JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 1990, 54, 173-184 NUMBER 3 (NOVEMBER)

THE UTILITY OF VERBAL AND BEHAVIORAL ASSESSMENTS OF VALUE

DANIEL J. BERNSTEIN AND RENEE L. MICHAEL

UNIVERSITY OF NEBRASKA-LINCOLN

Subjects lived in a laboratory apartment for up to 30 days, engaging in ordinary activities such as reading, sewing, and artwork. The amount of time devoted to each activity was recorded and compared with periodic verbal ratings of the amount of time devoted to the activities. The verbal and observational assessments of the time distribution were very similar, but there were some discrepancies. Based on self-reports and on observation of time actually devoted to the activities, contingencies were arranged in which time devoted to one activity produced time available for a second activity. When the contingency relation was based on behavioral assessment, predictions of time redistribution were more accurate than when the relations were based on verbal assessment. The close correspondence between observed distributions of time and verbally assessed distributions was probably due to the well-specified situation and rigorous assessment methods. Contrary to some cognitive-behavioral accounts, the contingency results suggest that verbal assessment is not necessarily preferable to observation when the two make discrepant predictions. It is suggested that verbal reports might be used more often in behavior analysis in place of lengthy or difficult observations, and attention is drawn to a personality model that parallels important components of behavior analysis.

Key words: reinforcement, time distribution, observation, verbal assessment, correspondence, personality, response deprivation, humans

A behavior-analytic approach to the relationship between verbal and nonverbal measures of behavior assumes no necessary correspondence between them. Skinner's (1945) analysis of the origin of self-descriptive verbal behavior suggests that verbal behavior will correspond to prior nonverbal behavior only if the language community uses agreement with nonverbal behavioral referents as the criterion for reinforcing verbal responding. Similarly, nonverbal behavior will correspond to prior verbal behavior only if agreement with verbal behavior is the criterion used to reinforce nonverbal responding.

Applied behavior analysts have explored the use of language in behavior change procedures (e.g., Brodsky, 1967; Lovaas, 1964). Early studies (e.g., Israel & O'Leary, 1973; Risley & Hart, 1968) focused on the conditions under which correspondence would be maintained, and conceptual development in this area now recognizes potential relations between many different parts of the parallel verbal and nonverbal chains of behavior (Paniagua & Baer, 1982; Rogers-Warren & Baer, 1976). Catania, Matthews, and Shimoff (1982) found that shaped verbal responses maintain better control over nonverbal behavior than does instructed verbal behavior, so a complete account of correspondence must analyze the establishment or synthesis of the verbal responses (cf. Catania, 1983). For example, Risley and Hart (1968) demonstrated that explicit correspondence training generated better correspondence than naturally occurring contingencies, and de Freitas Ribeiro (1989) found the same relationship when reporting a larger class of activities.

Other areas of psychology have also addressed this issue. Correspondence between saying and doing is a central requirement for the effective use of data from attitude surveys. Many social psychologists (e.g., Ajzen & Fishbein, 1977; Wicker, 1969) have reported that verbal behavior often has a poor correspondence with both past as well as future actions, and there has been substantial theoretical speculation on the reasons for the lack of correspondence (e.g., Nisbett & Wilson, 1977).

This research was supported by Grant MH29806-01 from the National Institute of Mental Health to D. Bernstein and by NIH Biomedical Sciences Support Grants RR-07055-09 and RR-07055-10 to the University of Nebraska–Lincoln. Portions of the data from 4 subjects were summarized previously in Bernstein (1986). Correspondence should be addressed to Daniel Bernstein, 209 Burnett Hall, University of Nebraska–Lincoln, Lincoln, Nebraska 68588-0308.

Smith and Miller (1978) suggested, however, that it is more appropriate to identify the conditions under which there will and will not be correspondence than it is to argue whether or not correspondence holds in general. For example, when the object of the question and the time frame are carefully specified, there is ample evidence of good agreement between aggregated self-reports and observation of both past and future actions (Ajzen & Fishbein, 1977; Fishbein, 1980).

Research on the measurement of psychological traits has also generated data on correspondence between self-report and observational measurement. Mischel (1968) concluded that global personality measures have little predictive utility for specific actions related to the traits assessed. In a manner parallel to the change in attitude measures, personality inventories tailored to individual situations and activities demonstrated improved predictive utility (e.g., Goldfried & D'Zurilla, 1969; Goldfried & Sprafkin, 1974). There has been extensive development of measurement strategies among behavioral researchers interested in assessment, and there are situation-specific measures not based on the trait concept (e.g., Ciminero, Calhoun, & Adams, 1977; Cone & Hawkins, 1977; Goldfried, 1976).

