The Behavior Analyst 1988, 11, 51–58 No. 1 (Spring)

Laboratory Lore and Research Practices in the Experimental Analysis of Human Behavior: Designing Session Logistics—How Long, How Often, How Many?[1]

Daniel J. Bernstein
University of Nebraska–Lincoln

My laboratory research has attempted to generalize principles derived from research with discrete, brief responses such as key pressing to more natural examples of human behavior such as reading, sewing, and artwork. As a part of this process, some variations on conventional experimental practice have emerged in my research and in that done by others who analyze human behavior. This essay presents strategies for adapting procedural conventions to whatever human behavior is the object of study. Criteria for selecting length, frequency, and number of sessions are discussed in that order, along with some limitations inherent in human research. Some of the implications of this lore are most useful with naturalistic repertoires, but suggestions for all human operant research are made as well.

HOW LONG SHOULD EACH SESSION BE?

The general goal of designing session logistics should be to assure a behavioral sample that is large enough to be both reasonably stable across sessions and representative of the response repertoire studied. No simple statement can define session logistics that will maximize the likelihood of stable data.

Current Conventions and Practices

Virtually all laboratory research with human participants is done in short sessions of approximately one hour. Within the experimental analysis of human behavior this tradition matches practices typical of both research with nonhuman subjects and social psychological studies with humans. Researchers often assume that participants can only respond with much intensity for an hour or so, and this practice is also consistent with the required research participation plans included in psychology classes. In most cases the participants engage in activities with brief, discrete topographies, and an hour of behavior represents a large sample of individual instances. In addition, the reinforcement schedules typically used in human research make contact with the participant's behavior many times during each session. Perhaps because the practice is nearly universal, there is little common lore on the utility of short repeated sessions.

Experience from Research on Reinforcement Value

For behavior having a more complex topography typical of responding outside the laboratory, however, longer sessions may be needed. If a typical burst of each activity is 20 or 30 minutes in length, then an adequate sample of that responding will require many hours of observation. If more than one response class is being observed, the time necessary to collect a sample of responding increases with each additional response. In addition, researchers commonly observe warm-up effects within sessions, as subjects adapt to experimental conditions. Some response repertoires have systematic cycles of preference, so any session smaller than an entire cycle will not adequately sample the range of behavior produced by the experimental conditions.

My own work has involved the observation of ordinary human activities such as reading, sewing, handwork, or exercise (e.g., Bernstein & Ebbesen, 1978), and the first experiment was to be completed

[1] Reprints may be obtained from the author, Psychology Department, 209 Burnett Hall, University of Nebraska at Lincoln, Lincoln, NE 68588-0308.

in thirty hours (two 15-hour observation sessions on consecutive days). A baseline of six hours was to be followed by six hours of a contingency and three hours of baseline on each day, with the second day providing a replication of the results. The subject's distribution of time among the activities during the second three hours of baseline was entirely different from the distribution during the first three hours, so the baseline was extended. Ultimately, the baseline condition continued through the entire thirty hours, as the only stability to be found was across entire fifteen-hour days of observation. Each day's activities had a distinct and similar pattern, which no smaller sample could accurately represent. Preferences within a repertoire of natural responses are not constant from hour to hour, so session length must be large enough to average across several occurrences of the target activities if point to point stability is important to the analysis.

A systematic empirical strategy can identify an appropriate session length for a particular response repertoire. Using repeated observations of the subject's behavior, individual session duration can be varied to identify a session length sufficiently long that all response categories have occurred with an established minimum frequency (Bernstein, 1982). This procedure could eliminate short session lengths in which the person does not engage in the entire range of available behavioral options. Other criteria for an adequate sample can also be applied, such as when the proportion of time devoted to each of the activities (or the average duration of bursts of each activity) reach asymptotic values.

An alternative strategy is to make a very long observation record that likely includes several alternations of preference among the responses. Different sized portions of this recorded stream can be examined for the necessary properties of a sufficient sample of data as was done by Kraemer, Alexander, Clark, Busse, and Riss (1977) and recommended by Johnston and Pennypacker (1980).

Session length should not be selected by simply following a convention from laboratory social psychology or nonhuman operant research. When designing an experiment with human participants, the nature of the repertoire being studied should be considered. While the conventions from free-operant nonhuman research will sometimes hold, the decisions should come from an examination of the behavior being studied and the likely range of effects that might be observed.

HOW OFTEN SHOULD SESSIONS BE HELD?

