

## Specificity in Contrast Effects: Judgments of Psychopathology

MELVIN MANIS

*University of Michigan and Ann Arbor VA Medical Center*

AND

JOAN R. PASKEWITZ

*Ann Arbor VA Medical Center*

Received October 18, 1982

Three experiments were conducted to explore the specificity of the contrast effect in judgments of psychopathology. In the first two studies, respondents initially attempted to infer whether each item in a series of behavior samples (vocabulary definitions in one study, handwriting samples in the second) came from a schizophrenic or a nonschizophrenic patient. Some respondents were presented with highly pathological samples in this *induction* series, while others were presented with relatively nonpathological samples. These divergent experiences led to marked contrast effects in evaluating test stimuli from the *same* behavioral domain (e.g., additional vocabulary definitions), but had significantly *less* impact on the respondents' reactions to stimuli from a different domain (e.g., handwriting samples). A third experiment yielded similar results, using a paired-comparison methodology. In this study subjects first judged a series of high- vs low-pathology definitions. They were then presented with a series of matched stimulus *pairs*, each including one vocabulary definition and one handwriting sample. Respondents were to indicate the member of each pair that seemed more indicative of schizophrenia. People assigned to the high-pathology induction group typically chose the handwriting samples as being more indicative of schizophrenia, compared with respondents who were assigned to the low-pathology conditions. These results were interpreted as supporting a representational (perceptual) theory of contrast.

Studies of social judgment have repeatedly demonstrated the replicability of the contrast effect. In evaluating a given stimulus (e.g., *how heavy*

All statements are those of the authors, and do not necessarily reflect the opinions or policies of the Veterans Administration. Reprint requests should be mailed to Dr. Melvin Manis, Psychology Research, Ann Arbor VA Medical Center, 2215 Fuller Road, Ann Arbor, MI 48105.

*is this weight? Or, how serious is this crime?*), the stimulus being judged is normally displaced "away" from the bulk of the judge's recent experience. Thus, a respondent who has been assigned to an experimental condition that requires him or her to judge the weight of several predominantly *light* objects, will normally rate "midscale" stimuli as being heavier than a judge who has been assigned to a series of predominantly *heavy* weights. This contrast pattern has been observed in a wide range of studies, involving both psychophysical and social stimuli.

Despite its robust generality, however, the contrast phenomenon has been difficult to explain unequivocally. Two main interpretations have emerged. (a) Some investigators have followed Helson's lead (1964) in viewing contrast as a centrally mediated ("perceptual") phenomenon. According to this account, the "midscale" weight in our previous example simply *feels* heavier when one has been lifting a series of relatively light weights. (b) Other investigators (e.g., Anderson, 1975; Parducci, 1965; Stevens, 1958; Upshaw, 1969) have argued for a "semantic" interpretation, contending that contrast effects mainly derive from changes in the subject's use of the response vocabulary (i.e., the rating categories) provided by the experimenter. According to this account, then, contrast effects do not result from changes in cognitive representation (i.e., the respondent's "immediate experience" of the stimulus); instead, they are thought to reflect changes in the way this representation is *described* (rated).

Those who support the semantic explanation emphasize three important arguments:

1. The end points of a judgmental rating scale are normally difficult (if not impossible) to specify with precision. Thus, in a weight-lifting study, the respondent may legitimately be somewhat puzzled in trying to determine just what the experimenter means by "extremely light" and "extremely heavy" weights (if these are to function as the end anchors of the rating scale). In such a situation he/she may simply position the end anchors to fit the range of stimuli that the experimenter has provided; subjects who have been assigned to divergent experimental conditions may consequently use the response scale in different ways. This, in turn, implies that people who have assigned different ratings to a given stimulus need *not* differ in the cognitive representations that underlie these ratings.

2. Respondents in a judgment experiment may implicitly assume that they should use the available response categories with roughly equal frequency. Hence, a judge who has been exposed to a series of predominantly light weights may react to the infrequency with which he/she has used the response categories at the "heavy" end of the scale by establishing a lowered threshold for the "heavy" response alternatives when such judgments seem at all appropriate. A lowered threshold of

this sort could underlie the contrast pattern, even if the subject's cognitive representation (perception) of the test stimuli had been unaffected.