Whereas some research on verbal reports identified conditions for maximum correspondence with behavior, other psychologists argued that noncorrespondent verbal reports should provide superior assessments of both behavior and environmental events. Mischel (1973) articulated a cognitive personality model based clearly on social learning (or behavioral) principles, including both antecedents of behavior and contingencies between actions and consequences. The model is cognitive, however, because the predictor variables should be assessed entirely by self-report, without direct observation of behavior or environmental conditions. The use of self-reports might simply reflect an assumption of good correspondence, implying only that verbal measures are a convenient substitute for detailed observation, but Mischel clearly states that subjective perceptions can be different from observations of the same events. Drawing from Rotter's (1954) social learning analysis and Kelly's (1955) personal construct theory, Mischel argues that people transform past behavior and environmental interactions into idiosyncratic perceptions of that behavioral history, and verbal measurement is in principle a better source of these perceptions. In discussing this topic, he provides the following argument:

> On the basis of direct experience, instructions, and observational learning, people develop expectancies about environmental contingencies. . . . Since the expectancies that are learned within a given situation presumably reflect the objective contingencies in that situation, an expectancy construct may seem superfluous . . . [but an] expectancy construct is justified by the fact that the person's expectancies (inferred from statements) may not be in agreement with the objective contingencies in the situation. (Mischel, 1973, p. 269)

A similar position has been taken by Mahoney (1977) and other proponents of cognitive behavior modification.

The present research was an extension of previous studies that focused on contingencies based on the amount of time devoted to ordinary human activities (Bernstein & Brady, 1986; Bernstein & Ebbesen, 1978). Mischel (1973) suggested that the extended baseline observations typical of time-based reinforcement (e.g., Premack, 1965) could be replaced with verbal assessment of the hierarchy of value. If the correspondence is good, then both measures would make similar predictions about the outcome of contingencies. If the two assessments of a pair of activities are different, then each assessment might predict a different result of a contingency between the activities. By adding verbal assessment to the existing human reinforcement procedure, it was possible to determine whether verbal measures could be an accurate substitute for baseline observations and to determine whether verbal measures are preferable when the two disagree.

Operant research on the value of activities has focused on Premack's (1965) observation that the relative amount of time devoted to ordinary activities can predict the outcome of contingencies established between them. As elaborated by Timberlake and Allison (1974) and Timberlake (1980), a time-based model predicts reinforcement only when the contingency schedule is matched with the observed distribution to produce deprivation of the reinforcing activity. Some schedules should produce increases and others should not, depending on the interaction of the specific terms of

the schedule and the values of the two activities involved. Predictions based on such individualized value hierarchies have been confirmed in previous research using long-term observation of human behavior (Bernstein & Brady, 1986; Bernstein & Ebbesen, 1978).

The present research procedures were designed to identify the maximum correspondence between verbal and nonverbal measures that these subjects could generate without explicit correspondence training. Failure to find correspondence may reflect inability of people to make the judgments required or it may reflect poor assessment techniques. Accordingly, Anderson's (1970) functional measurement technique was used because it provides both scale values and a check of the internal validity of the response scale. The goal was to realize the subjects' full capabilities rather than to provide descriptions of typical performance under natural conditions.

In this experiment human subjects lived alone for 30 days in an isolated laboratory apartment, engaging in hobbies or other activities of their choosing. Observed distributions of time were compared with the distribution reported verbally by the subjects. Reinforcement contingencies were arranged between pairs of activities, in which the predictions of results based on the verbal hierarchy were different from the predictions based on the behavioral assessment. There are two central questions addressed by the procedure. First, what kind of correspondence exists between verbal estimates of the amount of time devoted to activities and the observed distribution of time? Second, when verbal and observed measures differ, does the verbal measure represent an idiographic version of the predictor variables that will make better predictions of performance under time-based contingencies?

METHOD

Subjects

Newspaper advertisements offering money for participation in a long-term psychology experiment were used to recruit 5 subjects. Four were female and 1 was male (aged 18 to 29 years), and each was paid $525 for 30 days. Only $10 per day would be paid if the subject left before the end of the experiment. Subjects were screened by a clinical psychologist to prevent potentially vulnerable subjects from entering the experiment; one person who reported hallucinations was excluded. A written participation agreement required observance of restrictions on certain activities, and engaging in a restricted activity was grounds for termination at the $10 per day rate. It was made clear that no other aspects of performance would influence either length of stay or payment.

Apparatus and Living Situation

Subjects lived 24 hr per day in a large (8 m by 8 m) comfortably furnished room with no windows to the outside and a private bathroom with shower. There was a table with two chairs, two desks with chairs, one couch, two arm chairs, two single beds, a refrigerator, and a full complement of cooking, eating, and drinking utensils. The lights were turned on at 9:00 a.m. and off at midnight. Although there was no clock, radio, TV, phone, or mail, all standard services necessary for living (e.g., fresh food, laundry) were provided, and there was always enough material for all activities (e.g., books for reading, paper for drawing). The living area could be viewed through one-way mirrors from an adjoining control room. A two-way intercom, ceiling lights, and temperature of the area were under control of the experimenter. A panel of labeled red lights in the subject's room was used to indicate restrictions on activities, and additional lights were used to signal the end of each day, the arrival of supplies, and "yes" or "no" answers to questions. Most answers could be given with a signal of "yes" or "no," but longer answers were given occasionally on the intercom.

Activity categories. Each subject selected several hobbies to engage in during the time in the laboratory, and time devoted to these activities was observed and recorded. Each response category was defined in terms of body position and contact with appropriate materials, as done in the study by Bernstein and Ebbesen (1978). Table 1 lists the categories for each subject; the categories were mutually exclusive for each subject, with the exception that all subjects could be eating or drinking while engaging in other activities.

