

CONTRAST EFFECT IN THE USE OF PUNISHMENT

MICHAEL HARRISON¹ AND ALBERT PEPITONE

University of Pennsylvania

Informal accounts of what has been termed a police riot indicate that more clubbing, prodding, and macing of protestors was done by men who carried a revolver than by those who did not. A contrast hypothesis was suggested whereby a strong but prohibited degree of punishment serves as an "anchor" in contrast with which lower degrees of permissible punishment are judged to be weaker than the case without the anchor. To compensate for the insufficiency generated when there is such a discrepancy between a strong punishment that cannot be used and a weaker level of punishment that can be used, trainers deliver more of the latter. In an experiment designed to test the contrast hypothesis, the subject trained a rat to press a bar by administering *mild* and *slightly painful* shocks during a 12-minute period. In a control condition, only these shock levers were present; in experimental conditions, a *moderately painful* or *extremely painful* lever was also present but was strictly forbidden. Results were in complete agreement with the hypothesis: In the first 6 minutes—within which the experimenter said other students had trained their animals—the subjects in the large discrepancy condition delivered significantly higher frequency and proportion of slightly painful shocks than the control subjects. In the second 6 minutes—beyond the normative training time—the difference in absolute frequency is maintained, but the higher proportion of slightly painful shocks was no longer significant. The relative increase in the use of mild shock is attributed to exogenous factors which combine with the contrast effect, namely, a "fine tuning" effect that occurs toward the completion of shaping or a shift to a normative level implied by the experimenter with a growing sense of ineffectiveness. Additional evidence for the hypothesis is presented, and alternatives are discussed.

In recent years, studies of human aggression have employed a paradigm in which one subject, in the role of a trainer, delivers punishment to another person (usually a confederate) who serves as the learner. In general, these studies deal with the relationship between various stimulus or environmental conditions and the instigation, facilitation, or inhibition of aggressive behavior as reflected in the administration of punishment. For example, violent film episodes have been shown to increase the subject's delivery of shock to a confederate who previously had angered him (Berkowitz, 1965). The well-known study by Milgram (1963) demonstrated that subjects will inflict severe pain on a learner in compliance with the directives of an authority figure, especially in the milieu of a prestigious scientific institution. Other environmental properties that affect the delivery

of punishment by the trainer include the apparent harmful effects of his or someone else's punishment and the presence of harm-producing physical objects which are symbolically associated with aggression. In the former case, for example, the electric shock delivered to the learner is attenuated when the subject is exposed to pain cues in filmed aggression (Feshbach, Stiles, & Bitter, 1967; Hartmann, 1969) or by the feedback of expressions of pain directly from the learner (Baron, 1971). As for the effect of symbolic objects, Berkowitz and LePage (1967) have shown that the presence of a gun on a table, without apparent relationship to the experiment, caused the trainer to administer larger doses of shock than in a control condition involving symbolically "neutral" objects. That this result represents more than the effect of demand characteristics has been seriously questioned (Page & Scheidt, 1971). In any case, harmful objects like weapons are not only symbols of aggression but the *actual means* by which aggression is delivered. The present

¹ Requests for reprints should be sent to Michael Harrison, Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, Pennsylvania 19104.

research deals with the effect of the levels of punishment in the individual's environment upon the actual delivery of punishment. The hypothesis of the experiment relates analogously to the use of weapons as implements of institutionalized roles such as policemen, National Guardsmen, and soldiers.

Consider the policeman. The regular officer on the beat usually carries a gun which he is theoretically restrained from using except when lives are threatened. He also carries a night stick—a considerably milder form of punishment. On certain occasions, policemen are armed with tear gas, truncheons, electric "prodders," and other weapons of assorted power. In mass protest situations like the 1968 presidential nominating convention of the Democratic Party, there were a number of "irregulars" and reserve personnel who carried riot sticks and other implements but who were generally not armed with revolvers. Newsreels and written reports of this event (see Walker report, 1969) seemed to suggest that those who carried revolvers used their sticks, prodders, etc., more frequently and more brutally than those who were not armed with lethal weapons. Although these observations are completely uncontrolled, what seemed to be indicated was that the amount of punishment with clubs and other relatively weak weapons is increased by the presence of a lethal but prohibited weapon. More theoretically, these impressions suggested that a "contrast effect" might have been influential in determining the amount of punishment delivered. That is, the punishment administered in order to accomplish a given effect, for example, to disperse, incapacitate, or inflict pain on protestors, was judged to be weaker in contrast with the extremely harmful weapon that could not be used. To make up for the "insufficiency," a greater amount of punishment was required to do the job. In effect, the phenomenon being proposed with respect to judgments about the administration of punishment is the same that occurs when a person makes psychophysical judgments (Helson, 1964), or categorizes attitude items (Sherif & Hovland, 1961). According to this so-called judgmental process theory, the contrast effect occurs when there is a discrepancy between an "anchor" stimulus and the stim-