3. Subjects may respond so as to convey maximal differentiation of the stimuli that are presented (Rosenberg & Cohen, 1966). This may sometimes lead to the inhibition of reactions that are *implicitly* evoked, because the response in question is associated both with the stimulus then being judged and with other (similar) stimuli that the respondent has previously evaluated. According to this account, then, a subject who has rated a series of predominantly light weights may for a time inhibit his/her subsequent use of "light" ratings, since such responses, if emitted, might *not* reflect the perceived difference between the earlier ("light") stimuli, and the stimulus then being judged. This shift in "labeling rules" would of course, depend upon the subject's prior experience in the judgmental situation, and might account for the contrast effects that are so ubiquitously reported.

Previous investigators addressing this controversy have sometimes taken special care to select response measures that would (hopefully) be *stably* related to the respondent's cognitive representation. For example, in a study by Krantz and Campbell (1961), two groups of respondents were to estimate the length of a series of lines, in *inches*. The assumption here was that "number of inches" is a familiar response language that would be firmly linked to the respondent's subjective experience, *regardless* of the experimental treatment to which he/she had been assigned. This study yielded significant contrast effects (Subjects exposed to long lines estimated that lines of intermediate length were shorter than these same lines seemed to respondents who had been exposed to predominantly short lines). While this experiment is often regarded as requiring a perceptual interpretation, a skeptic might worry about the possibility that the obtained results may have derived from the subjects' attempts to use a wide variety of responses (see 1 and 2, above).

Manis (1971) has reported several experiments that attempted to bypass the traditional rating scale methodology and many of its attendant problems. In one study (Manis, 1967), subjects read a series of written descriptions based on posed emotions (e.g., fear, surprise). The respondents' task was to read each description, and then indicate the particular photograph (out of several alternatives) on which it was based. In induction trials, some respondents read descriptions of relatively *unpleasant* emotions, while others read descriptions of *pleasant* emotions. A clear-cut contrast effect was observed on a series of intermittent "test trials," when the two groups were presented with descriptions of neutral emotions (neither pleasant nor unpleasant). On these trials, subjects assigned to the unpleasant induction condition selected referent photographs that were rated as more pleasant (by a norm group) than the photographs selected by the pleasant

group. While these data were interpreted as supporting the view that the induction passages affected the cognitive representations of the test stimuli, they do not adequately address the problems posed by the equal frequency response bias, and by Rosenberg and Cohen's differentiation hypothesis. More recent work by Schneiderman and Manis (1978) suggests, however, that the contrast effects observed in this situation may indeed derive (at least in part) from central mechanisms. Studies by Krupat (1974) and by Simpson and Ostrom (1976), using very different experimental techniques, support a similar conclusion.

A series of studies by Mellers and Birnbaum (1982) is consistent with a more complex account, however. Respondents in one experiment judged the darkness of various stimulus squares. While a common subset of "test" squares was presented to *all* respondents, these stimuli were presented in disparate contexts. Some respondents were presented with an overall set of squares that was positively skewed with respect to darkness, while for others the stimulus distribution was negatively skewed. The two groups of respondents (positive vs negative skew) showed clear evidence of contrast when judging the common test stimuli. Subsequent research suggested, however, that these results might derive from differences in the respondents' "judgment function"; i.e., from a "semantic" shift in response language, rather than from changes in the representations (subjective scale values) of the individual test stimuli. On the other hand, when the judgment task required subjects to react to stimuli from *two* modalities (e.g., on each of several trials, some subjects were instructed to compare the *size* of a certain circle with the *darkness* of a square), a representational effect was observed. The results of this bimodality experiment suggested, in essence, that the representations of the individual test squares (how dark they looked) were affected by the overall array of squares that were presented in the course of the experiment. A conceptual replication of this work by Mellers (1982, Experiment 4) yielded similar results. In this study respondents judged the fairness of the compensation awarded to hypothetical faculty members, by comparing their salaries and their merit ratings.

