target performance. Matching-to-sample probes, in which a just-completed long (or short) IRT served as the sample stimulus, were used to assess the degree to which the birds could report preceding behavior. Such studies illustrate similarities between self-reports and the remembering and reporting of external stimuli (e.g., matching-to-sample procedures using key colors, rather than responses, as the sample stimuli).

We located only one human operant study using procedures similar to those utilized in animal self-report investigations (but see de Freitas Ribeiro, 1989; Rabbitt, 1966, 1979). Hefferline and Perera (1963) used as the target response a tiny, left-hand thumb twitch that could be electromyographically recorded, but about which subjects apparently were unaware. Right-hand button presses within 2 s after a thumb twitch earned 2 cents. In a baseline session, button presses were no more likely following the thumb twitch than at other times. However, by introducing, then fading out, a brief auditory stimulus after each thumb twitch, Hefferline and Perera were able to bring button presses under discriminative control of thumb twitches. In effect, the button press served as a report of the occurrence of a thumb-twitch target response (which, interestingly, subjects apparently still could not describe).

The general purpose of the present study was to extend the experimental analysis of self-reports by human subjects. We had two specific objectives. First, we sought to collect some detailed descriptive data about environmental events that may influence self-report content and accuracy. Hefferline and Perera (1963), and many of the studies using animals, presented only global measures of self-report accuracy. Second, we wished to consider the possibility of reciprocal interaction between effects of target behavior on self-reports and effects of self-reports on target behavior.

Experiment 1 was conducted to establish, in the absence of self-reports, some characteristics of the target response that subjects reported on in Experiment 2. For this purpose we modified a delayed-matching-to-sample (DMTS) procedure described by Baron and Menich (1985a, 1985b). In this procedure, a brief display of a compound sample stimulus is followed by a delay interval during which the sample is not visible. Two comparison stimuli, one matching an element of the compound sample and one randomly generated, then appear. To earn money, the subject must identify the matching stimulus within a time limit. Baron and Menich demonstrated that response speeds are sensitive to the temporal contingencies, but not maximally so, with extreme time limits leading to performance decrements. In addition, the accuracy of the matching response bears no necessary relation to its speed. Thus, the procedure can produce a target performance with a number of measurable characteristics that might influence a self-report.

Our self-report procedure, employed in Experiment 2, involved the insertion, after each DMTS trial, of a yes–no query about the preceding choice of a comparison stimulus. The query was presented on a computer screen and reports consisted of button presses, so that no direct interaction between subject and experimenter was necessary. The reports were restricted to a binary choice so that we could place our analytical emphasis on antecedent variables affecting report content rather than on interpreting the referents of complex verbal utterances. Finally, we arranged for the target and reporting responses to occur sequentially, so that the occurrence of one could not directly interfere with emission of the other (e.g., see Epstein, Miller, & Webster, 1976).

EXPERIMENT 1

Baron and Menich (1985a, 1985b) found that response speeds in a DMTS procedure were systematically related to the stringency of a conjunctive speed contingency. The purpose of Experiment 1 was to investigate two modifications of Baron and Menich's procedure before arranging for subjects to report their DMTS performance in Experiment 2. First, Baron and Menich manipulated time limits on a between-session basis. Because we wished to change the limits within sessions in signaled blocks of trials, Experiment 1 was conducted to determine whether response speeds would remain differentiated in this type of arrangement. Second, although the tangible reinforcer in the Baron and Menich studies was money (subjects were paid at the end of their participation), the immediate consequence of each trial was a feedback message indicating the trial outcome. We wished to omit experimenter-provided feedback during one self-report phase of Experiment 2. Thus, an additional goal of Experiment 1 was to determine some of the characteristics of human speeded DMTS performance in the absence of trial-by-trial feedback.

METHOD

Subjects

Two male undergraduate students, 19 and 20 years old, volunteered to participate in a laboratory experiment on "Decision-Making Under Time Constraints." Both reported that they were in good health, and their responses to the digit-symbol substitution, digit span, and general information subscales of the Wechsler Adult Intelligence Scale—Revised (Wechsler, 1981) ranked them at or above average for the general population. The men signed an informed consent agreement explaining the payment procedure, which involved an hourly wage of $2.00 supplemented by earnings during experimental sessions. Payment of the hourly wage (but not session earnings) was contingent on completing all scheduled sessions. Payment occurred at the end of each man's participation, but a written record of earnings was provided at the end of each work day. On average, the men earned about $4.50 per hour.

Apparatus

The experiment took place in a room (3 m by 2 m) containing a table, chair, and the experimental apparatus. A window in the room was covered during experimental sessions. A response console (51 cm wide by 20 cm high by 60 cm long) was on the table, and a video monitor with a 12-in. (31-cm) green screen sat atop the console. The console's sloping front panel contained a small metal button centered 7 cm from the top of the panel and flanked by two white lamps (each 5 cm away); four 3-cm, back illuminable, round response keys were arranged horizontally 12 cm from the bottom of the panel. A small blue lamp was located 3 cm above each key. Neither the illuminable response keys nor the blue lamps operated during Experiment 1. At each side of the console, 34 cm from the work panel's front edge and raised 4 cm from the table top, was a telegraph key extending toward the subject from a small metal casing attached to the console. Each key moved about 1 cm when pressed.

Extraneous sounds were masked by white noise provided through stereo headphones; a wall fan provided additional masking noise. A microcomputer was used to control experimental events and collect the data.

Procedure

Before each trial, the message "HOLD SIDE KEYS DOWN" appeared centered inside a "work area" defined by a frame (17 cm wide by 7 cm high) in the upper two thirds of the video screen. By depressing both telegraph keys the subject produced a compound sample stimulus consisting of three geometric stimuli, displayed about 2 cm apart in the center of the work area for 1.5 s. The stimuli are described below. After an 8-s delay interval, during which the work area on the screen was blank, two comparison stimuli appeared, one 2 cm to the left and one 2 cm to the right of the center of the work area. One of the comparison stimuli was randomly generated and the other was identical to an element in the sample display. Release of the telegraph key on the same side as the matching stimulus was counted as correct. When a correct response occurred within the scheduled time limit (there were four in a session; see below), the subject earned 2 points (1 point = 1 cent).

Each element of the compound sample stimulus, and each comparison stimulus, consisted of a 4 by 3 matrix of rectangular cells, of which as few as 2 or as many as 12 could be illuminated (see Baron & Menich, 1985b). A stimulus could be as large as 10 by 13 mm, depending on how many cells were illuminated. As in Baron and Menich's research, the stimuli were drawn randomly on each trial from a pool of about 4,000 shapes, with the restriction that the sample compound and the nonmatching comparison stimulus consist of four unique stimuli. Also determined randomly on each trial was the element of the sample compound (left, center, or right) to be reproduced as a comparison stimulus and the side of the comparison display (left or right) on which the matching stimulus appeared.

Selection of a comparison stimulus produced feedback messages as described below. Subsequently, the work area on the screen was cleared for an intertrial interval that lasted a minimum of 2 s but otherwise was adjusted on each trial to maintain a relatively constant pace of about three trials per minute. The intertrial interval ended with the reappearance of the message "HOLD SIDE KEYS DOWN," but the next trial did not begin until the subject pressed the telegraph keys to produce the sample stimuli.

The men worked for a minimum of 10 hr per week, normally in sets of two sessions separated by a 10-min break during which the men could leave the work room. Sessions were divided into four blocks lasting 40 trials or 16 min, whichever came first. The blocks were distinguished by the time limit placed on selection of a comparison stimulus (no limit and 2,000, 1,000, and 500 ms). Sessions always progressed from the least to the most stringent time limit, and the blocks were separated by 40-s intermissions, but otherwise the time limits were not explicitly signaled. After the fourth block the screen displayed the subject's earnings for the session.

During feedback conditions, selection of a comparison stimulus produced three messages, presented simultaneously inside the work area on the screen. The first stated "You selected this shape:" and was followed by a reproduction of the comparison stimulus just selected. The second message stated that the stimulus just selected was "Correct" or "Wrong," and the third indicated that the selection had been "Fast Enough" or "Too Slow" relative to the time limit in effect. If the response was both correct and fast enough, the white lamps flanking the small metal button on the console were lighted and a fourth message, "Press the silver button to collect points," also appeared in the work area. Pressing the button cleared the work area and produced a 2-s message stating "You earned 2 points," after which the white lamps darkened and the screen was blank for the intertrial interval. If the response selecting a comparison stimulus was wrong, too slow, or both, the feedback messages were accompanied by an instruction to "Press the silver button to go on." Doing so produced a 2-s message stating "You lost 0 points" (this wording was chosen because initially we considered using point penalties in some phases).

During no-feedback conditions, selection of a comparison stimulus produced the message "Thank you for choosing" plus a prompt to press the silver button. Pressing produced a 2-s message stating that the program was "About to go on . . . ."

The DMTS procedure required that both telegraph keys remain depressed throughout the sample and delay portions of the trial. If a key was released prior to the appearance of the comparison stimuli, the work area was cleared and the message "Illegal action! You released too soon" appeared for the remainder of the 18-s trial. If both keys were released upon appearance of the comparison stimuli, the message "Illegal action! You released both keys" appeared for the remainder of the trial.

The latency to select a comparison stimulus was measured in milliseconds using a machine-language subroutine; in pilot work, measurements were accurate to within plus or minus 1% of the obtained latency. Other session events were timed using interrupt-driven software clocks with a resolution of $\frac{1}{30}$ of a second (Perone, 1985).

Instructions

Before the first session, the men read printed instructions covering the following points (quotations indicate exact wording; other portions are paraphrased for brevity): (a) "The apparatus in front of you is used to study how people make choices and decisions. It is up to you to decide how to operate it to your best advantage." (b) Pressing down the side keys produces three sample shapes on the screen. (c) When the two "test" shapes appear, "your job is to decide which one matches. You can indicate your decision by releasing the key on the same side as the shape you have chosen." (d) "One of our interests is in how quickly you can release the correct key. To maximize your earnings your decision may have to be both correct and prompt." (e) "At the beginning of each session you will have unlimited time to choose the correct shape. Later in the session you will be required to choose within a time limit." (f) Correct, prompt choices earn points exchangeable for money. (g) "You should never release a side key until you have chosen a shape. If you do, the trial will begin again, wasting time in which you could be earning points." (h) If you release both keys on a trial, your response will be counted as wrong. (i) Sometimes messages or questions will follow your selection of a shape. The basic decision-making procedure remains the same regardless of whether anything happens after your selection. (j) During a session, "you may do what you like but remember that your pay depends on what you do. If you should go to sleep, for example, your earnings for the session could amount to nothing."

