Journal of Experimental Psychology
1961, Vol. 62, No. 1, 35–42

SEPARATING PERCEPTUAL AND LINGUISTIC EFFECTS OF CONTEXT SHIFTS UPON ABSOLUTE JUDGMENTS[1]

DAVID L. KRANTZ AND DONALD T. CAMPBELL
Northwestern University

In accordance with the usual strategy of parsimony in science, the effects of extreme contexts upon absolute judgment were interpreted initially as illustrating a *single* process of relativity of judgment or adaptation level. The most economical interpretation of the data to be presented here necessitates the postulation of *two* such processes, both producing contrast effects in the same direction. One of these processes, transient and reversible, is found in both of two judgmental languages, and is designated *perceptual.* The other, irreversible and specific to an unfamiliar, experimenter-defined response language, is designated *semantic.*

Two papers published in 1958 independently suggest dividing the judgmental phenomena currently grouped under Helson's concept of adaptation level (e.g., Helson, 1947, 1948, 1959) into semantic or linguistic effects on the one hand, and perceptual or end-organ effects on the other. One of these papers (Campbell, Lewis, & Hunt, 1958) argued that the "situationally relative," "novel," "arbitrary" and "restricted" response language ordinarily used in the method of single stimuli could well have produced the contrast effects resulting from extreme anchors or extreme contexts without generating distortions in identity judgments (had they been called for) or any perceptual illusion. That study then examined the effect of a shift in context of stimulation upon the judgment of a common stimulus, employing a judgmental language which was "absolute, extensive, and extra-experimentally anchored." Finding the usual contrast or adaptation level effects, the paper concluded that such phenomena could not be explained away as semantic artifacts. However, it did not raise the possibility that both semantic and perceptual effects are present in the typical judgmental assignment.

The other paper suggesting a duality of processes was that of Stevens (1958) in which he argued that different phenomena were involved in sensory adaptation (in which end-organ receptor-cell adaptation takes place) on the one hand, and in judgment (where redefinition of categories is involved) on the other, although previous approaches treat the two as if the same process were involved. Here again, two processes were postulated with no explicit assumption that both could be present in the same judgmental task. The outcomes of the present study seem to require such an assumption.

The setting of the present study is the judgment of the lengths of white lines projected on a screen. A common "tracer" stimulus recurs in all phases of the experiment, in one phase interspersed among longer lines, in another phase occurring among shorter lines. One set of *S*s judges in "inches," the other uses an *E*-defined

[1] This study is based upon an MA thesis done by the first author under the second author's direction. Certain analyses beyond those reported in the thesis (Krantz, 1959) have been financed by ONR Project 7onr-450(11), W. A. Hunt, Principal Investigator. The authors are also indebted to Allen Parducci for editorial suggestions.

response language for which the value "100" is "average." The effects of the shift in context upon judgments of the common tracer stimulus are examined.

Method

Slides were prepared which projected white lines of 13 different horizontal lengths, for a .2-sec. duration, with a 2-sec. interstimulus interval. The slides were made by using black tape on clear glass $3\frac{1}{4} \times 4$ in. slides. Projected on the screen, the shortest line (No. 1) was approximately 6 in., the longest 36 in., the middle "common" or "tracer" stimulus (No. 7) 20 in. (the exact values of the stimuli are presented in Table 1) and the vertical widths $3\frac{3}{4}$ in. The slides were organized in groups of seven (here called "trials") within each of which the common tracer Stimulus 7 occurred once, the order of stimuli within a trial being random. For the Low Phase, the stimuli in each trial ranged from Stimulus 1 to Stimulus 7. For the High Phase, the range was from Stimulus 7 to Stimulus 13. Trials in the Transition Phase also spanned seven steps, graduating between the Low and High Phases. The three transition trials ranged between Stimuli 3–9, 4–10, 5–11 for Low to High and vice versa for a High to Low transition. Both a High-Low and a Low-High group were run for each combination of the other three experimental variables.