This report is concerned with the specificity (vs generality) of the contrast phenomenon. Consider a situation in which respondents are provided with behavior samples that are to be used in estimating the psychopathology of the various individuals who produced these samples. Suppose moreover, that in an initial series of (context) judgments, some respondents are presented with predominantly "high-pathology" vocabulary definitions, while others receive "low-pathology" definitions. What will be the effect of these divergent contexts on subsequent judgments of psychopathology, when the respondents are presented with (a) additional vocabulary items and (b) with handwriting samples? A serendipitous finding in a previous experiment suggested that such judgments might

be essentially unaffected by the respondent's assigned "context condition" (high vs low pathology), unless the test stimuli were similar to the context stimuli; e.g., both vocabulary definitions.

The present experiments were designed to explore the robustness of that earlier finding. Parducci, Knoble, and Thomas (1976) have reported what appears to be a related result, using geometric forms (circles and squares) as stimuli. Their respondents rated the size of various stimuli from an intermingled series that included relatively small squares and somewhat larger circles. Subjects were instructed to establish a *separate* scale of size for the two types of stimuli, so that a square might be judged as "very large" (relative to the squares), and a circle as "very small" (relative to the circles), even though the circle in question was larger than the square. The results of this experiment indicated that the skewness of the stimuli within a given domain (e.g., squares) had a systematic effect on the judgments that were elicited by the other stimuli in that domain, with positive skewing leading to higher ratings than negative skewing (contrast). On the other hand, the judgments observed in one domain were *independent* of the skewness of the stimulus distribution in the other. In essence, then, in judging a particular stimulus (e.g., a square), the subjects in this study displayed contrast-like effects that depended on the overall array of squares that were included in the stimulus series. However, as in our own preliminary research (see above), the results observed by Parducci et al. were "domain specific," for judgments elicited by one type of stimulus (squares) were not affected by the distribution of stimuli in the other domain (circles).

The present report presents the results of three experiments in which respondents who had previously been exposed to divergent induction conditions rated the cognitive distortions implied in two types of behavior samples: vocabulary definitions and handwriting samples. In Experiments 1 and 2 the respondents rated each behavior sample separately. Experiment 3 involved a paired-comparison procedure; that is, in a series of test trials, subjects compared the psychopathology implied by a given definition with that implied by a matched handwriting sample.

## EXPERIMENTS 1 AND 2

Experiments 1 and 2 were very similar in design. They were undertaken to assess the replicability of the data pattern that we had previously obtained (see above). In both cases, subjects were initially exposed to either a high-pathology or a low-pathology series of behavior samples (vocabulary definitions in one experiment, handwriting samples in the other), each of which was supposedly obtained from a different "patient." For each stimulus the respondents were to indicate if they thought that the patient in question was schizophrenic or not. Following this induction phase, all respondents were presented with additional behavior samples

(the *test* series) which were said to be "randomly selected" from the same hospital as the initial series. Each test item was to be judged in terms of (a) the amount of thought distortion that it implied, (b) the patient's diagnosis (schizophrenic vs not schizophrenic), and (c) the respondent's *confidence* in his/her answer to item (b).

There were two types of test items in each experiment: handwriting samples and vocabulary definitions. Experiment 1 involved a between-subjects design; during the test phase of the experiment, half of the subjects from each induction condition (high vs low pathology) were presented with the vocabulary definitions and the other half with the handwriting samples. Experiment 2 involved a within-subjects design in which all respondents evaluated both types of test items, presented in an unpredictable sequence. The main difference between Experiments 1 and 2 lay in the *type* of material that was presented during induction. In Experiment 1 all respondents received an induction series that was composed of vocabulary definitions (high vs low pathology); in Experiment 2 the induction series involved a variety of handwriting samples (each patient had supposedly written the words "father" and "mother"). In both Experiments 1 and 2 there were two main groups of respondents. One group received an induction series that was relatively pathological (as evaluated by a norm group); a second group was presented with a nonpathological induction series.

### *Method*

*Subjects.* Experiments 1 and 2 were both conducted using students at the University of Michigan as respondents. These people were paid for their participation. In each experiment, half the respondents assigned to a given experimental condition (e.g., high-pathology induction) were male and half were female ( $N_1 = 72$ ,  $N_2 = 36$ ). Subjects were normally run in small groups (2-7), with some people tested individually.

