The Category Effect in Social Judgment: Experimental Ratings of Happiness

Douglas H. Wedell and Allen Parducci
University of California, Los Angeles

Three experiments demonstrated the applicability of a range–frequency analysis to social judgments. Subjects rated the happiness of either (a) schematic drawings of faces or (b) life events as expressed in short verbal descriptions. The relative frequency of these stimuli was manipulated experimentally, as was the number of rating categories. Consistent with psychophysical research, ratings became less sensitive to differences in the frequencies of contextual stimuli as the number of categories increased (the category effect). With more categories, ratings also showed less adjustment to the range of stimuli actually presented. The reduction in adjustment was greater when stimuli were presented successively and when the experimental set covered a limited range. These effects of varying the number of categories were interpreted as reflecting changes in the effective context for judgment: With more categories, the differences between the effective frequencies of contextual stimuli are reduced and new, more extreme comparison values are evoked. The implications of using coarse versus fine scales of judgment are discussed in terms of the dynamics of social judgment as well as choice of the appropriate number of categories in social research. We argue that selection of the number of rating categories should be guided by research objectives rather than by a search for "true" judgments.

Social scientists use category ratings as a gateway to the mind, to assess attitudes, perceptions, feelings, opinions, and preferences. One reason for their popularity is that they represent the way people commonly express subjective evaluations. Statements such as "The movie was *excellent*," "I am *slightly confused*," and "He is a *staunch conservative*" may be understood as evaluations that locate the stimulus or event along an implicit rating scale.

Although rating scales appear to provide a tantalizingly direct avenue to the inner "psyche," this roadway is fraught with difficulties. In addition to the fundamental problem of determining the relation between explicit responses and implicit beliefs (cf. Hovland, Janis, & Kelley, 1953), ratings shift dramatically with shift in context (Helson, 1964). The contextual dependency of category ratings is not necessarily a liability, however, because the lawfulness of these effects (as demonstrated in laboratory studies) provides a basis for inferring the context for judgment, which may well be of fundamental interest.

The specific context retrieved at the time of judgment may sometimes depend on subtleties in the wording of questions or instructions (see Kahneman, Slovic, & Tversky, 1982, for a review). In addition to the more typical framing effects associated with manipulations of phrasing, features of the rating scale itself can influence the context. For example, the research of Schwarz and his associates (Schwarz & Hippler, 1987; Schwarz, Hippler, Deutsch, & Strack, 1985) has demonstrated how category labels can provide survey respondents with norms or standards with which they can compare themselves with others. In these experiments, manipulation of labels affected not only the responses to the item itself, but also the responses to other items that apparently relied on comparisons to the label-generated norms.

The *number* of rating categories can also dramatically affect the context upon which judgments are based. It has commonly been believed that the number of categories included in the rating scale does not affect the *processes* underlying the judgments so that, for example, ratings made using 6- or 9-point scales may be linearly related (cf. Parducci & Perrett, 1971). However, recent psychophysical research has demonstrated that as the number of categories increases, ratings are less sensitive to the relative frequencies of contextual stimuli (Parducci, 1982; Parducci & Wedell, 1986).

Because of the wide use of category ratings in social science research, it is important to determine whether this *category effect* found for psychophysical judgments extends to social judgments. Beyond the methodological implications of such a finding, the category effect could itself prove an important mediator of social and self-perception. For example, judgments of happiness or satisfaction are known to exhibit strong contextual

The authors acknowledge the helpful comments of the reviewers of this article as well as those of Robert Wyer, Ulf Bockenholt, Richard Smith, and James Austin. Preparation of this article was completed in part while Douglas H. Wedell was at the University of Illinois on a postdoctoral traineeship, ADAMHA National Service Support Award MH14257, Lawrence E. Jones, Training Director.

Correspondence concerning this article should be addressed to Douglas H. Wedell, Department of Psychology, University of Illinois, 603 East Daniel Street, Champaign, Illinois 61820.

Journal of Personality and Social Psychology, 1988, Vol. 55, No. 3, 341–356

dependencies (Brickman & Campbell, 1971; Parducci, 1968). According to the category effect, an individual who naturally utilizes just two categories (e.g., *pleased* or *displeased*) in evaluating satisfaction with everyday events will be more sensitive to the shifts in frequencies of related events than an individual who utilizes a more fine-grained scale to judge events. Furthermore, because the category effect has been linked to a particular feature of the contextual distribution (the relative stimulus frequencies), testing for the effect with social stimuli provides an indirect method for determining how contextual information is represented in memory.

We present three experiments that test for the category effect with social judgments. Because this effect represents an interaction between the number of categories and the distribution of contextual stimuli, we begin by describing how category ratings are related to the contextual distribution.

The Context-Dependent Nature of Category Ratings

The context for judging a particular stimulus may be conceived as the set of stimuli with which it is compared, in other words, the set affecting how it will be judged.[1] Perhaps the most prominent feature of category ratings is their relativity: The rating assigned any particular stimulus reflects its position within the distribution of stimuli. Manipulation of the experimental set of stimuli produces what is generally referred to as a *contrast effect* (Beebe-Center, 1932): Judgments are displaced away from the values of contextual stimuli. For example, the same square may be rated *small* when the set of stimuli consists mostly of larger squares but *large* when the set consists mostly of smaller squares. Similar contextual effects have been reported for a wide variety of social domains including judgments of happiness (Brickman, 1975; Parducci, 1968), performance (Mellers & Birnbaum, 1983), fairness (Mellers, 1983a, 1986), degree of guilt (Pepitone & DiNubile, 1976), attractiveness (Kenrick & Gutierres, 1980; Wedell, Parducci, & Geiselman, 1987), apprehensiveness (Krupat, 1974), and degree of psychopathology (Campbell, Hunt, & Lewis, 1958; Manis & Paskewitz, 1984). The pervasiveness of these effects suggests that comparisons with recent contextual events constitute an integral part of the social judgment process.

Several theories describe the relation between category ratings and the distribution of stimuli either as deviations from a measure of central tendency (Helson, 1947, 1964; Johnson, 1944; Thibaut & Kelley, 1959) or as locations within the range of stimuli (Upshaw, 1962; Volkmann, 1951). However, careful manipulations of contextual distributions in the psychophysical laboratory have produced effects that are inconsistent with such theories (cf. Birnbaum, 1974b; Parducci & Perrett, 1971). Instead, these experiments support a more complex account of judgment in which the evaluation of the stimulus depends on both its location in the range as well as its rank in the distribution of contextual stimuli. These contextual dependencies are embodied in Parducci's (1963, 1965) range–frequency theory.

A Processing Model for Contextually Based Judgment

In Figure 1 we summarize the range–frequency model in the form of an informational flow chart (where processes for which sequential ordering is not well established are shown at the same level). This type of processing framework makes clear the distinction between the set of stimuli experimentally presented and the effective set of contextual stimuli retrieved and used at the time of judgment. In previous applications of range–frequency theory, it was generally assumed that all of the experimentally presented stimuli were included in the contextual set influencing judgments (e.g., Birnbaum, 1974b; Parducci & Perrett, 1971). However, our recent studies of the category effect (Parducci & Wedell, 1986; Wedell & Parducci, 1985) suggest that the effective context consists of a subset of the experimental stimuli and that the stimulus values included in that subset depend crucially on the number of rating categories.

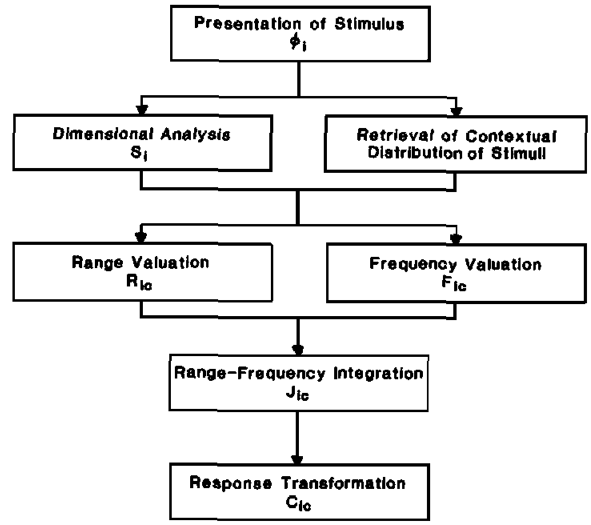

Figure 1. A process model for range–frequency judgment.

As described in Figure 1, the value of the presented stimulus is first located along the specified dimension of judgment. For psychophysical judgments, dimensional analysis corresponds to a simple sensory evaluation, but for more complex social stimuli, the analysis requires retrieval and integration of semantic information along the relevant dimension. In either case, the scale value of the stimulus along the chosen dimension is assumed to be invariant across contexts. The assumption of invariance of underlying scale values has been tested and supported by the good fits of the model to both psychophysical and psychosocial data (e.g., Birnbaum, 1974b; Mellers, 1986; Mellers & Birnbaum, 1982, 1983; Parducci & Perrett, 1971; Riskey, Parducci, & Beauchamp, 1979; Wedell et al., 1987), although inherently ambiguous stimuli may give rise to different scale

[1] The term *context* has been used in at least two different senses in the general literature: (a) the set of stimuli presented in the experimental situation and (b) the set of hypothetical or inferred stimulus values necessary and sufficient for predicting the ratings. In this article we use the modifier *effective* to distinguish the second meaning from the first. In some simple experiments (e.g., number judging), the context is the effective context; but in some of our research, the effective context has been inferred to be an unrepresentative sample from the context, sometimes including hypothetical stimulus values not actually presented. It is the effective context that is generally of theoretical interest to psychologists.

values in different contexts (cf. Hamilton & Zanna, 1974; Herr, Sherman, & Fazio, 1983; Higgins, Rholes, & Jones, 1977; Srull & Wyer, 1979; Wyer, 1974; Wyer & Srull, 1981, 1986). Because these invariant scale values are not ordinarily accessible to the subject, contextual effects may still have a strong phenomenological impact.

In order to judge the stimulus, a set of contextual stimuli is retrieved for comparison. This set is referred to as the effective context. Retrieval is assumed to take place unconsciously, automatically, and in parallel with the values of the retrieved stimuli rapidly aggregated along the dimension of judgment. This assumption of rapid aggregation of stimulus values is shared by other theories of contextual processing (e.g., Helson, 1964; Hintzman, 1986; Kahneman & Miller, 1986; Restle, 1978). We assume that only a limited number of contextual stimuli are retrieved, an assumption consistent with the finding that sequential dependencies do not generally extend beyond the last seven presentations (Lockhead & King, 1983; Ward & Lockhead, 1971). Further support for a limited retrieval set comes from a study of transfer following shifts in the distribution of psychophysical stimuli: The retrieval set was estimated to consist of only the 10 to 20 most recent stimulus presentations (Wedell, 1984/1985).