Behavioral and Verbal Time Estimates

Observed records of time on activities. Four subjects' responses were recorded using an Es-

Table 1
Activities for long-term subjects.

Subject DM	Subject MH	Subject JP	Subject DB	Subject HH
Reading	Reading	Reading	Reading	Reading
Artwork	Artwork	Artwork	Artwork	Artwork
Slides	Typing	Knitting	Candlemaking	Sewing
Writing	Playing guitar	Needlepoint	Chess problems	Exercise
Sewing	Playing banjo	Writing	Exercise	Crocheting
Puzzles/games	Making a rug	Miscellaneous	Quilting	Writing
Miscellaneous	Needlepoint		Playing drums	Wood printing
	Miscellaneous		Miscellaneous	Miscellaneous

terline-Angus event recorder and a timer that accumulated the total time spent on the responses; for the 5th subject a microcomputer with custom software (Bernstein & Livingston, 1982) was used to generate equivalent data. An observer was present 24 hr per day, but recording was conducted only during the 15 hr when the living area was illuminated. The recording was done by 15 observers in shifts of 2 to 4 hr. Whenever the definition for a response category was met, an observer activated both the event recorder and the timer by pressing a switch; the computer system was activated through the keyboard.

Reliability of observation was assessed between observers using reviews of selected videotaped observation periods. Reliability coefficients were calculated as the percentage of 1-min intervals in which two observers agreed on the activity to be scored. Percentages of agreement for all observers and all responses were above 93% and most were 99%.

Verbal estimates of time on activities. At break points in the procedure, an experimenter entered the living area for about an hour, and subjects sat facing the experimenter at the table in the center of the room. Ratings were made with a sliding scale mounted on a wood frame 45 cm long. A 3-cm board suspended from the top of the frame held a clear 4-cm slide-rule indicator with a red line down the center. The indicator moved along a 20-cm white segment of the top board that was labeled "none of the time" at the left end of the scale and "all of the time" at the right end. Equally spaced 1-mm gradations marked from 0 to 100 on the back of the scale were visible to the experimenter.

Subjects used the sliding scale to rate how much of their time was devoted to all possible pairs of their activities. The activities were presented in pairs so the task could include a combinatorial judgment, which is required for the scaling method. The subjects' task was to add the amounts of time devoted to the two activities and provide a rating of that sum. By using a factorial design, it was possible to verify that subjects were actually adding the time on the activities. If the activities combined additively (i.e., there was no interaction among pairs of activities) then the ratings were valid and the functional measurement design provided scale values from the factorial matrix. An additive data pattern is possible only if the subjects are both following the combination rule and using a valid response scale (see Anderson, 1970, for a complete explanation).

The test of additivity was made on a factorial submatrix made up by combining six or seven activities engaged in by each subject. Once the marginal means from the matrix were calculated, an additive constant derived from the data was subtracted from each. The constant represented the sum of all the comparison activities, so that each remainder was an algebraically pure representation of only one activity. In this manner the scale values retained the ratio properties derived from the labels at the ends of the rating scale and the instructions given for use of the scale. The scale value of each category was divided by the total of all categories, yielding normalized percentage values that summed to 100.

Procedure

The evening before the observation period formally began, each subject moved into the laboratory and learned the procedures for communication. Assurance was given that participation could be terminated at any time, and informed consent for participation was obtained.

Before retiring the first evening, each subject completed a verbal estimate of the amount of time that would be devoted to each of the activities available in the laboratory. Subjects were shown a pair of cards with the names of two of their activities, and they were told to rate the total time devoted to both activities using the sliding scale. To be certain that the instructions were understood, there were three practice trials in which the eight activities were given in four pairs. The total time on the four pairs should sum to "all the time" or 100 scale points, and subjects were given feedback during practice trials until they were within a score range of 90 to 110. Note that feedback was given only on the total amount of time, not on any individual activity. A complete set of ratings consisted of all possible pairs of activities, repeated three times each and given in random order.

Free operant baseline. An assessment of the relative time devoted to the various responses was made during an initial baseline condition lasting 6 to 8 days. There were no restrictions on how subjects could spend their time, and the percentage of time devoted to the responses during baseline was used as the behavioral hierarchy of value. There was also a baseline period following each experimental condition.

Second self-report. Following the initial baseline period, each subject completed another full set of the verbal assessments of the hierarchy using the same instructions. The assessment was done during the first waking hour of the day following the baseline, before the subject was informed of the nature of the next condition.

Restriction. In accord with a time-based account of reinforcement (e.g., Premack, 1965), a high-probability activity was designated as a reward for each subject during the contingency phase of the procedure (the contingent response). To identify any response substitution that might contribute to the instrumental increase (cf. Bernstein & Ebbesen, 1978), the proposed contingent response was restricted and the distribution of the newly available time among the remaining responses was observed for 2 to 4 days. Responses that increased during the restriction were considered substitutable, and these activities were not used as instrumental responses in the contingency conditions. For 4 subjects the activity was totally restricted, and for the 5th (DM) the activity was limited to short periods of access separated by a required pause. The experimenter explained over the intercom that the red light would be on continuously (or periodically for 1 subject), indicating that the designated activity would not be available until the light went off.