uli that are compared with it. When the discrepancy is large, judgments are contrasted. Thus, for example, if an initial anchor weight is markedly heavier than subsequent comparison weights, the latter are judged to be lighter than they would be in the absence of an anchor. Small discrepancies presumably result in judgments that are less contrasted. What makes prediction difficult, however, is that such judgments may actually reflect an "assimilation" tendency. For example, when the difference between a heavy anchor and comparison weights is relatively small, the latter may be judged *heavier* than would be the case without the anchor. Accordingly, the contrast hypothesis with respect to the administration of punishment may be stated only in terms of a large discrepancy between anchor and comparison stimuli: When there is a large discrepancy between a strong but prohibited level of punishment and weaker permissible levels, subjects will deliver a greater amount of the latter than in the absence of strong but prohibited punishment.

In considering the empirical implication of the contrast hypothesis, the question arises as to the meaning of "amount." It is theoretically possible for a large discrepancy to cause an increase in the frequency or the intensity of permissible punishment, or both. The hypothesis does not specify how the judged weakness or insufficiency of the punishment being delivered is to be made up. Consequently, the subjects in the present experiment were given an opportunity to increase the absolute frequency as well as the intensity of the punishment they administer.

The present experiment grew out of an earlier pilot study which apparently supported a diametrically opposed hypothesis. In this early study, the subjects were placed in the role of training rats through the administration of shocks. In the experimental condition, a lever capable of delivering an extreme shock was present but the subjects were strongly discouraged by the experimenter from using it. The results revealed an apparent "inhibition" rather than a contrast effect. That is, compared with a control group in which no extreme shock possibility was present, the experimental subjects made less frequent use of the permissible relatively weak

shocks. We believe that this outcome was due to the method of discouraging the use of extreme shock, namely, by telling the subject that it would induce convulsions in the animal. From what was gathered informally, it appeared that such a devastating and counter-productive consequence, coupled with uncertainty created by the absence of an absolute ban, created a generalized inhibition against the use of shock. Furthermore, the subjects may have responded to the demand characteristics of the situation and may have inferred that the experimenter wanted them to keep all levels of shock to a minimum. The pilot study, in other words, did not permit an unconfounded test of the contrast hypothesis. In the present experiment, the strong levels of shock were absolutely proscribed, not by the experimenter's personal discouragement, but by the requirements of the experimental design. Additionally, the implication was given that, even when employed by other groups of subjects, the strong shock does not incapacitate the animal.

METHOD

Subjects

The subjects were 64 male volunteers from introductory psychology classes. Four subjects were eliminated from the analysis for various reasons.² The remaining 60 subjects were randomly distributed across three conditions with 20 subjects in each

² Two subjects were eliminated because they succeeded in training the rat to criterion; a third was eliminated due to apparatus failure; and a fourth subject was dropped because he misunderstood the instructions.

group. The subjects were contacted by telephone and asked to participate in an experiment involving "teaching methods."

Procedure

The subjects were told that the study was concerned with the rat's learning process and involved an approach to the problem from a new perspective. By asking naive persons to try to train the rat under various conditions, and by interviewing successful and unsuccessful trainers, it was hoped that we would discover the optimal training principles.

After the subject had been given this orientation, the experimenter said:

Before beginning the experiment, I would like to measure your shock threshold. You will not be shocked during the experiment, but you will need a standard of comparison for judging the shock levels you will be administering to a rat during the training period. Is that all right with you?