Two valuation processes (corresponding to the range and frequency principles of judgment) locate the stimulus within the retrieved distribution of contextual stimuli. According to the range principle, the judgment of a stimulus corresponds to the proportion of the contextual range lying below it (cf. Parducci, 1983; Volkmann, 1951). The range value, R_{ic}, of Stimulus i in Context c is given by

$$R_{ic} = (S_i - S_{min})/(S_{max} - S_{min}), \tag{1}$$

where S_{min} and S_{max} are the minimum and maximum values included in that context. The range principle accounts for differences in ratings between contexts defined by different end-stimuli (e.g., other things being equal, the same event is more satisfying the closer it is to the most satisfying event in the effective context).

According to the frequency principle, the judgment of a stimulus is equal to the proportion of the total number of stimuli lying below it (i.e., its percentile rank divided by 100). The frequency value, F_{ic}, is given by

$$F_{ic} = (r_{ic} - 1)/(N_c - 1), \tag{2}$$

where r_{ic} is the rank of Stimulus i in Context c and N_c is the total number of contextual stimuli. The frequency principle accounts for the effects of any nonuniformity in the contextual distribution such as skewing or normality (e.g., with the same pair of contextual endpoints, an event will be more satisfying when a higher proportion of contextual events fall below it on the dimension of satisfaction).

The outputs of range and frequency processes are then integrated into an internal judgment. Thus, the subjective evaluation of the stimulus, J_{ic}, reflects a compromise between range and frequency principles that can be represented algebraically as their weighted average:

$$J_{ic} = wR_{ic} + (1 - w)F_{ic}, \tag{3}$$

where w is the relative weighting of these principles. In many judgment situations, range and frequency principles appear to be weighted almost equally, in other words, w approximates 0.5 (e.g., Birnbaum, 1974b; Parducci, 1963, 1965; Parducci & Perrett, 1971; Riskey et al., 1979).

There has been long controversy as to whether contextual stimuli alter the phenomenological evaluation of the event or merely the overt response (e.g., Krantz & Campbell, 1961). The location of the judgment at a subjective level within the model is supported by the fact that contextual effects occur across a wide variety of response modes, for example, category ratings (Parducci & Perrett, 1971), magnitude estimations (Mellers, 1983b; Parducci, 1963), matching tasks (Manis, 1967; Mellers & Birnbaum, 1982), unconstrained written descriptions (Simpson & Ostrom, 1976), comparative judgments (Manis, Paskewitz, & Cotler, 1986), and even physiological correlates to the experience being measured (Krupat, 1974; for counterarguments see Anderson, 1982; Poulton, 1979; Sherman, Ahlm, Berman, & Lynn, 1978; Stevens, 1971; Upshaw, 1978).

The final stage of the judgment process is to transform the internal judgment to an overt response. In the case of category ratings, a linear response transformation is assumed:

$$C_{ic} = bJ_{ic} + a, \tag{4}$$

where C_{ic} is the overt category rating, b is the range of categories (e.g., 8 for a 9-point scale), and a corresponds to the number assigned the lowest rating category.[2] The generally good fit of the range–frequency model to a wide variety of contextual manipulations of both psychophysical (Birnbaum, 1974b; Parducci, 1963, 1965; Parducci & Perrett, 1971; Riskey et al., 1979) and social stimuli (Mellers, 1983a, 1986; Mellers & Birnbaum, 1983; Wedell et al., 1987) provides support for the proposed integration processes, as outlined in Equations 1 through 3, and also for the assumption of a linear transformation function, Equation 4.

The Category Effect

The category effect refers to an *interaction* between the number of rating categories and the effects of varying frequencies of contextual stimuli: The magnitude of the contrast effect decreases as the number of categories increases (Parducci, 1982; Parducci & Wedell, 1986). A simple way to characterize the category effect within range–frequency theory is as a reduction in the weight of the frequency principle ($1 - w$) with an increase in the number of categories. In fits of the range–frequency model to psychophysical judgments, Parducci and Wedell (1986) found that the empirically estimated frequency weightings varied systematically from as much as .87 for two categories to as little as .07 for a 100-point scale. This means that ratings on the two-category scale corresponded closely to the percentile ranks of the stimuli (following the frequency principle alone), whereas ratings on the 100-point scale were virtually independent of manipulations of the relative frequencies of contextual stimuli (the range principle). For example, subjects us-

[2] When the number of categories is small (i.e., only two or three categories), the assumption of a continuous rating scale is clearly violated and therefore a transformation that takes into account the stepwise nature of the rating function must be used (cf. Parducci, 1965; Parducci & Wedell, 1986).

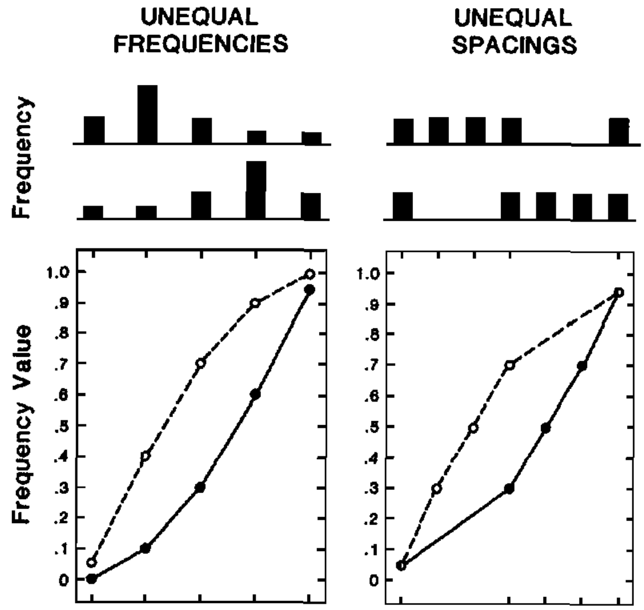

Figure 2. Histograms illustrating the difference between skewed frequencies and skewed spacings, and corresponding frequency functions. (Open circles denote positive skewing; solid circles denote negative skewing.)

ing the two-category scale rated one of the squares *large* 80% of the time in a positively skewed context (where small squares occurred most often) but only 25% of the time in a negatively skewed context with the same endpoints (a highly significant statistical difference); however, that same square size received mean ratings on a 100-point scale that were not significantly different (30 vs. 26) in the respective contexts.

The category effect depends upon the method of producing nonuniformity in the distribution of stimuli. Figure 2 illustrates two different ways to skew the contextual distribution of stimuli. The histograms on the left depict a skewed-frequencies manipulation; the same stimuli occur in both negatively and positively skewed contexts, differing only in relative frequency. The histograms on the right portray a skewed-spacings manipulation; each stimulus occurs equally often, but some of the stimuli appear in only one distribution. The corresponding cumulative frequency functions (i.e., the F_{ic} values of Equation 2) are shown below the histograms. These two manipulations were designed to create approximately the same degree of skewing, and the area between the functions is approximately the same for each method.

Parducci (1982) initially reported the category effect for distributions of squares whose frequencies were skewed. Although increasing the number of different stimuli included in the experimental set (from 5 values to 9 or 24) increased the effects of skewing (referred to as the *stimulus effect*), the category effect remained. Subsequent research (Mellers & Birnbaum, 1982; Parducci & Wedell, 1986) has shown that the category effect and stimulus effect disappear when skewing is achieved by varying the *spacings* of stimuli along the dimension of judgment rather than by varying their relative frequencies. This difference between the effects of skewed spacings and those of skewed frequencies is found whether the stimuli are presented successively or simultaneously (e.g., printed on a single page).

The complete disappearance of the category effect with skewed spacings rules out a number of different interpretations. For example, interpretations based on the idea that with as few as three categories any change in rating must be big but that with nine or more categories very small changes are possible cannot explain why there is no category effect with skewed spacing. Other manipulations that did not modify the category effect also seem to rule out interpreting it as an artifact of scaling. Tabulating nine-category data as though there were just three categories (i.e., tabulating ratings of 1, 2, or 3 as 1, etc.) did not increase the effects of skewing. Thus, the category effect does not seem to be an artifact of how the ratings are transformed to a common scale. Nor does forcing the subjects to use a particular scale seem crucial to the category effect. When subjects were free to choose their own number of categories, those choosing to use fewer than five categories were much more sensitive to differences in frequencies than subjects choosing to use five or more categories (see Parducci & Wedell, 1986, for a fuller account of these open scales and other tests).

A theoretical interpretation of the category effect (Parducci & Wedell, 1986) that is consistent with the pattern of results is that as the number of categories increases, the effective context for judgment becomes less representative of stimulus frequencies actually presented. This change in the effective context is interpreted as reflecting a compromise between the frequency principle and a principle of consistent identification. Strictly following the frequency principle will result in each of the rating categories being used equally often (Parducci, 1965). This means that repetitions of the more frequent stimulus values would be assigned to a number of different categories when the number of categories is large relative to the number of stimuli. Thus the frequency principle would be in conflict with consistent assignment of stimulus repetitions to the same category. One way to reduce the conflict is to progressively discount repetitions of the more frequent stimuli as the number of categories increases. This interpretation suggests that with more categories, the effective (subjective) frequency distribution is more nearly uniform despite variation in the objective frequency distribution. The reason the category effect disappears when it is stimulus spacings rather than frequencies that are skewed is that there are no differences in frequencies to discount.

Pursuing this approach, we developed a processing model in which the number of repetitions counted for any particular stimulus (referred to as depth of search) is limited by the number of categories in the rating scale (with the assumption that the set of stimuli retrieved, referred to as the search set, is limited to the last 12 presentations).[3] The fit of the model to the psychophysical data was encouraging. Not only did it account for the elimination of the category effect with skewed spacings

[3] A full description of the model and its assumptions is presented in Parducci and Wedell (1986). Because the model was developed and tested by using psychophysical judgments, the specific parameter estimates (e.g., search set equal to the last 12 trials) may not be applicable to social judgments. However, we believe the general processing framework has interesting implications for a broad range of judgment tasks (cf. Wedell & Parducci, 1985).

