# Contextual Effects in Salary Satisfaction

Michael H. Birnbaum[1] and Julien Rouvere[2]

[1]California State University, Fullerton

[1]mbirnbaum@fullerton.edu

[2]California State University, Fullerton

[2]rouvere@gmail.com

February 27, 2023

word count: 11,500

# Abstract

This article reports a series of studies of judgments of satisfaction with salary, manipulating the distribution of salaries of others doing the same work. The experiments were designed to compare six theories of contextual effects in judgment, including adaptation level theory, correlation-regression theory, inferred distribution theory, decision by sampling, ensemble theory, and range-frequency theory. Manipulations of the frequency distribution using cubic density functions produces a double crossover of curves relating judgments to salaries; this double crossover violates implications of four of the theories but remains consistent with decision by sampling and range-frequency theories. Manipulation of the endpoints produces changes in the heights and slopes of the curves, which are not explained by decision by sampling and are partially inconsistent with ensemble theory. Ensemble theory implies no effect of the rank of a salary and assumes that endpoints only affect judgments of salaries on the same side of the mean, contrary to the results. Range-frequency theory implies that ratings of stimuli holding the same ranks in two contexts with differing endpoints should be linearly related, and the data appeared consistent with this implication. Inferred distribution theory assumes that rank is inferred from the mean and endpoints, so it fails to describe the double crossover. Range-frequency theory is the only theory that gives a consistent account of all of the results. Range-frequency theory can be extended in order to estimate the effective context, which appears to differ systematically between people according to their full-time incomes.

Keywords: judgment of satisfaction, salary equity, range-frequency theory, ranking, context effects, ensemble theory.

Acknowledgments: Thanks are due to Allen Parducci for useful discussions.

# 1 Introduction

Psychologists have long known that "absolute" judgments such as "tall" or "short", "hot" or "cold", "moral" or "immoral", or "happy" or "unhappy" are relative (Helson, 1947, 1964; Parducci, 1968; Slovic, 1995). Contextual effects occur not only in perception and judgment, but also affect other behaviors, including choice (Ronayne & Brown, 2017; Wollschlaeger & Diederich, 2020), cognitive effort (Otto & Vassena, 2021), equity (Mellers, 1982, 1986), learning (Hayes & Wedell, 2022, 2023), memory (Wedell, Hayes, & Kim, 2020), similarity (Yearsley, et al., 2022), and temporal discounting (Stevenson, 1992, 2019).

Although one might argue that rational economic actors should care only about their own incomes, it has been reported that when people learn about the salaries earned by their peers, they can become dissatisfied with their job if they are paid less than the median of others in the same institution (Card, Mas, Moretti, & Saez, 2012). Boyce, Brown, and Moore (2010) concluded that it is the rank of one's income that largely determines satisfaction with one's salary (see also Brown, Gardner, Oswals, & Qian, 2008).

Putnam-Farr and Morewedge (2021) reported a series of studies to investigate which social comparisons affect satisfaction with one's salary. They argued against rank-based accounts and for an "ensemble" theory, which they described as follows: "A person making an above average salary would then compare her salary to the group mean and highest salary, for instance, whereas a person making a below average salary would compare his salary to the group mean and lowest salary.... our ensemble representation account implies that people should be insensitive to other properties of groups, ... such as their relative rank in the group." In one of their studies they failed to detect a significant effect of rank, which was interpreted as evidence in favor of the ensemble theory and not with rank-based theories such as decision by sampling (DbS), as in Stewart, Chater, and Brown (2006) or Boyce, Brown, and Moore (2010).

However, the studies of Putnam-Farr and Morewedge (2021) were not designed to provide a powerful test of the effects of rank as implied by DbS or by range-frequency (RF) theory (Parducci, 1965, 1968, 1995). One should not draw strong inferences from failure to reject the null hypothesis in a study not designed to provide a powerful, diagnostic test. The present study will provide such a powerful test.

