

ATTENUATION AND ENHANCEMENT OF COMPLIANCE WITH EXPERIMENTAL DEMAND CHARACTERISTICS

DOUGLAS J. NAVARICK
California State University, Fullerton

College students chose between reinforcement schedules in which a key press produced a cartoon video followed by nonreinforcement. The experimenter introduced a demand characteristic by stating a preference for choosing one schedule while indicating that the other schedule was also acceptable. With identical schedules (25 s of reinforcement, 5 s of nonreinforcement), compliant choices approached 100%. When the designated schedule presented a 5-s reinforcer followed by a 25-s timeout, participants complied initially, and then chose the other schedule on most trials. The aversiveness of the timeout, combined with the predominantly reinforcing alternative, deterred compliance. Elaborating the instructional context by stating that choosing the designated schedule would “help us with the data analysis” significantly increased compliant choices but did not alter the rise-and-decline pattern found with the simpler instructions. A salient relation between an experimenter’s preference and a scientific benefit thus promotes compliance with demand cues but does not insulate it from attenuation by relatively mild aversive consequences. Demand characteristics are not a uniquely powerful source of behavioral control.

The *demand characteristics* of a study (Orne, 1962, 1970) are uncontrolled cues that prompt participants to perform the kinds of behavior that the researcher has hypothesized. In addition to complicating the data analysis, demand characteristics potentially undermine the generalizability of research to behavior in the natural environment where similar cues are absent and may raise concerns about what Orne (1970) called the “ecological validity of the psychological experiment.” The present experiment employed operant conditioning methods to investigate compliance with demand characteristics as a dynamic process influenced both by instructional context and by consequences.

Demand characteristics are generally treated as implicit cues that

The author thanks Kelly Deegan for her invaluable assistance. Correspondence concerning this article should be addressed to Douglas J. Navarick, Department of Psychology, California State University, Fullerton, CA 92834-6846. (E-mail: dnavarick@fullerton.edu).

convey the researcher's expectations about the results (e.g., Intons-Peterson, 1983). A unique feature of the present approach is the operational definition of demand characteristics as an explicitly stated experimenter preference as to what behavior participants should perform (Navarick, 2004).

This explicitness allows the researcher to control the time when the biasing cue is detected so that changes in the degree of compliance can be tracked over trials, each of which is associated with prearranged consequences. Since the experimenter's preference is explicitly stated within the instructions, the effects of the instructional context may be investigated by manipulating the phrases surrounding the request.

Operant research on instructional effects has led to a distinction between two classes of variables that jointly influence a participant's tendency to comply with instructions: variables that determine the initial tendency to comply and variables that occur after the instructed behavior has been performed. In the former category would be social factors in the individual's preexperimental history of the sort that Orne emphasized, such as the prestige accorded to scientists and scientific research. In the latter category would be reinforcing or punishing consequences for the instructed behavior administered by the researcher's procedures. These have been referred to as *collateral consequences* (Cerutti, 1989), and they may serve to either strengthen or weaken the initial tendency to follow the researcher's instructions. For example, Galizio (1979) found that compliance with instructions in an avoidance conditioning procedure decreased after this behavior was punished by monetary loss.

Applying this model, Navarick (2004) operationally defined a demand cue as an explicitly stated experimenter preference as to what choice participants should make between two schedules of reinforcement. The basic strategy was to observe participants' initial preference on repeated trials and then ask them to choose their nonpreferred schedule, with the provision that they could continue to make either choice. This approach differs from one taken in many studies of rule-governed behavior in which instructions inaccurately describe programmed contingencies so that following the instructions results in a loss of reinforcement (Galizio, 1979; Hackenberg & Joker, 1994). Typically, behavior initially conforms to the instructions, but it then shifts in the direction of optimal performance under the contingencies. In Navarick's (2004) study, the instructions expressed an experimenter preference without describing a contingency, and the degree of compliance depended on the hedonic properties of the consequences rather than on the accuracy of the instructions. The reinforcer was a cartoon video, and the amount and delay of reinforcement were manipulated in a self-control procedure. Amount of reinforcement was the period for which the cartoon played before it was interrupted on a trial. Delay of reinforcement was the time between a key press and the onset of the video. On 20 trials, participants chose between a relatively small, immediate reinforcer (15 s of video followed by 75 s of waiting without any picture or sound) and a larger,

delayed reinforcer (55 s of waiting, 25 s of video, 10 s of waiting). Choice of the small, immediate reinforcer on at least 70% of trials represented impulsivity; choice of the large, delayed reinforcer on at least 70% of trials represented self-control. Approximately 50% of participants met the criterion for impulsivity and 27%, the criterion for self-control. In a second session consisting of 40 choice trials, participants in both subgroups were instructed, "In this part of the experiment, we would prefer that you press the disk on the [nonpreferred] side. However, you can press whichever disk you want." It was found that the majority of impulsive participants proceeded to make predominantly self-controlled choices and the majority of self-controlled participants predominantly impulsive choices. The experimenter's preference overrode the collateral consequences arranged by the schedules, a result consistent with the power usually ascribed to demand characteristics and an indication that the operational definition was a realistic and potentially useful one. (Cf. Hackenberg & Joker, 1994, p. 380 for an analogous case involving compliance with inaccurate instructions.)