*Induction booklets.* In both experiments, subjects were told that the study was concerned with "social judgments." In study 1, they were initially shown a series of 6 midrange definitions (as a warm-up), followed without delay by 20 "extreme" definitions, one to a page; these definitions were to be used in deciding if the patient in question was schizophrenic or not—a simple two-category choice. Respondents were given 10 sec to react to each of these items. They were told that the definitions in the booklet came from "a representative cross-section" of the patients in a state hospital. Unknown to the respondents, there were two different induction booklets, one containing relatively pathological definitions (scale positions ranged from 8.00 to 10.68 on an 11-point scale), while the other contained less-pathological definitions (1.50 to 3.95 on the rating scale). These items and their associated ratings were taken from an earlier study by Arnhoff (1953).

In study 2, the induction items were handwriting samples; each "patient" was represented by the words *mother* and *father*, which had ostensibly been taken from a biographical inventory that had been completed at admission. There were 6 midrange items followed by 20 "extreme" handwriting samples in each induction booklet. These had previously been rated for psychopathology by a norm group, whose mean ratings were used to determine the amount of thought disorder that each sample implied. The high-pathology induction booklet included 20 handwriting samples whose mean ratings ranged from 7.8

to 10.31 on an 11-point scale; the low-pathology booklet included 20 handwriting samples that ranged between 1.69 and 3.00.

When the induction booklets had been completed, the respondents in both experiments were told that they would receive a second series of stimuli which were also to be rated for psychopathology. These *test items* were presented as deriving from another "representative sample" of the patients in the hospital from which the induction series had been drawn. Each test item was presented by means of a slide projector. The experimenter tried to move to a new test slide each 10–15 sec, but waited until each subject in a given test session had completed his/her response. In the test series, subjects provided a more complete evaluation than was called for in the induction series. For each item, the subjects were to indicate: (a) the amount of thought distortion that it implied, using a 7-point scale that ranged from 1 (normal) to 7 (highly distorted/disturbed); (b) whether the patient seemed schizophrenic or not; and (c) their confidence with respect to the "diagnosis" provided in response to (b), using a 7-point scale that ranged from "guess" (1) to "very certain" (7).

Both Experiments 1 and 2 involved two types of test items: vocabulary definitions and handwriting samples. In both experiments, the vocabulary test items were located between 4.1 and 7.8 on the 11-point scale that was used to "calibrate" these materials. In Experiment 1, the handwriting samples fell between 1.7 and 7.1<sup>1</sup>; in Experiment 2, the handwriting samples in the test series were located between 4.1 and 6.6 on our 11-point scale.

In Experiment 1, each subject responded to 16 test items, the first of which was discarded as a "warm-up"; half the subjects assigned to each induction booklet were presented with handwriting samples, while the other half were presented with additional vocabulary definitions. In Experiment 2, following the induction series, all respondents were presented with a "mixed series" of 38 test items. Half of these items were handwriting samples and half were vocabulary definitions, presented in an irregular sequence; the test series included two warm-up items (one definition and one handwriting sample), which were discarded for purposes of analysis.

## Results

We initially analyzed our data by looking separately at (a) the subjects' ratings of *thought pathology*, and (b) their *schizophrenic judgment scores*, an index that combined the respondents' "diagnoses" concerning schizophrenia and their confidence ratings into a single 14-point scale that ranged from (1) "very certain" the patient is *not* schizophrenic, to (14) "very certain" that the patient *is* schizophrenic. Since measures (a) and (b) (above) proved to be highly correlated, they were combined additively into a single score (after standardization, to ensure that they would be equally weighted). This procedure was followed in both Experiments 1 and 2. Figure 1 presents the results from both experiments. The vertical axis in this figure represents a *difference score*, comparing the mean responses of our two induction groups (low pathology minus high pa-

<sup>1</sup> The handwriting samples in the test series of Experiment 1 extended across a somewhat broader range of the psychopathology continuum than the vocabulary definitions with which they were compared. However, subsequent research in an additional study (not fully reported here) indicates that the results of Experiment 1 are essentially unchanged when the test series includes a narrower, truly "midrange" set of writing samples, as in Experiments 2 and 3.