Wort, Walasek, & Brown (2022) commented on Putnam-Farr and Morewedge (2021) to caution that the effects of ranking had not been ruled out. They noted that Putnam-Farr and Morewedge (2021) did not take into account the substantial body of empirical research testing spacing and frequency effects in RF theory, which provide strong evidence of effects of ranking in related judgment domains. Indeed, because RF theory developed as an alternative to Adaptation-Level (AL) theory (Helson, 1964), and because one of the main ways to distinguish RF from AL theory was to manipulate frequency independent of the mean, a substantial body of evidence has been amassed to show significant effects of rank in many judgment tasks (Parducci, 1965, 1995; Parducci & Perrett, 1971; Birnbaum, 1974).

To model the results of Putnam-Farr and Morewedge (2021), Wort, et al. (2022) proposed inferred distribution (ID) theory, in which people infer a normal distribution from the mean and endpoints of the salaries presented, and people are assumed to base their judgments on the ranks implied by that inferred distribution.

The next sections provide a brief review of the relevant theories of contextual effects as they apply to the analysis of salary satisfaction. Following the introduction, we present results of a series of experiments to compare the ensemble theory with the predictions of earlier theories of contextual effects, finding that the ensemble theory can be rejected because there are significant effects of stimulus rank and of the endpoints, as implied by RF theory that are not compatible with ensemble theory or the model of inferred distribution (ID) theory proposed by Wort, et al. (2022).

All six theories in the next section allow that judgments of satisfaction do not depend

solely on one's salary but also on the amounts paid to others, but they differ in how the context affects judgments.

## 1.1 Adaptation-Level Theory

Helson (1947, 1964) proposed Adaptation-Level (AL) theory to provide a mathematical account of frame of reference effects in judgments. This theory predicted quantitatively the effects of the focal stimuli, anchors, background stimuli, and the residual context attributed to prior experience. The basic idea of AL theory is that all stimuli, past and present, real or imagined, pool to form the AL, which is a remembered representation of prior stimuli and which forms the frame of reference for judgment of new stimuli.

The AL is theorized to be a weighted average of all of these stimuli. Each participant is assumed to bring in his or her prior context (aka "residual" context) that represents the participant's memories of stimuli relevant to the task. For example, in a study of salaries, people are presumed to already have ideas about what salaries would be satisfying or unsatisfying.

This theory was developed and tested initially with psychophysical stimuli, but many studies have shown that the principles apply to a broader domain of stimuli, tasks and judgments (Helson, 1964); Edwards (2018) reviews the legacy and extensions of AL theory in the field of behavioral economics.

The AL is that stimulus whose subjective value equals the weighted average of the subjective values of all of the relevant stimuli in the context. For psychophysical stimuli theorized to follow Fechner's law, that subjective values are a logarithmic function of physical values, the AL is the antilog of the weighted average of the logs of the stimuli; therefore, AL is a weighted geometric mean of the physical stimuli. The stimulus that is called "average" is thus the average stimulus, and all other stimuli are judged in relation to it (Helson, 1947, 1965; Birnbaum, 1974).

Because a stimulus designated as an "anchor" is averaged with other stimuli to form

the AL (Helson, 1947), and because any averaging model is equivalent to an anchoring and adjustment model, the term "anchoring and adjustment" has been used (Tversky & Kahneman, 1974) to refer to a simplification of AL theory in which the residual context is ignored.[1]

The importance of residual context has been demonstrated in a number of papers (Helson, 1964), For example, Rethlingshafer and Hinckley (1963) asked people of different ages to judge how "old" or "young" people are. At what age is an adult neither young nor old but "middle" in age? According to the children tested (aged about 10), a middle aged person is 36 on average; according to college-aged participants, middle is 41; and for an older group in their seventies, middle-age is about 49. Rethlingshafer and Hinckley were able to fit these values via AL theory, in which the AL is a weighted average of the ages of the participants combined with the values of the stimuli.[2]

In this article, we examine one correlate of the residual context by examining the relationship between judgments of satisfaction with specified salaries and participants' incomes.

## 1.2 Correlation-regression theory

Johnson and Mullally (1969) proposed correlation-regression (CR) theory. In this theory, the standard deviation of the stimuli in a context and the mean of the stimuli determine how a stimulus relates to its context. Let $\mu_k$ and $\sigma_k$ represent the mean and standard deviation of the subjective values of stimuli in context $k$; let $s = u(x)$ represent the subjective value of stimulus $x$, where $u(x)$ is the psychophysical function (utility function) of physical value. The formula for a standard score ($z$-score) is as follows:

[1]The idea that anchors receive greater weight than other stimuli, sometimes called "insufficient adjustment", was stated as principle No. 2 in Helson (1947, p. 28). Tversky and Kahneman (1974) did not cite Helson, which led some authors to write that Tversky and Kahneman had proposed "anchoring and insufficient adjustment" as an original theory.