Precontingency self-report. At the end of the free-access baseline period preceding each contingency, another full set of verbal assessments was completed. All contingencies were established using the immediately preceding verbal and behavioral hierarchies. For Subject JP's first contingency, the self-report following the first baseline was used as the verbal hierarchy. For Subject DM the precontingency verbal assessment occurred in the middle of a 2-day baseline.

Reinforcement contingencies. A low-probability activity was selected as the instrumental response for each subject, with the condition that it did not substitute for the high-probability contingent response during the restriction. Priority in selection was given to responses that had disparate values in the behavioral and verbal assessments so that the two assessments would make different predictions about the results of the contingencies. For example, suppose that the ratio of the instrumental response over the contingent response is 1:2 by a behavioral assessment and 1:1 by a verbal assessment. If the contingency schedule were also 1:1, then the verbal assessment would predict no reinforcement because the schedule is the same as the operant level ratio (no deprivation). According to the behavioral assessment, however, there would be deprivation and reinforcement should occur. For every one unit of instrumental performance, the subject would gain access to only one unit of the contingent response instead of the two units that were typically used during free access (Timberlake, 1980; Timberlake & Allison, 1974). These contrasting predictions can be made only if the verbal and behavioral assessments provide appropriate disparities, and it was not possible to find an equal number of examples that favored each assessment method. A total of nine contingencies were conducted; in six cases the behavioral assessment predicted reinforcement and the verbal assessment predicted no increase, and in three cases the predictions were opposite.

The amount of time available for the con-

Table 2

Order and duration (in days) of conditions and timing of self-report assessments for each subject.

Conditions	Subjects				
	DM	MH	JP	DB	HH
Self-report 1					
Baseline 1	8	6	8	8	8
Self-report 2					
Restriction 1	4	2	2	4	2
Baseline 2	4	4	4	4	4
Self-report 3					
Contingency 1	6[a]	5	3	4	4
Baseline 3	2	4	3	3	2
Self-report 4					
Contingency 2	3	4	3	3	4
Baseline 4	3	4	4	2	3
Self-report 5					

[a] Subject DM experienced a procedural deviation during this phase. Her data from this phase are not used in the analyses.

tingent response was a constant proportion of the time devoted to the designated instrumental response, so that increases in time on the first activity produced proportional increases in access to the restricted response. When a restriction was in force, a minimum amount of instrumental responding was required to remove the restriction and turn off the red light. Access to the contingent response did not have to be used immediately or all at one time. The subject could engage in the response at several different times for short durations or use the allotment of time all at once. Time devoted to the instrumental response after the red light went off accumulated additional time for the contingent response, and the subject could return to the instrumental response at any time and accumulate further access to the contingent response. As long as the total time on the contingent response did not exceed the amount earned, each subject could alternate between the two responses in any pattern. There was no limit on the amount of time that could be accumulated for the contingent response. When the earned time was used, the red light was turned on and the minimum instrumental performance was again required to remove the restriction.

Whenever the red light was turned on or off, the overhead room lights were dimmed for 1 or 2 s to alert subjects to the change. At the beginning of each contingency condition, the experimenter told the subjects over the intercom that they could have access to a given activity after engaging in another activity. The two activities were described and examples were given. Subjects were not told the duration of the instrumental requirement or the proportion of contingent responding earned for instrumental responding. They were told that the light would go off when they had satisfied the instrumental requirement and that it would go on again when they had used the earned amount of the contingent response. Finally, the subjects were told that they were not required to remove the restriction and could engage in other activities instead. Table 2 gives the order and duration of the conditions for each subject and the timing of the self-report assessments of the hierarchies of activities.

Final self-reports. Following the final free-access baseline period, each subject completed a full set of the verbal assessments during the last day of each subject's stay in the laboratory. When all self-reports were completed each subject was thoroughly debriefed, the experiment was discussed in detail, and all questions were answered.

RESULTS

The first form of analysis involved deriving scale values from the verbal assessment and comparing them with estimates of time distribution from observations. The second form of analysis involved examining the results of the contingencies comparing different forms of assessment.

Verbal Assessment

Scale values. Table 3 gives the verbal and observed time distributions for the active categories of all subjects during free-access baselines. The matrix of ratings for each subject yielded a normalized hierarchy of verbal estimates of time for each free-access baseline that summed to 100%. For all subjects, the verbal ratings of the "miscellaneous" category (cleaning, cooking, bathing, and relaxing) were consistently lower than the amount of time devoted to that class of activities. Even with the generally inaccurate preexperiment ratings excluded, the mean of the subjects' 19 self-reports of miscellaneous activities was 17% of each day (range, 12% to 22%). The mean percentage of time actually devoted to miscellaneous activities during 19 free-access base-

Table 3

Percentages of time devoted to each activity in free-access baseline for all verbal and behavioral assessments. PreSR is the self-report measure taken prior to Day 1 of the experiment. B1, B2, B3, and B4 are the observed measures of behavior for each of the four baselines, and SR1, SR2, SR3, and SR4 are the self-report measures taken immediately following each of the four baselines, respectively. N/A refers to missing data.