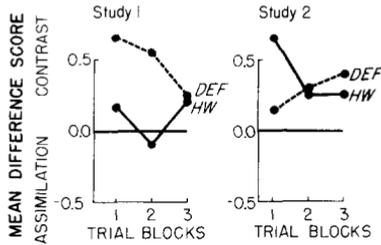


FIG. 1. Differences in judged psychopathology as a function of type of stimulus (definitions versus handwriting samples) and test blocks.

thology). Data are presented separately for evaluations of the definitions and the handwriting samples on three successive blocks of test trials. Difference scores that are *greater than zero* (i.e., positive differences) are indicative of a contrast effect. As in our earlier study, these results show clear evidence of contrast when the test items came from the *same* domain as the material in the induction series. Thus the respondents in Experiment 1 who were “induced” on vocabulary definitions show positive difference scores (contrast) in evaluating other vocabulary definitions; a similar contrast pattern appears in response to the handwriting samples in Experiment 2. In both cases, the contrast effect is most marked at the beginning of the test series and then tails off gradually. Now consider the “mismatched” test items, those from a *different* domain than the materials in the induction series. Here there is substantially less evidence that the induction series affected our respondents’ judgments, for the group differences are small and inconsistent from one study to the next.

To assess the replicability of these patterns, the data from each experiment were subjected to an appropriate analysis of variance. Experiment 1 involved a mixed design in which the respondent’s *induction condition*, *sex*, and type of *test item* (handwriting vs definitions) were treated as between-subject variables, with *test blocks* constituting a within-subject variable. Experiment 2 was analyzed similarly, save for the fact that in this case the type of test item was treated as a within-subject variable (since each subject received both handwriting samples and definitions). The results of these analyses supported the following conclusions:

1. In both studies, the effect of the respondent’s assigned induction treatment (high vs low pathology) depended on the type of test item he/she was evaluating. In Experiment 1 this interaction between treatment conditions and item types yielded an  $F(1, 64)$  of 4.31,  $p < .05$ ; this effect was qualified, however, by a significant three-way interaction,  $F(2, 128) = 4.70$  (induction level  $\times$  item type  $\times$  trial block), deriving from the fact that the group differences that were initially observed in response to the two types of test items had virtually disappeared by trial block 3 (see Fig. 1). Experiment 2 also resulted in a significant three-way interaction

between induction, item type, and trial block,  $F(2, 64) = 6.75, p < .005$ . Examination of Fig. 1 shows that as in Experiment 1, the first block of test trials yielded a relatively clear separation between the two types of test items in the amount of contrast that they elicited. A subsequent analysis of just the block 1 data showed a significant interaction between the respondent's induction level (high vs low) and the type of item being evaluated;  $F(1, 32) = 15.86, p < .001$ ; in blocks 2 and 3, on the other hand, the two types of test items did not differ reliably in the amount of contrast that they elicited ( $F < 1.00$ ).

2. In both studies, test items from the same domain as the induction series yielded significant evidence of contrast. Thus, in Experiment 1, the vocabulary items (considered alone) led to a significant main effect for level (high vs low pathology);  $F(1, 32) = 11.78, p < .005$ . A similar pattern was observed in Experiment 2. Here, the overall contrast effect for the handwriting samples yielded an  $F(1, 32)$  of 16.71,  $p < .01$ .

3. In Experiment 1, respondents assigned to the high- and low-pathology groups did not differ ( $F < 1.00$ ) in their evaluations of the "mismatched" test items (i.e., the handwriting samples). In Experiment 2, on the other hand, for the *overall* set of vocabulary items (considered separately) there was a modest but reliable contrast pattern,  $F(1, 32) = 5.50, p < .05$ . The replicability and generality of this last effect is unclear, however, for this pattern has not been found in other studies within the present series, nor was it observed in an analogous experiment by Parducci et al. (1976), using geometric stimuli.