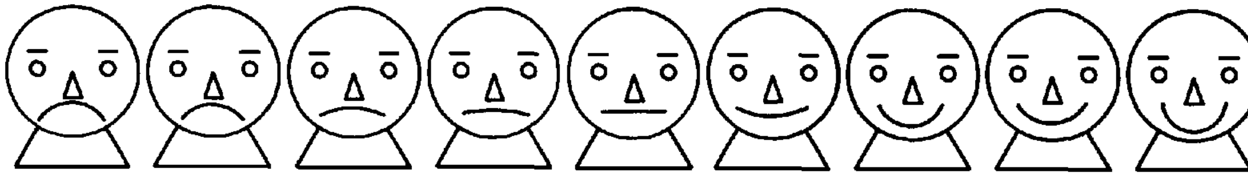

Figure 3. The nine test faces.

but also for the stimulus effect, all without empirically derived constants. It also predicted the results of a crucial test in which stimulus frequencies were manipulated independently of skewing.

The explanatory success of this elaborated range–frequency model for psychophysical data raises the question of whether the category effect also occurs for social judgments. People bring their own standards into experiments using social stimuli so that their contexts are less restricted to the particular values presented in the research session. It also seems possible that because of the relative complexity of social stimuli, the same effective value never occurs twice (cf. Heraclitus, 5th century B.C.). If this were the case, any nonuniformity in the distribution of contextual stimuli would be characterized by unequal spacing (even when the same small set of stimuli were presented with unequal frequencies); hence, there should be no category effect. However, limitations on the transmission of information along a single dimension of judgment (Miller, 1956) might result in highly similar stimuli being grouped together to create, in effect, a distribution with unequal frequencies (and thus the category effect).

The present experiments were designed to test for the category effect using two different tasks: (a) a *person-perception* task in which subjects judged the degree of emotion portrayed in schematic faces and (b) a *self-perception* task in which subjects judged how happy or sad they would have felt if different specified events had happened to them. An effort was made to introduce some of the variation normally encountered in these types of judgment situations, hence, contextual stimuli were generally altered slightly for different presentations. However, if during the judgment process highly similar values are grouped together as, effectively, repetitions of the same value, then the category effect should occur in these situations as well.

Experiment 1: Ratings of Perceived Happiness of Schematic Faces

Contextual contrast has been demonstrated repeatedly in the literature on person perception across a variety of domains, for example, judgments of attractiveness (Kenrick & Gutierres, 1980; Wedell et al., 1987), emotional intensity (Manis, 1967; Thayer, 1980a, 1980b), guilt (Pepitone & DiNubile, 1976), hostility (Herr, 1986), psychopathology (Campbell, Hunt, & Lewis, 1958; Manis & Paskewitz, 1984), merit (Mellers, 1983a, 1986), morality (Marsh & Parducci, 1978; Parducci, 1968), and performance (Mellers & Birnbaum, 1983). Experiment 1 tests whether the category effect (viz., decreased contextual effects with more categories) occurs for ratings of the perceived happiness of faces. Simple schematic faces (resembling those often used in developmental research as response scales) were used in an effort to retain some of the experimental control characteristic of psychophysical research. Because the category effect has been found only when stimuli are presented with unequal frequencies, the primary contextual manipulation consisted of varying the frequencies of faces with different degrees of smiles or frowns to create bell- and U-shaped distributions. These distributions were selected in anticipation of a tendency for subjects to show reduced contextual effects at the neutral point of the scale, in this case the face with a flat line for a mouth (see Marsh & Parducci, 1978, for a discussion of neutral-point anchoring). In order to introduce some of the variation characteristic of social stimuli, the angle of the eyebrows was also manipulated for the contextual faces.

Faces were either presented simultaneously on a single page or projected successively onto a screen. Although this variation in the method of presentation has not affected the magnitude of the category effect in previous research (Parducci & Wedell, 1986), it was thought that the slight variations in the eyebrows might convert the frequency manipulation into what is effectively a spacing manipulation. This seems more likely with simultaneous presentation. Because successive presentation places greater reliance on memory for the contextual faces, the variations in the eyebrows may be largely ignored so that faces with similar frowns or smiles are grouped together in memory. If these assumptions are correct, the category effect should be greater with successive presentation.

Method

Design. Experiment 1 used a $3 \times 2 \times 2 \times 9$ factorial design with three between-subjects factors: number of categories (3, 7, or 100), distribution (bell- or U-shaped), and presentation mode (simultaneous or successive). The one within-subjects factor was stimuli (nine target faces). The dependent variable was the mean rating assigned to each face.

Stimuli. The schematic faces, varying only in the arc of mouth and angle of eyebrows, were either printed on a single sheet of paper or projected (magnified 10×) onto a screen. As shown in Figure 3, each face consisted of a circle representing the head (diameter = 35 mm), an isosceles triangle for a nose (base = 6 mm, height = 8 mm), two small circles for eyes (diameter = 4 mm), a line for each eyebrow (5 mm long), and a straight or curved line for a mouth. The nine different mouth types varied in the degrees of a circle described by each arc: −116, −100, −60, −30, 0, 60, 104, 152, and 180, with the distance between the endpoints of the mouth held constant (sign indicates whether the arc is turned down [−] or up [+]). The two different contexts were created by varying the frequency with which each of the nine mouth types occurred: 1, 1, 3, 5, 5, 5, 3, 1, and 1 for the bell set; and 5, 4, 2, 1, 1, 1, 2, 4, and 5 for the U set. Position on the page and sequence in the slide presentation followed the same random order.

Figure 3 shows the nine target faces, each face representing one of the mouth types. These target faces appeared in the same position on the page or in corresponding positions in the slide sequence for both bell and U sets. All target faces had flat eyebrows.

Each of the additional 16 contextual faces was randomly assigned one of three eyebrow angles: flat (0°), slightly angled (15° upward), or very angled (25° upward). As well as introducing more variability of facial expressions, the variation of eyebrows made the independent variable less obvious.

Instructions. The experimenter told subjects that the study concerned how people judge the happiness of schematic faces. Their task was to record the number corresponding to each judgment, either directly below the face (simultaneous presentation) or on the corresponding line of the response sheet (successive presentation). Fractions were prohibited. For each of the three numerical scales, the lowest number was labeled *very unhappy;* the middle number, *neutral;* and the highest, *very happy.* The 3-point scale ranged from 1 to 3, the 7-point scale from 1 to 7, and the 100-point scale (actually 101 points) from −50 to +50.

Procedure. Each subject received printed instructions that were also read aloud by the experimenter. For simultaneous presentations, each subject received an 8½ × 11-in. page on which the 25 faces were drawn. Subjects rated each face once. For successive presentations, response sheets were handed out and the lights dimmed before presentation of the slides. Subjects viewed the slides from a distance of 1 to 3 m, a new slide appearing every 5 s with a 0.5 s switching interval. Subjects made a total of 50 ratings with each of the 25 slides shown twice, once in forward and once in reverse order.

Subjects. Subjects were 318 undergraduates who participated in partial fulfillment of a psychology course requirement at the University of California, Los Angeles (UCLA). Subjects were tested in groups of 8 to 12, with between 25 and 30 subjects randomly assigned to each condition.

Results

Figure 4 presents the mean ratings and the theoretical fit of the range–frequency model. The crossing-over of the functions for bell and U sets reflects the differences in percentile ranks for stimuli in these sets. As entailed by the theory, the degree of difference between the two rating functions in each panel represents the magnitude of the contextual effect. This difference decreases with increase in the number of categories (the category effect). The category effect appears to be stronger for successive than for simultaneous presentation, although the overall effects of context are much smaller for the former.

An analysis of variance (ANOVA) was conducted on the rat-

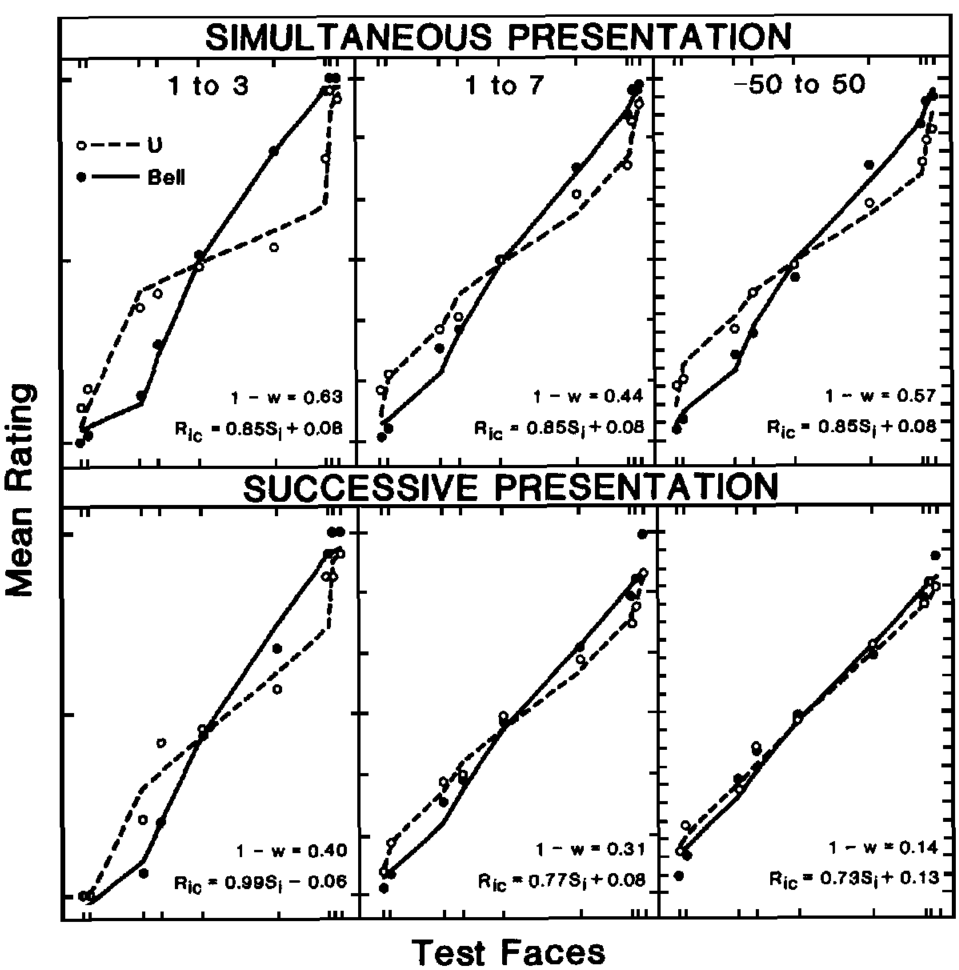

Figure 4. Category effect for ratings of happiness of schematic faces: differences between the ratings for the bell and U sets decrease with more categories. (Lines represent fit of range–frequency model, with stimulus spacing derived from the model.)

ings after linear transformation to a common scale using Equation 4 (i.e., the lowest possible rating, 1 or −50, was transformed to 0; and the highest, 3, 7, or 50, was transformed to 1.0). The effects of context are represented by the interaction between distribution and stimuli, which was statistically significant, $F(8, 2432) = 41.69$, $p < .001$. Of direct relevance to the contextual effects entailed by range–frequency theory (i.e., the crossing of the functions and their different curvatures) are the interactions of distribution with the linear and cubic components of the stimulus variable, both of which were highly significant ($p < .001$). For all tests reported, the effects of context are represented by the interaction of distribution with the combined linear and cubic trends of the stimulus variable.