Ideally experimental procedures should be continuously in effect to maximize the percentage of variability accounted for by systematic variables. Nonhuman subjects have lived twenty-four hours per day under experimental conditions (e.g., Collier, Hirsch, & Kanarek, 1977; Kavanau, 1969), and most conventional operant experiments implicitly include continuous experimental control (e.g., pigeons do not leave the laboratory to roost, eat, and peck between sessions). Total access to activities and food are constrained twenty-four hours per day, even though data are collected during only a brief portion of that time.

Current Conventions and Practices

Most research is conducted using short sessions but without any control over the participants' activities outside the laboratory. This practice formally mimics the conventions of nonhuman research, but it fails to include a substantive part of the procedure, namely control of the "home cage." The convenience of introductory student subject pools tends to maintain this strategy. An additional problem with intermittent participation is that people require repeated adaptation to the experimental context due to long periods between sessions. Warm-up effects can occur at the beginning of sessions, and it is more difficult to create laboratory variables that can compete with the large impact of uncontrolled variables in the natural environment. Keeping intersession intervals as short as possible minimizes this problem.

Experience from Research on Reinforcement Value

My own research sessions typically run continuously, with fifteen hours of observation and nine hours of sleep per day. After several years of research using this format, I attempted to increase the laboratory's productivity by having two people in the lab for six hours each. With two complementary participants each day the amount of time (and the cost) to complete an experiment would be cut in half. During the eighteen hours a day that each person was not in the experiment, she went about her life in her natural environment. One of these women was the first person I encountered whose instrumental performance was completely indifferent to deprivation of the reinforcer (reading magazines) during a contingency, and I discussed this issue with her during the debriefing following the experiment. When I asked if her normal routine had changed during her home time, she mentioned that sometimes she read more books than usual. She had not been aware of substituting one kind of reading for another before I probed, but the timing of her increased reading at home corresponded exactly to the deprivation conditions of the experiment.

Now when I want participants to spend less than fifteen hours per day on a particular set of activities, they still remain in the lab twenty-four hours per day. The day is divided between the activities they chose to bring into the laboratory and a standard set of activities I provide. The time division is adjusted to fit the needs of the experimental design, and in some cases two separate but parallel experiments have been conducted on the two sets of activities. This procedure allows for noncontinuous sessions without adding the problem of either repeated adaptation or extraneous effects due to activity outside the laboratory between sessions.

In general, human operant research will not be done with populations available for participation or control twenty-four hours per day, even though that would be ideal. Researchers should schedule sessions to minimize the problems that may result from intermittent participation. When sessions are scheduled as closely together as possible (e.g., multiple sessions per day), the experimental variables will likely have greater continuity and suffer less from disruption by extraneous variables such as adaptation to the setting or interference by activities from outside the laboratory.

HOW MANY SESSIONS IN A CONDITION OR AN EXPERIMENT?

Having many sessions in each condition is essential for providing stable data for the within-subject experimental designs used by researchers. Human research designs are logically equivalent to designs for nonhuman subjects, and design conventions (e.g., Hersen & Barlow, 1976; Johnston & Pennypacker, 1980; Sidman, 1960) recommend many sessions per condition to demonstrate the replicability of effects. In addition, the nature of human performance during the acquisition of a task is often qualitatively different from performance on the task at steady state, providing an additional reason for having many experimental sessions in a research plan.

Current Conventions and Practices

Most human operant research programs continue experiments beyond the duration of the one-hour human procedures typical of social psychology, demonstrating the difference between analysis of acquisition and steady state performance. For example, Burgess (1968) examined group communication patterns with a procedure that ran for fifteen hours instead of the usual one hour. Whereas most researchers found differences in problem solving as a function of communication patterns, Burgess noted that these effects disappeared after several sessions. At steady state there were no differences in performance, suggesting that conducting too few sessions led the researchers to mistake transitory acquisition effects for fundamental group phenomena. Research on cooperation and

competition by Hake (e.g., Hake, Olvera, & Bell, 1975; Hake & Schmid, 1981) and Schmitt (e.g., Marwell & Schmitt, 1975; Schmitt, 1976) extended the duration of experiments from the one hour prisoner's dilemma games typical of social psychology to multiple-session operant procedures. These research programs went beyond brief transitory effects, and in many cases the variables most important in short experiments were not those that controlled behavior over many sessions.