### EXPERIMENT 3

Experiments 1 and 2 suggest that the presentation of a biased induction series, consisting (for example) of distorted, poorly organized vocabulary definitions, leads to contrast in estimating the amount of psychopathology implied by other stimuli of the same type (e.g., other vocabulary definitions); when, however, stimuli from a different domain are evaluated with respect to the psychopathology that *they* imply, the preceding induction experience does not have a consistent impact. While this pattern of results seemed compatible with an interpretation emphasizing a central (representational) mediator, Experiment 3 was designed to provide a more direct assessment of this possibility.

Unlike Experiments 1 and 2 where the individual handwriting samples and vocabulary definitions of the test series were rated *separately*, Experiment 3 involved *paired-comparison* judgments. Each stimulus pair consisted of a vocabulary definition and a (matched) handwriting sample that implied about the same amount of psychopathology. The underlying rationale for the experiment can perhaps be explained most readily by referring to Fig. 2, which depicts the situation confronting a respondent who had previously been exposed to a series of well-formed, *nonpath-*

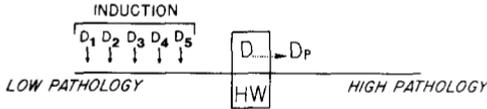


FIG. 2. A schematic representation of the theoretical basis for Experiment 3.

ological vocabulary definitions (symbolized by arrows  $D_1, D_2, \dots, D_5$ ). A representational account of our earlier results suggests that such an experience should at least temporarily lead the respondent to "displace" subsequent definitions toward the pathological end of the continuum. This is symbolized by the arrow running from  $D$  to  $D_p$  (the *perceived* location of our hypothetical test definition). The HW in Fig. 2 represents a handwriting sample that is typically seen (by norm subjects) as implying about the same degree of psychopathology as the definition with which it has been paired. If the representational account of our earlier results is correct, we would anticipate that when asked to indicate the more pathological of these two behavior samples our hypothetical subject should select the definition ( $D$ ). By analogous reasoning, we would anticipate that subjects who are presented with a *high-pathology* series of definitions (during induction) would normally choose the handwriting sample as the more pathological of the pair.

### Method

Experiment 3 was conducted with paid volunteers who were students at the University of Michigan. These people were run in small groups, ranging up to 9, with some respondents completing the experiment individually. There were 32 subjects, half men and half women (8 per cell of each sex).

*Procedure.* The subjects' instructions and initial induction experience were closely modeled after Experiment 1. After responding to 20 induction definitions that were presented in a booklet (2 midrange "warm-ups" and 18 polarized definitions), each respondent was presented with a series of 18 stimulus pairs, each pair consisting of a "midscale" vocabulary definition and a handwriting sample that had received about the same "pathology rating" when presented to a norm group. For each pair, the respondents were instructed to indicate which person was schizophrenic, the "top" person or the "bottom" person of the pair, and how confident they were about that choice, using a 7-point scale that ranged from "guess" to "very certain." These test pairs were presented on slides with a subject-paced presentation rate (as in Experiments 1 and 2). The definition was presented above its matched handwriting sample for half of the test pairs; this ordering was reversed on the remaining test items.

### Results

The results of Experiment 3 are presented in Fig. 3, which compares the mean test choices of our two respondent groups (those initially assigned to the high- vs the low-pathology definitions). The dependent variable in this graph reflects the mean number of choices (out of the six test pairs in each trial block) in which the definition was chosen as indicative

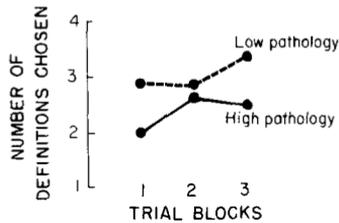


FIG. 3. Paired comparison judgments as a function of induction (high or low pathology) and test blocks.

of more pathology than its matched handwriting sample. Results are presented separately for the two induction conditions.

These results are quite compatible with Experiments 1 and 2. That is, in evaluating the psychopathology implied by the paired behavior samples, there was a clear tendency for respondents in the divergent induction groups to differ in their choices. When presented with a test series consisting of matched definitions and handwriting samples, those who had previously responded to high-pathology definitions (in the induction series) were *less* likely than those in the low-pathology group to select the vocabulary definitions as being more “schizophrenic” than the handwriting samples with which they were paired.