Fundamental questions about the effects of demand characteristics have received little attention. For example, what factors influence a participant's resistance to an experimenter's demands? Orne (1962) maintained that participants complied with demand characteristics because they were motivated to help the researcher obtain scientifically useful results (the "good subject effect") and toward this end would tolerate considerable discomfort, provided it was "perceived as essential for the research" (Orne, 1970, p. 221). For example, Orne noted that participants objected to being stuck a second time in a venipuncture procedure after the first attempt failed due to ostensible investigator error. In this context, the results of Navarick's (2004) experiment suggest that discomfort, and not just a lesser degree of preference, may be necessary to deter compliance. However, the intense discomfort represented by the venipuncture example may exaggerate the power of demand cues. A well-supported generalization from both animal and human research is that the suppressive effects of aversive consequences are often enhanced when an alternative source of positive reinforcement is available that does not include punishment (Lerman & Vordran, 2002). With such an alternative available, it is possible that even a relatively mild timeout from positive reinforcement could deter compliance with demand characteristics. Such a demonstration would suggest that demand characteristics are not a uniquely powerful source of behavioral control.

In the present experiment, one group was presented with a choice between identical schedules, 25 s of video followed by 5 s of waiting without reinforcement, and after 20 trials the experimenter expressed a preference for choosing the side that received fewer choices (usually the side opposite the hand preference) to ascertain the degree of compliance when it entailed no loss of reinforcement. In another group, the participants chose between 25 s of video followed by 5 s of waiting and 5 s of video followed by 25 s of waiting, and after 20 trials the experimenter expressed

a preference for choosing the latter schedule. Since choice of the former schedule permitted almost continuous viewing, choice of the latter imposed a virtual timeout from positive reinforcement, a condition that would be expected to function as an aversive stimulus and promote escape or avoidance responding (e.g., Rilling, Askew, Ahlskog, & Kramer, 1969). In addition, resistance to the experimenter's request could be produced by the very brief period of reinforcement preceding the timeout. Technically, for reasons to be discussed under "Discussion," the choice procedure could not provide direct evidence on whether the schedule, 5 s of reinforcement followed by 25 s of nonreinforcement, reduced compliant choices because of punishment or because participants preferred the larger amount of reinforcement available on the other schedule. Therefore, the term *attenuation* will be used instead of *punishment*. However, closely paraphrased debriefing reports (to be presented under "Results") support the aversive character of the timeout and its relevance to the hypothesis that relatively mild discomfort can promote noncompliance with demand cues.

The degree of resistance to aversive or minimally reinforcing consequences may depend on the instructional context. Demand characteristics could be conceptualized as consisting of two components: an experimenter's preference and a scientific benefit. In the foregoing conditions a preference was stated but not a benefit, and an association between the two may not have been made by some participants. To strengthen this association, another group was studied with a variation of the instructions, in which the experimenter's preference was justified as a way in which participants could "help us with the data analysis." An enhanced good subject effect could be seen in two ways: as an increase in the proportion of compliant choices on the initial trials after the request was made (as compared with the group that received the simpler instructions) and as a sustained or increasing level of compliance on later trials.

Method

Participants

A total of 90 introductory psychology students at California State University, Fullerton, participated in partial fulfillment of a research hours requirement. The participants chose among available studies partly on the basis of brief, descriptive titles. The present study was entitled "Cartoon Viewing" as a way of maximizing the number of participants for whom cartoons would be an effective reinforcer. There were 15 participants in each of 6 groups. No participant withdrew from the experiment. However, data from 11 additional participants were excluded from the analysis because the participants did not meet a preference criterion (described below) in groups choosing between the 5-s and 25-s reinforcers.