All of the subjects agreed to this procedure. The experimenter proceeded to measure the subject's sensation and pain thresholds and told the subject to consider these as "mild" and "slightly painful" shocks, respectively. The subject was then told to watch a television monitor for the rest of the instructions. It was explained that his objective was to train the rat to press the lever in the Skinner box by applying shocks that would last one-tenth of a second. Following some general remarks on aversive operant conditioning, the subject was shown a videotaped demonstration of a rat pressing a bar for water to assure him that rats can learn that response, and a second demonstration of a rat in a Skinner box being shocked by someone using mild and slightly painful shocks. Thus, the subject became familiar with the effects of shock on the rat. A mimeographed form was presented to the subject describing the range of shock intensities employed in the series of experiments. The control subjects were assigned two relatively weak levels of shock inten-

TABLE 1
MEAN FREQUENCY OF MILD AND SLIGHTLY PAINFUL SHOCKS AND PERCENTAGE OF SLIGHTLY PAINFUL SHOCK DELIVERED IN FIRST AND SECOND TRAINING PERIODS BY SMALL DISCREPANCY, LARGE DISCREPANCY, AND CONTROL SUBJECTS

Condition	First 6 minutes			Second 6 minutes		
	Mild	Slightly painful	% slightly painful	Mild	Slightly painful	% slightly painful
Control	40.0 ^a	15.0 ^b	23.7 ^c	53.5 ^d	20.1 ^e	23.9 ^f
Small discrepancy	54.3	21.6	28.2	64.5	33.1	32.2
Large discrepancy	48.2 ^a	33.7 ^b	38.6 ^c	73.5 ^d	35.7 ^e	32.9 ^f

Note.—All tests are based on the two-tailed Mann-Whitney *U* statistic. *n* = 20 for each of the three conditions.

^a *U* = 177, *ns*.

^b *U* = 97, *p* < .02.

^c *U* = 101.5, *p* < .02.

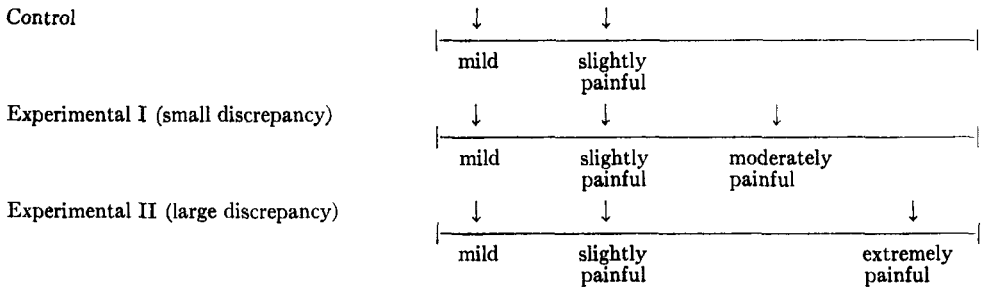
^d *U* = 141, *ns*.

^e *U* = 131, *p* < .10.

^f *U* = 141.5, *ns*.

sity: mild and slightly painful.³ The two experimental groups were assigned both of these alternatives plus a third: a moderately painful shock or an

extremely painful shock, respectively. The scales describing the options given to the subjects in each condition are as follows:



It can be seen that between the slightly painful and the moderately painful shock there is a *small discrepancy*, while between the slightly painful shock and the extremely painful shock there is a *large discrepancy*.

After examining the scale, and reading a description of each of the shock levels on it, the subjects in the experimental conditions read a second communication telling them that in the particular group to which they had been assigned only two of the three levels of shock intensity—the mild and slightly painful shock—could be used. In other words, the moderately painful shock level in the small discrepancy condition and the extremely painful shock level in the large discrepancy condition were prohibited. Thus, the only difference between experimental and control conditions was the presence in the former of moderately or extremely painful shock levers which could not be used.

All of the subjects were then told that according to results of a similar experiment with University of Wisconsin freshmen, *they should be able to train the rat in 6 minutes*. The subjects were also told that the effects of mild shock should be sufficient to train the rat.⁴ Training was defined as a 30-second period of repetitive bar pressing. Bar pressing was signaled to both the subject and the experimenter by a buzzer activated by movement of the bar.