Activity	PreSR	B1	SR1	B2	SR2	B3	SR3	B4	SR4
Subject DM									
Reading	16	30	19	53	20	43	22	22	22
Slides	25	9	15	9	12	1	10	44	12
Writing	10	22	17	24	20	19	18	4	18
Sewing	17	8	16	0	9	0	8	0	9
Puzzles/games	7	7	7	3	8	8	12	4	12
Artwork	10	4	5	0	7	0	6	0	7
Subject MH									
Typing	17	6	17	7	15	7	21	6	12
Playing guitar	11	8	6	2	8	5	9	5	8
Playing banjo	6	8	14	1	6	2	5	0	4
Making a rug	5	14	15	5	12	21	16	7	11
Needlepoint	4	0	1	0	2	0	1	1	1
Artwork	2	13	15	6	11	1	7	21	7
Reading	15	17	4	46	15	40	17	19	12
Subject JP									
Reading	19	33	16	29	N/A	36	23	32	16
Knitting	12	12	22	20	N/A	20	22	23	17
Artwork	20	5	7	3	N/A	4	11	10	13
Needlepoint	8	24	21	23	N/A	14	20	0	14
Writing	19	8	14	5	N/A	5	10	3	9
Subject HH									
Reading	4	2	4	2	11	1	9	5	7
Sewing	20	16	8	16	15	16	12	10	12
Exercise	8	4	5	3	5	2	6	3	5
Crochet	12	20	12	18	14	19	13	21	14
Artwork	6	5	6	5	6	7	8	8	8
Woodcutting	7	7	7	7	7	4	7	2	6
Subject DB									
Candlemaking	6	4	6	7	6	9	5	4	7
Chess problems	8	11	7	9	7	2	6	2	6
Exercise	4	2	4	5	4	3	4	1	4
Sewing	5	3	5	2	5	4	7	3	5
Artwork	16	9	11	4	14	10	10	7	11
Drum playing	18	15	15	18	16	5	11	8	12
Reading	7	16	12	15	8	20	12	24	14

line periods was 34% of each day (range, 17% to 47%). The mean self-report was only half of the actual value, and there was little overlap in the two distributions. Because the sum of the verbal assessments of time devoted to the active categories was too high, the sum of the verbal assessments of the activities was matched to the observed behavioral sum, preserving the proportions of the time devoted to each of the active categories. This transformation does not alter any predictions made about reinforcement contingencies, because those predictions are based on ratios among the activities. The values reported reflect both the forced normalization of the total and the constraint on the miscellaneous category.

Correspondence with observed time. Comparisons were made between each verbal hierarchy and the actual time devoted to the activities during the free-access baseline period just completed. In addition, the first baseline period was also compared with the verbal assessment taken before the experiment began. Table 4 gives the Pearson's correlations be-

Table 4

Correlations between verbal and behavioral measures of time spent on activities. Values are placed between headings to indicate the two measures correlated. The final column of each row shows the total correlation between all self-report measures and all behavioral observations for each subject. PreSR is the self-report measure taken prior to Day 1 of the experiment. B1, B2, B3, and B4 are the observed measures of behavior for each of the four baselines, and SR1, SR2, SR3, and SR4 are the self-report measures taken immediately following each of the four baselines, respectively. N/A refers to missing data.

	Pre SR	B1	SR1	B2	SR2	B3	SR3	B4	SR4	Total
Subject HH	.79	.94	.92	.79	.76	.90	.87	.93	.81	
Subject DM	−.10	.49	.55	.89	.86	.94	.19	.29	.65	
Subject MH	.12	.26	−.31	.60	.64	.62	.31	.48	.50	
Subject JP	−.26	.46	N/A	N/A	N/A	.87	.63	.72	.65	
Subject DB	.59	.89	.76	.52	.15	.64	.79	.84	.71	
All subjects	.33	.64	.31	.70	.64	.76	.45	.54	.64	

tween the verbal and behavioral measures of time distribution. The correlations between the preexperimental assessment and the first baseline were quite good for 2 subjects and poor for the other 3. Three subjects did not generate an accurate verbal hierarchy before having specific behavioral experience. In contrast, the correlations between the first baseline and the immediately following verbal assessment were uniformly good. All 5 subjects' first retrospective verbal assessment corresponded with behavior better than did their preexperimental assessments.

This general pattern continued throughout the experiment, such that each verbal assessment was likely to have a higher correlation with the baseline observations that preceded it than with the nearest baseline following it. Table 4 shows 13 instances of a correlation between an individual subject's verbal assessment (SR1, SR2, or SR3) and baseline observations both preceding and following, and in 10 of those cases the correlation is higher for retrospective verbal assessment. The combined correlations for all subjects show the same pattern; SR1 correlates .64 with preceding and .31 with following baselines, for SR2 the equivalent correlations are .70 and .64, and for SR3 they are .76 and .45. Finally, the overall correlation for all verbal assessments, including prospective and retrospective baselines, was .64, indicating that the behavioral hierarchy of value was well represented by the verbal assessment.

Note that the correlations represent comparisons of verbal and observed hierarchies of value for many responses combined, sometimes observed over several baseline periods, and occasionally with all subjects combined into a single correlation. Lord (1953) and Hays (1973) note that independence assumptions are critical for hypothesis testing but not for descriptive statistics. Given the consistency of individual subjects' data, aggregation of data can be used to describe the overall order in the correlations.