Setting and Apparatus

All of the experimental sessions and debriefings were conducted by the same person, a female graduate student. After participants signed an

informed consent statement, which explicitly permitted withdrawal without any penalty, they were seated in a 2.7×2.1 -m unlit room facing a response panel, 53 cm high and 48 cm wide, that contained two translucent disks spaced 30 cm apart. At the beginning of a trial, both keys were lit by rear-mounted stimulus projectors. The participant had an unobstructed view of a Panasonic 21-in color video monitor that was placed 178 cm from the chair on the top shelf of a video cart. A Hitachi, 4-head VHS videocassette recorder was on a shelf below the monitor and was operated remotely with programming equipment. The tape was started at the beginning of a reinforcement period and stopped at the end by signals to the pause control. During nonreinforcement periods following cartoon presentation, the monitor was dark and silent. When reinforcement started, the tape resumed from the point where it was previously stopped. A total of 26 classic and contemporary cartoons was recorded and made available for selection.

Procedure

In each half of the 1-hr session there were 20 free-choice trials preceded by 4 forced-choice trials for which the participant had been instructed to press the keys in the sequence, left-right-left-right. A single press on a lit key darkened both key lights and produced the corresponding schedule of reinforcement. Schedule presentation lasted 30 s, and when it was completed the keys were lit again for another trial without an additional intertrial interval. For the large reinforcer the schedule consisted of immediate presentation of the cartoon for 25 s followed by 5 s of waiting. For the small reinforcer the schedule consisted of immediate presentation of the cartoon for 5 s followed by 25 s of waiting. The basic measure of performance was a choice proportion calculated within each block of 10 free-choice trials (number of responses for a schedule divided by 10). At the beginning of the session the experimenter read the following instructions:

First, we need to know whether you prefer to pick up things with your right hand or your left hand. We have made some tapes showing a variety of cartoons. Before we start, you will have an opportunity to select a cartoon, but first I want to tell you something about how the equipment works.

From time to time, these disks will be lit red. For most of the experiment you can press the one on the left or the one on the right. Be sure to press in all the way. Do not press both at the same time. The first four times the disks are lit, we want you to press them in the following manner: The first time, press left; the second time, press right; the third time, press left again; the fourth time, press right again. Then you can press whichever disk you prefer. As an extra signal that it is time to make a choice, these three yellow lights at the top will come on at the same time that the disks are lit.

Here is some more information. After you make your choice, both disks will go dark. They will stay dark for a period of time and during part of this time the video will be shown. After awhile, the video will

be turned off and the disks will be lit again. The disk you press may affect how long the tape will play before it is interrupted.

The experiment will be in two parts, each lasting about 15 minutes. After the first part is over, I will return and you will have the option of staying with the same cartoon or switching to another one. That is all I can tell you now about the procedure.

To review, the first four times the disks are lit, we want you to press them in the following way: The first time, press left; the second time, press right; the third time, press left again; the fourth time, press right again. Then you can press whichever disk you prefer. Remember, the disk you press may affect how long the tape will play before it is interrupted.

The participant was then handed a sheet listing 26 cartoons and was asked to make a selection. The experimenter advanced the tape to the appropriate position, and then returned to the reception room to start the trials. Upon completion of the schedule chosen on the 20th free-choice trial, the experimenter rejoined the participant and read additional instructions that varied according to the participant's group.

Group 25-25. The 25-s reinforcer schedule was associated with both keys in Parts 1 and 2 of the session. The instructions for Part 2 stated twice the experimenter's preference that the participant press a designated key, which was whichever key the participant pressed on fewer than half the trials in Part 1. For 12 of 15 participants this was the side opposite the stated hand preference. If equal responses occurred in Part 1 (2 cases), then the side opposite the hand preference was designated in Part 2. The instructions were as follows (the lines related to the experimenter's preference are italicized):

We'll now start the second part of the experiment. The procedure is the same as before. The first time the disks are lit, press left; the second time, press right; the third time, left; the fourth time, right. Then press whichever disk you prefer. *In this part of the experiment, we would prefer that you press the disk on the _____ side. However, you can press whichever disk you want.*

Would you like to stay with the same cartoon or switch to another one? [The experimenter made any necessary changes.]

Again, the first time the disks are lit, press left, then press right, then left, then right. After that, you can press whichever disk you prefer. *We would prefer that you press the disk on the _____ side but you can press whichever disk you want.*

Group 25-25 Control. The lines related to the experimenter's preference were deleted.

Group 5-25. Participants chose between 5- and 25-s reinforcer schedules in both parts of the session, with the 25-s reinforcer presented on the side opposite their hand preference so that a preponderance of choices on that side would more plausibly be attributable to the

schedule than to position preference. Of the 15 participants, 13 had choice proportions below .50 for the 5-s reinforcer across the 20 trials of Part 1. (In all groups choosing between 5-s and 25-s reinforcers, participants were excluded who had choice proportions greater than .50 for the 5-s reinforcer in Part 1 so that it was more likely that the analysis in Part 2 would include participants who discriminated between reinforcer durations.) For Part 2, the experimenter read the same instructions as for Group 25-25 but stated a preference for choosing the side on which the 5-s reinforcer was presented.