The presence of the category effect is clearly indicated by the significant interaction of number of categories and effects of context, $F(4, 608) = 7.55$, $p < .001$. Planned comparisons (at the $p < .05$ level) showed that contextual effects for the three-category scales were significantly greater than for either 7- or 100-point scales, although the latter did not differ significantly. The category effect was significant for both simultaneous and successive presentation modes, taken separately, $F(4, 608) = 3.61$ and $F(4, 608) = 4.69$, $p < .01$, respectively. Despite the appearance of a stronger category effect for successive presentations, this interaction was not significant ($F < 1$).[4] The effects of context were greater for simultaneous than for successive presentations, $F(2, 608) = 9.72$, $p < .001$.

Beyond the expected contextual effects and hypothesized category effect, the data follow a rather complicated pattern of interactions. These may best be understood in terms of the underlying parameters of the range–frequency model. The focus here is not so much on the fitting procedure itself (for more details, cf. Parducci & Wedell, 1986; Wedell et al., 1987) but on the use of the model as a basis for interpreting the results.

The theoretical functions shown in Figure 4 represent the fit of the model to the data. The shapes of these functions reflect a compromise between range and frequency values. Because the same extreme facial expressions defined the experimental range for both distributions, range values for corresponding target faces were equated for bell and U sets so that all differences in ratings between the sets were attributed to differences in frequency values. Following this logic, an estimate of $1 - w$ was calculated for each target face by dividing the difference in obtained ratings (transformed to a 0 to 1 scale) by the difference in frequency values from Equation 2.[5] A weighted average of these estimates was then used to determine a single value of $1 - w$ for each panel of Figure 4 (with each individual estimate weighted by the difference in frequency values for that face).

The ratings, frequency values, and estimates of w were then substituted into Equation 3 to solve for range values. In the most constrained case, a single range function would be used to construct each of the 12 theoretical functions of Figure 4. However, results of the ANOVA indicated that a single range function would not fit the data. A significant main effect of presentation, $F(1, 304) = 15.90$, $p < .001$, reflected lower overall ratings of faces when presented in succession. The only way to model this effect within range–frequency theory was to estimate separate range functions for the different presentation conditions; the range function for successive presentations reflected greater extension of the range beyond the happiest of the target faces. A significant three-way interaction among number of categories, presentation mode, and stimuli, $F(16, 2432) = 11.93$, $p < .001$, indicated the necessity for additional variation in the range functions. Specifically, the Linear × Linear component of the Category × Stimulus interaction (for the middle five faces) was significant ($p < .05$) for successive but not for simultaneous presentations, implying that the range function varies with number of categories only for successive presentation. Thus, a single range function was estimated for the simultaneous conditions, and three range functions, one for each number of categories, were estimated for successive conditions. These functions are described by the linear equations in the panels of Figure 4, in which the stimulus scale values, S_i's, are the range values that would be expected if the subjective range was defined by the stimulus set. (Spacing of these values was determined by averaging range values from all conditions.)

The generally close adherence of the theoretical functions to the empirical points provides support for a range–frequency interpretation of the results. Altogether, 22 parameters were estimated to fit 108 data points (8 parameters reflecting the spacing of stimuli, a slope and an intercept parameter for each of the four range functions, and a weighting parameter for each panel of Figure 4); however, the difference in curvature between the two functions in each panel was captured by variation in a single parameter, $1 - w$. In assessing the model's fit, it is useful to consider how well the data would be fit by adaptation-level theory, range–frequency theory's closest competitor. Because each empirical adaption level (stimulus rated neutral) is nearly the same for all 12 conditions, adaptation-level theory entails a single function. The statistically significant interaction effects clearly disprove adaptation-level theory, as it has been disproven by nearly 30 years of research on the effects of context on psychophysical judgments.

Discussion

Experiment 1 confirms the category effect previously demonstrated using simpler, psychophysical stimuli (Parducci & Wedell, 1986). Ratings of the happiness portrayed by schematic faces became less sensitive to difference in stimulus frequencies

[4] In the overall analysis, the four-way interaction was significant, $F(16, 2432) = 1.85$, $p < .05$; however, no linear, quadratic, or cubic trends approached significance ($F < 1$). Thus, this interaction reflects some higher order differences in the rating functions that are not of particular relevance to the present discussion.

[5] Because the assumption of a continuous scale is clearly inappropriate when subjects are limited to just three categories, the counting algorithm, used for psychophysical experiments, was used instead of Equation 2 to calculate frequency values (cf. Parducci & Wedell, 1986; Wedell, Parducci, & Geiselman, 1987). The counting algorithm calculates what the stimulus would have been rated if subjects literally used each category equally often while maintaining a perfectly ordinal scale. Because the counting algorithm includes the number of categories in the calculation of frequency values, frequency functions become more steplike when there are just a few categories. This tends to increase the differences between frequency values for the same stimulus. Consequently, the value of $1 - w$ is lower when fewer categories are used. For example, the value of $1 - w$ for the three-category, simultaneous-presentation condition is .63 using the counting algorithm but .77 using Equation 2 (the values of $1 - w$ for 100-point scales are identical for the two methods).

with an increase in the number of categories. These results suggest that judgments in person perception tasks may depend critically on the number of categories brought to mind at the time of judgment.

The greater effects of context with simultaneous presentation are perhaps not surprising because the stimulus frequencies are right there for comparison. The variations in eyebrow may also have been made more apparent so that faces with the same mouth may have been processed as different. This would imply a greater category effect for successive presentations, where these slight variations might more readily be grouped together in memory. The fit of the range–frequency model tends to support this implication, although the interaction between categories, context, and presentation was not statistically significant.

A second type of category effect is represented by the systematic decline in the slope of the rating functions for successive presentation with an increase in the number of categories (the three-way interaction among categories, presentation, and stimuli). Within range–frequency theory, the slope of the range function indicates the degree to which the range of the effective context is determined by the endpoints of the set actually presented. A slope of unity would represent complete determination by these endpoints. The decline in slope with more categories can be interpreted as a tendency to extend the effective contextual range beyond the experimental set. A large number of categories may encourage the subject to retrieve (for the effective context) values more extreme than any actually presented.

We have observed this same effect in a recent experiment in which subjects judged the physical attractiveness of faces in photographs (Wedell et al., 1987). The present results, however, suggest that the mode of presentation may be of crucial importance. The slope did not vary with the number of categories when stimuli were simultaneously present. Because the range of variation is clearly defined when all relevant stimuli are present, it may not seem necessary (from a communication point of view) to reserve extreme categories for stimuli of more extreme value than those presented. Only when stimuli are presented successively is there a sense in which the complete set of stimuli is not yet defined so that categories must be saved for values more extreme than any thus far presented. However, this saving of extreme categories would prevent discrimination among the stimuli when there are only a few categories.

With social judgments, the set of contextual stimuli is typically not simultaneously present and therefore extension of the effective range with more categories should be expected. In the present experiment, the lower ratings for faces in the successive presentation condition indicated that there was greater extension of the subjective range above the experimental set than below it. The extension was in the opposite direction for ratings of physical attractiveness (Wedell et al., 1987). The degree of extension of the subjective range in either direction may be interpreted as indicative of where the presented stimuli lie within the broader range of possible values. For example, if the set is near the top of the possible range, extension should be predominantly downward. Because this extension is greater when there are more categories, many categories should be used when the objective is to determine the judgment of the stimuli within the broadest possible range of experiences.

Experiment 2: Ratings of Happiness of Life Events

Although Experiment 1 suggests that the category effect also occurs with social judgments, one may question whether subjects were really judging happiness or only the curvature of the schematic mouth. The relatively simple nature of these faces leaves open the question of whether the category effect occurs with more complex social stimuli. If the contextual distribution for complex social stimuli is more properly characterized by unequal stimulus spacings than by unequal frequencies, then one should not expect to find the category effect with such stimuli. Experiment 2 addresses this issue by testing for the category effect with happiness ratings of hypothetical life events.

A number of psychological treatments of everyday value experiences are based on the relativistic assumption that the degree of happiness or satisfaction associated with a particular event depends upon the experiences with which it is compared (e.g., Brickman, 1975; Brickman & Campbell, 1971; Helson, 1964; Parducci, 1968, 1984; Thibaut & Kelley, 1959). Experimental research has supported such relativity. For example, ratings of the satisfaction associated with monetary rewards follow a range–frequency compromise (Marsh & Parducci, 1978; Parducci, 1968). Similar contrast effects have been demonstrated for judgments of life satisfaction (e.g., Gutek, Allen, Tyler, Lau, & Majchrzak, 1983; Strack, Schwarz, & Gschneidinger, 1985) and even for the incentive value of food given to rats (e.g., Crespi, 1942).

In Experiment 2 we explored whether the relativism typically found in studies of self-perception would diminish with an increase in the number of judgmental categories. We asked subjects to read a series of descriptions representing events that might occur during a 5-day school week and to imagine how they would feel if these events happened to them. In order to skew the contextual distribution, the descriptions of the first 4 days consisted of events that were all either highly positive or highly negative. The relativistic approach to happiness assumes that events described for the last day, which were the same for both contextual distributions, would be rated more positively when preceded by negative events than when preceded by positive ones.

To produce large contextual effects, both the range and relative frequencies of events were manipulated. The extreme events described for the first "4 days" anchored one end of the range, with the other less extreme end of the range not experienced until the "last day." Although no description was repeated exactly, many were similar, changing only a word or two. It was assumed that these similar statements would be grouped together on the dimension of judgment so that the contextual distributions, in effect, would consist of a relatively small set of events occurring with unequal frequencies.

Although Parducci and Wedell (1986) found no evidence that the extremity of labels assigned to the different categories had any significant effect on the magnitude of contextual effects, it seems possible that social judgments might be more sensitive to the particular labels prescribed. This possibility is encouraged by the finding in Experiment 1 that more categories evoked a more extended subjective range; extreme category labels could have a similar effect. Experiment 2 tested this hypothesis, using one set of moderate labels and one set of extreme labels.