Test for additivity. A 3 × 3 factorial subset of each of the 5 subjects' functional measurement ratings of the time devoted to the activities was checked for additivity. Separate analysis of variance for each subject's data matrix indicated that 4 subjects' ratings were additive, providing evidence for the internal validity of the scale values derived. Subject MH's ratings had an interaction between the two sets of activities, $F(9, 169) = 7.47$. A graph of the interaction revealed that one activity was so low in value that all combinations with it were rated with a nearly constant low total time. Because the overall pattern of the data for the 5 subjects was clearly additive, the ratings were taken as an internally valid verbal scale of time devoted to each activity. Given the pattern of correlations with observations of time devoted to the activities, the ratings were also taken as having substantial external validity.

Outcome of the Behavioral Contingencies

For each of the contingencies that were actually implemented on the subjects' behavior, the behavioral and verbal hierarchies made opposite predictions about the results—one predicting an increase, the other predicting no change. In general, the behavioral hierarchies

Table 5

Results of contingencies placed on behavior of long-term subjects. % change is the increase in instrumental performance as a percentage of baseline.

Subjects and predictions	Verbal baseline ratio	Behavior baseline ratio	Contingency ratio	Baseline instrumental performance	Contingency instrumental performance	% change
Behavioral hierarchy predicts increase						
MH-1	.57/1	.07/1	.60/1	.03	.12	300
JP-1	.25/1	.07/1	.16/1	.03	.08	166
JP-2	.5/1	.22/1	.5/1	.04	.12	200
DB-2	.97/1	.25/1	1/1	.05	.07	40
HH-1	.63/1	.06/1	.31/1	.02	.23	1,050
DM-2	.33/1	0/1	.33/1	.00	.11	N/A
Behavioral hierarchy predicts no increase						
MH-2	.83/1	4/1	4/1	.26	.31	23
DB-1	.6/1	3.5/1	1.5/1	.14	.27	93
HH-2	.65/1	1.5/1	1.5/1	.18	.18	0

made more accurate predictions of the contingency results than did the verbal hierarchies. In seven of the nine cases, the results were in the direction predicted by the behavioral hierarchy. Table 5 gives the verbal and behavioral baseline ratios, the schedule ratio actually used, and the performance of the instrumental response for each contingency. During contingencies expected to produce an increase according to the behavioral hierarchy, subjects devoted an average of 8.3% (range, 2% to 21%) more of their day to the instrumental responses. During the contingencies in which the verbal hierarchy predicted an increase, subjects devoted an average of 6% (range, 0% to 13%) more of their day to the instrumental responses.

The absolute size of the increases is slightly misleading, however, because the two sets of instrumental responses had very different baseline levels. Because there were few large discrepancies between verbal and observed value, it was difficult to find contingency ratios that made differential predictions; there were often few choices of activities available for contingencies. When the instrumental increase is calculated as a percentage of the baseline level, the contingencies based on behavioral baselines increased an average of 265%, whereas those based on verbal baselines increased an average of 33%. It is interesting to note that both of the errors made by the behavioral prediction were increases in the instrumental response observed when none was expected, and overall eight of the nine contingencies produced increases in the instrumental response, regardless of the predictions.

DISCUSSION

In general, there was very good correspondence between the verbal measure of value and the actual distribution of time to activities. A structured psychophysical procedure within a specific self-report context produced reasonably high correlations between subjective estimates and observed times. This result suggests that there is more correspondence training inherent in interactions within the language community than many behavior analysts normally assume, although this view is congruent with the successful verbal measurement strategies developed by researchers in behavioral assessment. The capacity for accurate reporting may not be part of the typical verbal repertoire of most people, but it is possible to describe one's recent actions accurately when the response format offers a manageable frame of reference for the questions. One implication of this finding is that behavior analysts should be more cautious about rejecting research results simply because subjects' behavior is assessed with verbal measures.

The results reveal a few systematic discrepancies between the verbal and observed hierarchies, and those results fit a consistent pattern related to the development of a self-descriptive verbal repertoire. Skinner (1945) suggested that people use their own behavior as the primary referent for descriptions of pri-

vate events because the language community uses those same public actions as the criterion for reinforcing accuracy. Bem (1967, 1972) elaborated this point of view in social psychology, arguing that many forms of self-perception are driven by observation of public action. One relevant form of discrepancy in the present data is the consistent verbal underestimation of the amount of time devoted to the miscellaneous category. A conventional explanation based on the study by Kahneman and Tversky (1973) holds that subjects give less weight to miscellaneous activities in their predictions because there were no representative instances of the category. A behavioral account of self-perception is also congruent with this finding, because speakers would not likely receive precise feedback on labeling a category without clear referent instances.

A second interesting discrepancy in the data was the inability of 3 subjects to make an accurate estimation of their time distribution before they actually engaged in the activities in the laboratory. A behavioral account assumes that accurate labeling results from reinforcement from the language community for correspondence between self-descriptions and prior actions, so a person who has not behaved in a particular situation would be less likely to give an accurate estimate of how time would be distributed among alternatives. Bem's (1972) self-perception account suggests that verbal reports reflect previous actions rather than cognitive plans for future action, and that notion is supported by a third pattern in the present data. Correlations between each verbal assessment and the preceding baseline were consistently higher than the correlations with the following baseline, suggesting that people use prior actions more than expectations about future conditions as the referent for self-perception.