Group 5-25 Control. The italicized lines related to the experimenter's preference were deleted.

Group 5-25 Elaborated Instructions. Lines were added to Group 5-25's instructions to link the experimenter's preference to a research objective: assisting with the data analysis. This elaboration was performed in each of the two places where the experimenter's preference was stated: "In this part of the experiment, we would prefer that you press the disk on the _____ side. *This will help us analyze the data from Part 1...*", and, "*To help us with the data analysis, we would prefer that you press the disk on the _____ side but you can press whichever disk you want.*"

Group 5-25 Required Choice. Orne (1962) distinguished between demand characteristics and procedural variables, explicit requirements that permitted no alternative behavior except withdrawal from the research. As will be discussed, he treated these factors as functionally equivalent and equally powerful sources of behavioral control. A comparison was made with the operationally defined demand characteristic by including a required choice group in which there was no statement of experimenter preference and no statement that participants could press whichever key they preferred, although both keys were lit and operative. In accordance with the informed consent statement, participants could choose to leave at any time without a penalty. Following the two statements about what to do on the first 4 trials, participants were instructed, "After that, please press the disk only on the _____ side."

Debriefing. Informally, participants were asked several questions related to their performance in Parts 1 and 2. Of particular interest were questions asking why a particular schedule was chosen, especially after the experimenter expressed a preference. Answers were recorded immediately after the participant left and were either closely paraphrased or summarized with the use of some of the participants' terms. Data from Groups 5-25 and 5-25 Control will be presented that are related to the aversiveness of the 25-s time-out period and the participants' stated reasons for their choices.

Results

Figure 1 presents the mean choice proportion for the experimenter's preferred schedule in each of the six groups as a function of blocks of 10 trials. Blocks 1 and 2 represent a baseline phase preceding the statement

of the experimenter's preference, and Blocks 3 and 4 represent the phase in which the preference (or for one group, the required choice) was stated. In Groups 25-25 Control and 5-25 Control, a preference was not stated but choice for the corresponding schedules was tracked. Changes from Block 2 to Block 3 measure initial compliance with the request, whereas changes from Block 3 to Block 4 measure the maintenance of compliance. Significance was assessed using two-tailed Mann-Whitney tests for between-groups comparisons and two-tailed Wilcoxon signed-rank tests for within-groups comparisons. Trends within Blocks 3 and 4 were further analyzed through division of the trials into four successive blocks of five trials and assessment of the significance of changes in choice proportions with a Friedman two-way analysis of variance by ranks, a type of chi-square test.

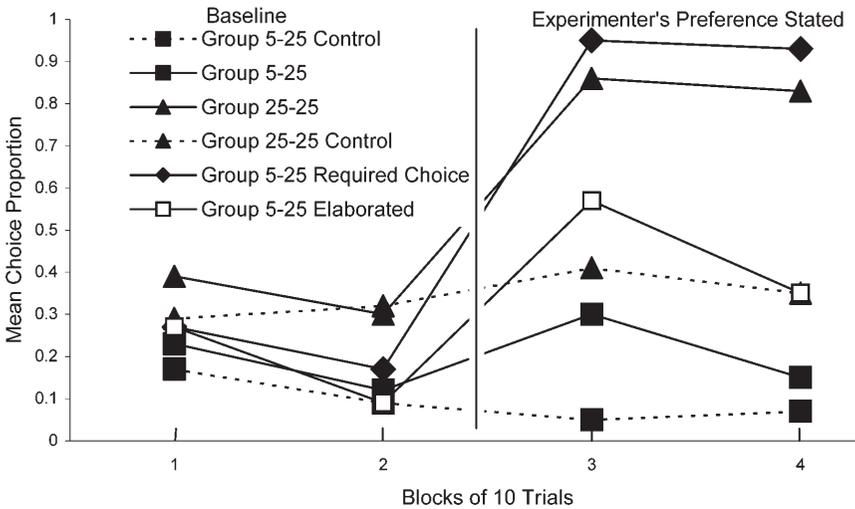


Figure 1. Mean choice proportions for the experimenter's preferred schedule in successive blocks of 10 trials during the baseline phase, before a preference was stated, and during the postrequest phase. The schedule designated for choice was either the 5-s reinforcer schedule (5-s video presentation followed by 25-s timeout) or, in the case of Group 25-25, the schedule that received fewer choices during the baseline (usually the side opposite the hand preference). In the Required Choice group, the request was stated as a procedural requirement. In Groups 25-25 Control and 5-25 Control, a preferred schedule was not stated but choice for the corresponding schedule was tracked.