The rats were male baby albinos weighing between 150 and 250 grams. Rats were discarded after being used once. The mild shock actually delivered to them was .2 milliamps; the slightly painful shock was 1.2 milliamps. Shock was delivered to the floor grid of the Skinner box through a scrambler.

When the basic instructions were completed, the experimenter turned on a large laboratory clock which sat in front of the subject and told him to begin training. On the basis of previous observations, we were confident that the subjects would not be able to train the rat within the time limit of the

experiment. After 12 minutes, the experimenter stopped the training and gave the subject brief questionnaires on the subject's mood (the Multiple Affect Adjective Check List; Zuckerman, Lubin, Vogel, & Valerius, 1964), on his attitudes toward his experience as a trainer, and his hypotheses and suspicions regarding the experiment. Finally, the subject was told about the purpose of the experiment, his questions were answered, and he was requested not to discuss the experiment with other students.

RESULTS

Since it was implied by the experimenter that the animal could be trained in 6 minutes, the shock data from the 12-minute training period are divided into two equal periods representing the amount of punishment delivered within the normatively prescribed time and the period which exceeds this standard. Table 1 shows the mean absolute frequency of mild and slightly painful shocks and the mean percentage of slightly painful shock delivered in each of the periods by the subjects in experimental and control conditions.

Looking first at the data within the normative period, it can be seen that the large discrepancy subjects delivered a higher average frequency of both slightly painful and mild shocks than the control subjects. In the case of mild shocks, the difference is not significant, while the difference in the frequency of slightly painful shocks delivered is both large and significant ($p < .02$). It can also be seen that the percentage of slightly painful shocks administered by the average large discrepancy subject is significantly greater than it is for the average control subject. These results would appear to be in complete accord with the statement of the contrast hypothesis

³ Two levels of shock in the lower part of the range were used so as to provide a measure of the relative intensity of shock used.

⁴ The implied recommendation that the animal could be trained with the use of the mild shock was made in order to bias the subjects against the exclusive use of the slightly painful shock.

proposed above. The subjects who experienced a large discrepancy underestimated the punishment their rat was receiving and as a result delivered a greater amount of shock than the control subjects for whom there was no discrepancy between prohibited and permissible shock levels. The contrast effect is particularly manifested both in greater absolute frequency and proportion of slightly painful shock. Finally, it should be noted that the average frequency and proportion of slightly painful shock delivered by the small discrepancy subjects fall in between the control and large discrepancy subjects. This result suggests that the small discrepancy made for judgments that were less contrasted on the average than judgments made under large discrepancy conditions.

Turning to the second half of the training period—the part that reflects more training time being taken than the experimenter believes is necessary—we see that the large discrepancy subjects were still delivering more frequent slightly painful shocks than controls ($p < .10$). The large discrepancy subjects still administered a greater average proportion of slightly painful shocks, but this difference is no longer significant. It is again noted that the small discrepancy subjects fall in between on all three measures. The notable item in the second 6-minute period is with respect to mild shock. Here, on the average, the large discrepancy subjects delivered 20 more mild shocks than the control subjects. What has happened becomes clearer when the average change in shock delivery from the first to the second period is considered. The average percentage of slightly painful shock by the large discrepancy subjects *decreased* (a drop of almost 6%), while the average frequency of mild shock *increased* by 25.3 ($p < .01$, two-tailed Wilcoxon signed-ranks test). In other words, the large discrepancy subjects increased the proportion of mild shock in the second 6 minutes. The picture in the second 6 minutes, then, is that the contrast effect is still evident in the large discrepancy-control difference in frequency of slightly painful shock, but at the same time there has been a relative shift by the large discrepancy subjects to increased employment of mild shock. Insofar as this increased emphasis on mild shock represents a reduction in shock inten-

sity, it would seem to be inconsistent with the contrast hypothesis. However, it is also possible that the shift to increased mild shock was brought about by constraints associated with the second 6 minutes of training and that what occurred is not at all contradictory to the hypothesis. There are two plausible interpretations of increased mild shock by the large discrepancy subjects:

1. It was observed that as time progressed the subjects succeeded in moving their animals closer to the lever in the Skinner box. When the subjects became aware of this, they tended to increase their use of mild shock so as not to produce gross escape movements in the animal. Although the subjects in all three conditions showed this "fine tuning" effect, the large discrepancy subjects were prone to judge the mild intensities as relatively weaker. Hence, they compensated more on this level.