To assess the reliability of these results, the data from Experiment 3 were subjected to a 2 (induction condition)  $\times$  2 (sex)  $\times$  3 (trial blocks) analysis of variance, with the last of these being a within-subject variable. The difference between the two induction groups (high vs low pathology) was significant [ $F(1, 28) = 6.80, p < .025$ ]. Similar results were obtained from an analysis in which the respondents’ choices were weighted by their confidence ratings, yielding a 14-point scale that ranged from (1) very certain that the handwriting sample was schizophrenic, to (14) very certain that the vocabulary definition was schizophrenic. For this weighted measure, the overall contrast effect yielded an  $F(1, 28)$  of 7.68 ( $p < .01$ ).

### Discussion

Perhaps the most important finding in these studies is the relative *specificity* of the contrast phenomenon—the fact that it appears quite clearly when respondents evaluate stimuli of the same type as those presented in the induction series, but does not appear with comparable clarity and consistency when stimuli from a different domain are presented for evaluation. Judgmental theories that emphasize the shifting character of the subject’s response scale do not provide a convincing account for these experiments, particularly Experiment 3, where respondents were simply asked for paired-comparison judgments.

Paired comparisons seem less ambiguous than “absolute” ratings in many situations. If Smith is rated as *very tall* by one rater, and simply

as *tall* by another, we cannot be confident that these two respondents differ in their private (phenomenal) impressions; they may well hold to a common view of Smith, but may describe that impression in rather different terms. On the other hand, if we learn that one judge regards Smith as taller than Jones, and that a second judge favors the reverse ordering (Jones taller than Smith), we can feel reasonably confident that these judges disagree in their cognitive representations. The underlying basis for our confidence in this second instance derives, we believe, from the plausible idea that for skilled speakers of English, the contrasting statements, "Jones is taller than Smith" vs "Smith is taller than Jones," are unlikely to reflect a *common* (shared) view of the world. Considerations of this sort lead us to give special attention to the results of Experiment 3, where a biased induction series affected our respondents' subsequent reactions to a paired-comparison task.

If we assume (as above) that the present results reflect systematic changes in cognitive representation, it is interesting to speculate as to underlying mechanisms that may be involved:

1. Following in the tradition established by Helson (1964), it might be contended that the results of all three experiments reflect representational effects, in which the presentation of a biased induction series influenced the respondents' phenomenal reactions to subsequent test items from that domain (e.g., other definitions). An account of this sort seems most plausible if we assume, for example, that a respondent who had been presented with a low-pathology series of definitions during induction, would recognize that the midscale "test" definitions were more pathological than those that had come before. This recognition (in turn) might lead to the inference that the test definitions might reflect a substantial degree of psychopathology in some *objective* sense (e.g., schizophrenia). An account of this type emphasizes the possibility that ordinal judgments regarding the pathological character of the test definitions relative to the induction items may, through an "inferential leap," affect the respondent's immediate impressions, given that the test items come from the same domain as the induction items.

2. An alternative account focuses primarily on Experiment 3, where the evidence for a representational effect seems clearest because of the paired-comparison methodology that was used there.<sup>2</sup> This account starts with the assumption that in considering a given test pair, respondents

<sup>2</sup> The specificity pattern that was observed in Experiments 1 and 2 may, in this view, derive from the sort of end-anchoring effects that Upshaw found in his work on congeneric scales (1978). In these studies, persuasive manipulations that affected the scale boundaries of one response scale were ineffective in affecting other (congeneric) scales. Analogously, the manipulation of induction materials in Experiment 1 may have affected the response scale used for labeling (rating) test definitions, without influencing the response scale that was used to evaluate handwriting samples.

may evaluate the constituent behavior samples separately, assigning each item an "implicit rating" based on its standing relative to other items of that type (e.g., other definitions); the judge's choice as to the more pathological of the two elements in a pair would then depend on these "implicit ratings." If, for example, the definition seemed relatively pathological compared to other recently seen definitions and the handwriting seemed average, the respondent might conclude that the definition implied a more severe state of psychopathology than the handwriting sample.<sup>3</sup>