In Group 25-25, the mean choice proportion for the experimenter's preferred schedule rose significantly from .30 in Block 2 (reflecting a preference for the side corresponding to the preferred hand) to .86 in Block 3 ($T = 120$, $N = 15$, $z = 3.39$, $p = .0007$), and it remained at virtually the same level (.83) in Block 4. For the 20 trials after the experimenter's preference was stated, the mean choice proportions in blocks of 5 trials were .89, .83, .83, and .81. The Friedman test indicated that this slight downward trend was not significant ($\chi^2_r = 1.44$, $df = 3$, $N = 15$, $p = .6962$). According to Mann-Whitney tests the choice proportions in the two 10-trial

blocks were significantly higher than the ones in the corresponding blocks for Group 25-25 Control (with $N = 30$, Block 3: $U = 5$, $z = 4.44$, $p < .0001$; Block 4: $U = 15$, $z = 4.02$, $p = .0001$). Although the experimenter's request produced compliant choices on most trials, compliance was significantly higher in Group 5-25 Required Choice, with means of .95 and .93 in Blocks 3 and 4, respectively (with $N = 30$, Block 3: $U = 61.5$, $z = 2.09$, $p = .0366$; Block 4: $U = 64$, $z = 1.99$, $p = .0446$). Thus, even when compliance with the experimenter's request entailed no loss of reinforcement but only the possible inconvenience of reaching to the side opposite the hand preference, instructional control was weaker than it was when the request was expressed as a procedural requirement and compliance resulted in a loss of reinforcement.

In Group 5-25, choice proportions for the experimenter's preferred schedule rose significantly from Block 2 to Block 3 ($T = 91$, $N = 13$, $z = 3.16$, $p = .0016$), an increase found in 13 of 15 participants, but then decreased significantly in Block 4 ($T = 78$, $N = 12$, $z = 3.04$, $p = .0024$), a decrease found in 12 participants, and became statistically indistinguishable from choice proportions in Block 2 of the baseline phase. Across the 20 trials after the experimenter's preference was stated, the mean choice proportions in blocks of 5 trials were .44, .16, .19, and .12. The Friedman test indicated that this steep downward trend was significant ($\chi^2_r = 18.98$, $df = 3$, $N = 15$, $p = .0003$). In the two 10-trial blocks, the choice proportions were significantly below those in Group 25-25 (with $N = 30$, Block 3: $U = 24$, $z = 3.65$, $p = .0003$; Block 4: $U = 16$, $z = 3.98$, $p = .0001$). Altogether, 11 of 15 participants showed the complete, up-and-down pattern across Blocks 2 to 4 illustrated in Figure 1. Comparisons with Group 5-25 Control further support these effects. In the first 10 trials after the experimenter's preference was stated (Block 3), the mean choice proportion for the designated schedule in Group 5-25 was .30, whereas in Group 5-25 Control the mean choice proportion for the analogous schedule was only .05, a significant difference on a Mann-Whitney test ($U = 29.5$, $N = 30$, $z = 3.42$, $p = .0006$). However, in Block 4 this initial compliance effect disappeared: a mean of .15 in Group 5-25 versus .07 in Group 5-25 Control ($U = 112$, $z = 0$, $p = 1.00$).

Debriefing statements suggested that for participants in Group 5-25 the schedule combining minimal reinforcement and extended timeout had aversive effects and did not function solely as a source of a smaller amount of positive reinforcement. Trial-by-trial choices of individual participants in this group are presented in Table 1. The most commonly used term in Group 5-25 was "annoying" (Participants 31, 36, 38, 39, 40, 43, Table 1). Participant 32 said she "couldn't take it after awhile, it was driving me nuts." (This participant made compliant choices on the first 4 trials and then noncompliant choices on the remaining 16 trials.) Participant 41 said, "I tried to press the right button more because you said you preferred me to. But I didn't like sitting in the dark, waiting, so it was hard . . . it was better to watch something than to sit around with nothing to do." (This participant made 4 compliant choices in Block 3 and

3 compliant choices in Block 4). In summarized, partially paraphrased terms, Participant 35 said “he didn’t like to sit around in the dark for too long” and Participant 37 said “she didn’t like the right disk because she sat and waited longer.” Altogether, 10 of the 13 participants who had choice proportions below .50 for the 5-s reinforcer schedule during the baseline phase characterized this schedule in terms of discomfort.

