CONTEXTUAL SKEWING AND OVERALL MEAN JUDGMENT

Allen Parducci
University of California, Los Angeles <aparducc@ucla.edu>

Abstract

Systematic manipulation of the context for category ratings of psychophysical stimuli has isolated two dominating factors: 1) the range of contextual stimuli, and 2) the relative frequencies of stimuli within this range. The effects of these two factors upon the rating or judgment of any particular stimulus can be described as a weighted average between the proportion of the contextual range and the proportion of all contextual stimuli below the stimulus, implying that the overall mean of the judgments is proportional to contextual skewing. This characterization of the effects of skewing is illustrated by an experiment in which participants rated how satisfying monitary outcomes were in different contexts. A negatively-skewed context resulted in a greater overall mean judgment in this experiment and in each of a sample of more traditional psychophysical experiments. These results are described by a physical model, the teeter-totter, which also suggests how a decision that extends the range upward can lower the overall mean judgment. When the judgment is taken as a measure of pleasantness, a basic distinction between pleasure and utility is illuminated by contextual considerations.

Because psychophysical experiments permit control of the set of stimuli affecting the rating of each of the presented stimuli, such experiments have proven particularly useful for understanding the effects of the stimulus context upon category ratings. In the simplest cases, this context is constituted of just those stimuli presented in the experimental session. For example, if presented with a series of lifted weights, subjects might rate the lightest of those presented "1—very light," the heaviest "7—very heavy" on a typical seven-category scale, applying intermediate categories for weights intermediate between these two extremes. By systematically manipulating different features of the frequency distribution of weights presented, such as its endpoints or skewing, the experimenter can determine how these features affect the resulting ratings.

The same principles of judgment have been found to apply whenever the context can be controlled experimentally: for lifted weights, numerousness of dots, sizes of squares, sweetness of lemonades, and even for the pleasantness of facial expressions (Parducci, 1995). In all of these cases, individual ratings are determined by the place of each stimulus in the contextual range and its percentile rank in the frequency distribution of contextual stimuli, with each rating a compromise between these two determinants:

$$J_i = wR_i + (1 - w)F_i, \qquad [1]$$

where R*i* , the range value of Stimulus *i*, represents what its judgment would have been if determined solely by the position of *i* in the stimulus range, specifically by the proportion of the range below it. The frequency value of this same stimulus, F*i*, represents what the judgment would have been if it had been determined solely by its rank in the frequency distribution of contextual stimuli, specifically by its percentile rank (divided by 100). The relative weighting of the range and frequency values is given by *w*, a value between 0 and 1. The overt category rating is then a linear transformation of this judgment, depending upon the numerical values assigned to represent the categories (usually their ranks).

This paper concentrates on the effect of skewing the distribution of contextual stimuli on the grand, overall mean of the judgments. It is well known that the same stimulus receives a lower judgment when presented in a negatively-skewed distribution than in a positively-skewed distribution with the same endpoints, as predicted by Equation 1. Nevertheless, the overall mean of the judgments is actually higher for stimuli presented in a negatively-skewed distribution because the higher among its stimulus values are more frequent. This effect of skewing is expressed more precisely by Equation 2 (which derives algebraically from Equation 1):

$$\bar{J} = .5 + w(\bar{S} - MP)/Range, \qquad [2]$$

where $\bar{J}$ is the overall mean of the judgments, $\bar{S}$ is the mean of all stimulus values,
and MP is their midpoint (halfway between the two endpoints of the contextual range). Equation 2 is a measure of skewing that correlates almost perfectly with more conventional measures, though with algebraic sign reversed. In conventional terms, the more negatively skewed the distribution of contextual stimuli, the higher the overall mean of the judgments.

As a concrete example, consider an experiment (Parducci, 1968) that is easy to describe because its stimuli came with numerals already attached to them. Each of the participants "won" a series of monetary outcomes by turning over one of the three cards presented on any given trial and rating how satisfying it was. These outcomes ranged from 1 to 21 cents in the negatively-skewed condition, with the larger values coming more frequently--as in the representative sample diagrammed in Figure 1:

Neg. Skew: $\bar{J}$ = 4.5 (on scale from 1 to 7)

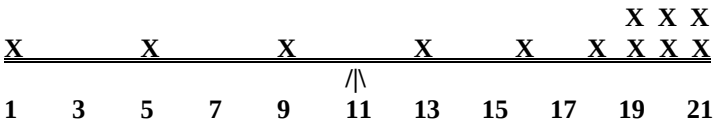

Figure 1. Negatively-skewed distribution of outcomes ranging from 1 to 21 cents.

In the positively skewed condition, the outcomes ranged from 7 to 27 cents, with the smaller among the outcomes coming more frequently (as shown in Figure 2).