[2]Although Rethlingshafer and Hinckley referred to participants' ages as a type of "background" stimulus, we prefer to use the terms "residual" or "prior" context for experiences that differ among participants, and we reserve "background" for stimuli that are presented in the experiment but fixed in value.

$$z = \frac{s - \mu_k}{\sigma_k} \tag{1}$$

where $z$, the standard score, describes the relationship between stimulus $x$ and its context, represented by mean and standard deviation of subjective values. The key idea of this theory is that apart from error, people would choose a response such that that the standard score of the response relative to the response distribution matches the standard score of the stimulus relative to its distribution.

When there are errors in perceptions or memories of the stimuli or in the assignment of responses to stimuli, there will be regression that can be described by the correlation coefficient between stimuli and responses. Indeed, the least-squares regression (prediction) formula states that the $z$ score of the predicted response is the product of the correlation coefficient and the $z$ score of the stimulus.

This CR theory is more general than AL theory because the response to a stimulus depends on both the mean and the variance of stimuli in a context, whereas in AL theory, the response to a stimulus depends only on its relation to the AL.

## 1.3 Inferred Distribution Theory

Wort, et al. (2022) proposed that the memories of stimuli are sampled to infer a normal distribution, and the response to a stimulus depends on its rank in that inferred distribution (ID).

The response to a stimulus is assumed to be a linear function of the rank of a stimulus in the normal distribution, where:

$$r_k = N[\frac{s - \mu_k}{\sigma_k}] \tag{2}$$

where $N$ is the cumulative standard normal distribution function, $r_k$ is the rank of stimulus

$s$ in Context $k$, as a cumulative probability on a scale from 0 to 1. The response is assumed to be a linear function of $r_k$.

This ID theory can be viewed as a modification of the decision by sampling (DbS) theory of Stewart, Chater, and Brown (2006), described in the next section, and it can also be interpreted as a modification of CR theory, because the ranking is a function of the standard score of the stimulus in its distribution. In DbS, the response to a stimulus is a function of the rank of a stimulus in the sampled distribution of the context, whereas in ID theory, the distribution is assumed to be normal and so the distribution can be summarized by the mean and standard deviation, which are inferred from the mean and endpoints of the sampled distribution.

The theory differs from CR theory in that it assumes that responses are linearly related to rank, rather than linearly related to the standard score, but at the heart of ID theory is the same $z$ score that appears in CR theory to express the relationship of a stimulus to its context.

## 1.4 Decision by Sampling

Stewart, et al. (2006) proposed Decision by Sampling (DbS), which is based on two main ideas: (1) When makimg judgments about stimuli, people sample from memory and rank the stimuli in the sample, and (2) when comparing two stimuli, people only compare stimuli an ordinal scale; that is, people can say which is more or better, for example, but cannot or do not relate them on a metric scale. In this theory, what has been labeled as a metric utility or psychophysical function is instead a relative ranking of the stimuli in the sampled context, which includes prior memories.

Let $k$ index the context, and suppose there are $n$ stimuli in the sample. The stimuli are ranked from 1 (lowest or worst) to $n$ (highest or best), where $r_{xk}$ is the absolute rank of stimulus $x$ in Context $k$, then the relative rank of stimulus $x$ is given as follows:

$$F_k(x) = \frac{r_{xk} - 1}{n - 1} \tag{3}$$

where $F_k(x)$ is the relative rank value of $x$ in Context $k$, which ranges from 0 to 1. According to DbS theory, a person's satisfaction with salary depends only on the relative rank of the salary in the sampled distribution (Brown, Gardner, Oswald, & Qian, 2008). The rating is assumed to be linearly related to this relative rank value; for example, on a 7 point scale, it would be $6F_k(x) + 1$.

This DbS theory does not assume a normal distribution as in ID theory, so it is more general than ID theory in this regard; however, DbS does not explicitly account for experimental manipulations of the endpoints, which ID can accommodate via their assumed effects on the inferred value of $\sigma$.