Method

Design. Experiment 2 used a five-way design, with number of categories (three, five, or nine), skewing (positive or negative), test events (posi-

Table 1
Representative Event Descriptions Used in Experiment 2

Domain	Level	Event
Misc.	HN	You wake up late, feeling sick, and don't have time for a shower.
Misc.	N	You wake up and find the shower won't rise above lukewarm.
Misc.	M	Today you wake up but only have time for a quick shower.
Misc.	P	You wake up early and take a long, hot shower.
Misc.	HP	Today you sleep in late, wake up refreshed, and take a luxurious shower.
Meals	HN	You have no money for lunch today.
Meals	N	You get one of your favorite lunches on campus, but the food isn't up to par.
Meals	M	You eat the sack lunch you prepared last night.
Meals	P	You eat lunch with a friend.
Meals	HP	At lunch time you have a picnic with friends in the sculpture gardens.
Drive	HN	Your car dies several times on the way home.
Drive	N	On your way home, you dislike all the music that's played on the radio.
Drive	M	The drive home is uneventful.
Drive	P	Traffic isn't too bad on the way home.
Drive	HP	On your way back, traffic is fairly smooth and you arrive at home quickly.
Acad.	HN	You are late to your discussion section and get a zero for the quiz that was given.
Acad.	N	During your second class, your pen breaks, spilling ink all over your notes.
Acad.	M	You have a guest lecturer for your afternoon class.
Acad.	P	In your second class, your T.A. extends the due date for the next assignment.
Acad.	HP	In your first class, you get your paper back with several good comments on it.
Social	HN	Your friends go to a movie you want to see, but you have to stay home and study.
Social	N	In the evening you watch a bunch of reruns on TV by yourself.
Social	M	Later you go window shopping in Westwood.
Social	P	You get a wedding invitation in the mail.
Social	HP	You and some of your friends then go to the beach and relax in the sun.

Note. HN = highly negative, N = negative, M = moderate, P = positive, and HP = highly positive. Misc. = miscellaneous, Acad. = academic, and T.A. = teaching assistant.

tive, moderate, or negative), and verbal labels (extreme or moderate) as between-subjects variables. The only variable manipulated within subjects was domain (academic, driving, meals, social, miscellaneous, overall day, and overall week). Dependent measures were the ratings of the individual descriptive statements, overall rating of the last day, and overall rating of the week.

Event descriptions. Statements describing different kinds of events were written in the second person to encourage subjects to imagine experiencing the events themselves. There were 13 numbered descriptions to a page, each page representing a different day. In Table 1 are examples of different types of events. For each domain, events were selected to elicit ratings varying from highly negative to highly positive. In a pilot study, 30 subjects rated the statements in random order using a 9-point scale. On the basis of these ratings, statements were assigned to one of five groups: highly negative (HN), negative (N), moderate (M), positive (P), and highly positive (HP). Because items within different domains spanned different evaluative ranges (e.g., subjects were less likely to be very happy or unhappy about the weather than about grades received in a class), the criteria for assigning statements to the different groups had to vary for each domain. This procedure introduced some variation into the evaluative levels of the events nominally at the same level (e.g., HN); however, because statements at the same level within a domain were matched in value, the manipulation was predominantly one of skewed frequencies rather than skewed spacings.

Eleven different pages of descriptions were then constructed: four with all HN descriptions, four with all HP descriptions, and one each with all N, M, or P descriptions. The different pages were used to construct six different booklets. Three booklets were positively skewed, consisting of four HN days followed by either an N, M, or P day; the other three booklets were negatively skewed, consisting of four HP days followed by either an N, M, or P day. Thirteen numbered statements were double-spaced on each page, with the number of the day (1–5) printed in the top right corner. For each booklet, the four contextual pages appeared in one of four sequences. The test page always appeared last.

Instructions. The experimenter read the instructions aloud, stating that the study was concerned with how people evaluate daily events. The 13 sentences on each page of the booklet were said to describe a day in the life of a UCLA student. Subjects were told to imagine that they were that student—to imagine how they would feel if they actually experienced the events described. They were instructed to use a 3-, 5-, or 9-category scale to rate their immediate reactions to each event, that is, how happy or unhappy they would have felt if the events described had happened to them. Each point on the scale was labeled with a verbal category and a number (except for the 9-category, moderate-label condition in which only five category labels were used, one for every other number). The middle category was always labeled *neutral.* Table 2 shows the range of category labels used for moderate- and extreme-label conditions.[6]

Subjects recorded the numeral representing their judgment on the appropriate line of their response sheet. The response sheet consisted of five columns, each headed by a label indicating the day (e.g., Day 1, Day 2, etc.). Under each column were 13 numbered lines and a last line labeled *overall.* After rating all the events on a particular page, subjects rated their overall impression of the day before moving on to the next day. Finally, when subjects had worked through all 5 days, they rated their overall impression of the week.

[6] Because labels are not the same for different categories within the same labeling condition, there is a potential difficulty in interpreting these effects. However, the general lack of effects of the manipulation obviates this problem.

Table 2
Range of Category Labels for Different Labeling Conditions

Scale	Moderate	Extreme
3-pt.	*Unhappy* to *happy*	*Very unhappy* to *very happy*
5-pt.	*Unhappy* to *happy*	*Very very unhappy* to *very very happy*
9-pt.	*Very unhappy* to *very happy*	*Very very unhappy* to *very very happy*

Note. pt. = point.

Subjects. The experiment used 1,004 undergraduates randomly selected from the same pool sampled for Experiment 1. Subjects first participated in one of several short psychophysical judgment experiments. Four subjects neglected to rate all 5 of the overall days and 24 subjects neglected to rate the overall week.

Results and Discussion

Figure 5 plots the mean ratings of the events described for Day 5, overall ratings of that day, and of the week. Supportive of the relational approach to social judgment, ratings of events and of overall day show the usual contrast. However, the magnitude of this contextual effect is again inversely related to the number of categories prescribed by the experimenter. This is the category effect. The difference in ratings of the overall week reflects the preponderance of highly positive events for the negatively skewed distributions. Ratings of overall week became less extreme as the number of categories increased, consistent with the idea that subjects evoke a more extreme range of comparison events when required to use more rating categories.

Separate four-way ANOVAs were run, using subjects' ratings of day, week, and the means of their event ratings as dependent variables in the respective analyses (ratings were first transformed to a 0 to 1 scale as in Experiment 1). Results were completely parallel for events and overall day, and both showed highly significant effects of test events and skewing ($F > 100$). The interaction of number of categories and skewing was also significant, $F(2, 968) = 6.12$, $p < .01$ (for event ratings), and $F(2, 957) = 10.72$, $p < .001$ (for ratings of overall day).[7] Planned comparisons (at $p < .05$) revealed that in both analyses the effects of skewing were greater for three and five categories than for nine categories. In addition, with ratings of overall day, the effects of skewing were greater for three than for five categories. Thus, the same category effect found for ratings of simple, perceptual stimuli was also found with happiness ratings of complex verbal statements.

The only other statistically significant effect in these two analyses was the interaction between number of categories and events, $F(4, 968) = 6.07$, $p < .001$ (for event ratings), and $F(4, 957) = 3.02$, $p < .05$ (for ratings of overall day). As with the ratings of successive faces in Experiment 1, this interaction reflects a decrease in the slopes of the rating functions with an increase in the number of categories. Once again, this effect can be interpreted as a tendency for subjects to include in their subjective contexts events that were more extreme than any actually described when they used more categories.

Overall ratings of week. Ratings of the week differed conceptually from ratings of events and day in two important ways. First, the same stimulus was not rated in the two skewing conditions: Either positive or negative events made up 4/5 of the week's events; thus the highly significant main effect of skewing ($p < .0001$) was hardly surprising. Second, there was no manipulation of context, because only 1 hypothetical week was presented. Consequently, the significant Categories × Skewing interaction, $F(2, 937) = 81.54$, $p < .001$, is not readily interpreted as the category effect, but may be explained in terms of greater extension of the subjective range with more categories. When subjects were given just three categories to rate how happy they were with the week just experienced, the contextual stimuli retrieved for comparison may have been limited to events that were descriptively similar but representative of different evaluative levels; for example, subjects might have retrieved happy and unhappy events typical of other "school weeks." However, when nine categories were available, the subjective range may have been extended to include events that were descriptively quite dissimilar and evaluatively more extreme, for example, winning the jackpot in the lottery or thoughts of a loved one dying.

The present study also sheds some light on how stimulus information is combined to arrive at an overall rating of the week. The significant interaction between skewing and test events, $F(2, 937) = 7.74$, $p < .001$, contradicts a constant-weight averaging rule that would predict parallel effects of test event across skewing conditions. Instead, the greater slopes of the rating functions for the positive-skewing conditions indicated that negative events received less weight than positive events: *positivity weighting.* This type of asymmetrical weighting is just the opposite of that normally found for ratings of likability (e.g., Birnbaum, 1974a), in which negative traits receive greater weight. One possible explanation of positivity weighting for ratings of happiness is in terms of an ego-protecting mechanism: It pays (i.e., makes one happier) to discount the importance of negative events in one's life.

When happiness is defined as the sum of all individual value experiences (as determined by a range–frequency judgment process), then overall happiness depends on events being negatively skewed (Parducci, 1968, 1984). However, the present finding of positivity weighting suggests that when people judge their overall happiness, they may rate themselves as happy even when events are positively skewed, as long as the weight assigned to happy experiences is large enough to overcome the unhappy experiences. However, some caution should be exercised in generalizing this result, because the format for recording judgments allowed the subject to view all previous ratings of the individual events. In most realistic situations, values for prior experiences must be retrieved from memory; hence, the experiences retrieved for integration will depend on a number of memory factors among which mood should be particularly important (cf. Bower, 1981; Schwarz & Clore, 1983).

More generally, the differences between ratings of overall week and the mean of ratings of individual events across the week illustrate a distinction that is easily blurred in the literature on self-avowals of happiness. The surveys typically define happiness with respect to the overall rating for life as a whole or for a recent period of one's life (Diener, 1984). However, the

[7] In separate analyses of variance run for each event domain, the category effect was significant ($p < .05$) for three of the five domains and marginally significant ($p < .10$) in a fourth.