Design

The experiment consisted of two conditions arranged in A-B-A fashion. In the baseline phase, feedback messages followed each trial. In the second phase the no-feedback messages followed each trial. The third phase was a return to baseline. Table 1 shows the number of sessions in each phase.

Phases were changed when DMTS performance was stable over the most recent four sessions. Three performance variables (percentage of responses that were correct, fast enough to meet the time limit, and reinforced) were considered in each of the four time-limit conditions, for a total of 12 variable-condition combinations. For each such combination, the difference between the means of consecutive sets of two sessions was considered as a proportion of the four-session mean. Phases were changed when at least 11 of 12 proportions were less than .15. Retrospectively, in 69 of 72 cases (encompassing three phases for each of 2 subjects) the proportion was less than .15 and, in 58 of 72 cases, less than .10.

RESULTS

Analyses are based on the final four sessions of each condition. Latency data have been converted to speed scores (1,000/latency). For example, a latency of 500 ms converts to a speed score of 2.0, and a latency of 1,000 ms to a speed score of 1.0. Speed scores have several advantages over latency measures, not the least of which are normalization of distributions and ease of comparison with response rates (e.g., a speed score of 2.0 is roughly equivalent to a rate of 2 responses per second; see Baron, 1985).

Table 1 shows the number of trials at each time limit during the three experimental phases, summed across the last four sessions per phase. Although some variability is evident, the number of completed trials was always within 10% of the maximum of 160. Table 1 also lists mean choice speeds at the four limits, showing the degree of control by the temporal contingencies. Speeds in the no-limit and 2,000-ms conditions were not systematically different from each other, but otherwise the rank order of mean speeds corresponded to the stringency of the contingencies. Manipulating feedback did not affect this rank order; however, it did affect the absolute speed at the strictest limit: In the 500-ms condition, both men responded more slowly during the no-feedback phase.

Figure 1 shows the percentage of trials on which DMTS choice responses were correct, fast enough to meet the time limit, and reinforced, in each of the three phases. Each variable is expressed as a function of the time limit (converted to speed), with data shown separately for the feedback and no-feedback phases.

Table 1

Experiment 1: Number of sessions per phase, plus number of trials and mean DMTS choice speeds (1,000/latency) in the last four sessions per phase. Data are shown separately for each of the four time limit conditions. Numbers in parentheses are standard deviations.

	Phase		
	Feedback 1	No feedback	Feedback 2
Subject Y1			
Sessions	10	11	5
Trials			
No limit	147	150	157
2,000 ms	152	150	148
1,000 ms	151	154	149
500 ms	145	150	149
Mean choice speed			
No limit	0.95 (0.31)	1.21 (0.36)	1.15 (0.35)
2,000 ms	1.02 (0.27)	1.18 (0.38)	1.21 (0.35)
1,000 ms	1.36 (0.33)	1.34 (0.32)	1.55 (0.38)
500 ms	2.41 (0.54)	2.09 (0.67)	2.28 (0.52)
Subject Y2			
Sessions	11	12	7
Trials			
No limit	148	150	157
2,000 ms	151	146	153
1,000 ms	155	143	148
500 ms	148	143	146
Mean choice speed			
No limit	0.94 (0.33)	1.40 (0.39)	1.46 (0.39)
2,000 ms	0.99 (0.32)	1.37 (0.37)	1.52 (0.44)
1,000 ms	1.70 (0.78)	1.58 (0.46)	1.71 (0.61)
500 ms	3.29 (1.03)	2.17 (0.67)	3.19 (1.12)

The figure shows that increasing the stringency of the time limit had the effect of decreasing all three aspects of performance. Responding was most successful in meeting the conjunctive speed–accuracy contingencies in the no-limit and 2,000-ms conditions and least successful in the 500-ms condition. Intermediate values occurred when the time limit was 1,000 ms. The percentage of reinforced trials reflects combinations of these speed and accuracy effects.

Figure 1 also permits comparison of DMTS performance in the presence versus absence of feedback. The top panels show that removal of feedback had no systematic effect on the accuracy of DMTS responses, as compared to performance in the presence of feedback. The middle two panels show that the withdrawal of feedback reduced the percentage of trials with prompt responding in the 500-ms condition, but otherwise had no effect. Again, reinforcement frequency reflects a combination of speed and accuracy effects.

Discussion

Experiment 1 investigated two properties of responding on the DMTS task. Response speeds were found to be sensitive to the within-session manipulation of temporal contingencies, and the absence of feedback after each trial had only a limited effect on performance. In the former case, speeds were lowest in the most lenient conditions (no limit and 2,000 ms), and increased systematically as the time limits became stricter (1,000 ms and 500 ms). Although responding was fastest in the 500-ms condition, responses most often failed to meet the time limits in this condition as well. These results compare well with those of Baron and Menich (1985a, 1985b), who manipulated temporal contingencies on a between-session basis.

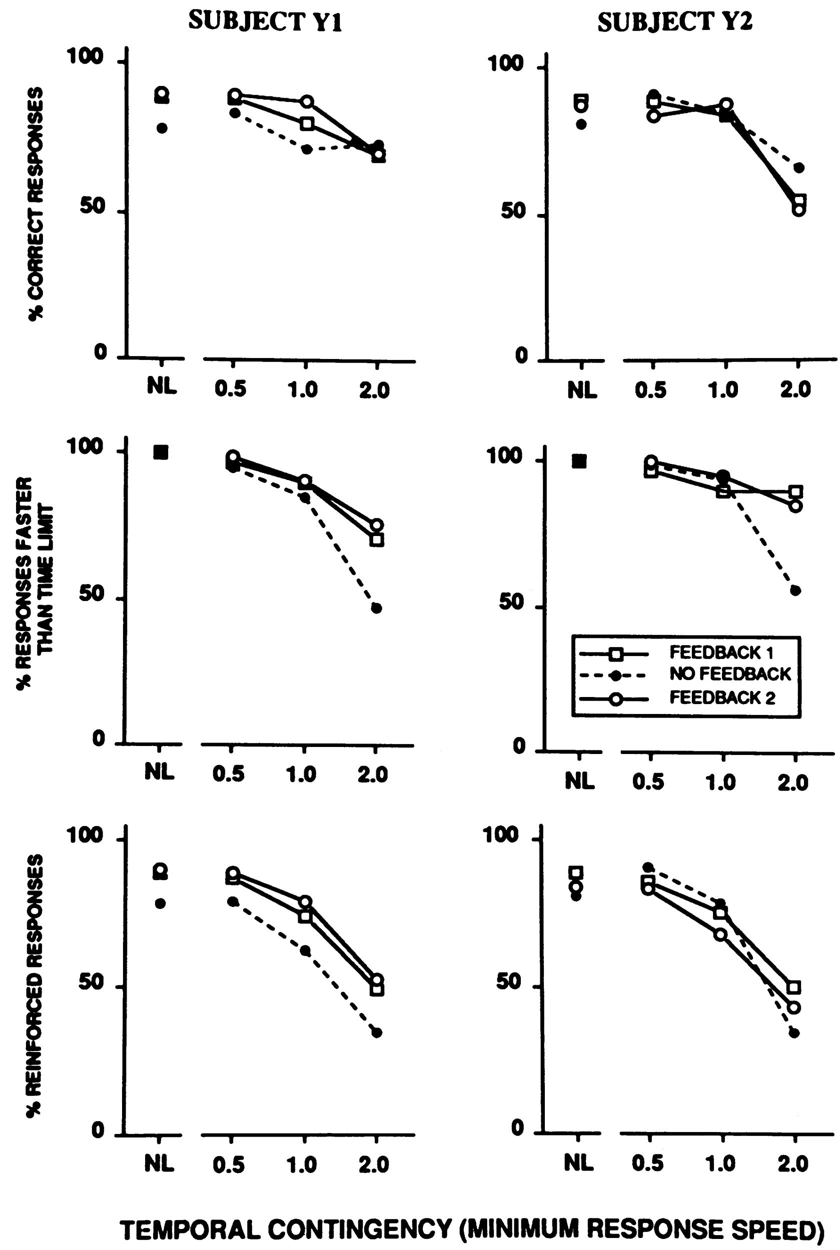

Fig. 1. Experiment 1: Percentage of DMTS responses that were correct, faster than the time limit, and reinforced, in the last four sessions per phase. Each variable is shown as a function of the time limit on responding. Time limits are expressed in terms of speed (1,000/latency): thus, 0.5 = 2,000 ms; 1.0 = 1,000 ms; 2.0 = 500 ms; NL = no limit.

The fact that time limits always occurred in the same order in a session raises the possibility that changes in response speeds across the conditions (see Table 1 and Figure 1) reflect a warm-up effect. But we observed no systematic changes in speeds across sessions conducted one after another on the same day, suggesting that the temporal contingencies, rather than warm-up, accounted for the differences in response speed.

When compared with a baseline in which feedback about response outcomes followed each trial, the absence of feedback appeared to reduce speeds in the 500-ms condition (Table 1 and Figure 1) but otherwise had no consistent effect. Some speed reduction is not surprising given the well-documented effects of "knowledge of results" on reaction time. In studies of those effects, response speed has sometimes been facilitated by the addition of response-produced messages describing the absolute speed of the response, or the relation of response speed to some criterion (e.g., Church & Camp, 1965; Luce, 1986). The removal of feedback in this experiment merely reverses the traditional order of conditions in knowledge of results investigations.

EXPERIMENT 2

Experiment 1 demonstrated that contingencies could be arranged to establish a target response that varied, within a session, in terms of its speed, accuracy, and relation to a contingency. In Experiment 2, we used similar contingencies to produce a target response (the DMTS response) about which subjects could provide a simple self-report on a trial-by-trial basis.

Because we wished to examine self-reports as operant behavior, we considered each term of a three-term contingency involving the response of self-reporting, its consequences, and potential discriminative events. As previously described, to reduce ambiguity we restricted reporting to a yes/no response to an experimenter-provided query. We also arranged a generalized consequence for reporting that was independent of report topography. This strategy allowed us to focus mainly on the antecedent (discriminative) control of self-reports while holding other variables relatively constant, and appears to be a logical approach given Skinner's (1957) description of self-reports as largely under discriminative control. In Skinner's analysis of verbal behavior, self-reports have properties of the *tact,* a verbal operant strengthened by antecedent stimuli and reinforced by generalized social consequences. Our preparation appears to synthesize conditions like those specified for the tact by Skinner.