## 1.5 Ensemble Theory

Putnamn-Farr and Morewedge (2021) proposed ensemble (EN) theory, which assumes that people summarize a contextual distribution by the statistics of mean and endpoints, and that the upper endpoint is applicable when the stimulus exceeds the mean whereas the lower endpoint is applicable when the stimulus falls below the mean.

Putnamn-Farr and Morewedge (2021) did not state EN theory as a mathematical model. To express their ideas mathematically, we combined their statements about the theorized effects of mean and endpoints with some assumptions that seem implicit in their presentation. We assumed that judgments should be a monotonically increasing function of salary, that the response will be at the middle of the scale when salary is equal to the mean, that it will be minimal and maximal when equal to the lower and upper endpoints, respectively, and that each segment of the function is linear. The following equations then express these ideas:

$$e_k = \begin{cases} (s - \mu_k)/(s_{mk} - \mu_k), & \text{if } s > \mu_k \\ (s - \mu_k)/(\mu_k - s_{0k}), & \text{if } s \leq \mu_k \end{cases} \tag{4}$$

where $e_k$ is the ensemble value of stimulus $x$ in Context $k$ having a subjective value of $s = u(x)$; $s_{0k}$ and $s_{mk}$ are the minimum and maximum in the context; and the final rating is assumed to be a linear function of $e_k$. For example, on a 7 point scale, the response is assumed to be $3e_k + 4$ because $e_k$ ranges from $-1$ to 1; in this case, the response would be 1 when $s$ is the minimum, it would be 7 when the the stimulus is maximal, and it would be 4 when equal to the mean.[3]

## 1.6 Range-Frequency Theory

Range-Frequency (RF) theory (Parducci, 1965, 1968, 1995) was proposed as an alternative to Helson's AL theory. In RF theory, the context is represented as a probability distribution rather than as a single value, as in AL. Although the theories differed in how context affects judgments, Parducci (1995, Chapter 3) retained and elaborated Helson's conception of the context as a combination of residual, background, and experimental stimuli. RF theory was developed to understand human happiness, but RF theory has been tested mostly with psychophysical stimuli because of the better control over context available with such stimuli compared to social or hedonic stimuli where people might bring vastly different contexts to the experiment. But studies with social, moral, and hedonic stimuli have been consistent with findings with psychophysical stimuli (Helson, 1964; Parducci, 1968, 1995; Birnbaum, 1982; Mellers & Birnbaum, 1983; Wedell & Parducci, 1988; Tripp & Brown, 2016).

Whereas in Helson's AL theory, the effects of all stimuli pool to form a single value,

[3]Other representations might be possible for an ensemble of mean and endpoints, but Expression 4 seemed the most plausible of those we considered. An alternative assumption that comparisons of salary with the mean and appropriate endpoint have additive contributions easily leads to responses that are not a monotonic function of salary.

the AL (average), in RF theory, the effects of experimental manipulations and experience combine to produce a distribution, and judgments are represented as a compromise between how each stimulus compares to the cumulative frequency (rank) and the position of that stimulus relative to the endpoints of the distribution (range).

For this paper, a special case of Parducci's (1965, 1995) RF theory will be presented for judgments of satisfaction with one's salary. More general statements of RF theory are available in Birnbaum (1974, 1982), Mellers and Birnbaum (1982), and Wedell, Hayes, and Kim (2020).

Range-frequency theory posits that one's happiness with a salary depends in part on a context-independent utility function and in part on the context for judgment. In RF theory, context can be thought of as a mental representation of a distribution of salaries that form the *frame of reference for judgment*. This distribution depends on the participant's experiences, real or vicarious that represent what other people earn or might earn.

Thus, the *effective* context for judging salary satisfaction is an aggregation that depends on the residual (prior) context that a participant brings to the lab, background factors produced by the experimental materials in a given study, and the distribution of salaries earned by others who do the same work and are equally deserving in Context $k$. The context provided by the experimenter in a study thus combines with the participant's prior context to form a new distribution that is the effective context for judgment.