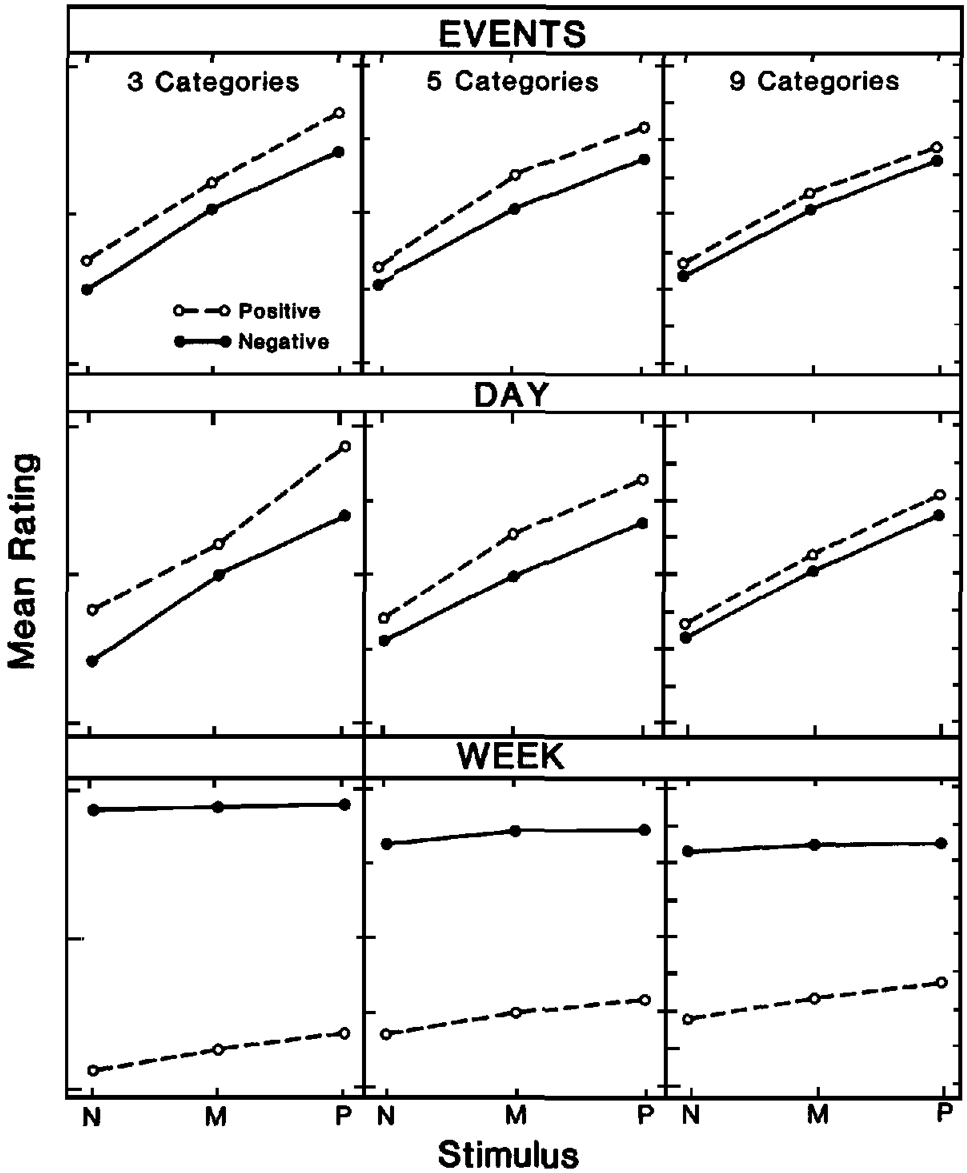

Figure 5. Category effect for ratings of happiness of events, overall day, and overall week. (Top two rows: differences between the rating functions for the positively and negatively skewed sets decreased with more categories. Bottom row: ratings became less extreme with more categories.)

utilitarian conception of happiness defines it in terms of the average of separate experiences. Parducci (1984) has argued for the utilitarian definition on the grounds that the overall rating can have little meaning unless people could accurately average countless separate ratings and unless context of different lives is specified.

Category labels. The effect of the label manipulation was minimal. Of the 24 possible effects involving label, the only one to reach statistical significance was the three-way interaction among categories, skewing, and label in the analysis of overall week, $F(2, 937) = 5.05, p < .01$. This interaction, however, was not readily interpretable (the effects of skewing were slightly larger for extreme labeling with three and nine categories, but were slightly smaller for five categories). Thus, it appears that the context retrieved for judgment is not so much dependent upon the extremity of verbal labels but upon the number of categories. However, when labels represent specific values as in the Schwarz et al. (1985) study, they may well affect the standards retrieved for comparison as described earlier.

Experiment 3: Happiness Ratings for Full-Range Sets

One difficulty in interpreting the category effect found in Experiment 2 is that both range and frequencies of contextual stimuli were varied. Although Parducci and Wedell (1986) attributed the category effect to differential sensitivity to stimulus frequencies, an alternative interpretation of Experiment 2 can be formulated strictly in terms of changes in subjective range.

The interactions between number of categories and test stimuli found in both Experiments 1 and 2 are consistent with the notion that there is a greater tendency to extend the subjective range beyond the experimental set as the number of categories increases. Thus, when nine categories are used to judge satisfaction with a particular grade (a typical judgment in the academic domain), subjects may extend the range to include the best and worst possible grades (A and F) even when only one of these extremes is presented in the experiment. However, when presented with just three categories, subjects may be more likely to match the subjective range to the range of hypothetical events actually presented (e.g., grades varying from A to C in a negatively skewed distribution). This type of differential adjustment would itself produce a category effect, because differences in range values for the moderate test stimuli would be greater when fewer categories are used. Experiment 3 was designed to test this range interpretation of the category effect by using contextual distributions that are asymmetrical only with respect to stimulus frequencies and not to stimulus range. If a category effect is obtained, then one can reject an interpretation of Experiment 2 solely on the basis of greater extension of the range with more categories.

Method

The major difference between Experiments 2 and 3 was that the events described for the 2nd day were the opposite value of those described for the 1st, 3rd, and 4th days. Thus, the positively skewed set consisted of HN, HP, HN, and HN days, whereas the negatively skewed set consisted of HP, HN, HP, and HP days. Only the moderate labels of Experiment 2 were used. Subjects were 489 students randomly drawn from the same source as in Experiment 2, with between 25 and 30 subjects randomly assigned to each of the 18 experimental conditions generated by the Categories (three, five, or nine) × Skewing (positive or negative) × Test Event (N, M, or P) between-subjects factorial design. Three subjects failed to rate overall day and 22 subjects failed to rate overall week.

Results and Discussion

The pattern of results shown in Figure 6 are generally consistent with the category effect: Contextual contrast (although naturally weaker) was still evident for ratings of events and overall day for both three- and five-category scales but was virtually absent for nine-category scales. Furthermore, the slopes of the rating functions for these domains did not appear to differ much across categories, consistent with the notion that presenting the full range of stimuli lessened differences in the effective contextual ranges across category conditions. This means that the category effect demonstrated in Experiment 2 did not depend solely on greater extension of the effective range with more categories. However, ratings of the overall week, in which the context was not manipulated, did suggest greater extension of the effective range of possible weeks with an increase in the number of categories.

We used separate three-way ANOVAS to analyze ratings of events, day, and week. Once again, analyses for events and day were essentially parallel. Both analyses revealed significant effects of context, $F(1, 471) = 5.57$, $p < .01$ (for event ratings), and $F(1, 468) = 4.84$, $p < .05$ (for ratings of day). However, neither analysis revealed a significant interaction between number of categories and skewing, $p > .10$. The failure of this statistical test of the category effect may reflect, in part, reduced power to detect the effect: Both the effects of skewing (as measured by differences in frequency values) and the number of subjects were cut in half for Experiment 3 (as compared with Experiment 2). Planned comparisons showed significant effects of skewing for three- and five-category scales ($p < .05$) but not for nine-category scales ($F < 1$), which seems consistent with the category effect. The lack of significant interactions between number of categories and test events ($p > .50$ in both analyses) indicates that the greater sensitivity of the three- and five-category scales to differences in contextual skewing cannot be explained in terms of greater restriction of the effective range.

The analysis of ratings of the entire week followed the same pattern as in Experiment 2. Positivity weighting was again indicated by the significant interaction between skewing and test events, $F(2, 449) = 3.88$, $p < .05$. As in Experiment 2, the significant Categories × Skewing interaction, $F(2, 449) = 29.51$, $p < .001$, can be interpreted in terms of consideration of a greater range of possible comparison values as the number of categories increases. If this interpretation is correct, then presumably the effect could be eliminated by presenting comparison weeks that span the full range retrieved by subjects using the nine-category scales (parallel to the manipulation for events and day).

General Discussion

The present experiments demonstrate that the effects of the distribution of recent contextual events depend upon the number of rating categories prescribed for judgment. Increasing the number of categories produced two types of effects: (a) reduction of the contextual effects associated with unequal stimulus frequencies and (b) extension of the effective context to include more extreme stimuli.

Sensitivity to Frequencies

Consistent with the category effect found in psychophysical experiments, increasing the number of rating categories resulted in reduced sensitivity to stimulus frequencies for ratings of happiness in both person perception and self-perception tasks. For example, in the successive-presentation condition of Experiment 1, the difference in the judged happiness of Face 4 across contextual conditions was highly significant when subjects used just three categories, but was not significant when subjects used the 100-point scale. Because psychophysical research demonstrates that the category effect does not occur when contextual differences are based on skewed spacings, the occurrence of the category effect suggests that the present manipulations produced effective variations in the frequencies of the stimuli rather than in their spacings.