One purpose of Experiment 2 was to determine whether self-reports changed systematically with characteristics of the target behavior. However, we also wished to consider the possibility that the act of self-reporting might influence target responding. The clinical self-monitoring literature describes one such effect: When clients report on a target behavior and maintain written records of their reports, the target behavior almost always changes in a socially desirable direction (for a review, see Nelson, 1977).

Several mechanisms of report-induced change are possible. First, prompts for self-reported information that occur during the target event may constitute an implied demand, amounting to what is essentially instructional control over the target behavior (Hayes, 1986; see also Adair, 1973; Nisbett & Wilson, 1977). Second, self-report procedures may occasion previously absent acts of self-observation. If the newly observed target behavior has (or lacks) some critical feature, a number of hypothesized processes could then produce behavior change (Carver & Scheier, 1981; Duval & Wicklund, 1972; Festinger, 1957; Nelson & Hayes, 1981). Third, it has been suggested that self-reports share functional properties with external feedback, and that both self-reports and experimenter-provided feedback may produce behavior change under similar circumstances (Hayes & Nelson, 1983; Prue & Fairbank, 1981; Winett, Neale, & Grier, 1979). Each of the above accounts assumes a generally beneficial form of reactivity to self-reports, but there remains the alternative prospect that self-reports can actually disrupt target behavior—for example, by competing for limited attentional resources that normally would be focused exclusively on the target response (see Anderson, 1985; Baron, Myerson, & Hale, 1988; for other potential sources of disruption, see Nisbett & Wilson, 1977; Schneider & Shiffrin, 1977). Although we did not seek to distinguish among these theoretical accounts, a second purpose of Experiment 2 was to detect any changes in DMTS performance that might be attributed to the presence of self-reports.

The experiment was structured to allow three kinds of comparisons. First, self-reports about target behavior under different degrees of time pressure were studied by varying the time limits across blocks of trials within each session. Second, target-response patterns were observed in the presence and absence of self-reports by including no-report phases similar to those in Experiment 1. Third, the effects of target-response feedback were investigated by arranging self-report phases in the presence and absence of feedback.

Although the DMTS task retained most of the essential features described in Experiment 1, several aspects were modified on the basis of pilot work and the previous experiment. In particular, for the sake of simplicity a session now arranged only two time limits; 2,000 ms and 500 ms were selected because in the first experiment they produced distinctly different patterns of responding. Other procedural modifications are described below.

METHOD

Subjects and Apparatus

Two male undergraduates, 21 and 20 years old, participated. Details of subject screening, informed consent, subject payment, and apparatus were as described in Experiment 1, with the exception that reinforcers were worth slightly more (see below) and the men earned, on average, about $5.00 per hour.

Procedure

The typical trial was identical to that of Experiment 1, in that a compound sample was followed by a delay, comparison stimuli, and choice of a comparison stimulus. Sessions consisted of two 40-trial blocks, with time limits on selection of a comparison stimulus set at 2,000 ms in the first block and 500 ms in the second. Each block was preceded by a message stating "During this period you will have a LONG (or SHORT) amount of time to decide on each trial." To increase the likelihood of subjects attending to the message, a press on the silver button was required to start the blocks, which were further distinguished by characteristic patterns in the frame around the work area on the video screen. A 40-s intermission separated the blocks.

Each session lasted about 25 min, and sets of two sessions were separated by a 10-min break during which subjects could leave the work room. Normally, the men completed four sessions in a work day. To ensure the timely completion of sessions, a maximum duration of 15 min was established for each block. If target-response or report speeds became too low, it was possible for a block to end before 40 trials were completed.

When feedback was scheduled following a DMTS trial, a message appeared indicating to the subject that "You earned 2 points" (1 point = 1.15 cents) or "You lost 0 points," as appropriate to the preceding target response. This was accompanied by an instruction to press the silver button to collect the points or to proceed, respectively. These messages differed from those of Experiment 1 in that they did not provide specific information about the stimulus chosen or the speed or accuracy of the response. A silver-button press then produced a 2-s message indicating that any earned points had been added to the subject's total. The no-feedback message, when scheduled, was identical to that of Experiment 1. It thanked the subject for choosing and prompted a press on the silver button. Pressing produced a 2-s message stating that the program was "About to go on. . . ." Unlike Experiment 1, in no-feedback phases the subjects were not informed of their earnings at the end of each session. Rather, information about earnings was provided at the end of the phase. These changes in feedback were intended to simplify the interpretation of any effects of feedback on self-reports by minimizing the potential instructional functions of the feedback.

Self-reports, when scheduled, occurred immediately following selection of a comparison stimulus; in such cases presentation of the feedback (or no-feedback) message was deferred until completion of the self-report. If no self-report was scheduled, the trial advanced immediately to the feedback (or no-feedback) message.

When self-reports occurred, the work area inside the frame on the screen went blank, and the question "Did you earn points?" appeared centered in the area below the frame. At the bottom of the screen were the words "Yes," printed 3 cm from the left edge of the screen, and "No," printed 3 cm from the right edge of the screen. At the same time, the two outermost keys on the front panel of the console were illuminated, the left key corresponding to "Yes" and the right corresponding to "No." A response on one of the lighted keys cleared the report area of the screen, extinguished the report keys, and recorded the latency from query to report in milliseconds, accurate to plus or minus 1%. Responses on any other keys cleared the report area of the screen, extinguished the report keys, and produced the message "ILLEGAL RESPONSE" for 8 s. The query was not repeated following an illegal

Table 2

Experiment 2: Sequence of phases and number of sessions in each, number of trials, number of self-report trials, and mean DMTS target-response speeds (1,000/latency). Data are shown separately for the two time limit conditions and represent the last 10 sessions per experimental phase. Numbers in parentheses are standard deviations.

	Phase						
	Shaping	FB	NFB	SR + NFB	NFB	FB	SR + FB
Subject Y3							
Sequence	1	2	3	4	5	6	7
Sessions	6	19	28	19	14	19	19
Trials							
2,000 ms	—	400	400	400	400	400	400
500 ms	—	400	400	400	400	400	400
Self-report trials							
2,000 ms	—	—	—	398	—	—	400
500 ms	—	—	—	398	—	—	399
Mean choice speed							
2,000 ms	—	1.60 (0.35)	1.57 (0.43)	1.51 (0.43)	1.76 (0.42)	1.45 (0.45)	1.70 (0.49)
500 ms	—	2.81 (0.52)	2.31 (0.52)	2.04 (0.59)	2.25 (0.51)	2.76 (0.48)	2.70 (0.48)
Subject Y4							
Sequence	1	2	3	4	7	5	6
Sessions	11	25	32	19	18	21	14
Trials							
2,000 ms	—	400	400	374	400	400	400
500 ms	—	396	400	382	398	400	400
Self-report trials							
2,000 ms	—	—	—	367	—	—	396
500 ms	—	—	—	376	—	—	397
Mean choice speed							
2,000 ms	—	1.28 (0.38)	1.43 (0.44)	1.50 (0.48)	1.26 (0.44)	1.49 (0.41)	1.40 (0.44)
500 ms	—	2.74 (0.52)	3.00 (0.58)	2.91 (0.85)	3.19 (0.53)	2.82 (0.38)	2.66 (0.39)

Note. FB = feedback; NFB = no feedback; SR = self-report.

response. In either case, the session advanced to the next scheduled event (either a feedback or a no-feedback message). The only contingency favoring reporting responses over illegal ones was that pressing an illegal key delayed the ensuing feedback or no-feedback message for 8 s. No differential point consequences were contingent on any aspect of the self-reports (e.g., accuracy, speed, etc.).

Instructions

The written instructions given before the first session were similar to those of Experiment 1, with modifications to reflect the change in the feedback messages. In particular, the modifications stated that (a) a message indicating no point earnings "may mean that your decision was inaccurate, too slow, or both." (b) "Sometimes you will be told at the end of the session how much money you earned. At other times, you will not be told right away how much you earned." (c) When you are not told your earnings, the experimenter will keep careful records and tell you how much you made when it is possible to do so.

Design

Table 2 shows the sequence of phases and the number of sessions in each. Because DMTS feedback did not describe response speed, a shaping procedure was used to differentiate speeds under the two time limits. In the shaping phase, feedback followed all DMTS trials and no self-reports were scheduled. Initially, time limits in the two blocks were set at 2,000 and 1,500 ms. Over the next several sessions, the stricter time limit was gradually lowered to 500 ms, using the following guidelines. When a subject's responses met the time limit on more than 50% of the trials in a session, the limit was lowered to 80% of the mean latency of the just-completed session. When a subject failed to respond quickly enough on 50% of the trials in two consecutive sessions, the time limit was raised to 110% of the mean latency of the just-completed session.

Subsequently, phases differed in terms of the events that followed the target response. Two phases involved no self-reports. In one, the target response immediately produced the feedback message (feedback phase). In the other, the target response immediately produced the no-feedback message (no-feedback phase). These phases essentially replicate those of Experiment 1, with the procedural modifications described above. There also were two self-report phases. In one, a self-report, and then the no-feedback message, followed each target response (self-report + no-feedback phase). In the other, a self-report, then the feedback message, followed each target response (self-report + feedback phase).

Stability calculations were based solely on DMTS performance. Separate calculations were performed for four variables (percentage of trials correct, fast enough, and reinforced, plus mean choice speed) in each of the two time-limit conditions. For each of the eight variable-condition combinations, the difference between means of consecutive five-session sets was considered as a proportion of the 10-session grand mean. A condition was terminated when seven of the eight possible proportions were less than .15 and visual inspection of graphed data revealed no trend in seven of eight functions. Retrospectively, in 94 of 96 cases (encompassing six phases for each of 2 subjects) the proportion was less than .15 and, in 92 of 96, less than .10.

Results

Analyses, based on the final 10 sessions per phase, are presented below for DMTS target behavior and self-reports. Characteristics of the target behavior will be discussed first to establish the behavioral context in which self-reports took place.

Target Behavior

Table 2 shows that the men usually completed all scheduled trials (40 per block). Table 2 also shows that, as in Experiment 1, mean choice speeds corresponded to the relative stringency of the temporal contingencies. For both subjects in all six phases, mean choice speed was greater in the 500-ms condition than in the 2,000-ms condition. Mean choice speeds did not vary systematically as a function of the experimental phases, however.