The problems due to a small number of experimental sessions are especially acute when dealing with the interaction between language and behavior. Subjects bring an undocumented history of verbal behavior with regard to the general issue of contingencies and consequences, and we provide them with brief exposure to a specific set of contingencies and consequences. Not surprisingly, however, laboratory variables sometimes have less effect than an extensive verbal history. Some researchers run more sessions so that the experimental variables can compete with the initial verbal control.

Experience from Research on Instructed Performance

Currently many operant researchers are examining the relative contributions of rule-governed and contingency-shaped responding in humans (Bentall, Lowe, & Beasty, 1985; Catania, Matthews, & Shimoff, 1982). For an appropriate comparison of the two classes of responding, however, subjects must have a significant history with the contingencies (cf. Weiner, 1964) so they can compete with existing verbal control. Some of the experiments that compare rule-governed and contingency-shaped behavior are run with relatively few sessions. For example, subjects in the research of Catania, Shimoff, and Matthews (1982; Matthews, Catania, & Shimoff, 1985; Shimoff, Catania, & Matthews, 1981) typically participate for only 6 or 8 sessions. These researchers demonstrated powerful and interesting effects of prior verbal histories, and their subject's responding was insensitive to substantially different schedules in the two components of a multiple schedule.

During an extended visit in Catania et al.'s lab, I was involved in many discussions of participants' apparent insensitivity to experimental variables. Eliot Shimoff (personal communication) described repeated failures to obtain any decrease in pressing during an extinction component of a multiple schedule, even after as many as 6 one-hour sessions. In those discussions, Terje Sagvolden and I urged that the procedure should be run longer to see if the verbal control would hold up.

I recently replicated that study to see how long the verbal history would dominate current contingency control. I used a multiple schedule procedure identical to that of Catania, Matthews, and Shimoff (1982), initially comparing ratio and interval components. After three 45-minute sessions, one person showed identical high rates in both components, replicating Catania et al.'s result. Next, she was exposed to seven 45-minute sessions of a multiple schedule with random ratio and extinction components, and gradually the response rates separated until the rate on extinction was one-third the rate for the ratio component. Finally, there were two sessions with the original multiple schedule, and although the rates were more variable, the interval rate was only one-third the ratio rate. A second participant was exposed to twelve sessions of the original multiple schedule, and the rates separated to the same level even without an extinction history.

With a minor extension of the original procedure, the experimental variables competed effectively with the verbal history brought into the laboratory. The power of rule-governed behavior is clearly great, but its role should be studied with extended experimental analysis (cf. Baron, Kaufman, & Stauber, 1969; Bentall et al., 1985; Lowe, 1979). It is unlikely that the *Journal of the Experimental Analysis of Behavior* would publish research in which naive pigeon subjects were run in several conditions during a six-hour experiment, but other human studies (in addition to the research on

rule-governed responding) have appeared with very few sessions (e.g., Case, Fantino, & Wixted, 1985; Fantino & Case, 1983). Human research is greatly needed, but it will be of limited value if we ignore the conventions of experimental design because our participants are human. If anything, a case could be made that human experiments should run longer than nonhuman studies because of instruction effects and the absence of control over the home environment. If practical difficulties limit research on instructional control to relatively brief experiments, we may overestimate the impact of instructions and underestimate the importance of environmental variables such as contingencies.

In my long-term laboratory, one subject inadvertently demonstrated the importance of the duration of the procedure. At the beginning of a contingency condition, this subject performed an instrumental response that could produce access to a restricted activity, and then she ignored both of the activities in the contingency. Since there were several additional activities available, this was one of her options. She proceeded quite well for a while, but after ten hours she started changing activities more and more frequently. Finally, after fourteen hours, she engaged in the contingent response until it was restricted again, and then she returned to the instrumental response. Her performance over the next forty-five hours of that condition was quite regular, and the recovery of baseline was as expected. It is worth noting that the first session of the contingency condition was longer than entire experiments in some studies of rule-governed behavior, and the experimental variables operating in this study only began to take effect at the end of the first session.

Human behavior is under the control of a more complex combination of variables than is the behavior of our captive experimental animals. Progress in the analysis of human behavior will not likely be made with procedures that are simpler and shorter than those used in the analysis of nonhuman behavior. Studies with humans from Baron's laboratory (e.g., Baron et al., 1969; Baron & Galizio, 1976; Baron, Menich, & Perone, 1983; Perone & Baron, 1980) have demonstrated that precise analysis of human operant behavior can be accomplished through patient and persistent work. It may seem unexciting to run so many sessions to answer a simple question in human operant behavior, but that may be necessary for real progress in the experimental analysis of human behavior.