In the second experimental treatment variable, the "Inches" condition was presented to *S*s with these instructions, "You are to judge these lines in terms of how long, in inches, they appear to be." The "Rating" condition was introduced this way:

You are to judge these lines in terms of a rating scale which you will construct during the course of the experiment. When you see a line which appears "average in length" to you, judge that line with the number 100. If a line appears less than average, assign the number 95. If another line appears even smaller than the average line and smaller than the 95 one, use the number 90 to designate "less, less than average."
In the same way, if a line is above average in length, assign the number 105. If another line is larger than the 100 average line and larger than the 105 line, assign the number 110. In other words, you can build your rating scale above and below the average rating of 100. You are not restricted to the numbers I have given you. The numbers I have given you on the reference sheet are only examples. In-between numbers, like 93 or 106 are appropriate. You can use numbers less than 90 or greater than 110. In other words, you are constructing a scale which shows the differing lengths of lines presented. Any system of numbers, such as in-between numbers or numbers above 110 or below 90 are usable. The only stipulations are that you use 100 as average and that if you assign a number to a particular length of line, the next time that line is presented you assign it the same number. Let me repeat, always use the same number for the same length of line; do not change your judgment of the same line.

Let me review the categories or numbers on your reference sheet. The number 100 stands for the average line; 95 for less than average; 90 for less, less than average; 105 for more than average; 110 for more, more than average.

In choosing this particular form of rating, there was an effort to eliminate the effects of "vocabulary exhaustion" (Campbell, Lewis, & Hunt, 1958) such as occurs when all of a restricted number of ratings have been employed prior to a shift of context, and to exclude a deliberately relative or comparative usage by *S*s which would allow them consciously to redefine values in the course of the experiment.

The third experimental variable employed was Preshift Training, i.e., the length of the initial phase, presumably manipulating the persistence of the adaptation level established therein. "Phase 1" indicates the initial set of trials given, irrespective of the context employed. This period was followed by the transition series ("Phase 2"). A final series of trials after shift of context is designated "Phase 3." In the Short condition, five Phase 1 trials were used; in the Long condition, 15.

The fourth variable was the degree of illumination, which was presumed to affect stimulus discriminability. Under the Bright condition, the projected lines registered .73 apparent ft-c. Under the Dim condition a neutral density filter reduced the brightness to .32 apparent ft-c. The background brightness of the screen was .26 apparent ft-c for both illumination conditions.

These four dichotomies generate 16 experimental treatments. For each of these, 20 *S*s were obtained, a total of 320 *S*s. The *S*s were drawn from introductory classes in psychology. They were tested in groups of four to six, seated in two rows, 13 and 16 ft. from the screen. Experimental sessions were randomly assigned to treatments, with the restriction that all 16 treatments be evenly represented during all periods of the three month course. In two successive quarters the experimental design was replicated with 10 *S*s in each cell.[2]

The *S*s recorded their own judgments in a

---

[2] One session of 6 *S*s (in treatment High-Low, Rating, Short training, Dim illumination) was discarded because of extremely aberrant results. This session was subsequently rerun twice with different *S*s to confirm the aberrance, attributable to intraexperimental history (Campbell, 1957). It is felt that this substitution of the rerun group data augments rather than reduces the replicability of the study's findings.

TABLE 1

MEAN JUDGMENT OF EACH STIMULUS DURING FIRST FIVE TRIALS OF PHASE 1

Stimulus Number	Low Context			Stimulus Number	High Context		
	Length (In.)	Mean Judgment: Inches	Mean Judgment: Ratings		Length (In.)	Mean Judgment: Inches	Mean Judgment: Ratings
1	6.1	3.74	79.90	7	20.2	12.74	90.98
2	9.0	5.37	88.99	8	23.0	14.38	94.00
3	11.0	6.75	92.48	9	25.0	16.95	97.78
4	12.4	7.80	96.09	10	28.1	20.90	101.74
5	15.8	10.77	102.55	11	31.3	23.32	103.72
6	18.7	13.32	106.29	12	33.1	27.16	108.21
7	20.2	17.21	111.77	13	35.8	31.59	112.63
Mean		9.28	96.87	Mean		21.01	101.29

Note.—Preshift training and discriminability conditions pooled, $N = 80$ *S*s per mean.

booklet of slips with each page allowing 10 judgments. Twelve *S*s were discarded for getting out of step with the stimulus presentation and for incomplete records. Excess *S*s over the 20 required for each treatment were discarded by the use of random numbers.

RESULTS AND DISCUSSION

While the statistical analyses to be presented have been done separately for each judgment language, in Fig. 1 and 2, there is an effort to show both sets of results on the same graph. This necessitates computing a conversion factor between the two judgment languages, which in turn necessitates examining data on judgments of stimuli other than Stimulus 7, data not otherwise used. Because of their relevance to the attempted translation between the two languages, and to the more general problem of the relationship of stimulus and judgmental distances, these data are presented in Table 1.