Extension of Range

A second effect of the number of categories revealed in Experiments 1 and 2 is the tendency for the effective range of contextual stimuli to extend further beyond the set actually presented when ratings are made by using more categories. This tendency is reflected by the mean ratings of end-stimuli becoming less

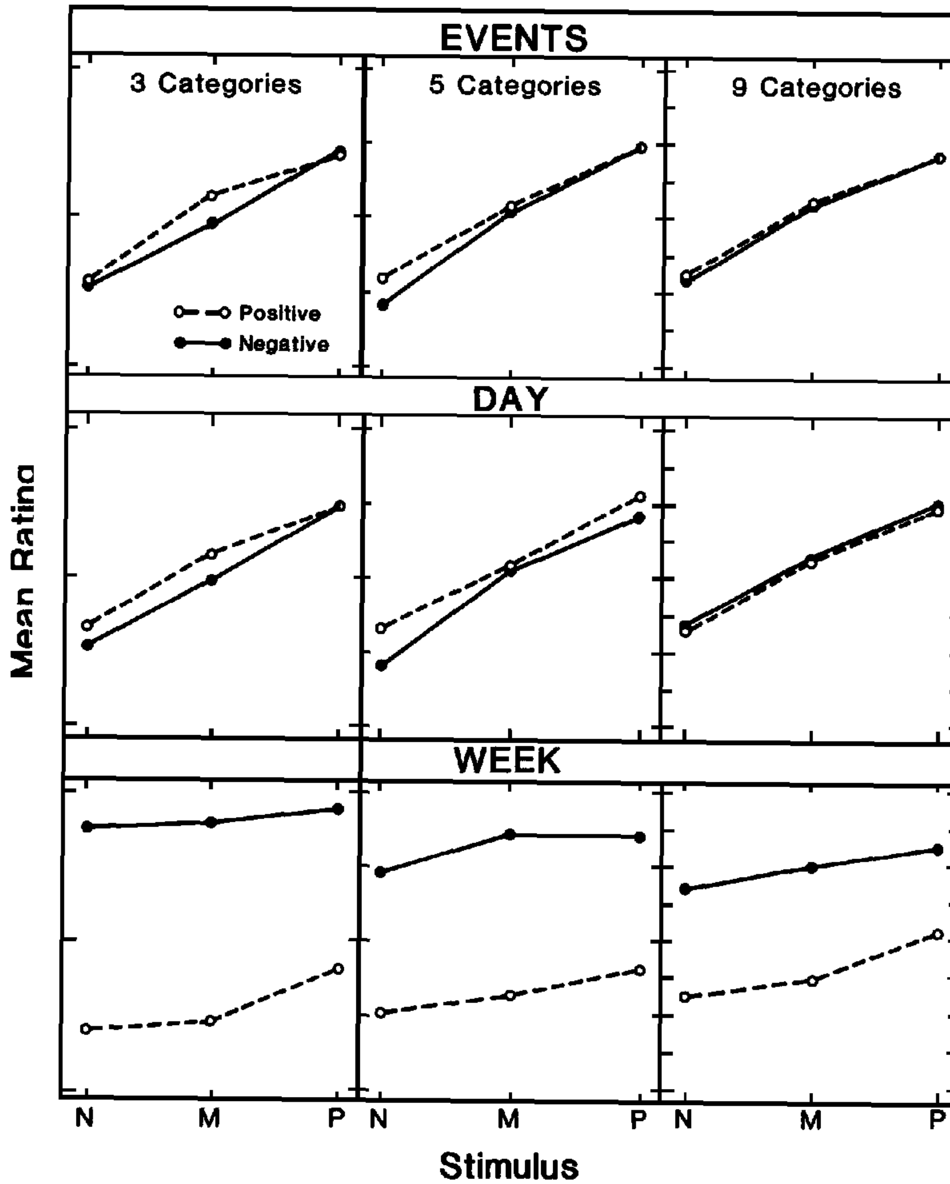

Figure 6. Category effect for ratings of happiness of events and of overall day obtained for full-range sets.

extreme as the number of categories increased. Experiment 1 demonstrated that this effect was strongest when stimuli were presented successively and that it was eliminated when all stimuli were simultaneously present. This effect was also minimized when the set of recent contextual events already included a wide range of variation, as in Experiment 3.

Features of the Effective Context for Judgment

In recent years there has been expanded interest in the cognitive processing underlying social judgments and decisions. Research in social cognition has focused on the effects of priming a conceptual category such as reckless or adventurous on subsequent judgments of ambiguous stimuli (e.g., Higgins et al., 1977). This priming generally results in an assimilation effect; in other words, judgments of the stimulus are displaced toward the value of the primed category. However, when the stimulus is not ambiguous or its value is distant from the value of the primed category, contrast effects may occur (Herr, 1986; Herr et al., 1983). These contrast effects are consistent with the notion that primed values are more readily activated at the time of judgment and may serve as part of the context with which the target stimulus is compared.

Range-frequency theory provides a detailed account of the nature of the comparison processes underlying contrast effects. The present article places the range-frequency judgment process within an information-processing framework and considers how the number of rating categories may mediate which values enter into the effective context influencing judgment. One can explain the category effect with this theory by assuming that the frequencies of recent events are less likely to be included in the context when there are more categories, thereby reducing the skewing of the effective distribution (Parducci & Wedell, 1986). Furthermore, increasing the number of categories evokes extreme contextual values, perhaps more extreme than any recently presented. This account of the effects of varying the number of categories suggests that the rating scale itself may be conceived as an effective prime, evoking values outside the set actually presented. Kahneman and Miller (1986) have described other possibilities for evoking contextual values, some of which may never have been experienced themselves.

Dynamics of Social Judgment

The processing framework summarized by Figure 1 assumes that contextual effects have a strong phenomenological impact. This implies that the number of categories underlying everyday evaluations may play an important role in social and self-perception. Consider an individual who uses only three implicit categories to evaluate events: good, bad, and indifferent. Such a judgmental style should lead to greater dependency of judgments on recent contextual events. On the other hand, an individual who uses a more fine-grained evaluative scale (e.g., consisting of nine possible categories) will evoke a context that is less tied to the range and relative frequencies of recent contextual events. Use of the more discriminating scale should then lead to greater consistency of judgment of the same event across varying contextual circumstances.

The differential impact of using coarse versus fine scales of judgment may bear some relation to the recent evidence that self-complexity provides a buffer against depression (Linville, 1987). If we assume that the stress introduced by recent negative events may lead to depression, the broader contextual range evoked by using more categories may make these events seem less extreme and hence easier to handle.

A central assertion of the present framework is that the judgment process is an ongoing *constructive* act: A distribution of contextual stimuli is rapidly retrieved or evoked, and the stimulus is located within this distribution via range and frequency principles. Our work with the category effect for psychophysical stimuli (Parducci & Wedell, 1986) has demonstrated that (a) it does not depend on whether judgment time is limited to 10, 5, or just 2 s and (b) it does not depend upon how the stimulus distribution is encoded (incidentally or under instructions for judgment); rather, the effect occurs at the time of judgment.

In contrast to this constructive view is the *associative* view that social judgment consists of retrieving a pre-encoded categorical representation of the stimulus (e.g., Higgins et al., 1977; Wyer & Srull, 1986). According to this view, judgment follows a process similar to cued recall. Because stimuli are complex, a number of different judgmental categories may be associated with a stimulus so that the particular representation retrieved may depend on which conceptual categories have recently been activated or primed. There seems to be strong evidence supporting both of these types of judgmental processes. Indeed, the final judgment may represent some interplay between evaluations determined by these two processes. An experiment by Higgins and Lurie (1983) nicely illustrates this. Subjects in that experiment apparently retrieved a previously associated judgmental category and then used the recently presented stimulus distribution to construct a stimulus value corresponding to that judgment.

Interpretation of Rating Scale Data

Because judgments are less determined by the range and relative frequencies of recent events when more categories are used, interpretations of rating scale data should take into account the number of rating categories. Suppose that in a well-conducted poll, a particular public figure received an approval rating of 60%, eliciting choices of 60% approve and 40% disapprove, but that another, equally reliable poll elicited a mean rating of only 4 (*slightly disapprove*) on a scale from 1 (*strongly disapprove*) to 9 (*strongly approve*). Application of the theoretical framework presented here suggests that the discrepancy in these hypothetical results could reflect two types of differences in the effective contexts for judgment. For example, we could conceive of the recent context as a set of other public figures, some of whom are in the news more often and are therefore represented with greater frequency. If these individuals are embroiled in a public scandal, the approval rating of this particular public figure may benefit by comparison. However, this upward displacement should be greater when ratings are made on the two-category scale, which is more sensitive to differences in frequencies. The 9-point scale may also tend to evoke a range of comparison stimuli that includes individuals of much stronger character than are represented in current affairs and next to whom this public figure may elicit a rating of *slightly disapprove*. Insofar as judgments are based on an implicit distribution of contextual stimuli, some retrieved from memory and others evoked by what is being judged, it does not seem farfetched to expect these category-mediated effects in surveys of public opinion, or in any other situation in which value judgments are expressed, whether overtly to others or privately to ourselves.

Choosing the Appropriate Number of Categories

How many rating categories are optimal for social research? Previous discussions of the optimal number of rating categories have generally focused on the reliability of such scales. Although some researchers have argued for the use of seven or more categories to maximize the transmission of information (e.g., Garner, 1960), others have suggested that little relevant information is gained by extending the number of categories beyond two (e.g., Peabody, 1962). Results of the present experiments suggest that apart from the effects on transmission of information, varying the number of categories may systematically affect which contextual standards are used at the time of judgment.

Our opinion is that there is no optimal number of categories, but rather that the choice depends on the researcher's objectives. One objective may be to determine algebraic processes of integration. Within this framework, there has been an attempt to minimize contextual effects in order to reduce nonlinearities in the rating scales that might in turn reduce the power of the test. To this end, Anderson (1982, pp. 6–9) has advocated the use of large numbers of categories as well as procedures that define the end-categories using the extreme stimuli. Both of these techniques make sense from the viewpoint of the present research, but we emphasize that they do not guarantee the elimination of contextual effects. Mellers (1983a, 1986) found very large contextual effects when judgments of merit were made in the form of salary allocations, a virtually continuous scale. Her findings do not contradict the present results, however, because that research used a manipulation of stimulus spacings, not frequencies, and therefore the category effect was not to be expected.

Much of the research involving rating scales is conducted through surveys using widely varying numbers of categories. For example, one popular measure of life satisfaction uses three categories (Gurin, Veroff, & Feld, 1960) and another, nine categories (Cantril, 1965). The results of Experiments 2 and 3 suggest that these two instruments may evoke different effective contexts, the three-category scale being more sensitive to recent past experiences. Which is the better instrument? The question begs the answer by assuming that there is a single true judgment for any event, a judgment that is somehow distorted by the quirks of the rating instrument. We believe that these differences reflect differences in effective contexts; the judgments obtained with either instrument may be phenomenologically valid. It has long been known within survey research that slight changes in wording can invoke radically different ways of thinking about a question and, hence, can result in very different responses (cf. Sudman & Bradburn, 1974, for a review). The current line of research suggests that the number of rating categories prescribed by the researcher may operate in a similar fashion.

Finally, these systematic effects suggest indirect avenues for investigating individual processing styles and their consequences for behavior. Returning to our hypothetical political opinion poll, it would be of interest to determine which scale is a better predictor of subjects' behavior (e.g., voting, campaign contributions, etc.). If cognitions mediate such decisions, then the important question becomes: What is the implicit number of categories people use at the time of their decisions? One way to tap this would be to use open scales that permit subjects to select their own sets of categories. In our psychophysical research (Parducci & Wedell, 1986), we have found that some subjects generate as few as two categories, and others more than nine, with a median close to five. Reliable scales were obtained that exhibited the usual range–frequency effects and also the category effect; in other words, subjects generating fewer categories were more sensitive to the skewing of the immediate stimulus context. Although the research on open scales used psychophysical materials, the present demonstration of the category effect for social judgments encourages generalization to social research.