ALTERNATIVE GOALS OF THE EXPERIMENTAL ANALYSIS OF HUMAN BEHAVIOR

Although a strong case can be made for extended experimental analysis, good arguments can also be made against gaining tight experimental control by running many long and frequent sessions. To make clear experimental inferences we typically analyze stable, steady-state performance, but transitory effects are sometimes very interesting and worth studying. For example, the work on rule-governed and contingency-shaped behavior (e.g., Catania et al., 1982) is fascinating because that distinction has great potential for education and socialization. The phenomenon of insensitivity to contingencies varies widely under circumstances that seem to differ only slightly. As a result, the selection of a research strategy will depend on one's goals as a scientist. If you pursue the model of psychologist as basic scientist, you will study the limits of the phenomenon, conducting experiments to find steady-state effects of the relevant variables. Research on instructions should compare old verbal histories with both extensive nonverbal experimental variables and experimentally created verbal repertoires. Experimental sessions will be long and frequent, and they will continue until the effects of all the strong variables have been assessed. If you pursue the model of psychologist as behavioral engineer, you will study the phenomenon under conditions that are relevant to the context in which you work. The specific responses and variables studied will be matched to the interventions typical of

the environment of interest, and the experiment will continue only as long as people normally remain in that one context.

Both of these research strategies have useful places in behavior analysis. During some cooperative research with colleagues at another institution, we had to compromise about procedures for our work. In one case, my procedure gave subjects a great deal of flexibility in daily activities, while their procedure had a rigid daily schedule that was matched to the needs of the funding source. I felt that natural patterns of performance should be observed, and a fundamental property of behavior would be obscured by a strong constraint. My colleagues felt that my procedure did not match the environment to which the results were to be generalized, and it would not likely produce stable data with the time we had available. In another case, we were studying the effect of a drug on performance under a strong contingency, and performance was not disrupted by the drug. That time I suggested that we had selected levels of the two variables typical of the natural environment, so we should continue the procedure. My colleagues took the basic science position that we had failed to create a sensitive preparation, and we should weaken the contingency control and increase the drug dose.

The last example points out the difficulty of offering a fixed prescription for how to conduct sessions. In general, more long sessions will produce better experimental control and test the limits of the effects of variables. That kind of design is most appropriate when the research goal is to make broad conceptual statements about substantial classes of behavior. In contrast, fewer short sessions will likely preserve the small or transitory effects of naturally occurring variables such as instructions, group structure or participants' expectations. When the research goal is to provide specific interventions for a known context, the relative strengths of experimental variables should be matched to their prevalence in the environment to which the results are to be generalized. If extended exposure to one variable overwhelms the effect of another relevant variable, then a brief experiment would be appropriate. In some cases both procedures will produce the same conclusion, but maintaining a conceptual separation of methods is certainly worthwhile.

SOME INHERENT LIMITATIONS ON THE EXPERIMENTAL ANALYSIS OF HUMAN BEHAVIOR

Even when the goal of the research is clear and an ideal strategy can be planned, the type and number of sessions that can be run with human subjects are constrained by other factors. First and perhaps foremost is the cost of human research. Institutions do not maintain captive populations of people available at low cost and for use whenever needed. Even if the research would benefit from 15 sessions per condition, there may be no practical way to accomplish that goal. Much excellent human research is conducted by students without grant support, and a limited number of sessions can be supported by available resources.

A second problem arises from the analysis of any kind of a natural repertoire. Such research will include either very long experiments or much more variability than is typical of research with nonhuman subjects. My residential experiment has been conducted as long as 49 consecutive days to obtain conventionally stable data, and that is approaching the limits of my participants' tolerance for confinement.

A third set of limits emerges from consideration of research ethics. Some human research ethics committees have raised objections to any contingent monetary payment of participants in experiments, and the next APA ethics guidelines will recommend abolishing that practice (Stanley & Melton, in preparation). In addition, seven weeks is about the maximum experimental length that my institutional review board is prepared to approve. At that point the potential extra cost for participants may be greater than the additional research benefit of

extending the duration of individual experiments.

One potential problem is that our conventional stability criteria may come in conflict with the amount of variability found in some studies of natural human repertoires that can be observed for an unlimited time. We will have to consider balancing the benefit of an increase in external validity with the cost of some loss of precise experimental control. Preservation of reporting of individual subject analyses is essential to an experimental analysis. To make the order in the data apparent to readers, however, it may be necessary to report condition averages rather than daily session points. Systematic replicability of changes in condition means can meet the usual logical criteria for ruling out extraneous causes without obscuring the effects in session to session variability. Perhaps at that level we will make some compromises to obtain the added value of research with human participants, while continuing within the conventions of experimental analysis.