The data do not lead to any univocal conversion ratio. If ratings be plotted against stimuli, the plots are satisfactorily linear, but of differing slope for the two contexts, the High context being the flatter. If judged inches be plotted against stimuli, the slopes depart slightly from linearity in the unexpected direction of positive acceleration for both High and Low contexts, with the High context being the steeper in this case. When rating means are plotted against judged inches, the Low context shows a slight bowing, and a steeper slope than the High context. As an approximate average of the slope of the two contexts, a ratio of 1 in. = 1.50 rating points has been chosen for the purpose of graphing Fig. 1 and 2. The graphs have been centered on the mean for the two contexts of the column means of Fig. 1, i.e. 15.64 in. and 99.08 rating points. Because of the differing slopes for the two contexts, this represents a more arbitrary decision than had been hoped for.

In Fig. 1 and 2, all data represent judgments of the same tracer stimulus, No. 7, and all connected points represent judgments by the same 40 *S*s. From the Phase 1 differences between the corresponding High-Low and Low-High groups, one can see that the expected contrast effects emerge. The striking shifts in each line between Phase 1 and Phase 3 are likewise uniformly in the expected direction. But perhaps most dramatic is the difference in overall pattern between the Inches and Ratings conditions, a difference which provides the central problem of interpretation.

A summary of the results of six independent analyses of variance, shown in Table 2, provides statistical

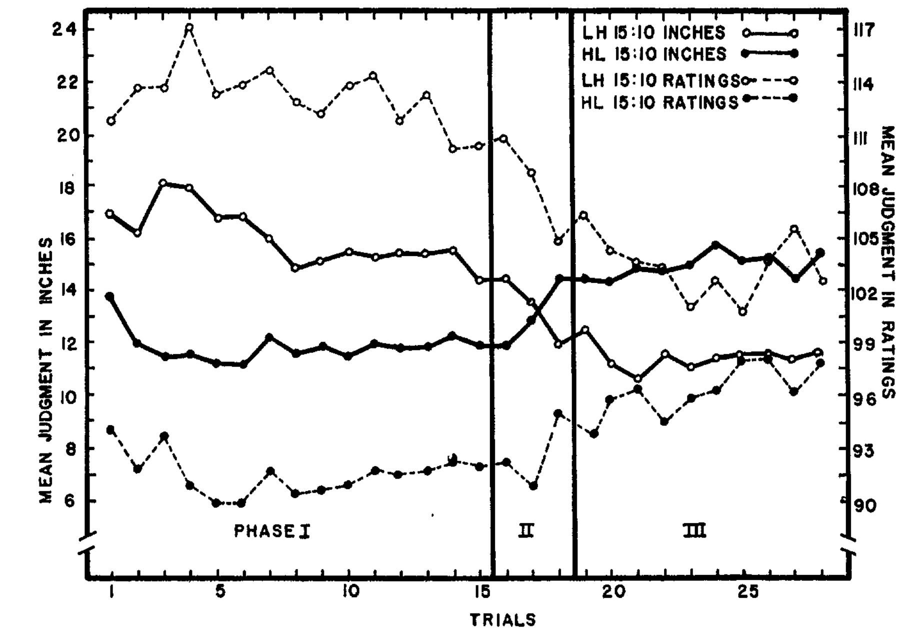

FIG. 1. Mean trial-by-trial judgment of tracer Stimulus 7 for Long training condition. (Discriminability conditions pooled, $N=40$ for each point.)

confirmation of the overall pattern for Inches and Ratings conditions. In the Phase 1 analyses, the mean of each *S*'s judgment of Stimulus 7 during the last four trials of Phase 1 has been used, since this set of responses is comparable across Preshift Training conditions. For both response languages, the main effect of context is highly significant and is the only significant effect. As graphically illustrated in Fig. 1 and 2, the effect is greater for the Ratings. Discussion of outcomes such as these frequently imply that the difference between the *F* ratios is itself significant. Rarely, however, is this higher-order significance tested. Bradley and Schumann (1957) have recently provided an *F* test among *F*s appropriate for parallel experiments such as the present one. By their criteria ($w'=21.32$, $a'=33.47$, $b=76$, $P<.05$, minimum $w'_{.05}=1.85$),[3] the effect is significantly greater for the Ratings condition. This is one of the major findings of this study.