Figure 2 shows the percentage of correct target responses in each phase. As in Experiment 1, choices were correct more often in the 2,000-ms condition than in the 500-ms condition, and accuracy did not appear to be systematically related to the experimental phases.

Figure 3 shows the percentage of target responses that occurred within the time limit in each phase. The responses of both men were almost always fast enough in the 2,000-ms condition, regardless of the experimental phase. In the 500-ms condition, the men showed somewhat different patterns, with one key similarity. Consider first the data for Subject Y3. Comparison of the initial feedback and no-feedback phases shows that removing feedback led to reduced response speeds. When self-reports were introduced (self-report + no-feedback phase), speeds deteriorated further. Speeds then recovered in orderly increments when self-reports were removed (second no-feedback phase) and feedback messages were reinstated (second feedback phase). However, when self-reports were reintroduced in conjunction with feedback (self-report + feedback phase), no disruption of target response speed was evident.

Subject Y4 met the time contingency equally often in the initial no-feedback and feedback phases, but like Y3, showed deterioration when self-reports occurred in the absence of feedback (self-report + no-feedback phase). The deterioration was less dramatic than that shown by Subject Y3, however. Performance levels from the feedback and no-feedback phases were recovered in replications, and, as was the case with Subject Y3, no speed disruptions were apparent in the self-report + feedback phase. For both men, then, self-reports appeared to disrupt target-response speed when feedback was absent but not when it was present.

The effects described above are illustrated in better detail in Figure 4, which shows the relative frequency of choice speeds during the 500-ms condition as a function of the experimental phases (distributions have been collapsed across replications of feedback and no-feedback phases). Each class interval in the distributions represents a speed score range of

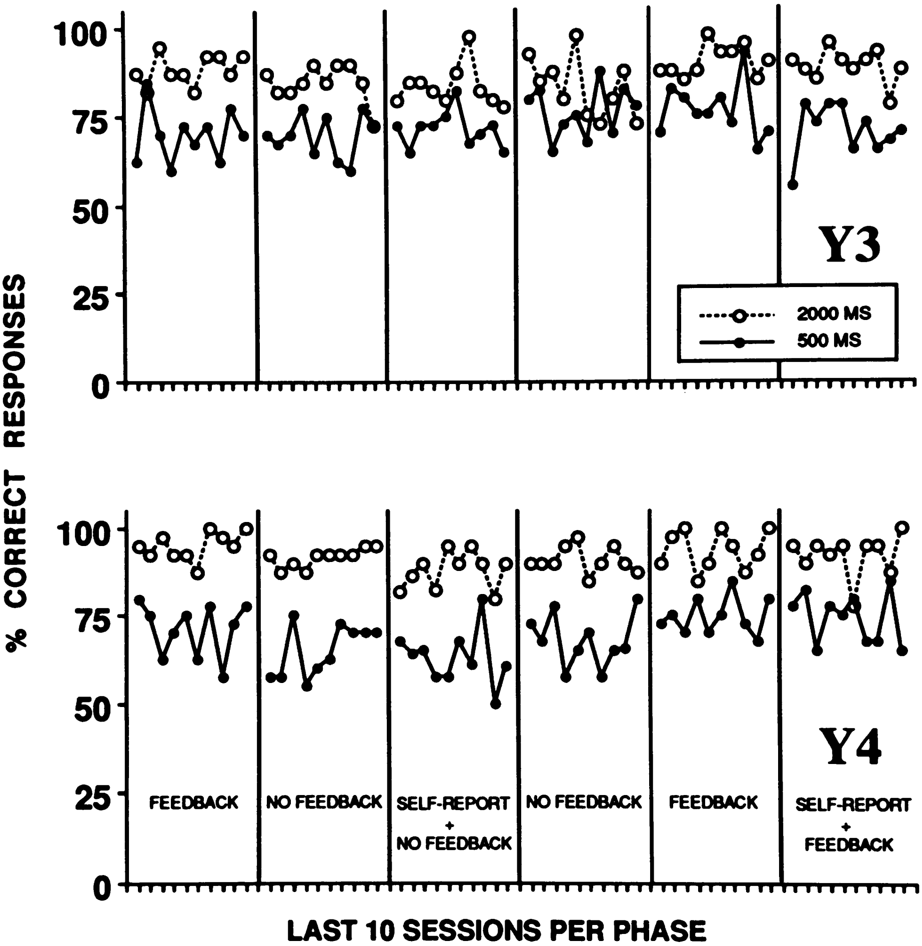

Fig. 2. Experiment 2: Percentage of DMTS target responses that were correct in the last 10 sessions per phase.

0.2. Of particular interest are the darkened bars, depicting choices that failed to meet the 500-ms time limit. Rarely were target responses too slow in the feedback phases. In the no-feedback phases, Subject Y3 showed a substantial increase in the frequency of responses that were slower than the time limit, whereas Subject Y4 showed no increase. The third row of panels (self-report + no-feedback phase) reveals the disruption precipitated by self-reports. For both men the darkened bars cover a greater area of the distribution in this phase than in any other, indicating a greater proportion of target responses slower than the time limit. The distributions clarify the different patterns of disruption shown in Figure 3. For Subject Y3, the modal target-response speed in this phase was slower than the time limit. For Subject Y4, most target responses remained faster than the time limit; however, unlike in other phases, speeds were occasionally well below the time limit. The bottom row of panels shows that when feedback was provided, self-reports did not disrupt target-response speed (e.g., compared to the top row of panels).

Table 3 lists the percentage of DMTS responses that were correct and fast enough to meet the time limits. (These percentages summarize data presented in Figures 2, 3, and 4.) Table 3 also lists the percentage of DMTS responses that were reinforced, providing an additional perspective on the speed decrements observed in the self-report + no-feedback phase. Both men lost a substantial number of reinforcers in this phase as compared with the

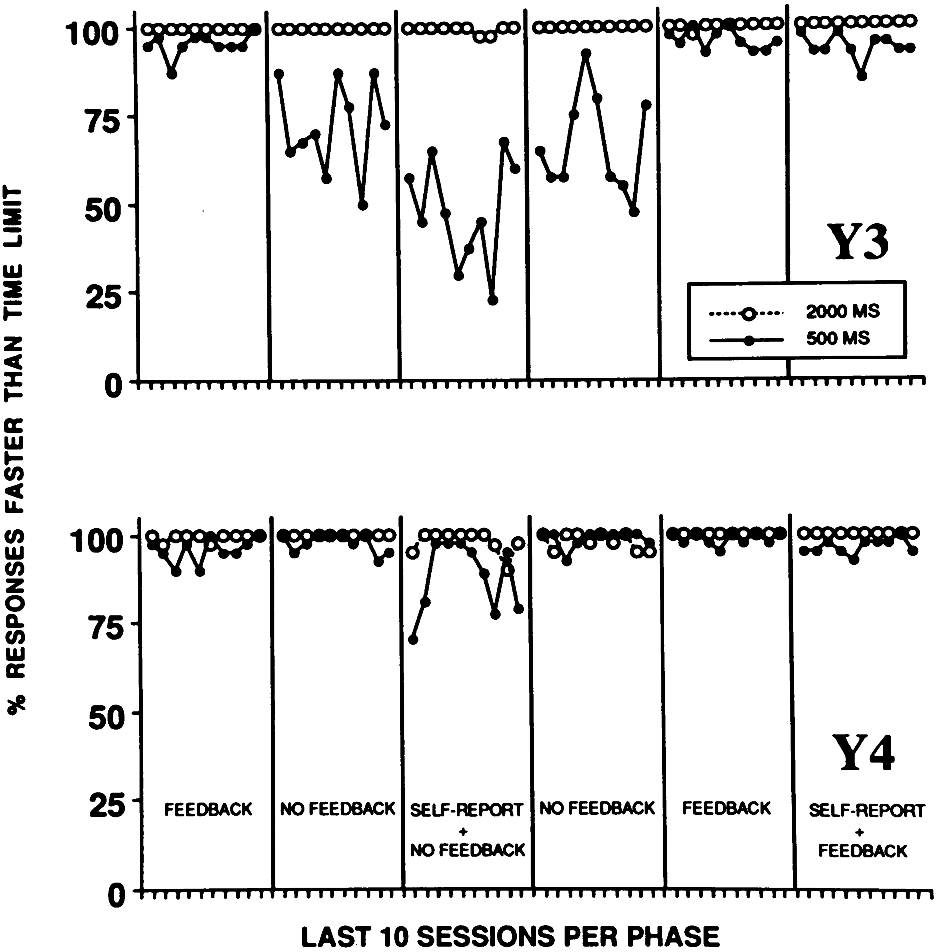

Fig. 3. Experiment 2: Percentage of DMTS target responses that were fast enough to meet the time limits in the last 10 sessions per phase.

no-feedback phases, indicating that they did not adapt to the variables responsible for slower responding even when it reduced their earnings.

Self-Reports

Report accuracy. Figure 5 shows the accuracy of self-reports as a function of the time limits on, and feedback about, the target performance. Overall, report accuracy ranged from 56% to 95% and exceeded 80% in seven of the eight combinations of time limit and feedback represented in the figure. Comparison of results across the two time limits shows that reports tended to be less accurate when the target performance was under greater time pressure. Comparison across experimental phases also shows that reports were less accurate when experimenter-provided feedback about the target response was withheld.