Factors that affect the residual context and thus the effective context might include a participant's own income, the salaries of one's friends and family, and vicarious experiences from media and other sources of information about salaries. For example, a person who earns \$150,000 per year and associates with others earning the similar values would likely judge a salary of \$50,000 per year to be unsatisfying, whereas a person who is currently earning \$30,000 per year might consider \$50,000 to be very satisfying.

For simplicity, predictions of RF theory will be initially calculated as if the context

for judgment is produced entirely by the stimuli presented within the experiment, ignoring individual residual contexts outside the lab, but a method for using RF theory to estimate the effective context, reflecting prior context, will be presented in a later section. Therefore, the next sections assumes that prior context can be ignored, and the predictions are calculated as if these were judgments of abstract numbers, as in Birnbaum (1974). In addition, the context-free psychophysical function for salary, $u(x)$, will be assumed to be linear to further simplify the presentation.[4]

Let $x_{0k}$ and $x_{mk}$ represent the minimum and maximum salaries presented in Context $k$, and let $F_k(x)$ = the cumulative probability (relative rank) of $x$ in Context $k$; by definition, $F_k(x_{0k}) = 0$ and $F_k(x_{mk}) = 1$.

Range-frequency theory posits that judgments are a compromise between two systems of judgment: the range principle, which transforms judgments linearly relative to $u(x)$ and the endpoints of the distribution, and the frequency principle, which evaluates stimuli relative to their cumulative probabilities (relative ranks).[5]

### 1.6.1 The Range Principle

Let $H_k(x)$ be the range value of salary $x$ in Context $k$, which is defined as follows:

$$H_k(x) = \frac{u(x) - u(x_{0k})}{u(x_{mk}) - u(x_{0k})} \tag{5}$$

where $u(x)$ is the utility function for salary. $H_k(x)$ will range from 0 to 1, as $x$ ranges from $x_{0k}$ to $x_{mk}$.

[4] Birnbaum(1974) showed how one can estimate the $u(x)$ function using RF theory from empirical data.

[5] In Parducci's (1965) theory, the frequency principle is equivalent to a tendency to use the response categories with equal frequency; that is, a tendency to assign an equal number of stimuli to each category. In Birnbaum's (1974, p. 94-95) more general extension of RF theory, the judge may have another target distribution of responses besides the uniform distribution; for example, when assigning grades, a teacher might have tendencies to grant fewer A than B or C grades, and to assign fewer D and F than B and C.

### 1.6.2 The Frequency Principle

The frequency value of salary $x$ in Context $k$ is $F_k(x)$. When $n$ stimuli have been ranked by successive integers from the lowest, $r_{0k} = 1$ to the highest $r_{mk} = n$, and $r_{xk}$ is the rank of salary $x$ in Context $k$, $F_k(x)$ is given by the following:

$$F_k(x) = \frac{r_{xk} - 1}{n - 1} \tag{6}$$

The frequency value also ranges from 0 to 1.

### 1.6.3 Range-frequency Compromise

The range-frequency compromise is an average between the position of a stimulus relative to the range and relative to the frequency (ranking) of the stimuli.

$$RF_k(x) = (1 - w)H_k(x) + wF_k(x) \tag{7}$$

where $w$ is the weight of the frequency principle.[6]

### 1.6.4 Response Scale

The transformation from the subjective range-frequency value, $RF$, to the overt response, $R$, will depend on the subjective values of response values, the spacing and frequency of example responses, the number of categories, and on the psychophysics of the response mechanism (Birnbaum, 1982; Parducci, 1982). In psychophysical studies, participants are sometimes

[6]Parducci (1982) and Wedell and Parducci (1988) examined factors that affect the relative weighting of the range and frequency principles. Tripp and Brown (2016) fit individual participant data for conditions with fixed endpoints and found that most people had weights between 0 and 1, compromising range and frequency principles, but a few people could be fit with weights of 0 or 1. Hayes and Wedell (2023) summarize studies showing $w$ is about 0.5. In Decision by Sampling (DbS) theory (Stewart, Chater, & Brown, 2006), it is argued that only the ranking term is needed. The distinction between RF theory and DbS theory will be further explored in Experiment 2 of the present article.

instructed to assign the lowest response to the smallest value and the highest response to the highest stimulus (sometimes these stimuli are called "end anchors"), and it is often assumed that a uniform distribution of one-digit integers are equally spaced. Let $R_0$ and $R_m$ represent the minimum and maximum response on an equally spaced rating scale.[7] With these simplifying assumptions:

$$R_k(x) = (R_m - R_0)RF_k(x) + R_0 \tag{8}$$

where $R_k(x)$ is the predicted rating of salary $x$ on an equal interval scale from $R_0$ to $R_m$ in Context $k$.