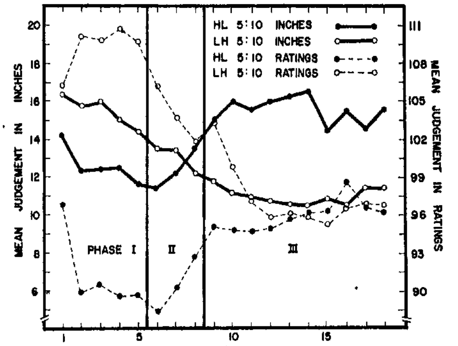

FIG. 2. Mean trial-by-trial judgment of tracer Stimulus 7 for Short training condition. (Discriminability conditions pooled, $N=40$ for each point.)

[3] The tables supplied by Bradley and Schumann provide w' values only at 5% level of significance. These minimum w' values will be presented in parenthesis following the obtained values. Note that the obtained values greatly exceed these minimums.

The Phase 3 analyses are based upon the means of each *S*'s judgments of Stimulus 7 in the 10 Phase 3 trials. Here the effect of context is highly significant for the Inches condition. (That the *F* of Phase 3 Inches seems more significant than that of Inches in Phase 1 is due primarily to reduced variability which is presumably attributable to the greater amount of practice and to scores based upon 10 judgments rather than 4.) As seen in Fig. 1 and 2, the effect of context on Ratings is weak and in the opposite direction from that on Inches. This main effect ($P < .05$) is graphically clear only for the Long Preshift Training. The difference in Phase 3 context effects *F*s between Ratings and Inches is highly significant by the Bradley and Schumann (1957) test ($w' = 7.94$, $a' = 4.23$, $b = 76$, $P < .05$, minimum $w'_{.05} = 3.18$).

Note the apparent inconsistency of the Phase 1 and Phase 3 findings. For Phase 1 both response languages show context effects in the same direction with Ratings being larger. For Phase 3 the context effects are in different directions, with Inches being greater. In the graphic presentation of trial-by-trial results of Fig. 1 and 2, the Inches judgments show complete reversal with the reversal of contexts, the lines crossing in the transition phase. In contrast, the Ratings never cross over, the direction of difference in Phase 3 remaining the same as in Phase 1.

Minor Findings.—In the Phase 3 analysis, the effect of Preshift Training is significant at the marginal 5% level for the rating groups. Those groups with shorter Phase 1's give lower ratings in Phase 3. This main effect is almost wholly contributed by the Low-High groups with the Short Phase 1.

The findings of the Phase 1-Phase 3 Shift analysis are not to be confused with a test of the significance of the Phase 1-Phase 3 shift, the uniformly high significance of which is reported below. Rather, the Phase 1-Phase 3 shift analysis shown in Table 2, examines the effects of treatments other than Phase upon the amount of shift. The score employed for each *S* is the difference in mean judgment between the last 4 trials of Phase 1 and the 10 trials of Phase 3, the direction of subtraction being reversed between High-Low and Low-High groups so that all changes in the expected contrast direction are positive. The analysis was designed primarily to demonstrate the expected effects of Discriminability and Preshift Training on stability of the adaptation level, or upon the differential occur-

TABLE 2

F VALUES FROM ANALYSES OF VARIANCE

Source	*df*	Phase 1 Judgments		Phase 3 Judgments		Phase 1-Phase 3 Shift	
		Inches	Ratings	Inches	Ratings	Inches	Ratings
Order (context) (O)	1	11.95**	254.79***	40.99***	5.16*	2.42	18.48***
Preshift training (T)	1	.79	2.27	.29	5.12*	1.99	4.16*
Discriminability (D)	1	1.40	1.34	.47	3.47	1.22	.22
O × T	1	.05	.02	.52	2.66	.54	1.85
O × D	1	.55	1.02	.01	.43	1.14	1.08
T × D	1	.54	.34	.23	.37	.97	.57
O × T × D	1	.00	.25	.20	.62	.02	.01
Within groups (*MS*)[a]	152	(23.20)	(60.35)	(21.52)	(84.27)	(8.72)	(42.98)
Inhomogeneity of variance	1/20	1.60	1.85	2.42	1.24	2.79	4.87*

[a] Mean squares shown in parentheses, rather than *F*s.
* $P < .05$.
** $P < .01$.
*** $P < .001$.

rence of assimilation and contrast illusions among *S*s. Of the 320 *S*s, only 20 had negative scores and these were individually insignificantly different from no shift at all. Thus, this setting produced no persons who were significant assimilators, unlike two other settings in which these effects have been examined (Campbell, Hunt, & Lewis, 1957, 1958; Campbell et al., 1958).

The significant main effect for Order for shift scores is conceptually an interaction with the counterbalancing treatments. The shift is greater under the Low-High order than under the High-Low. This is very significant for Ratings and the trends for Inches are in the same direction. Note that the significant heterogeneity of variance casts suspicion upon this one analysis.