Figure 6 shows the probability of an accurate self-report as a function of the obtained probability of a reinforced target response (point trials). Each data point represents one combination of the time limit and feedback manipulations (about 400 trials). The inset depicts the function for a hypothetical reporter whose reports are always accurate regardless of the probability of a reinforced target response. For this ideal reporter, data would fall along a horizontal line with a y intercept of 1.00. The functions for the men in Experiment 2 (fitted by least squares linear regression) deviate from ideal largely in terms of a down-

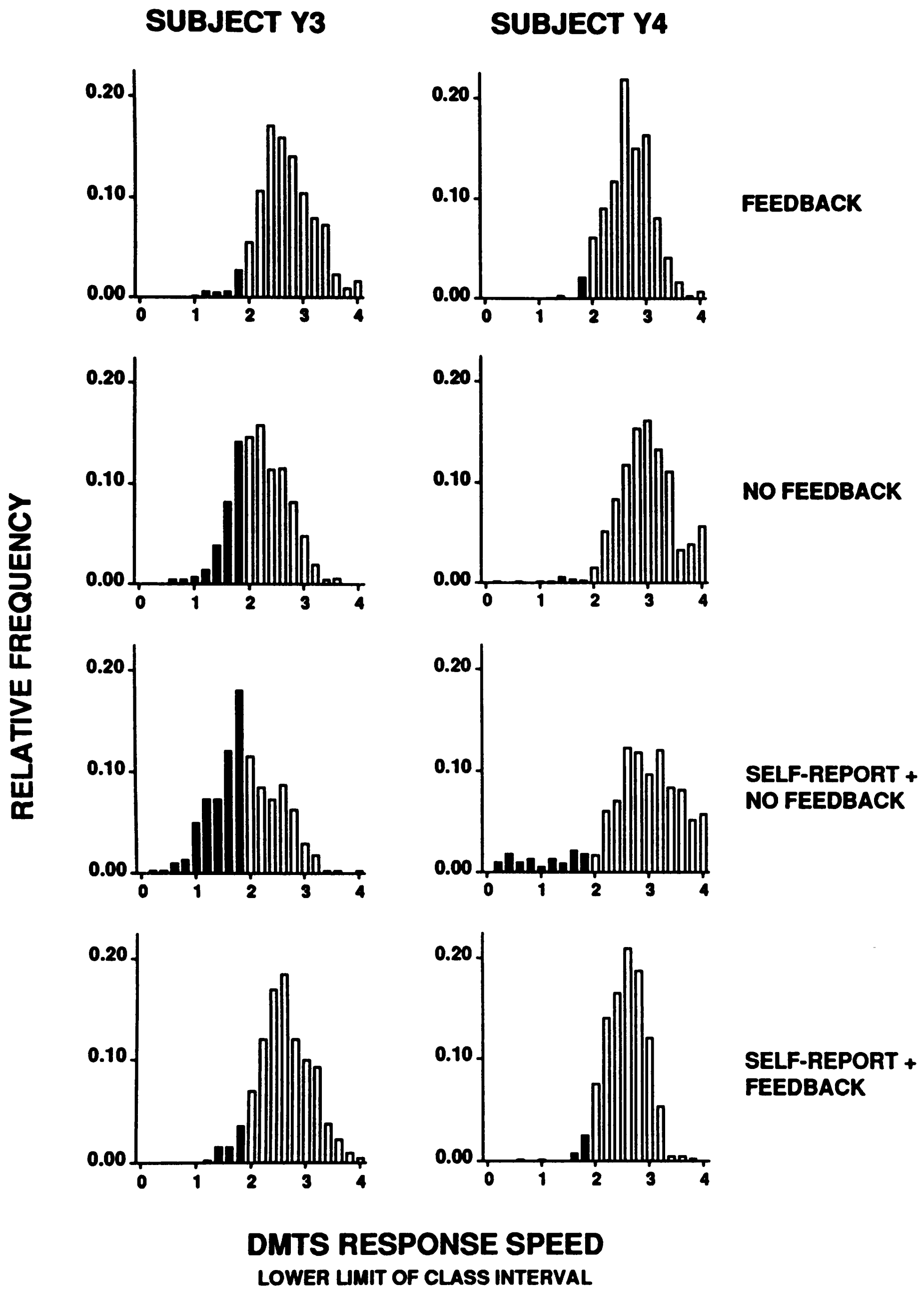

Fig. 4. Experiment 2: Relative frequency of DMTS target-response speeds, in the 500-ms condition, during the last 10 sessions per phase. Feedback and no-feedback distributions have been collapsed across replications. Class intervals encompass a speed score range of 0.2. Darkened bars represent responses that were slower than the time limit.

ward shift on the *y* axis, indicating that as target-response reinforcement became less likely, self-reports became less accurate. As can be seen in the slopes of the best fitting lines, this pattern was more pronounced for Subject Y3 than for Subject Y4.

Report content. Another way to examine report accuracy is to consider relations between the content of self-reports and the characteristics of target responses that preceded them. A global view of this relation is shown in Figure 7. Here the probability of a report of point earnings (point reports) is plotted as a function of the obtained probability of reinforced target responses (point trials). The inset shows the hypothetical function for an ideal reporter whose self-reports are always accurate. For the ideal reporter there is a match between the probability of a point report and the probability of a reinforced target response, producing a function with a slope of one and a y intercept of zero.

The actual functions (fitted by least squares linear regression) differ substantially for the 2 men. Subject Y4 reported much like the ideal, accurate reporter. Subject Y3's reports showed evidence of a bias in which report content was somewhat insensitive to decreases in the frequency of reinforced target responses. Casually speaking, when reinforcement was infrequent, Y3 tended to overestimate his success.

Figure 8 examines in greater detail the relations between self-report content and target-response characteristics. The obtained probability of point reports is shown as a function of the speed of the preceding target response. Class intervals encompass target-response speeds in a range of 0.5. Each panel in the figure represents a combination of one time limit contingency and one feedback condition. Each also depicts two functions, one for reports following choices of the correct DMTS comparison stimulus, and one for reports following incorrect choices. No data points were plotted in either function for class intervals containing fewer than five self-reports. Shadowing indicates responses that were slower than the time limit.

Two types of control over report content are evident in Figure 8. First, all panels show that point reports were more likely following correct target responses than following incorrect ones, although both men emitted such reports at least occasionally following incorrect target responses. Comparing across adjacent panels, it can be seen that Subject Y3 generally was more likely to report reinforcement after an incorrect target response than was Subject Y4, a finding also illustrated in Figure 7.

Second, reports also were sensitive to the speed of the target response, at least when that response was correct. In every panel of Figure 8, the probability of a point report increased as the speed of the preceding correct target response increased. In this display, maximal sensitivity of the reports to target-response speed contingencies would be indicated by zero percentages in shaded class intervals (recall that shading indicates speeds below the time limit). However, point reports still occurred at least occasionally following tardy target responses.

At least one effect of feedback on self-reports can be seen in the solid functions of Figure 8. The flatter functions in panels from feedback phases, compared to equivalent no-feedback phases, show that in a given class interval point reports were more likely when feedback was provided.

The gradient described above for reports following correct target responses was not apparent in self-reports following incorrect target responses. These data should be interpreted cautiously, however. As Table 3 shows,

Table 3

Experiment 2: Percentage of DMTS target responses that were correct, fast enough to meet the time limit, and reinforced, over the last 10 sessions per phase. Data are shown separately for the two time limit conditions.

	Phase					
	FB	NFB	SR + NFB	NFB	FB	SR + FB
Subject Y3						
Correct						
2,000 ms	89	85	84	85	90	88
500 ms	70	70	72	76	76	70
Fast enough						
2,000 ms	100	100	100	100	99	100
500 ms	96	72	48	67	96	93
Reinforced						
2,000 ms	89	85	84	85	89	88
500 ms	67	54	37	51	73	66
Subject Y4						
Correct						
2,000 ms	95	92	88	92	94	93
500 ms	69	65	63	68	75	75
Fast enough						
2,000 ms	100	100	98	98	100	100
500 ms	95	98	88	99	99	96
Reinforced						
2,000 ms	94	92	87	90	94	92
500 ms	66	64	53	67	74	71

Note. FB = feedback; NFB = no feedback; SR = self-report.

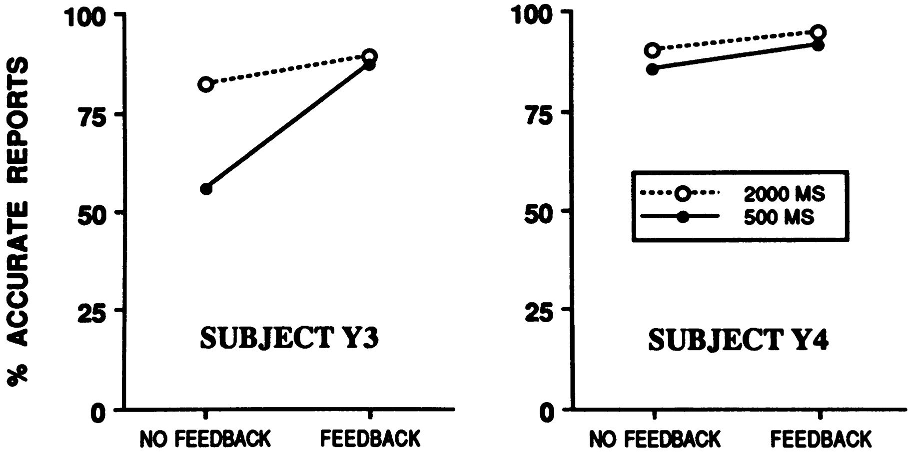

Fig. 5. Experiment 2: Percentage of accurate self-reports, as a function of DMTS time limit and the presence versus absence of experimenter-provided feedback about the target response.

incorrect DMTS responses occurred no more than 6% to 37% of the time for either man, providing a limited sample of trials for analysis in any given class interval. Thus, it is unclear whether the dotted function lines in Figure 8 depict relations truly different from those of the solid lines or merely the variability that can occur in small samples.

Report speed. In general, reports occurred quickly, with rarely more than 1 s elapsing between appearance of the query and emission of the report itself. Nevertheless, the speed of the self-reports was related to characteristics of both the target and reporting responses. Table 4 lists mean report speeds as a function of the DMTS temporal contingencies (2,000- or 500-ms time limit), DMTS outcome (points or no points), feedback manipulation (feedback or no feedback), and the accuracy of the self-reports. Four patterns are evident. First, in seven of eight comparisons in which points were earned, reports were faster when feedback was present than when it was absent. When points were not earned there was no systematic effect of feedback on report speeds. Second, in 13 of 16 comparisons overall, self-reports were emitted more quickly following 500-ms trials than 2,000-ms trials. Third, in 15 of 16 comparisons, reports were emitted more quickly when points had been earned in

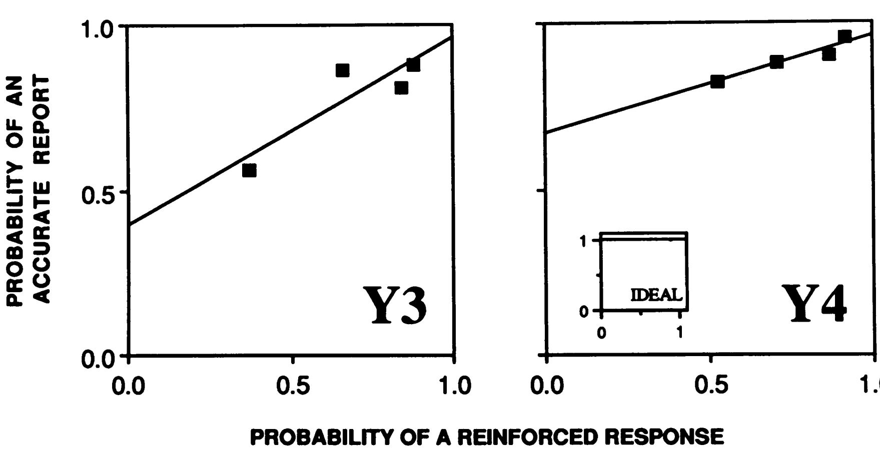

Fig. 6. Experiment 2: Probability of an accurate self-report as a function of the probability of a reinforced DMTS target response. Each point represents trials employing one DMTS time limit for the last 10 sessions of an experimental phase (about 400 trials). Inset shows the function for an ideal reporter whose self-reports are always accurate.