Although there was generally good correspondence between verbal ratings of time and the actual time distribution, the few discrepancies that did occur provide contrasting predictions about the outcome of contingency arrangements. Mischel's (1973) personality model anticipates discrepancies between environmental events and their subjective representations because of subjects' use of personal constructs in forming perceptions, and an idiosyncratic, verbal assessment should provide a better estimate of each individual's reactions to a given contingency situation. The present contingency results suggest that behavioral observation is at least as good a predictor of contingency performance as the subjective representation of value, and one might argue that observed baselines made better predictions. Under the conditions of this experiment, the untransformed observation of the controlling variable makes an accurate prediction even when it differs from the perceived value of the same variable.

This conclusion must be considered tentative because there were instrumental increases in two of the three contingencies in which observed baseline values predicted no response deprivation. Variables other than response deprivation may also have contributed to instrumental performance in all of the contingencies. The overall magnitude of the effect of the contingencies was larger under one set of conditions, however, and a reasonable assumption is that the difference was due to a larger contribution from response deprivation when observed behavior was used for the estimate of baseline value. It should also be noted, however, that the comparison was based on the few cases in which there were discrepancies, and verbal assessment would have made accurate predictions in the many cases in which the two measures were similar.

Given that systematic verbal assessments of time distribution are reasonably accurate, they could be useful in practical settings as a substitute for lengthy observation of behavior. For example, Welsh, Bernstein, and Luthans (in press) used verbal assessment as a substitute for baseline observations of job preference when applying response deprivation contingencies in a business setting. It is also possible that verbal measures could be used in complex circumstances in which there is no observable behavioral or environmental referent available. One might view Mischel's (1973) personality model as a verbally assessed substitute for a behavioral analysis of the person's reinforcement history. By tapping the person's recognition of the combined influences of previous history, the difficult and often impossible task of identifying that history might be finessed. To the extent that there is natural or planned correspondence training, such reports could give a behavior analyst a good bit of information about the environment in which the target person's behavior occurs. This model of

personality draws heavily on behavior-analytic elements (e.g., contingencies, consequences, elicitors, rule-governed behavior) and is closer to an operant account of individual differences than is Harzem's (1984) discussion based on types of people who respond differently to contingencies.

Mischel's (1973) focus on idiosyncratic perception of environmental events is not without precedent in behavior analysis. An idiosyncratic hierarchy of value based on percentage of time devoted to activities is also central to Premack's (1965) theory of reinforcement. Instead of rejecting the notion that individuals perceive the environment differently, research from a behavioral perspective could identify the conditions that influence the perception of behaviorally relevant variables. For example, Wasserman, Chatlosh, and Neunaber (1983) studied the relations between actual contingencies and subjects' perceptions of contingent relations. Mischel's (1973) personality model adopts a behavioral framework to guide the identification of verbal measures that can predict behavior, and verbal assessment will certainly be a desirable and important tool in behavior analysis. Although the model is expressed in mentalistic terms, it is a useful first step in the analysis of the determinants of individual performance. Whenever individual verbal accounts of controlling variables differ from observations of those variables, it will be necessary to analyze the sources of the verbal behavior as a replacement for Mischel's unspecified cognitive transformations. Future research in this area could expand Wasserman et al.'s (1983) treatment of perception of contingencies, perhaps integrating it with recent work on rule-governed behavior.

In principle, the relationship between behavioral and verbal assessments can vary from total independence to complete isomorphism. In the area of relative value of activities, most people have many experiences involving time devoted to one activity (work) that leads to access to other, more valued activities (fun). With this natural training as background, it was possible to use well-developed verbal assessment procedures to produce results that parallel behavioral data. The results do not imply that all people can provide valid self-reports on any topic. Instead, given careful assessment procedures and proper training, people may be able to provide more valid self-reports than has been assumed in behavioral research. When conditions maximize correspondence between verbal and nonverbal measures of reinforcement value, reinforcement operations function as well as or better than subjective perceptions of the same operations. Verbal reports can be an orderly reflection of behavior under most circumstances and need not result from idiosyncratic transformations that qualitatively alter the nature of the controlling event. To the extent that this finding holds in general, it represents an important tool for the analysis of behavior that is difficult to observe directly.

REFERENCES

Ajzen, I., & Fishbein, M. (1977). Attitude–behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin,* **84,** 888–918.

Anderson, N. H. (1970). Functional measurement and psychophysical judgment. *Psychological Review,* **77,** 153–170.

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review,* **74,** 183–200.

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 2–62). New York: Academic Press.

Bernstein, D. J. (1986). Correspondence between verbal and observed estimates of reinforcement value. In P. N. Chase & L. W. Parrott (Eds.), *Psychological aspects of language: The West Virginia lectures* (pp. 187–205). Springfield, IL: Thomas.