Table 1

Trial-by-Trial Choices on the Left and Right Keys, and Choice Proportions for the 5-s Reinforcer Schedule, for Participants in Group 5-25*

Participant	Phase	Choice Proportion	Trial-by-Trial Choices
31F	Baseline	.25	lrrllrll rlllllll
	Postrequest	.10	lrrlllll llllllll
32F	Baseline	.10	llrlllll rlllllll
	Postrequest	.20	rrrrllll llllllll
33F	Baseline	.50	llrllrrll lrrllrrrr
	Postrequest	1.00	rrrrrrrrrr rrrrrrrrrr
34F	Baseline	.15	llrllrll llllllll
	Postrequest	.05	llrlllll llllllll
35M	Baseline	.05	llllllll lllrllll
	Postrequest	.30	lrrllllr rllrllllr
36F	Baseline	.00	llllllll llllllll
	Postrequest	.00	llllllll llllllll
37F	Baseline	.20	rrllrlll llrllrll
	Postrequest	.05	rlllllll llllllll
38M	Baseline	.05	lrllllll llllllll
	Postrequest	.05	rlllllll llllllll
39M	Baseline	.05	rlllllll llllllll
	Postrequest	.05	llrlllll llllllll
40M	Baseline	.30	lrrrrllr llllllll
	Postrequest	.05	lrllllll llllllll
41F	Baseline	.30	rllrllrll lrrllrll
	Postrequest	.35	lrrllrll rllrllrll
42F	Baseline	.05	llrlllll llllllll
	Postrequest	.10	rllrllll llllllll
43F	Baseline	.10	lrllllrll llllllll
	Postrequest	.15	lrllrllll llllllll
44F	Baseline	.50	lrrllrll lrrllrtr
	Postrequest	.80	rrrrrrrrrr rrrrrrllll
45F	Baseline	.10	rlllllll llllllll
	Postrequest	.15	llrlllll lrllllll

M, SE Baseline .18, .16
M, SE Postrequest .23, .29

*Gender is represented by M or F. All participants in this group had a right-hand preference, and the 5-s reinforcer schedule was presented on the right side.

In contrast to these statements, 12 of 14 participants in Group 5-25 Control (no statement from 1 participant) characterized the schedule in terms of a shorter video time or a longer waiting time, or both, without reference to discomfort or other feelings. Examples are as follows: “In the beginning I figured out that the right side was a lot shorter than the left and there was more time in between turns. After that I picked the left side because it was longer”; “I chose the left side more because it played

longer”; “Not only was the left side longer, the right side was a lot shorter and it had a long wait in between, so that’s really why I chose the left side over the right”; “I chose the left side more for both parts because it was longer than the right side. I picked the right side a few times because I was curious to see if it would change.”

The contrasting statements from these two groups suggest that characterizations of the nonpreferred schedule were influenced partly by the schedule and partly by the instruction-induced increase in Group 5-25’s choices for this schedule in Block 3. In other words, the self-reports depended on the context in which the schedule was presented (cf. Davies & Best, 1996). Despite the complexity of the underlying processes, the debriefing reports indicate an association between expressed discomfort and noncompliance with demand cues.

In Group 5-25 Elaborated, in which the experimenter indicated that choice of the designated schedule would help with the data analysis, the mean choice proportion rose from .09 in Block 2 to .57 in Block 3, a steeper rise than that found in Group 5-25 (from .12 to .30). This increase was significant on a Wilcoxon test ($T = 120$, $N = 15$, $z = 3.39$, $p = .0007$), and the choice proportions in Block 3 were significantly higher than those in Group 5-25 on a Mann-Whitney test ($U = 55$, $N = 30$, $z = 2.36$, $p = .0183$). However, choice proportions decreased significantly in Block 4 ($T = 66$, $N = 11$, $z = 2.91$, $p = .0036$), and they were no longer significantly different from those in Group 5-25 ($U = 70$, $N = 30$, $z = 1.74$, $p = .0819$). In blocks of 5 trials, the mean choice proportions after the experimenter’s preference was stated were .65, .48, .36, and .35. This decreasing trend was significant according to the Friedman test ($\chi^2_r = 11.42$, $df = 3$, $N = 15$, $p = .0097$). The elaborated instructions thus enhanced only participants’ initial compliance with the experimenter’s request.

Discussion

When the experimenter expressed a preference for choosing a particular key, choices of that key initially increased in all of the groups. The size of this increase and the extent to which compliance was maintained across trials were dependent on consequences. With identical schedules (Group 25-25), the proportion of compliant choices was maintained at a relatively high level, although below the level produced by the instructional requirement to choose a less favorable schedule (Group 5-25 Required Choice). When responding in accordance with the experimenter’s preference resulted in a loss of reinforcement (Group 5-25), the proportion of compliant choices in the first 10 trials was much lower than it was in Group 25-25, and during the final 10 trials choices of the designated key dropped to the level that prevailed before the preference was stated. Relating the experimenter’s preference to a scientific purpose (Group 5-25 Elaborated) increased the initial level of compliance—an enhanced good subject effect—but during the final block of trials the proportion of compliant choices dropped to about the same level as that found in Group

5-25. In the context of research on rule-governed behavior, these are the first data to show attenuation of compliance with instructions based on the hedonic properties of the consequences rather than on instructions that inaccurately describe the consequences.