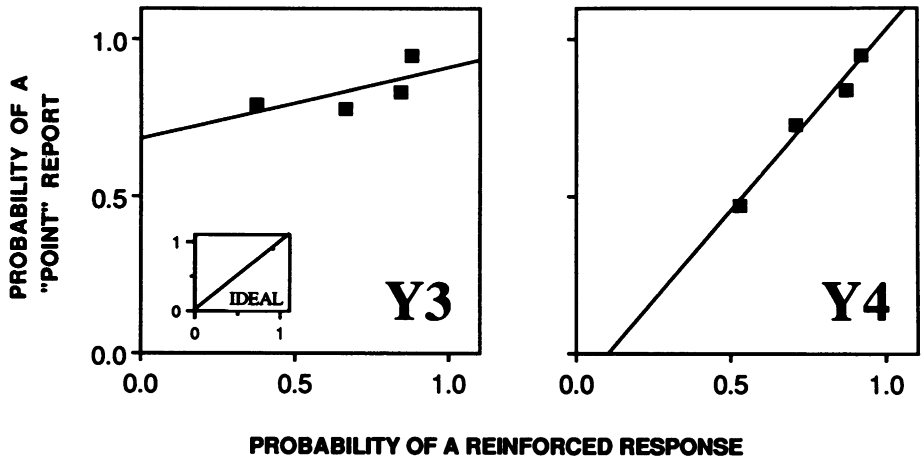

Fig. 7. Experiment 2: Probability of a self-report of reinforcement (point report) as a function of the probability of a reinforced DMTS target response. Each point represents trials employing one DMTS time limit for the last 10 sessions of an experimental phase (about 400 trials). Inset shows the function for an ideal reporter whose self-reports are always accurate.

the DMTS trial, regardless of the feedback condition or the accuracy of the report. Finally, in all 16 comparisons, the mean speed of accurate self-reports exceeded that of inaccurate self-reports.

Discussion

In Experiment 2, subjects were asked to self-report about relatively speeded and relatively unspeeded target responses, in the presence and absence of experimenter-provided feedback about the reinforcement of those responses. They reported when asked, and often did so accurately. This experiment was designed to examine events that influenced self-reports, as well as to assess any reciprocal effects of self-reports on the target behavior.

Events Influencing Self-Reports

If self-reports are to be viewed as behavior under stimulus control, then an analysis of self-report accuracy should identify events that exert discriminative control over reporting responses. In this experiment, the target behavior consisted of a DMTS response under a conjunctive contingency (see Catania, 1979) involving both speed and accuracy requirements for reinforcement. The self-report query asked about the occurrence of reinforcement and thus, by implication, about whether both aspects of the contingency had been met.

This method produced sufficiently diverse target responding that the men self-reported following both accurate and inaccurate responses of a variety of speeds. Both the speed and accuracy of the target response exerted clear discriminative control over self-reports. Figure 8 suggests an interactive relation between speed and accuracy as discriminative stimuli: Target-response speed systematically influenced reports only when the target response also was correct. Moreover, discriminative control was modulated by the presence versus absence of feedback; that is, reports were more accurate when feedback was provided (Figures 5 and 8).

In both self-report phases, the control exerted by speed and accuracy was imperfect in the sense that the men sometimes reported inaccurately about their target responses. For both subjects, the absence of feedback was a factor in reporting errors. For 1 subject (Y3), inaccurate reports also were found to reflect an apparent bias for reporting reinforcement rather than nonreinforcement (a similar effect has been reported in some studies of humans' ratings of response–reinforcer contingencies; Alloy & Abramson, 1979; Wasserman & Neunaber, 1986). This bias helps characterize Subject Y3's relatively low report accuracy in portions of Figures 5 and 6: When target-response reinforcement was infrequent, a rigid pattern of point reports necessarily would produce many reporting errors.

Regardless of the reasons for reporting errors, the extent to which the men were ap-

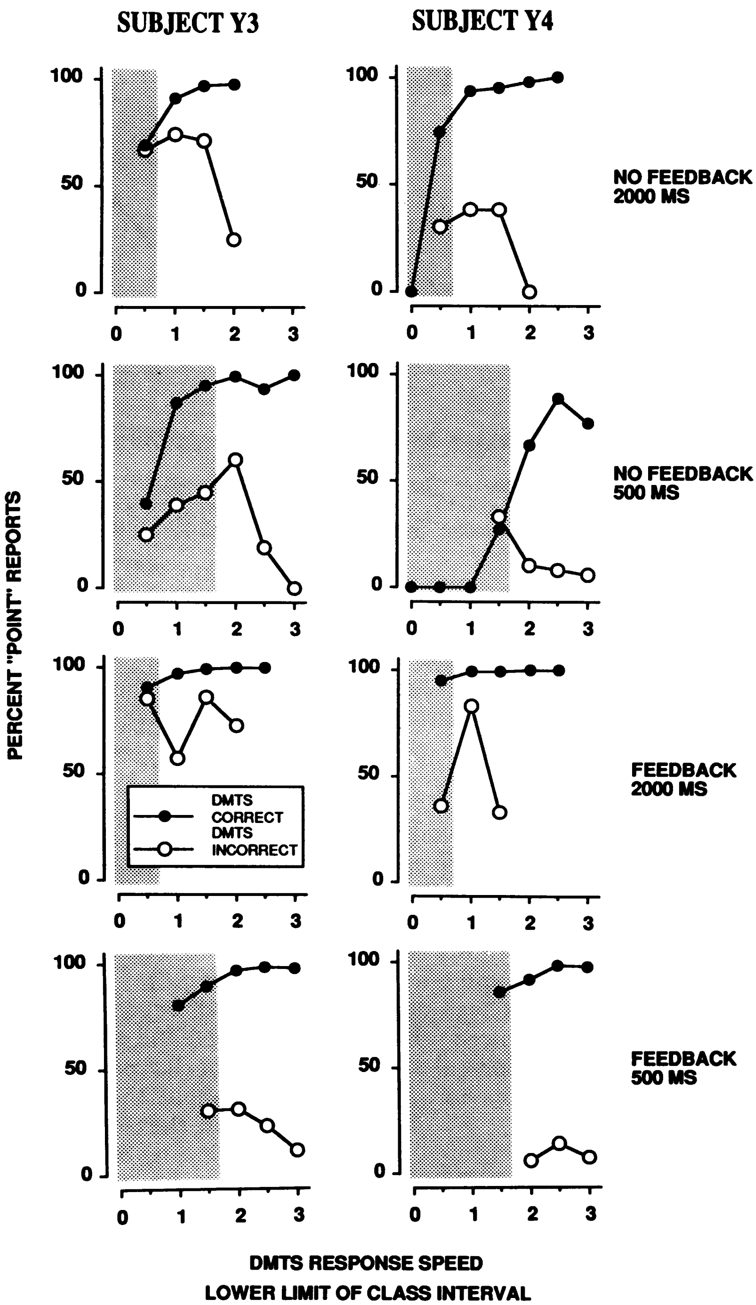

Fig. 8. Experiment 2: Percentage of self-reports indicating point earnings (point reports) as a function of the speed and accuracy of the preceding DMTS target response. Class intervals represent a speed score range of 0.5. Separate functions show cases in which correct and incorrect DMTS responses preceded the self-report. Shading indicates cases in which the DMTS response failed to meet the time limit.

parently unaware of their own behavior (5% to 44% of the time) may be surprising given the immediate and specific nature of the reports they emitted. Reports occurred in a distraction-free environment, just after a discrete and relatively simple target response (by contrast, in other settings, verbal reports may occur long after the target behavior or may de-

Table 4

Experiment 2: Mean self-report speeds (1,000/latency) as a function of self-report accuracy and DMTS time limit, point outcome, and feedback status. Data are shown separately for each time limit condition and represent the last 10 sessions per experimental phase. Numbers in parentheses are standard deviations.

	Points earned				No points earned			
	2,000 ms		500 ms		2,000 ms		500 ms	
	Trials	Mean speed	Trials	Mean speed	Trials	Mean speed	Trials	Mean speed
Subject Y3								
No DMTS feedback								
Accurate reports	305	1.80 (0.58)	144	1.99 (0.50)	21	1.54 (0.61)	81	1.79 (0.73)
Inaccurate reports	27	1.62 (0.58)	5	1.80 (0.73)	45	1.31 (0.68)	168	1.43 (0.50)
DMTS feedback								
Accurate reports	345	2.24 (0.54)	254	2.47 (0.61)	10	1.62 (0.64)	96	1.78 (0.81)
Inaccurate reports	8	1.96 (0.56)	8	1.84 (0.70)	37	1.15 (0.52)	41	1.08 (0.44)
Subject Y4								
No DMTS feedback								
Accurate reports	294	1.90 (1.11)	159	1.86 (1.13)	36	1.17 (0.58)	162	1.63 (0.82)
Inaccurate reports	26	1.10 (1.12)	41	1.16 (0.81)	11	0.89 (0.58)	14	1.39 (1.09)
DMTS feedback								
Accurate reports	359	2.63 (1.02)	269	3.00 (1.17)	15	1.11 (0.44)	94	2.12 (1.03)
Inaccurate reports	6	1.91 (0.86)	12	2.09 (1.20)	16	0.69 (0.22)	22	1.58 (0.82)

scribe diverse and complex responses or response sequences). In addition, based on their extensive exposure to the experimental procedures, the men could reasonably expect to be asked to report on each trial in the self-report phases. In short, conditions of the experiment seemed to be conducive to accurate reporting. Nevertheless, the men could not always say what they had just done.