2. When the large discrepancy subjects perceived that the animal was not being trained within the prescribed time period, their judgment of the relative weakness of their efforts was confirmed, and their sense of insufficiency enlarged. Thus, although the subjects in all three conditions felt that they were not succeeding with the rat, the large discrepancy subjects felt it more strongly. The pressure on them to shift to the level of shock intensity recommended by the experimenter as being adequate was stronger than for the control subjects.

In general, then, large discrepancy subjects increased their emphasis on mild shock either because they have not succeeded or because, in a sense, they did not want to undo their successful shaping of the rat. In the first case, their greater sense of failure led them to over-conform to the experimenter's implicit recommendation; in the second case, the control effect became localized at the mild level of shock intensity to which they shifted for fine-tuning purposes.

In either training period, the higher frequency and proportion of slightly painful shocks by the large discrepancy subjects raises the question as to whether the actual behavior of the animal was responsible for the effect. That is, was the greater use of shock due to the poorer performance of the large

discrepancy rats under their tutelage? A tabulation of the average number of buzzes in each condition (it will be recalled that a buzzer sounded whenever the rat pressed the lever) fails to show any systematic difference across conditions (control = 8.0, small discrepancy = 5.1, large discrepancy = 8.5). There is, then, no evidence that the large discrepancy subjects were obtaining relatively poorer results from their animals, or of the large discrepancy subjects being reinforced by success for shocking their animals.

We have seen that the mean frequency of mild and slightly painful shocks as well as the percentage of the latter administered by the small discrepancy subjects are, with one exception, in between the large discrepancy and control averages. This suggests that there was less contrast produced by a small discrepancy. However, this is not to say that assimilated judgments did not occur. The presence of the subjects who both overestimated and underestimated the effect they were having should make for greater variability in the small discrepancy condition. In line with this analysis, the variance of the small discrepancy distributions was larger in every comparison, with control and large discrepancy distributions with F ratios ranging from 1.14 to 3.26.

DISCUSSION

The results clearly support the contrast hypothesis: The presence of relatively strong but prohibited means of punishment causes individuals to increase the absolute frequency and relative intensity of a weaker permissible level of punishment.

Inevitably, there are some alternative interpretations, or at least refinements, that need to be delineated for further work on the problem. For instance, one could distinguish between a pure psychophysical contrast effect and a "rationalizing" effect. In the first case, as we have proposed, the presence of an extreme level of punishment functions as a scale anchor causing the punishing agent to underestimate the punishment he is delivering. According to the second effect, the presence of an extreme but prohibited level of punishment allows the punishing agent to justify to himself the use of increased punishment. The agent feels freer to increase punishment

when, for one reason or another, he does not use the extreme degree of punishment. It is as if he says: "This guy is pretty lucky—I could kill him if I were to use the *really* punishing intensity."

Presumably, the contrast effect is based on the anchoring function of the extreme punishment stimulus. In the present experiment, the extreme punishment was not only present but prohibited. One might propose an interpretation that centers on the prohibition. Thus, the more extreme the prohibited punishment, the more "deprived" the subject feels. It will be recalled that the subjects were told that the extreme shock level could not be used in the experimental condition to which they had been assigned. Prohibiting an extreme degree of punishment conceivably could have aroused a very strong sense of relative deprivation or discrimination and stimulate trainers to employ the shock levels that they are allowed to use with greater frequency. What casts doubt on this interpretation is the informal finding that, when questioned about it later, many of the subjects said they experienced *relief* upon hearing that they would not be allowed to use the extreme shock. Furthermore, there was no evidence in the questionnaire results for a greater sense of deprivation on the part of the large discrepancy subjects.

In discussing the generality of the contrast hypothesis, it is important to keep in mind not only that the moderate and extreme levels of punishment were prohibited but that the prohibition was not violated. It would be another matter if the prohibition were lifted. Indeed, if given the opportunity, it is quite likely that the large discrepancy subjects who encountered difficulty in training the animal in the prescribed period of time would have shifted to the more extreme level of punishment rather than to the implicitly recommended milder level. Similarly, where the person being "trained"—such as the protestor—strikes back at the agent, it is likely that pressure to violate the ban on extreme intensities will be built up.