### 1.6.5 Estimating the Effective Context via RF Theory

In RF theory, the effective context is not represented by a single number, as it is in AL theory, but instead by a probability distribution that combines the effects of the experimental stimuli with the person's prior experience. The third section of results in Experiment 1 (Section 2.3.3) introduces a method (that to the best of our knowledge is new) for estimating the effective contexts for groups of people who might reasonably be theorized to have different prior contexts.

Just as people of different ages might be anticipated to have different contexts for judging whether a person is young or old (Rethlingshafer & Hinckley, 1963), it seems reasonable that people who have different incomes would have different contexts for judging satisfaction with hypothetical full-time salaries. Therefore, we will examine judgments of salary satisfaction by people who work full time and have different levels of income. The method assumes RF theory and estimates the effective distribution for each income group as the frequency distribution that reconciles RF theory with their data.

[7] Methods for testing if ratings are equally spaced, and for analysis when responses are only assumed to be monotonic are discussed in Birnbaum (1974, 1982).

Table 1: Theories of Contextual Effects

| Abbrev | Theory | Relativity Factor |
|---|---|---|
| AL | Adaptation Level | $s - \mu_k$ |
| CR | Correlation-Regression | $(s - \mu_k)/\sigma_k$ |
| ID | Inferred Distribution | $N[(s - \mu_k)/\sigma_k]$ |
| EN | Ensemble, $s > \mu_k$ | $(s - \mu_k)/(s_{mk} - \mu_k)$ |
| | Ensemble, $s \leq \mu_k$ | $(s - \mu_k)/(\mu_k - s_{0k})$ |
| DbS | Decision by Sampling | $F_k(s)$ |
| RF | Range-Frequency | $wF_k(s) + (1 - w)(s - s_{0k})/(s_{mk} - s_{0k})$ |

### 1.7 Summary of Theories

Table 1 presents a summary of the theories of contextual effects, including their abbreviations along with expressions that express the key idea of each theory. All of the theories allow a psychophysical function, $s = u(x)$. The mean and standard deviation of the subjective values in Context $k$ are $\mu_k$ and $\sigma_k$, respectively; mimimum and maximum in Context $k$ are $s_{0k}$ and $s_{mk}$, respectively.

## 2 Experiment 1: Frequency/Ranking

### 2.1 Predictions for Experiment 1

In Experiment 1, we employ two distributions of salary in which there were 7 common levels of salary: \$40K, \$42K, \$44K, \$46K, \$48K, \$50K, and \$52K (where K indicates thousands of USD). In Condition C1, there were 5 additional contextual stimuli with values between \$40K and \$42K and 10 additional between \$46K and \$50K; whereas in Condition C2 there were 10 contextual stimuli between \$42K and \$46K and 5 between \$50K and \$52K. These

were based on the cubic distributions used by Birnbaum (1974) in a study of judgments of the magnitudes of numbers.

Predictions of the simplified RF theory are shown in Figure 1; they are calculated on a 7-point rating scale, as used in Putnam-Farr and Morewedge (2021) and in the present studies. Predictions are plotted in Figure 1 as a function of salary, with a separate curve for each context. RF theory implies that for these distributions, the curves should cross twice, at \$44K and \$48K.

Unlike RF theory or DbS, EN theory implies that rank of a stimulus has no effect and that endpoints only influence judgments on the same side of the mean.[8] Thus, EN theory cannot imply curves that cross twice. The implication of a double crossover in RF theory will be tested for judgments of salary satisfaction in Experiment 1.