Pos. Skew: $\bar{J}$ = 3.4

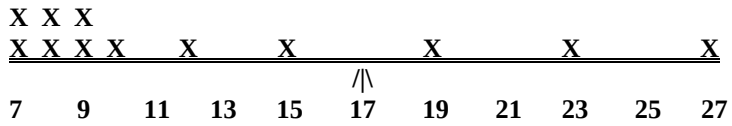

Figure 2. Positively-skewed distribution of outcomes ranging from 7 to 27 cents.

Regardless of which particular choices participants made (on any given trial, all three cards had the same value), their winnings always averaged to 14 cents in both conditions. This is above the midpoint for the negatively-skewed context, below it for the positively-skewed context. As predicted by Equation 2, the grand mean of all the judgments for the positively-skewed condition is a full category-step lower (on a scale from "1—Very Dissatisfying" to "7—Very Satisfying"), with none of the 30 subjects in this condition producing an overall mean as high as the 4.5 obtained for those in the negatively-skewed condition. The predictions from Equation 2 for these conditions, assuming w = .5, were 3.5 and 4.5, respectively, both very close to the obtained means.

That the effects of skewing produce higher overall mean judgments for negatively- than for positively-skewed distributions of stimuli is supported by tests using a wide variety of stimulus dimensions, mostly psychophysical. These overall mean differences are presented in Table 1 for various sets of published data available to me:

EFFECTS OF SKEWING UPON DIFFERENCE IN OVERALL MEAN OF RATINGS

Stimulus	Difference		Experimental	Source
Dimension	Overall	Mid-scale	Conditions	(Reference)
Money (cents)	.18	-.11	1-21 vs 7-27	Parducci (1968)
Square size	.15	-.20	High vs Low Mean	Parducci (1963)
	.13	-.16	Neg. vs Pos. 1-24	Parducci & Perrett (1971)
	.28	-.26	Neg(Neg) vs Pos(Pos)	Parducci & Perrett (1971)
	.12	-.26	Low vs High Midpoint	Parducci (1963)
Numerousness	.15	-.20	HL vs LH	Parducci (1963)
Sweetness	.13	-.31	Neg. vs Pos.	Risky et al.(1979)
Numbers	.19	-.21	-2A vs +2A	Birnbaum (1974)
	.11	-.42	Neg(F) vs Pos(D)	Parducci et al. (1960)
Pleasantness	.06	-.19	Normal vs U	Parducci (1989)

Table 1. *Overall* mean of judgments always higher for negative skewing (as predicted by Equation 2), difference in *midscale* judgment always in opposite direction (as predicted by Equation 1).

Because the scale of judgment is from 0 to 1, these differences may appear small. The differences in overt category ratings are many times larger, depending upon the number of categories (which varied from 5 to 9 in different experiments). To compute an *overall* mean judgment, all ratings obtained for a particular experimental condition were averaged; the resulting mean was then transformed linearly to the 0-to-1 scale. To compute a *midscale* judgment, category ratings were averaged across subjects for a single stimulus located near the midpoint of the stimulus range and then subjected to the same transformation. Thus, the comparison Neg(Neg) vs Pos(Pos) produced a difference in overall mean of more than one fourth of the scale, i.e., 1.25 category-steps on the six-category scale of size, with its difference in judgment of the midpoint stimulus almost as large but in the opposite direction (i.e., higher for the positively-skewed condition).

The smallest differences (bottom line of Table 1) were obtained for judgments of the pleasantness of different facial expressions displayed in photographs of an actress simulating different degrees of friendliness (described as "on a first meeting"). Pleasantness is not monotonic with friendliness in such a situation, the facial expression rated most pleasant being intermediate in friendliness (like one of Aristotle's golden means). When the faces are ordered by preference, the Normal and U contexts become negatively and positively skewed, respectively, and the judgments confirm the implications of Equations 1 and 2.

A Physical Model

A simple physical model, the teeter-totter, provides an intuitive representation of these effects of skewing on the overall mean of the judgments. In this model, the contextual range is represented by a weightless plank, with a fulcrum at its midpoint (as in each of the figures). The respective endpoints of the plank represent the extreme values of the context for judgment, whatever the absolute values of those extreme stimuli might be. Each stimulus in the context is represented as a weight (symbolized in the figures by an **X**), placed on the plank in accordance with its position relative to the endpoints of the contextual range. The physical tipping-moment of the plank, its tendency to tip down on one side or the other, is then proportional to the distance separating the mean location of these weights from the fulcrum. Equation 2 tells us that this tipping moment must then be proportional to the mean of all the judgments.