For Discriminability treatments, all effects are convincingly absent. For Preshift Training, effects are generally absent, with the 5% level findings for Ratings being in the direction of greater shift under Short Phase 1 treatments, a finding similar to that of Johnson (1949). While this effect is primarily contributed by the one Low-High Short Rating group, which also accounts for the Phase 3 Training main effect, slight trends are in the same direction for High-Low Ratings, and also for the two Inches comparisons.

Possible Interpretations.—To return to the initial findings presented, our preferred interpretation for the striking discrepancies between the Phase 1 and Phase 3 analyses is this: The symmetrical and reversible findings for Inches represent the operation of a single perceptual contrast process. This process is also present, and reversible, for the Ratings judgments, but in addition, the Phase 1 Ratings data contain a second process of defining perceptual equivalents for novel judgmental terms. That this operates in an adaptation level (Helson, 1947, 1948, 1959) fashion is indicated in Table 1 by the closeness with which the mean ratings of *all* stimuli approach the designated "average" value of 100 in all treatments. Of the Phase 1 effects of context on Ratings (of the tracer Stimulus 7) about one-half is attributed to this adaptation-level definitional process, the other half to the perceptual contrast effect. This definitional process is deemed irreversible without explicit retraining involving instructional feedback. This irreversibility is particularly strong under our instructions, "always use the same number for the same length of line," but would probably be so under usual single stimuli instructions, except in so far as "vocabulary exhaustion" (Campbell, Lewis, & Hunt, 1958) forced redefinition. In Phase 3, the experimentally reversed perceptual contrast effect and the perseverating definitional contrast effects tend to cancel each other just as they augmented each other in Phase 1.

This interpretation leads to the expectation that the magnitude of the Phase 1-Phase 3 shift for both Inches and Ratings should be the same, inasmuch as both are primarily affected by perceptual contrast effect, resulting from the experimental context reversal. While our choice of a common graphic scale for Fig. 1 and 2 was a somewhat unsatisfactory compromise, the method of selecting it was independent of the Phase 1-Phase 3 shifts. Hence, the degree to which the shift magnitudes turn out to be similar for the two judgmental languages encourages this interpretation. Note particularly in Fig. 1 the parallel course which the two High-Low lines take, and similarly the two Low-High lines. Graphic superposition (not shown here) shows them to be practically indistinguishable. We anticipate that were a more unequivocable psychophysical scale equivalence to be achieved, as through a study in which all 13 stimuli occur in each "trial" with an unchanging "phase," the magnitudes of our Phase 1-Phase 3 shifts would not significantly differ when translated into the common metric.

Some other indirect evidence for the equivalence of shifts under the two conditions comes from the t ratios for shift, computed by pooling all Discriminability and Preshift Training conditions. For Ratings and Inches, respectively, these were 11.92 and 11.33 for Low-High, and 10.95 and 11.34 for High-Low ($df = 79$, $P < .001$, each case). If the t ratios can be used as comparable measures of magnitude of effects (since the number of *S*s are equal), the effects are strikingly equivalent, with not the slightest trend toward statistical significance of difference. This as evidence for equivalence is of course quite inferential. However, the mode of inference is quite independent of that based upon the conversion formula employed in the graphs. The approximate agreement of these two independent forms of inference is taken as encouraging the retention of the hypothesized two-process interpretation.

The temporal weighting of experience.—The attractiveness of the two-process interpretation may be increased by considering an alternative uniprocess theory which would presume quite different weighting parameters in the temporal accumulation of experience under the two conditions. Three temporal weighting outcomes can be considered: a primacy effect, in which the first visual experiences would have greater effect than later ones in determining the adaptation level at any given time; a recency effect, in which the adaption level would be predominantly determined by very recent experience; and an equal temporal weighting outcome, in which all experiences (within the experimental session at least) would be equally weighted. Inches judgments, as seen in Fig. 1 and 2, clearly show a strong recency effect. The reversibility is almost complete, the two curves cross almost at the center of the transition phase. Apparently only the current trial stimuli or the ones just preceding are having much influence on the adaptation level against which the tracer stimulus is being judged.

In contrast, the Ratings fit perfectly a model of equal temporal weighting. Under this assumption, the judgments of Stimulus 7 for the High-Low and the Low-High groups should converge at that trial at which the total of experimental stimuli presented was the same, for at that point, each would be judged against the same adaptation level. For the Short treatment, this would be at Trial 14, i.e., when the number of Phase 1 and Phase 3 trials were equal. This is exactly what happens. For the Long groups, this would happen at a projected Trial 33. From Fig. 1 it seems plausible that this might occur although a slight primacy effect is also plausible.