This type of account is compatible with the results of an important study by Sherman, Ahlm, Berman, and Lynn (1978). These investigators demonstrated that the importance ratings assigned to the topic of *recycling* might be systematically influenced (in a contrastive direction) by the relative importance of *other* topics that the respondent had rated. The Sherman et al. study showed, moreover, that under appropriate conditions of "high saliency," ratings of the recycling topic were directly related to the respondents' subsequent behavior in volunteering to help a local recycling project. The authors concluded that the judgmental pattern they had obtained was probably semantic in character (Upshaw, 1969; 1978), but that this "labeling effect" had ultimately influenced the respondents' everyday actions, serving as a behavioral input for self-inference (Bem, 1972; Salancik & Conway, 1975). In a similar vein, the "representational effects" that were observed in the paired-comparison task (see Fig. 3), may depend on the subjects' implicit responses to the component behavior samples. Recent research by Mellers (1982) and Mellers and Birnbaum (1982) have prompted a similar model, derived from Parducci's range-frequency theory (1965); this general approach is also consistent with an article by Higgins and Lurie (1983).

## REFERENCES

- Anderson, N. H. (1975). On the role of context effects in psychophysical judgment. *Psychological Review*, *82*, 462-482.
- Arnhoff, F. H. (1953). *Some factors influencing the unreliability of clinical judgments*. Unpublished PhD dissertation, Northwestern University.
- Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press.
- Dixon, N. F. (1971). *Subliminal perception: The nature of a controversy*. New York: McGraw-Hill.
- Helson, H. (1964). *Adaptation-level theory*. New York: Harper & Row.
- Higgins, E. T., & Lurie, L. (1983). Context, categorization, and recall: The "change-of-standard" effect. *Cognitive Psychology*, *15*, 525-547.
- Krantz, D. L., & Campbell, D. T. (1961). Separating perception and linguistic effects of context shifts upon absolute judgments. *Journal of Experimental Psychology*, *62*, 35-42.

---

<sup>3</sup> The authors acknowledge the contributions of Harry Upshaw and Jim Sherman for independently proposing this general line of analysis.

- Krupat, E. (1974). Context as a determinant of perceived threat: The role of prior experience. *Journal of Personality and Social Psychology*, **29**, 731-736.
- Manis, M. (1971). Context effects in communication. In M. H. Appley (Ed.), *Adaptation-level theory* (pp. 237-255). New York: Academic Press.
- Manis, M. (1967). Context effects in communication. *Journal of Personality and Social Psychology*, **5**, 325-334.
- Mellers, B. A. (1982). Equity judgment: A revision of Aristotelian views. *Journal of Experimental Psychology: General*, **111**, 242-270.
- Mellers, B. A., & Birnbaum, M. H. (1982). Loci of contextual effects in judgment. *Journal of Experimental Psychology: Human Perception and Performance*, **8**, 582-601.
- Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, **72**, 407-418.
- Parducci, A., Knobel, S. & Thomas, C. (1976). Independent contexts for category ratings: A range-frequency analysis. *Perception & Psychophysics*, **20**, 360-366.
- Rosenberg, S., & Cohen, H. D. (1966). Referential processes of speakers and listeners. *Psychological Review*, **73**, 208-231.
- Salancik, G. R., & Conway, M. (1975). Attitude inferences from salient and relevant cognitive content about behavior. *Journal of Personality and Social Psychology*, **32**, 829-849.
- Schneiderman, W., & Manis, M. (1978). Response vs. perception: An enduring question revisited in the domain of context effects. *Motivation and Emotion*, **2**, 259-273.
- Sherman, S. J., Ahlm, K., Berman, L., & Lynn, S. (1978). Contrast effects and their relationship to subsequent behavior. *Journal of Experimental Social Psychology*, **14**, 340-350.
- Simpson, D. D., & Ostrom, T. (1976). Contrast effects in impression formation. *Journal of Personality and Social Psychology*, **34**, 625-629.
- Stevens, S. S. (1958). Adaptation level vs. the relativity of judgment. *American Journal of Psychology*, **71**, 633-646.
- Upshaw, H. S. (1969). The personal reference scale: An approach to social judgment. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 4). New York: Academic Press.
- Upshaw, H. S. (1978). Social influence on attitudes and on anchoring of congeneric attitude scales. *Journal of Experimental Social Psychology*, **14**, 327-339.