CONCLUSION

The length of experimental sessions can follow the usual conventions when humans engage in activities such as button pressing that resemble the repertoires studied in nonhumans. When more natural samples of human behavior are taken, the repertoire should be sampled to determine the necessary length of sessions to provide the kind of stable data our designs require. Problems with human research arise because participants are not under continuous experimental control, and these problems can be minimized by running sessions as frequently as possible. The number of sessions should not be shortened just because the participants are human. If anything, human experiments need more sessions for the experimental variables to compete with the history that people bring to the setting. This is especially important when verbal behavior is the object of experimental analysis. One can also pursue different goals in the course of psychological research, and researchers with applied interests may choose to match the experimental variables to a specific setting rather than pick variables that will provide the most stable data. Ultimately, the limitations inherent in running experiments with human participants may require flexibility in the traditional criteria for adequate behavior analysis. It is important, however, that the compromises are not accepted largely for convenience of the experimenter, and we must be certain that in return there are increases in the external validity of the research.

REFERENCES

Baron, A., & Galizio, M. (1976). Clock control of human performance on avoidance and fixed-interval schedules. *Journal of the Experimental Analysis of Behavior, 26,* 165–180.

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

Baron, A., Menich, S. R., & Perone, M. (1983). Reaction time of younger and older men and temporal contingencies of reinforcement. *Journal of the Experimental Analysis of Behavior, 40,* 275–287.

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.

Bernstein, D. (1982, May). Response probability drowns in the stream of human behavior. Paper presented at the meeting of the Association for Behavior Analysis, Milwaukee, WI.

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

Burgess, R. (1968). Communication networks: An experimental re-analysis. *Journal of Experimental Social Psychology, 4,* 324–337.

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

Collier, G. H., Hirsch, E., & Kanarek, R. (1977). The operant revisited. In W. Honig & J. E. R. Staddon (Eds.) *Handbook of operant behavior* (pp. 98–124). Englewood Cliffs, NJ: Prentice-Hall.

Fantino, E., & Case, D. A. (1983). Human observing: Maintained by stimuli correlated with

reinforcement but not extinction. *Journal of the Experimental Analysis of Behavior, 40,* 193–210.

Hake, D. F., Olvera, D., & Bell, J. C. (1975). Switching from competition to sharing or cooperation at large response requirements: Competition requires more responding. *Journal of the Experimental Analysis of Behavior, 24,* 343–354.

Hake, D. F., & Schmid, T. L. (1981). Acquisition and maintenance of trusting behavior. *Journal of the Experimental Analysis of Behavior, 35,* 109–124.

Hersen, M., & Barlow, D. H. (1976). *Single case experimental designs.* Oxford: Pergamon.

Johnston, J. M., & Pennypacker, H. S. (1980). *Strategies and tactics of human behavioral research.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Kavanau, J. (1969). The behavior of captive white-footed mice. In E. P. Willems & H. L. Rausch (Eds.) *Naturalistic viewpoints in psychological research* (pp. 221–270). New York: Wiley.

Kraemer, H. C., Alexander, B., Clark, C., Busse, C., & Riss, D. (1977). Empirical choice of sampling procedures for optimal research design in the longitudinal study of primate behavior. *Primates, 18,* 825–833.

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

Marwell, G., & Schmitt, D. R. (1975). *Cooperation: An experimental analysis* New York: Academic Press.

Matthews, B. A., Catania, A. C., & Shimoff, E. (1985). Effects of uninstructed verbal behavior on nonverbal responding: Contingency descriptions. *Journal of the Experimental Analysis of Behavior, 43,* 155–164.

Perone, M. & Baron, A. (1980). Reinforcement of human observing behavior by a stimulus correlated with extinction or increased effort. *Journal of the Experimental Analysis of Behavior, 34,* 239–261.

Schmitt, D. R. (1976). Some conditions affecting the choice to cooperate or compete. *Journal of the Experimental Analysis of Behavior, 25,* 165–178.

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

Sidman, M. (1960). *Tactics of scientific research.* New York: Basic Books.

Stanley, B., & Melton, G. B. (in preparation). *Ethics in research with human participants: A casebook.* Washington, D.C.: American Psychological Association.

Weiner, H. (1964). Conditioning history and human fixed-interval performance. *Journal of the Experimental Analysis of Behavior, 7,* 383–385.