While each of these uniprocess interpretations is plausible, it does not seem reasonable that so small a difference in conditions as the reporting language could make such a fundamental difference in the perceptual and memory processes as to provide an equal weighting in one instance and a recency weighting in another. Furthermore, a time-specific recency effect is probably the general rule for all nervous system processes, both end-organ adaptation and memory. For example, end-organ adaptation as involved in judgments of temperature, salinity, brightness, etc. all show recency effects so dominant that the primacy-recency problem, or the temporal weighting problem in any form, is almost never raised. In judgment studies, Parducci (1954, 1959) and Campbell, Lewis, and Hunt (1958) have found recency effects, and recency effects are to be inferred from Johnson's (1949) data. Recency effects likewise seem universal in learning, even if concomitant with and occasionally disguised by primacy effects (e.g., Miller & Campbell, 1959).

Accepting the dual-process interpretation would lead to the conclusion that inferences as to temporal weighting made from single-reversal studies using the method of single stimuli in its usual forms are erroneous. These would include Helson's early experiment (1947, pp. 16–17) used to justify equal temporal weighting, and the studies of Campbell, Hunt, and Lewis (1957, 1958) and Segall (1959). The two-process interpretation would predict that recency weighting is more general and that apparent equal weighting or primacy effects are a function of the semantic rather than perceptual effects. The evidence assembled here is not crucial for the two-process interpretation. A more definitive approach would be multiple reversal studies of a Low-High-Low-High nature. With increasing numbers of reversals of context, the semantic factors should decrease in salience, allowing the perceptual aspect to be manifest more directly.

Summary

The effects of divergent and shifting stimulus contexts upon the judgment of a common tracer stimulus were measured for the task of judging lengths of lines projected on a screen. For judgments made in inches, a reversible perceptual-contrast effect was demonstrated. Judgments made with an uncrowded numerical rating scale showed in addition a response-definitional contrast effect augmenting the apparent contrast effect in the initial phase and persisting into subsequent phases. These findings cast doubt upon previous estimations of the appropriate temporal weighting of stimuli in contributing to an adaptation level. Effects of stimulus discriminability and preshift training are also examined.

References

Bradley, R. A., & Schumann, D. E. The comparison of the sensitivities of similar experiments: Theory. *Ann. math. Statist.*, 1957, **28**, 902–920.

Campbell, D. T. Factors relevant to the validity of experiments in social settings. *Psychol. Bull.*, 1957, **54**, 297–312.

Campbell, D. T., Hunt, W. A., & Lewis, N. A. The effect of assimilation and contrast in judgment of clinical materials. *Amer. J. Psychol.*, 1957, **70**, 347–360.

Campbell, D. T., Hunt, W. A., & Lewis, N. A. The relative susceptibility of two rating scales to disturbances resulting from shifts in stimulus context. *J. appl. Psychol.*, 1958, **42**, 213–217.

Campbell, D. T., Lewis, N. A., & Hunt, W. A. Context effects with judgment language that is absolute, extensive, and extra-experimentally anchored. *J. exp. Psychol.*, 1958, **55**, 220–228.

Helson, H. Adaptation-level as a frame of reference for the prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 297–313.

Helson, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 1–29.

Helson, H. Adaptation-level theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959.

Johnson, D. M. Learning function for a change in scale of judgment. *J. exp. Psychol.*, 1949, **39**, 851–860.

Krantz, D. L. Stimulus and response relationships in absolute judgment. Unpublished master's thesis, Northwestern University, 1959.

Miller, N., & Campbell, D. T. Recency and primacy in persuasion as a function of the timing of speeches and measurements. *J. abnorm. soc. Psychol.*, 1959, **59**, 1–9.

Parducci, A. Learning variables in the judgment of single stimuli. *J. exp. Psychol.*, 1954, **48**, 24–30.

Parducci, A. An adaptation-level analysis of ordinal effects in judgment. *J. exp. Psychol.*, 1959, **58**, 239–246.

Segall, M. H. The effect of attitude and experience on judgment of controversial statements. *J. abnorm. soc. Psychol.*, 1959, **58**, 61–69.

Stevens, S. S. Adaptation level vs. the relativity of judgment. *Amer. J. Psychol.*, 1958, **71**, 633–647.

(Received April 30, 1960)