Insofar as the overt ratings of satisfaction in the gambling experiment are assumed to be valid expressions of the participants' internal judgments, i.e., what they are actually experiencing, the negatively-skewed distribution of payoffs is much happier. Following Bentham (1789/1948), this identifies happiness with the overall mean of the hedonic experiences, viz., the successive pleasures and pains averaged over the period in question (in this case the experimental session). However, contrary to Bentham's psychological hedonism, there is nothing in the model to imply that subjects exposed to both distributions of payoffs would choose the one that was negatively skewed. Pleasure and utility are not the same (Kahneman & Varey, 1991; Parducci, 1995), although they can be subject to the same contextual effects (Zaidel, 1971; Mellers and Cooke, 1994, 1996). The relationship between utilities and category ratings seems analogous to that between comparative and absolute judgments, with measurement of utility (e.g., the lottery method described by Raiffa, 1968) more analogous to the psychophysical method of adjustment. The relationship between utilities and hedonic judgments would seem a fruitful field of study for psychophysicists accustomed to experimentally manipulating the contexts for judgment.

Suppose that, in a misguided spirit of altruism, the experimenter decided to increase the happiness of the already happy participants receiving the negatively-skewed distribution of

outcomes. For example, it might be thought that by substituting payoffs of 30 cents for half of their previous top payoffs, viz., 21 cents, this new high would extend the right end of the plank--so that its new endpoint was further from the fulcrum. But that would be wrong. In order to represent the relational character of the range-frequency compromise, the length of the plank must remain constant. With the new, higher endpoint, the position of the previous high, 21 cents, is shifted almost one-third of the way toward the other end of the plank which is now evenly balanced as shown in Figure 3:

Pos. Extended: predicted $\bar{J}$ = 4.0

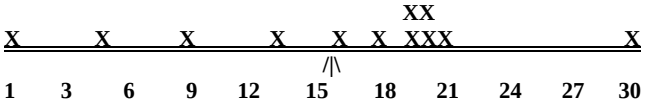

Figure 3. Previous negatively-skewed distribution of outcomes modified so that one of the 21-cent outcomes is shifted to 30 cents. Length of plank (representing the range of outcomes) remains unchanged, but skewing is eliminated so that plank is now evenly balanced**.**

Instead of increasing the overall mean rating of satisfaction above what it had been for the negatively-skewed distribution of payoffs (cf., Figure 1), the upward extension of the range has lowered this mean to 4.0 (on the 1-to-7 scale). The belief that an upward extension must raise the average level of satisfaction seems a common mistake, not just with respect to laboratory experiments but also in attempts to think about everyday satisfactions and dissatisfactions. This error can lead to bad choices when the alternatives would establish different contexts. A big part of the problem is that although outcomes with higher utility give more pleasure within any particular context, the high correlation between utility and pleasure *within* contexts does not hold *across* contexts. We may prefer any of the outcomes in one context to any of the outcomes in another context; and yet, if the skewing of the second is more negative than the skewing of the first, the second will yield more pleasure.

This may appear counterintuitive. Indeed, in a computerized game developed to explore this problem, players regularly get worse with practice (Parducci, 1995). One factor that may partially explain this unusual progression is that players are reinforced immediately for extending their range upward, i.e., by getting something they prefer to anything they had been getting; but the resulting reduction in satisfaction from lesser outcomes is delayed until these are experienced in the extended context. Another factor is that comparison between contexts takes place in a new, higher-order context in which the positively-skewed distribution of outcomes may be preferred—especially if it offers the possibility of more money. Learning that more may be experienced as less, i.e., as less pleasant overall, is one of life's more difficult lessons.

References

Bentham, J. (1948). *An introduction to the principles of morals and legislation.* New York: Hafner. (original work published 1789)

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

Mellers, B.A., & Cooke, A.D.J. (1994). Tradeoffs depend on attribute range. *Journal of Experimental Psychology: Human Perception and Performance, 20, 1055-1067.*

Mellers, B.A., & Cooke, A.D.J. (1996). The role of task and context in preference measurement. *Psychological Science, 7,* 76-82.

Parducci, A. (1963). The range-frequency compromise in judgment. *Psychological Monographs, 77* (2, Whole No. 565).

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

Parducci, A. (1989). Hedonic judgments for Aristotelian domains. In G. Canevet, B. Scharf, A.-M. Bonnel, & C.A. Possamai (Eds.), *Fechner Day 1989* (pp. 36-41). Cassis, France: International Society for Psychophysics.

Parducci, A. (1995). *Happiness, pleasure, and judgment: The contextual theory and its applications.* Mahwah, NJ: Erlbaum.

Parducci, A., Calfee, R.C., Marshall, L.M., & Davidson, L.P. (1960). Context effects in judgment: Adaptation level as a function of mean, midpoint, and median of the stimuli. *Journal of Experimental Psychology, 60,* 65-77.

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

Raiffa, H. (1968). *Decision analysis.* Reading, MA: Addison-Wesley.

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

Zaidel, D. (1971). A judgmental approach to decision analysis. *Dissertation Abstracts International,* 31, 10B. (University Microfilms No. 71-09,264).