Could the relatively greater punishment administered by the large discrepancy subjects be explained by the eliciting cue theory of Berkowitz (1964)? According to this theory, the cue value of a stimulus is a function of its association with a previous aggres-

sion-arousing situation. As already mentioned, a gun on a table stimulates more aggression (shock) toward a person who had previously angered the subject than does a badminton racket. Presumably, the gun has stronger symbolic associations with aggressive situations in the subject's past than the racket. Similar reasoning may be applied to experiments which show that more shocks are delivered to a confederate, who earlier had shocked the subject after the subject views a violent movie episode (the protagonist of which bore the same name as the confederate; Berkowitz, 1965). We find it difficult to apply this theory to the present data. For one thing, the punishment stimuli were placed along a single quantitative continuum. There is no a priori reason to assume that any one intensity level along this continuum has greater cue value in administering shock for training purposes than any other. Indeed, one might argue that (a) relatively weak intensities of punishment are more frequently associated with aggressive situations than extreme intensities which are rarely employed, and (b) extreme levels of punishment are just as likely to be associated with mass *inhibition* of aggression than with the elicitation of it. Furthermore, the experiments that have attempted to test the eliciting cue hypothesis have shown that an experimentally induced state of anger needs to exist for the elicitation of aggressive responses by external cues. There is no evidence that the trainers in the present experiment were aroused to anger in the first (normative) period when the contrast effect was most strongly in evidence.

Although we cannot think of any viable alternative to the contrast interpretation, it should be noted that critical evidence of an underestimation of punishment being delivered by the large discrepancy subjects is lacking. Postsession questionnaires, especially an item asking the subjects to estimate the amount of pain the animal had suffered, do not differentiate across groups. We have no way of knowing if the scales were insensitive or whether, because of the delivery of compensatory shock, the estimates of pain were, in fact, equalized.

The general implication of this study would seem to be that the frequency of punishment

by weapons is partially a function of the range of punitive power in the weapons being carried. Thus, to prevent a police riot involving the excessive use of clubs and other crowd-control weapons, revolvers should not be carried. Alternatively, one could substitute for the lethal revolver a weapon that is not so extremely powerful, for example, a tear gas or rubber pellet gun. If this moderately extreme weapon could then be severely restrained, a lower frequency of night stick punishment would be expected than in the case where a lethal weapon is carried.

REFERENCES

- BARON, R. A. Magnitude of victim's pain cues and level of prior anger arousal as determinants of adult aggressive behavior. *Journal of Personality and Social Psychology*, 1971, 17, 236-243.
- BERKOWITZ, L. Aggressive cues in aggressive behavior and hostility catharsis. *Psychological Review*, 1964, 71, 104-122.
- BERKOWITZ, L. The concept of aggressive drive: Some additional considerations. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. New York: Academic Press, 1965.
- BERKOWITZ, L., & LE PAGE, A. Weapons as aggression-eliciting stimuli. *Journal of Personality and Social Psychology*, 1967, 7, 202-207.
- FESHBACH, S., STILES, W. B., & BITTER, E. The reinforcing effect of witnessing aggression. *Journal of Experimental Research in Personality*, 1967, 2, 133-139.
- HARTMANN, D. P. Influence of symbolically modeled instrumental aggression and pain cues on aggressive behavior. *Journal of Personality and Social Psychology*, 1969, 11, 280-288.
- HELSON, H. *Adaptation level theory*. New York: Harper & Row, 1964.
- MILGRAM, S. Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, 1963, 67, 371-378.
- PAGE, M. M., & SCHEIDT, R. J. The elusive weapons effect: Demand awareness, evaluation apprehension, and slightly sophisticated subjects. *Journal of Personality and Social Psychology*, 1971, 20, 304-318.
- SHERIF, M., & HOVLAND, C. I. *Social judgement: Assimilation and contrast effects in communication and attitude change*. New Haven: Yale University Press, 1961.
- WALKER, D. *Rights in conflict*. New York: Bantam Books, 1968.
- ZUCKERMAN, M., LUBIN, R., VOGEL, L., & VALERIUS, E. Measurement of experimentally induced affects. *Journal of Consulting Psychology*, 1964, 28, 418-425.

(Received June 1, 1971)