Bernstein, D. J., & Brady, J. V. (1986). The utility of continuous programmed environments in the experimental analysis of human behavior. In H. W. Reese & L. J. Parrott (Eds.), *Behavior science: Philosophical, methodological, and empirical advances* (pp. 229–243). Hillsdale, NJ: Erlbaum.

Bernstein, D. J., & Ebbesen, E. B. (1978). Reinforcement and substitution in humans: A multiple-response analysis. *Journal of the Experimental Analysis of Behavior,* **30,** 243–253.

Bernstein, D. J., & Livingston, C. (1982). An interactive program for observation and analysis of human behavior in a long-term continuous laboratory. *Behavior Research Methods & Instrumentation,* **14,** 231–235.

Brodsky, G. (1967). The relation between verbal and non-verbal behavior change. *Behaviour Research and Therapy,* **5,** 183–191.

Catania, A. C. (1983). Behavior analysis and behavior synthesis in the extrapolation from animal to human behavior. In G. C. L. Davey (Ed.), *Animal models of human behavior* (pp. 51–69). Chichester, England: Wiley.

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.

Ciminero, A. R., Calhoun, K. S., & Adams, H. E. (Eds.).

(1977). *Handbook of behavioral assessment.* New York: Wiley.

Cone, J. D., & Hawkins, R. P. (1977). *Behavioral assessment: New directions in clinical psychology.* New York: Brunner/Mazel.

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.

Fishbein, M. (1980). A theory of reasoned action: Some applications and implications. In M. M. Page & H. E. Howe, Jr. (Eds.), *Nebraska symposium on motivation: Vol. 27. Beliefs, attitudes, and values* (pp. 65–116). Lincoln: University of Nebraska Press.

Goldfried, M. R. (1976). Behavioral assessment. In I. B. Weiner (Ed.), *Clinical methods in psychology* (pp. 281–330). New York: Wiley.

Goldfried, M. R., & D'Zurilla, T. J. (1969). A behavioral-analytic model for assessing competence. In C. D. Spielberger (Ed.), *Current topics in clinical and community psychology* (Vol. 1, pp. 151–196). New York: Academic Press.

Goldfried, M. R., & Sprafkin, J. N. (1974). *Behavioral personality assessment.* Morristown, NJ: General Learning Press.

Harzem, P. (1984). Experimental analysis of individual differences and personality. *Journal of the Experimental Analysis of Behavior,* **42,** 385–395.

Hays, W. L. (1973). *Statistics for the social sciences.* New York: Holt, Rinehart & Winston.

Israel, A. C., & O'Leary, K. D. (1973). Developing correspondence between children's words and deeds. *Child Development,* **44,** 575–581.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review,* **80,** 237–251.

Kelly, G. A. (1955). *The psychology of personal constructs* (2 vols.). New York: Norton.

Lord, F. M. (1953). On the statistical treatment of football numbers. *American Psychologist,* **8,** 750–751. (Comment)

Lovaas, O. I. (1964). Control of food intake in children by reinforcement of relevant verbal behavior. *Journal of Abnormal and Social Psychology,* **68,** 672–678.

Mahoney, M. J. (1977). Reflections on the cognitive-learning trend in psychotherapy. *American Psychologist,* **32,** 5–13.

Mischel, W. (1968). *Personality and assessment.* New York: Wiley.

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review,* **80,** 252–283.

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

Paniagua, F. A., & Baer, D. M. (1982). The analysis of correspondence training as a chain reinforceable at any point. *Child Development,* **53,** 786–798.

Premack, D. (1965). Reinforcement theory. In D. Levine (Ed.), *Nebraska symposium on motivation* (Vol. 13, pp. 123–180). Lincoln: University of Nebraska Press.

Risley, T. R., & Hart, B. (1968). Developing correspondence between non-verbal and verbal behavior of preschool children. *Journal of Applied Behavior Analysis,* **1,** 267–281.

Rogers-Warren, A., & Baer, D. M. (1976). Correspondence between saying and doing: Teaching children to share and praise. *Journal of Applied Behavior Analysis,* **9,** 335–354.

Rotter, J. B. (1954). *Social learning and clinical psychology.* Englewood Cliffs, NJ: Prentice-Hall.

Skinner, B. F. (1945). The operational analysis of psychological terms. *Psychological Review,* **52,** 270–277.

Smith, E. R., & Miller, F. D. (1978). Limits on perception of cognitive processes: A reply to Nisbett and Wilson. *Psychological Review,* **85,** 355–362.

Timberlake, W. (1980). A molar equilibrium theory of learned performance. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14, pp. 1–58). New York: Academic Press.

Timberlake, W., & Allison, J. (1974). Response deprivation: An empirical approach to instrumental performance. *Psychological Review,* **81,** 146–164.

Wasserman, E. A., Chatlosh, D. L., & Neunaber, D. J. (1983). Perception of causal relations in humans: Factors affecting judgments of response-outcome contingencies under free-operant procedures. *Learning and Motivation,* **14,** 406–432.

Welsh, D. H. B., Bernstein, D. J., & Luthans, F. (in press). Application of the Premack principle of reinforcement: Analysis of the impact on quality performance of service employees. *Journal of Organizational Behavior Management.*

Wicker, A. W. (1969). Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues,* **25**(4), 41–78.

Received December 22, 1989
Final acceptance June 13, 1990