Diverse sources warn of the potential for unreliability in self-reports in natural settings (e.g., Ciminero, Nelson, & Lipinski, 1977; Kagan, 1988; O'Farrell et al., 1984). However, the current data serve as a reminder that laboratory settings are not necessarily immune to reporting errors (see Shimoff, 1986). Although self-reports rarely provide the primary data in laboratory studies of human operant behavior, investigators often interview subjects following an experiment and interpret data in light of what subjects report about their thoughts and performance during the experiment. Such interpretations may include speculation about the status of unmeasured variables. For example, Case et al. (1985), in a study of conditioned reinforcement in human subjects, used instructions to circumvent baseline conditions involving direct exposure to the stimulus–reinforcer relations. Case et al. argued for the adequacy of this expedient on the basis of subjects' responses to a postexperimental questionnaire about the procedures. Increasingly, operant researchers use a similar basis to argue for the influence of private events, such as covert rules or self-instructions, on public behavior (e.g., Bentall, Lowe, & Beasty, 1985; Harzem, Lowe, & Bagshaw, 1978; Laties & Weiss, 1963). In some cases, this may involve the reanalysis of experimental data in terms of categories determined from responses to the postexperimental interview (e.g., Lippman & Meyer, 1967).

At issue is not whether events described in verbal reports exist, or even whether they can control behavior, but whether their influence can be ascertained solely on the basis of verbal reports (Perone, 1988; Perone et al., 1988). Given the occasional failure of the men in Experiment 2 to make accurate reports immediately after a single trial, one might wonder about the ability of subjects to report on remote events of several sessions past or to categorize overt or covert behavior spanning a number of sessions (as postsession or postexperimental interviews often require). The present data illustrate that task characteristics can influence report accuracy—so accuracy could vary across procedures—but to our knowledge there exist no empirically derived criteria for deciding in advance when, and whether, to trust self-reports whose putative referents are not observed independently.

Cognitive psychologists have addressed this problem more directly than their operant colleagues, generating conceptual frameworks to guide researchers in the interpretation of self-reports. In post-hoc analyses of published data, certain of these approaches appear to accommodate diverse experimental findings (Ericsson & Simon, 1984; Nisbett & Wilson, 1977). But as Hayes (1986) points out, these approaches often suffer from a reliance on nonmanipulable causes (such as memory mechanisms) and from a dearth of supporting data in which the referents of self-reports have been corroborated independently. All things considered, there may be no easy or immediate solution to the problem of when to trust a verbal self-report. The present research suggests that a productive first step will be the analysis of reports about public target behavior under the control of manipulable variables. Principles derived from such research should be of practical and theoretical interest to both operant and cognitive researchers.

The speed with which subjects emit their self-reports provides an additional focus for research with potentially broad appeal. "Reaction time" historically has been the province of cognitive psychologists, who may appeal to different cognitive processes when confronted with different speeds of execution for topographically identical responses (Baron, 1985; Luce, 1986). Different mental processes are believed to require different amounts of time to execute and to generate a response (e.g., see Salthouse, 1985). By contrast, from an operant perspective, response speeds presumably vary as a function of the environmental events that produce the response. It is beyond the scope of the present data and design to account for differences in the men's "reporting reaction times," but feasible studies can be imagined. Consider the fact that our subjects reported relatively slowly about unreinforceable DMTS responses (Table 4). Possibly, such reports were emitted because of the variables that motivate self-reporting in general, but were delayed because of aversive properties like those acquired by external feedback about nonreinforcing or aversive events or outcomes. A procedure that allows subjects to cancel self-reports without penalty would provide support for this hypothesis if subjects did so more often after unreinforceable target responses than after reinforceable ones.

Influence of Self-Reports on the Target Response

In one phase of Experiment 2 the self-report procedure disrupted target performance on the DMTS task. This effect is apparently at odds with the tendency of self-reports in clinical self-monitoring interventions to produce beneficial changes in target behavior (Nelson, 1977). Yet self-monitoring effects are poorly understood (Nelson & Hayes, 1981), and nonbeneficial outcomes have been noted occasionally (see Critchfield, 1989; Critchfield & Vargas, in press). Apparently, much remains to be learned about the influence of self-reporting on target behavior.

The present design did not permit further analysis or intrasubject replication of the disruptive effect of self-reporting. However, the fact that our subjects had extensive exposure to the target task (over 4,000 trials) before self-reports were introduced is suggestive of nonoperant accounts of "automatized" responses, which, as a result of repeated practice, are believed to occur without effort or conscious attention (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977; see also Mishkin & Petri, 1984). According to one hypothesis, once an act becomes automatic it may be especially vulnerable to disruptive effects of self-reports, because reporting can reestablish conscious cognitive processing. Moreover, because cognitive processing takes time to complete, the disruption should be observed in terms of the speed rather than the accuracy of the target response (for details, see Nisbett & Wilson, 1977). In Experiment 2, disruptive effects (in the self-report + no-feedback phase) were indeed apparent in terms of speed, but not accuracy, of the DMTS target response. However, similar disruption was not observed when self-reports were followed by feedback about the target response, and it is unclear how an explanation based on the "automaticity" of the target response would account for this discrepancy.

We prefer an interpretation in operant terms, based on the allocation of attentional (observational) resources to simultaneously occurring responses (Baron et al., 1988; Reynolds, 1961). Because in our procedure target and reporting responses occurred at different times, speed disruption (in the self-report + no-feedback phase) could not result from the physical incompatibility of these two responses. If, however, an act of self-observation is prerequisite to self-reporting (see Skinner, 1953, pp. 288–292, 1957, pp. 138–146), then it seems reasonable that subjects under time pressure sometimes would find it difficult to perform the DMTS task and self-observe simultaneously. In fact, one would predict that both self-reporting and the DMTS task might be adversely affected, which was exactly the case in our self-report + no-feedback phase.

The absence of disruption when feedback followed self-reports may be explained by either of two motivational properties of feedback. First, concordant with the findings of knowledge-of-results studies, feedback may engender faster target responding (recall that 3 of 4 subjects in our two experiments responded more quickly in the presence of feedback than in its absence; see Figures 1 and 4). Second, feedback may have also increased the speed of any private self-observation responses, just as it sometimes resulted in faster overt reporting responses (see Table 4). Faster target and self-observation responses would translate into less competition between the two when feedback was provided, and thus less disruption in both responses—just as we observed.

Worth noting in the suggestion above is the implication that self-observation is not a necessary concomitant of other responding (see Skinner, 1957). The DMTS disruption observed in our self-report + no-feedback phase was not apparent in the phases that preceded or followed. That self-observation is responsible for this effect is disputable, but it is clear that once prompted to self-report the men began doing *something* differently. This type of effect may have implications for concurrent verbalization ("think aloud") procedures, which require subjects to verbalize about private events (e.g., hypotheses, rules, or self-talk) while engaging in public responses (e.g., see Catania et al., 1982; Ericsson & Simon, 1984). The possibility that prompts for self-reports may induce rather than simply harness self-observation (Farber, 1963; Krasner & Ullmann, 1963; Nisbett & Wilson, 1977) complicates the interpretation of reported information, unless the absence of reactive effects can be demonstrated explicitly.

GENERAL DISCUSSION

The present data illustrate the possibility of fruitful inquiry into the environmental determinants of verbal self-reports by humans in a laboratory setting. As previously noted, an operant approach to this inquiry must place emphasis on each element of a three-term contingency involving the reporting response. In the present investigation, we held constant the topography and consequences of the reporting response and manipulated the antecedents. This approach may be more typical of cognitive than operant research (see Baron et al., 1988), but it is certainly amenable to an operant analysis.

An operant analysis also might examine the effects of self-report topography on other characteristics of the report, such as frequency or speed (see Baron & Journey, 1989, and Newhall & Rodnick, 1936, for some possibilities). However, consequences have been the primary focus in operant psychology, and it is here that operant researchers may be expected to contribute most to the analysis of verbal self-reports. Motivational variables almost certainly play a role in many self-reports (e.g., Rosen, 1966; Skinner, 1957; see Baron et al., 1988, for an interesting discussion of interactions between discriminative and motivational variables). For example, our subjects might have reported more accurately had their earnings depended on it. To explore this possibility, our procedure could be modified to allow independent scheduling of consequences for target and reporting responses, and a variety of contingencies could be arranged either to encourage or discourage accurate reporting (see de Freitas Ribeiro, 1989).

Although we neither shaped nor differentially reinforced self-reports, the men in Experiment 2 reported (instead of aborting the report procedure) and often did so accurately. This represents a state of affairs not encountered in studies employing animal subjects (e.g., Shimp, 1981, 1982, 1983), in which explicit consequences for accurate self-reports must be arranged. To account for the difference one may appeal to Skinner (1957) and other theorists who posit social origins of self-awareness (e.g., Carver & Scheier, 1981; Duval & Wicklund, 1972). In Skinner's analysis, self-awareness emerges from explicit training by a verbal community that not only shapes self-observation skills but in so doing also creates a lasting predisposition to provide information about self. Our procedure likely harnessed the products of this special history.

However, simply to state that animals lack the histories shared by most humans obscures the more general point that all self-reports, regardless of the reporter's species, reflect an interaction of prior history and current contingencies. As a case in point, recall the human subjects of Hefferline and Perera (1963), who initially could not report tiny thumb twitches even when their monetary earnings depended on it. Although these individuals might have easily reported about other responses, their preexperimental histories did not facilitate discrimination of the target response. Similarly, for Subject Y3 in our Experiment 2, discriminative control of target responses over self-reports was limited by a bias that may have been acquired preexperimentally. It is in part the diverse and unspecified histories of most humans that make their self-reports difficult to evaluate outside the laboratory. In a laboratory context, such histories have prompted some investigators to regard human subjects as not suitable for research in which basic processes may be at issue (e.g., Dinsmoor, 1983).

Whether the experimental analysis of self-reports by humans will reveal basic processes (or even generate answers to applied questions about their use) remains to be seen. To date, most operant laboratory research on self-reports has used animal subjects. Our data indicate that the self-reports of human subjects can be sensitive to laboratory manipulations. Other lines of research, on topics ranging from instructional control to stimulus equivalence, have demonstrated the value of studying the behavior of human subjects when the interest is in phenomena typically observed in humans (Baron & Perone, 1982; Hake, 1982). Given the nature and historical context of verbal self-reports, it seems logical to pursue a similar course in their analysis.