Punishment or Reinforcement?

Attenuation of compliance by the schedule that combined brief reinforcement with timeout is consistent with a punishment process, especially given the debriefing statements characterizing the timeout periods in terms of discomfort, but it is also consistent with a reinforcement process in which participants exhibited a preference for a longer over a shorter duration of reinforcement. The ambiguity stems in part from the use of a choice procedure in which a single response was required to produce a schedule: Reductions in responding on one key necessarily corresponded to increases in responding on the other. In addition, demonstrations of punishment usually require a decrease in responding below an unpunished baseline. In the present case, there was no unpunished baseline within groups and a more complex assessment would be required.

An alternative arrangement that separated responding on the two keys and straightforwardly compared punished responding with an unpunished baseline was employed by Thomas (1968) in a study of timeout in pigeons. During a baseline phase the animals responded on equal, concurrent variable-interval schedules without punishment. Then, during the punishment phase, responding on one of the variable-interval schedules produced a timeout on a fixed-ratio schedule. Performance was measured in two ways: by the absolute response rates on the punished and unpunished schedules and by a relative response rate calculated as $\text{punished rate}/(\text{punished} + \text{unpunished rate})$, which would be analogous to the choice proportions for the 5-s reinforcer schedule. As the fixed-ratio size decreased, making timeout more frequent, the relative rate on the punished schedule decreased. In addition, the absolute punished rate generally decreased and was below the baseline rate, whereas the unpunished rate in some cases increased and in other cases decreased. Punishment was shown by the decrease in punished responding below the unpunished baseline, which occurred independently of the response rate on the alternative schedule.

A possible baseline in the present case would be the high proportion of compliant choices in Group 25-25 during Blocks 3 and 4. But whether the reduction in choice proportions below this baseline was due to punishment by the longer timeout or to choice of the larger amount of reinforcement on the alternative schedule is still unresolved because of the requirement to choose one schedule or the other. Nominally, the 5-s reinforcer schedule imposed a timeout because it prevented participants from receiving almost continuous positive reinforcement on the other schedule. Context was relevant to categorizing the schedule. If the 5-s reinforcer schedule was the only one available, and participants' options

were to see nothing or to press the key and watch the cartoon for 5 s at 30-s intervals, then nominally there would be no cost for responding and this schedule may well have functioned as a source of positive reinforcement in the sense that the response rate would have exceeded the rate during a no-reinforcement baseline. But in the choice situation, the consequences of pressing this key also included the forfeiture of reinforcement. Given the extensive database on the suppressive and aversive effects of timeout and given participants' own statements that characterized the waiting periods as aversive, the position that the waiting periods were neutral and had no behavioral effect would seem difficult to support.

Two Stages or One?

There are also two possible interpretations of Group 25-25's high level of compliance. The operant model of instructional control discussed in the introduction proposes a two-stage process: Compliant choices were initiated by the instructions because of a preexperimental history of reinforcement for instruction following and were then maintained by positive reinforcement from the video, a collateral consequence. However, the only procedural change that preceded the change in behavior in Block 3 was the request to choose a designated schedule, so the evidence directly supports only the process of instruction following and not the additional process of reinforcement by collateral consequences. But this one-stage interpretation does not consider the covariation between choice proportions and the proportion of reinforcements received from a schedule, which are confounded in single-response choice procedures. Once choice of the designated schedule was initiated, the procedure arranged for a progressively higher relative rate of reinforcement for compliant choices than for noncompliant choices. Since the initiation process and the support process would move behavior in the same direction, the two stages would be difficult to discern. In contrast, with Groups 5-25 and 5-25 Elaborated, the choice proportions during the second stage (arguably involving punishment) moved in the opposite direction from those during the initial stage, so the two stages were separable. A plausible inference is that an analogous, two-stage process could have influenced performance in Group 25-25. The two-stage model also provides an explanation for why compliance in Group 25-25 was not maximal. (Group 5-25 Required Choice had significantly higher choice proportions.) As participants veered toward noncompliance, their choices would be reinforced and the relative rate of reinforcement for compliant choices would be reduced. An implication is that with equal schedules, the amounts of reinforcement could affect the degree of compliance. For example, a relatively low amount of reinforcement could be sufficient to sustain the compliant choices (given the added reinforcement from a preexperimental history of following instructions) but insufficient to sustain the noncompliant choices. The one-stage model would predict that if the schedules were equal, the proportion of compliant choices would be independent of the amounts of reinforcement.