References

Anderson, N. H. (1982). *Methods of information integration theory.* New York: Academic Press.

Beebe-Center, J. G. (1932). *Pleasantness and unpleasantness.* New York: Van Nostrand.

Birnbaum, M. H. (1974a). The nonadditivity of impression formations [Monograph]. *Journal of Experimental Psychology, 102,* 543–561.

Birnbaum, M. H. (1974b). Using contextual effects to derive psychophysical scales. *Perception & Psychophysics, 15,* 89–96.

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36,* 129–148.

Brickman, P. (1975). Adaptation level determinants of satisfaction with equal and unequal outcome distributions in skill and chance situations. *Journal of Personality and Social Psychology, 32,* 191–198.

Brickman, P., & Campbell, D. T. (1971). Hedonic relativism and planning the good society. In M. H. Appley (Ed.), *Adaptation-level theory* (pp. 287–302). New York: Academic Press.

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

Cantril, H. (1965). *The pattern of human concerns.* New Brunswick, NJ: Rutgers University Press.

Crespi, L. (1942). Quantitative variation of incentive and performance in the white rat. *American Journal of Psychology, 55,* 467–517.

Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95,* 542–575.

Garner, W. R. (1960). Rating scales, discriminability, and information transmission. *Psychological Review, 67,* 343–352.

Gurin, G., Veroff, J., & Feld, S. (1960). *Americans view their mental health.* New York: Basic Books.

Gutek, B. A., Allen, H., Tyler, T. R., Lau, R. R., & Majchrzak, A. (1983). The importance of internal referents as determinants of satisfaction. *Journal of Community Psychology, 11,* 111–120.

Hamilton, D. L., & Zanna, M. P. (1974). Context effects in impression formation: Changes in connotative meaning. *Journal of Personality and Social Psychology, 29,* 649–654.

Helson, H. (1947). Adaptation-level as frame of reference for prediction of psychophysical data. *American Journal of Psychology, 60,* 1–29.

Helson, H. (1964). *Adaptation-level theory.* New York: Harper & Row.

Herr, P. M. (1986). Consequences of priming: Judgment and behavior. *Journal of Personality and Social Psychology, 51,* 1106–1115.

Herr, P. M., Sherman, S. J., & Fazio, R. H. (1983). On the consequences of priming: Assimilation and contrast effects. *Journal of Experimental Social Psychology, 19,* 323–340.

Higgins, E. T., & Lurie, L. (1983). Context, categorization and recall: The "change-of-standard" effect. *Cognitive Psychology, 15,* 525–547.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13,* 141–154.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion.* New Haven, CT: Yale University Press.

Johnson, D. M. (1944). Generalization of a scale of values by averaging of practice effects. *Journal of Experimental Psychology, 34,* 425–436.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93,* 136–153.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics & biases.* New York: Cambridge University Press.

Kenrick, D. T., & Gutierres, S. E. (1980). Contrast effects and judgments of physical attractiveness: When beauty becomes a social problem. *Journal of Personality and Social Psychology, 38,* 131–140.

Krantz, D. L., & Campbell, D. T. (1961). Separating perceptual and linguistic effects of context shifts upon absolute judgments. *Journal of Experimental Psychology, 62,* 35–42.

Krupat, E. (1974). Context as a determinant of perceived threat. *Journal of Personality and Social Psychology, 29,* 731–736.

Lockhead, G. R., & King, M. C. (1983). A memory model of sequential effects in scaling tasks. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 461–473.

Linville, P. W. (1987). Self-complexity as a cognitive buffer against stress-related illness and depression. *Journal of Personality and Social Psychology, 52,* 663–676.

Manis, M. (1967). Context effects in communication. *Journal of Personality and Social Psychology, 5,* 326–334.

Manis, M., & Paskewitz, J. R. (1984). Specificity in contrast effects: Judgments of pyschopathology. *Journal of Experimental Social Psychology, 20,* 217–230.

Manis, M., Paskewitz, J. R., & Cotler, S. (1986). Stereotypes and social judgment. *Journal of Personality and Social Psychology, 52,* 663–676.

Marsh, H. W., & Parducci, A. (1978). Natural anchoring at the neutral point of rating scales. *Journal of Experimental Social Psychology, 14,* 193–204.

Mellers, B. A. (1983a). Equity judgment: A revision of Aristotelian views. *Journal of Experimental Psychology: General, 111,* 242–270.

Mellers, B. A. (1983b). Evidence against absolute scaling. *Perception and Psychophysics, 33,* 523–526.

Mellers, B. A. (1986). "Fair" allocations of salaries and taxes. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 80–91.

Mellers, B. A., & Birnbaum, M. H. (1982). Loci of contextual effects in judgment. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 582–601.

Mellers, B. A., & Birnbaum, M. H. (1983). Contextual effects in social judgment. *Journal of Experimental Social Psychology, 19,* 157–171.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Parducci, A. (1963). Range-frequency compromise in judgment. *Psychological Monographs, 77* (92, Whole No. 565).

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review, 72,* 407–418.

Parducci, A. (1968). The relativism of absolute judgment. *Scientific American, 219,* 84–90.

Parducci, A. (1982). Category ratings: Still more contextual effects. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 262–282). Hillsdale, NJ: Erlbaum.

Parducci, A. (1983). Category ratings and the relational character of judgment. In H. G. Geissler, H. F. J. M. Buffort, E. L. J. Leeuwenberg, & V. Sarris (Eds.), *Modern issues in perception* (pp. 89–105). Berlin: VEB Deutscher Verlag der Vissenschaffen.

Parducci, A. (1984). Value judgments: Toward a relational theory of happiness. In J. R. Eiser (Ed.), *Attitudinal measurement* (pp. 3–21). New York: Springer-Verlag.

Parducci, A., & Perrett, L. F. (1971). Category rating scales: Effects of relative spacing and frequency [Monograph]. *Journal of Experimental Psychology, 89,* 427–452.

Parducci, A., & Wedell, D. H. (1986). The category effect with rating scales: Number of categories, number of stimuli, and method of presentation. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 496–516.

Peabody, D. (1962). Two components in bipolar scales: Direction and extremeness. *Psychological Review, 69,* 65–73.

Pepitone, A., & DiNubile, M. (1976). Contrast effects in judgments of crime severity and the punishment of criminal violators. *Journal of Personality and Social Psychology, 33,* 448–459.

Poulton, E. C. (1979). Models for biases in judging sensory magnitude. *Psychological Bulletin, 86,* 777–803.

Restle, F. (1978). Relativity and organization in visual size judgments. In E. Leeuwenberg & H. Buffort (Eds.), *Formal theories of visual perception* (pp. 247–263). New York: Wiley.

Riskey, D. R., Parducci, A., & Beauchamp, G. K. (1979). Effects of context in judgments of sweetness and pleasantness. *Perception and Psychophysics, 26,* 171–176.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45,* 513–523.

Schwarz, N., & Hippler, H. J. (1987). *What response scales may tell your respondents: Informative functions of response alternatives.* Unpublished manuscript.

Schwarz, N., Hippler, H. J., Deutsch, B., & Strack, F. (1985). Response categories: Effects on behavioral reports and comparative judgments. *Public Opinion Quarterly, 49,* 388–395.

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. M. (1976). Contrast effects in impression formation. *Journal of Personality and Social Psychology, 34,* 625–629.

Srull, T. K., & Wyer, R. S., Jr. (1979). Category accessibility and social perception: Some implications for the study of person memory and interpersonal judgments. *Journal of Personality and Social Psychology, 38,* 841–856.

Stevens, S. S. (1971). Issues in psychophysical measurement. *Psychological Review, 78,* 426–450.

Strack, F., Schwarz, N., & Gschneidinger, E. (1985). Happiness and reminiscing: The role of time perspective, affect, and mode of thinking. *Journal of Personality and Social Psychology, 49,* 1460–1469.

Sudman, S., & Bradburn, N. (1974). *Response effects in surveys: A review and synthesis.* Chicago: Aldine.

Thayer, S. (1980a). The effect of expression sequence and expressor identity on judgments of the intensity of facial expressions. *Journal of Nonverbal Behavior, 5,* 71–79.

Thayer, S. (1980b). The effect of facial expression sequence upon judgments of emotion. *Journal of Social Psychology, 111,* 305–306.

Thibaut, J. W., & Kelley, H. H. (1959). *The social psychology of groups.* New York: Wiley.

Upshaw, H. S. (1962). Own attitude as an anchor in equal-appearing intervals. *Journal of Abnormal and Social Psychology, 64,* 85–96.

Upshaw, H. S. (1978). Social influence on attitudes and on anchoring of congeneric attitude scales. *Journal of Experimental Social Psychology, 14,* 327–339.

Volkmann, J. (1951). Scales of judgment and their implications for social psychology. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads* (pp. 279–294). New York: Harper & Row.

Ward, L. M., & Lockhead, G. R. (1971). Response system processes in absolute judgment. *Perception & Psychophysics, 9,* 73–78.

Wedell, D. H. (1985). A process model for psychophysical judgment. (Doctoral dissertation, University of California, Los Angeles, 1984). *Dissertation Abstracts International, 45,* 3102B.

Wedell, D. H., & Parducci, A. (1985). Category and stimulus effects: A process model for contextual memory in judgment. In G. d'Ydewalle (Ed.), *Cognition, information processing, and motivation* (pp. 55–70). New York: Elsevier.

Wedell, D. H., Parducci, A., & Geiselman, R. E. (1987). A formal analysis of ratings of physical attractiveness: Successive contrast and simultaneous assimilation. *Journal of Experimental Social Psychology, 23,* 230–249.

Wyer, R. S., Jr. (1974). Changes in meaning and halo effects in personality impression formation. *Journal of Personality and Social Psychology, 29,* 829–835.

Wyer, R. S., Jr., & Srull, T. K. (1981). Category accessibility: Some theoretical and empirical issues concerning the processing of social stimulus information. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario Symposium on Personality and Social Psychology* (pp. 161–197). Hillsdale, NJ: Erlbaum.

Wyer, R. S., Jr., & Srull, T. K. (1986). Human cognition and its social context. *Psychological Review, 93,* 322–359.

Received July 25, 1986
Revision received December 21, 1987
Accepted January 7, 1988 ■