REFERENCES

Adair, J. G. (1973). *The human subject: The social psychology of the psychology experiment.* Boston: Little, Brown.

Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General,* **108,** 441–485.

Anderson, J. R. (1985). *Cognitive psychology and its implications* (2nd ed.). New York: Freeman.

Baron, A. (1985). Measurement scales and the age-complexity hypothesis. *Experimental Aging Research,* **11,** 193–199.

Baron, A., & Journey, J. W. (1989). Reinforcement of human reaction time: Manual-vocal differences. *Psychological Record,* **39,** 285–296.

Baron, A., & Menich, S. R. (1985a). Age-related effects of temporal contingencies on response speed and memory: An operant analysis. *Journal of Gerontology,* **40,** 60–70.

Baron, A., & Menich, S. R. (1985b). Reaction times of younger and older men: Effects of compound samples and a prechoice signal on delayed matching-to-sample performances. *Journal of the Experimental Analysis of Behavior,* **44,** 1–14.

Baron, A., Myerson, J., & Hale, S. (1988). An integrated analysis of the structure and function of behavior: Aging and the cost of dividing attention. In G. Davey & C. Cullen (Eds.), *Human operant conditioning and behavior modification* (pp. 139–166). Chichester, England: Wiley.

Baron, A., & Perone, M. (1982). The place of the human subject in the operant laboratory. *Behavior Analyst,* **5,** 143–158.

Bentall, R. P., Lowe, C. F., & Beasty, A. (1985). The role of verbal behavior in human learning: II. Developmental differences. *Journal of the Experimental Analysis of Behavior,* **43,** 165–181.

Boring, E. G. (1953). A history of introspection. *Psychological Bulletin,* **50,** 169–189.

Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control-theory approach to human behavior.* New York: Springer-Verlag.

Case, D. A., Fantino, E., & Wixted, J. (1985). Human observing: Maintained by negative informative stimuli only if correlated with improvement in response efficiency. *Journal of the Experimental Analysis of Behavior,* **43,** 289–300.

Catania, A. C. (1979). *Learning.* Englewood Cliffs, NJ: Prentice-Hall.

Catania, A. C., Matthews, B. A., & Shimoff, E. (1982). Instructed versus shaped human verbal behavior: Interactions with nonverbal responding. *Journal of the Experimental Analysis of Behavior,* **38,** 233–248.

Church, R. M., & Camp, D. S. (1965). Change in reaction-time as a function of knowledge of results. *American Journal of Psychology,* **78,** 102–106.

Ciminero, A. R., Nelson, R. O., & Lipinski, D. P. (1977). Self-monitoring procedures. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), *Handbook of behavioral assessment* (pp. 195–232). New York: Wiley.

Critchfield, T. S. (1989). Self-recording mutually exclusive multiple responses. *Behavior Modification,* **13,** 361–375.

Critchfield, T. S., & Vargas, E. A. (in press). Self-recording, instructions, and public self-graphing: Effects on swimming in the absence of coach verbal interaction. *Behavior Modification.*

Danziger, K. (1980). The history of introspection reconsidered. *Journal of the History of the Behavioral Sciences,* **16,** 241–262.

de Freitas Ribeiro, A. (1989). Correspondence in children's self-report: Tacting and manding aspects. *Journal of the Experimental Analysis of Behavior,* **51,** 361–367.

Dinsmoor, J. A. (1983). Observing and conditioned reinforcement. *Behavioral and Brain Sciences,* **6,** 693–728. (Includes commentary)

Duval, S., & Wicklund, R. A. (1972). *A theory of objective self awareness.* New York: Academic Press.

Epstein, L. H., Miller, P. M., & Webster, J. S. (1976). The effects of reinforcing concurrent behavior on self-monitoring. *Behavior Therapy,* **7,** 89–95.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Farber, I. E. (1963). The things people say to themselves. *American Psychologist,* **18,** 185–197.

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Giorgi, A. (1975). Convergence and divergence of qualitative and quantitative methods in psychology. In A. Giorgi, C. T. Fisher, & E. L. Murray (Eds.), *Duquesne studies in phenomenological psychology, Volume II* (pp. 72-79). Pittsburgh, PA: Duquesne University Press.

Hake, D. F. (1982). The basic-applied continuum and the possible evolution of human operant social and verbal research. *Behavior Analyst,* **5,** 21–28.

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

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., & Nelson, R. O. (1983). Similar reactivity produced by external cues and self-monitoring. *Behavior Modification,* **7,** 183–196.

Hefferline, R. F., & Perera, T. B. (1963). Proprioceptive discrimination of a covert operant without its observation by the subject. *Science,* **139,** 834–835.

Kagan, J. (1988). The meanings of personality predicates. *American Psychologist,* **43,** 614–620.

Kramer, S. P. (1982). Memory for recent behavior in the pigeon. *Journal of the Experimental Analysis of Behavior,* **38,** 71–85.

Krasner, L., & Ullmann, L. P. (1963). Variables affecting report of awareness in verbal conditioning. *Journal of Psychology,* **56,** 193–202.

Laties, V. G., & Weiss, B. (1963). Effects of a concurrent task on fixed-interval responding in humans. *Journal of the Experimental Analysis of Behavior,* **6,** 431–436.

Lippman, L. G., & Meyer, M. E. (1967). Fixed interval performance as related to instructions and to subjects' verbalizations of the contingency. *Psychonomic Science,* **8,** 135–136.

Lowe, C. F. (1979). Determinants of human operant behaviour. In M. D. Zeiler & P. Harzem (Eds.), *Advances in analysis of behaviour: Volume 1. Reinforcement and the organization of behaviour* (pp. 159–192). Chichester, England: Wiley.

Lubinski, D., & Thompson, T. (1987). An animal model of the interpersonal communication of interoceptive (private) states. *Journal of the Experimental Analysis of Behavior,* **48,** 1–15.

Luce, R. D. (1986). *Reaction times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

Lyons, W. (1986). *The disappearance of introspection.* Cambridge, MA: MIT Press.

Maki, W. S., Moe, J. C., & Bierley, C. M. (1977). Short-term memory for stimuli, responses, and reinforcers. *Journal of Experimental Psychology: Animal Behavior Processes,* **3,** 156–179.

Mishkin, M., & Petri, H. L. (1984). Memories and habits: Some implications for the analysis of learning and retention. In L. R. Squire & N. Butters (Eds.), *Neuropsychology of memory* (pp. 287-296). New York: Guilford Press.

Nelson, R. O. (1977). Assessment and therapeutic functions of self-monitoring. In M. Hersen, R. M. Eisler, & P. Miller (Eds.), *Progress in behavior modification* (Vol. 5, pp. 263-308). New York: Academic Press.

Nelson, R. O., & Hayes, S. C. (1981). Theoretical explanations for reactivity in self-monitoring. *Behavior Modification,* **5,** 3–14.

Newhall, S. M., & Rodnick, E. H. (1936). The influence of the reporting-response upon the report. *American Journal of Psychology,* **48,** 316–325.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review,* **84,** 231–259.

O'Farrell, T. J., Cutter, H. S., Bayog, R. D., Dentch, G., & Fortgang, J. (1984). Correspondence between one-year retrospective reports of pretreatment drinking by alcoholics and their wives. *Behavioral Assessment,* **6,** 263–274.

Perone, M. (1985). A software system for real-time laboratory use of TRS-80 microcomputers. *Behavior Research Methods, Instruments, & Computers,* **17,** 119–121.

Perone, M. (1988). Laboratory lore and research practices in the experimental analysis of human behavior: Use and abuse of subjects' verbal reports. *Behavior Analyst,* **11,** 71–75.

Perone, M., Galizio, M., & Baron, A. (1988). The relevance of animal-based principles in the laboratory study of human operant conditioning. In G. Davey & C. Cullen (Eds.), *Human operant conditioning and behavior modification* (pp. 59–85). Chichester, England: Wiley.

Prue, D. M., & Fairbank, J. A. (1981). Performance feedback in organizational behavior management: A review. *Journal of Organizational Behavior Management,* **3,** 1–16.

Rabbitt, P. M. A. (1966). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology,* **71,** 264–272.

Rabbitt, P. (1979). How old and young subjects monitor and control responses for accuracy and speed. *British Journal of Psychology,* **70,** 305–311.

Reynolds, G. S. (1961). Attention in the pigeon. *Journal of the Experimental Analysis of Behavior,* **4,** 203–208.

Reynolds, G. S. (1966). Discrimination and emission of temporal intervals by pigeons. *Journal of the Experimental Analysis of Behavior,* **9,** 65–68.

Rosen, E. (1966). Self-appraisal, personal desirability, and perceived social desirability of personality traits. *Journal of Abnormal and Social Psychology,* **52,** 151–158.

Salthouse, T. A. (1985). Speed of behavior and its implications for cognition. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 400-426). New York: Van Nostrand Reinhold.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review,* **84,** 1-66.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review,* **84,** 127-190.

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

Shimp, C. P. (1981). The local organization of behavior: Discrimination of and memory for simple behavioral patterns. *Journal of the Experimental Analysis of Behavior,* **36,** 303-315.

Shimp, C. P. (1982). On metaknowledge in the pigeon: An organism's knowledge about its own behavior. *Animal Learning & Behavior,* **10,** 358-364.

Shimp, C. P. (1983). The local organization of behavior: Dissociations between a pigeon's behavior and self-reports of that behavior. *Journal of the Experimental Analysis of Behavior,* **39,** 61-68.

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

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

Wasserman, E. A., & Neunaber, D. J. (1986). College students' responding to and rating of contingency relations: The role of temporal contiguity. *Journal of the Experimental Analysis of Behavior,* **46,** 15-35.

Wechsler, D. (1981). *Wechsler adult intelligence scales—revised.* New York: Psychological Corporation.

Winett, R. A., Neale, M. S., & Grier, H. C. (1979). Effects of self-monitoring and feedback on residential energy consumption. *Journal of Applied Behavior Analysis,* **12,** 173-184.

Ziriax, J. M., & Silberberg, A. (1978). Discrimination and emission of different key-peck durations in the pigeon. *Journal of Experimental Psychology: Animal Behavior Processes,* **4,** 1-21.

Received April 14, 1989
Final acceptance December 10, 1989