Demand Characteristics Versus Experimental Variables

When choice of the 5-s reinforcer schedule was required rather than requested, 14 of 15 participants complied on virtually every trial. No participants exercised their option to withdraw from the experiment. Only when an alternative response could be made without withdrawing did participants consistently choose not to comply. In general, demand characteristics differ from procedural requirements (experimental variables) in that the former always allow participants to complete the study without complying whereas the latter provide only the option of withdrawal. In Orne's (1962; 1970) discussions, experimental variables and demand characteristics were represented as functionally equivalent components of the research environment that drew their power from participants' identification with the goals of science and their motivation to help the researcher. The influence of such a factor is shown by the significantly higher choice proportions during Block 3 in Group 5-25 Elaborated Instructions than in Group 5-25. However, the finding that these groups had much smaller compliance effects than the Required Choice group and that their initial compliance was not sustained implies that demand characteristics and experimental variables are not functionally equivalent. This lack of equivalence may seem unsurprising when one considers the different consequences associated with noncompliance: completing the study in the former case versus withdrawing in the latter. But if one interprets the effects of demand cues and experimental variables in terms of a common motivational state, as did Orne, then functional equivalence would be a reasonable expectation and one might assume that a demonstration of the power of experimental variables to gain compliance would equally represent the power of demand characteristics (such as the venipuncture example mentioned in the introduction, among other classic examples). The present results indicate that demand cues have much weaker effects than procedural requirements and are not uniquely powerful sources of behavioral control.

It was Orne's view that experimental demand characteristics were representative of a broader class of variables that operated in any interpersonal context: "Rather than relegating demand characteristics to the realm of artifact, they should be acknowledged as a pervasive influence upon all human interaction" (Orne & Whitehouse, 2000, p. 470). The importance of making demand characteristics the focus of investigation was underscored recently by Rosnow (2002). The present experiment furthers this goal and provides a methodology for integrating diverse theoretical perspectives.

References

- CERUTTI, D. T. (1989). Discrimination theory of rule-governed behavior. *Journal of the Experimental Analysis of Behavior*, 51, 259–276.
- DAVIES, J. B., & BEST, D. W. (1996). Demand characteristics and research into drug use. *Psychology and Health*, 11, 291–299.
- GALIZIO, M. (1979). Contingency-shaped and rule-governed behavior: Instructional control of human loss avoidance. *Journal of the Experimental Analysis of Behavior*, 31, 53–70.
- HACKENBERG, T. D., & JOKER, V. R. (1994). Instructional versus schedule control of humans' choices in situations of diminishing returns. *Journal of the Experimental Analysis of Behavior*, 62, 367–383.
- INTONS-PETERSON, M. J. (1983). Imagery paradigms: How vulnerable are they to experimenters' expectations? *Journal of Experimental Psychology: Human Perception and Performance*, 9, 394–412.
- LERMAN, D. C., & VORDRAN, C. M. (2002). On the status of knowledge for using punishment: Implications for treating behavior disorders. *Journal of Applied Behavior Analysis*, 35, 431–464.
- NAVARICK, D. J. (2004). Analysis of impulsive choice: Assessing effects of implicit instructions. *The Psychological Record*, 54, 505–522.
- ORNE, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17, 776–783.
- ORNE, M. T. (1970). Hypnosis, motivation, and the ecological validity of the psychological experiment. In W. J. Arnold & M. M. Page (Eds.), *Nebraska symposium on motivation* (pp. 187–265). Lincoln: University of Nebraska Press.
- ORNE, M. T., & WHITEHOUSE, W. G. (2000). Demand characteristics. In A. E. Kazdin (Ed.), *Encyclopedia of psychology* (pp. 469–470). Washington, D.C.: American Psychological Association and Oxford Press.
- RILLING, M., ASKEW, H. R., AHLKOG, J. E., & KRAMER, T. J. (1969). Aversive properties of the negative stimulus in a successive discrimination. *Journal of the Experimental Analysis of Behavior*, 12, 917–932.
- ROSNOW, R. L. (2002). The nature and role of demand characteristics in scientific inquiry. *Prevention and Treatment*, 5, Article pre0050037c. Retrieved August 20, 2005, from <http://journals.apa.org/prevention/volume5/pre0050037c.html>.
- THOMAS, J. R. (1968). Fixed-ratio punishment by timeout of concurrent variable-interval behavior. *Journal of the Experimental Analysis of Behavior*, 11, 609–616.