Four theories, AL, CR, ID, and EN, cannot imply that curves can cross both above and below the mean. Further, because the mean of the stimuli in C1 (\$45.7K) is slightly lower than the mean in C2 (\$46.3K), the judgment of \$46K should be equal or higher in C1 than in C2 according to AL, CR, EN, or ID, which is opposite of the prediction of RF and DbS. RF and DbS imply that the rating of \$46K should be higher in C2, due to the higher ranking of \$46K in C2 relative to C1. Thus, these cubic distributions provide a test of the effects of ranking and distinguish RF and DbS theories, which can imply the double crossover, from the other four theories.

## 2.2 Method

The participants read a list of salaries received by people doing the same job and judged how happy or unhappy they would be to receive each of those salaries. There were two between-subject conditions using different distributions of salaries, to which participants

[8]Putnam-Farr and Morewedge (2021) reported that the effect of rank was not significant, nor was the effect of the maximum on judgments of salary below the mean; however, failure to find statistical significance does not prove the null hypotheses.


Figure 1: Predicted judgments based on simplified Range-Frequency theory for two cubic contextual distributions of Experiment 1; Condition C1, shown with open circles and dashed curve, has 5 additional contextual stimuli between \$40 and \$42K and 10 between \$46K and \$50K; Condition C2 is shown with filled squares and solid curve and had contextual values between \$42K and \$46K and between \$50K and \$52K.

were randomly assigned. This study was not preregistered.

### 2.2.1 Instructions and Procedure

The instructions read (in part) as follows: "This is a study of satisfaction with salary and how it depends on comparisons of salary with salaries paid to others working in the same job.

"Imagine that you have worked for a company for 2 years and you learn for the first time that not everyone doing the same work is paid the same. You find a list of 22 people who are doing the same work and have been evaluated as equally qualified and productive....

"Your task is to rate how dissatisfied or satisfied, how happy or unhappy, you would be if you received each of those salaries, now that you know what other people are getting who are doing the same work. Please make your ratings on the 7 point scale ...to indicate how satisfied or dissatisfied you would feel about your salary: "

The experiment was conducted online. Those who volunteered to participate clicked a link, which randomly assigned them to one of two conditions. Complete instructions and materials for the conditions can be found at the following URLs:
https://konstanzworkshop.neocities.org/Salary22/salary_c1xy66a.htm
and https://konstanzworkshop.neocities.org/Salary22/salary_c2xy66a.htm

Participants were asked to read the list of salaries and to imagine how they would feel if they received each of the salaries. The list was then presented a second time, with the request to rate how happy or unhappy they would be to receive each salary, which they did by clicking on a seven button response scale, labeled from 1 = "Not at all Happy" to 7 = "Extremely Happy."

### 2.2.2 Stimuli and Design

Conditions C1 and C2 resemble two cubic distributions used by Birnbaum (1974), except there were only 22 values used here instead of 46. Condition 1: \$40K, \$40.2K, \$40.4K, \$40.5K, \$40.6K, \$40.7K, \$42K, \$44K, \$46K, \$47.1K, \$47.2K, \$47.5K, \$47.7K, \$47.8K, \$48K, \$48.1K, \$48.4K, \$48.5K, \$48.8K, \$49K, \$50K, \$52K.

Condition 2: \$40K, \$42K, \$43K, \$43.4K, \$43.6K, \$43.8K, \$43.9K, \$44K, \$44.1K, \$44.3K, \$44.4K, \$44.6K, \$45K, \$46K, \$48K, \$50K, \$51K, \$51.5K, \$51.7K, \$51.8K, \$51.9K, \$52K.

Note that there are 7 values common to both distributions: \$40K, \$42K, \$44K, \$46K,

$48K, $50K, and $52K. Salaries were displayed in American style; e.g., $40.2K was displayed as $40,200.

The questionnaire also requested participant's gender, age, highest level of education, nationality, total hours per week worked for pay, and yearly income, rounded to the nearest thousand USD.

### 2.2.3 Participants

There were 325 participants who were recruited via /r/SampleSize subreddit (URL = https://www.reddit.com/r/SampleSize/) and Twitter (URL = https://www.twitter.com). There were 164 and 161 in Conditions 1 and 2, respectively. Of the 318 who indicated gender, 166 responded male (52%). Age ranged from 18 to 61, with 39% aged 30 or older, and 18% were 22 or younger; 68% reported holding bachelor's degrees, including 7% with doctorates.

Of the 325 participants, 313 provided income information, reporting a median of $45K per year, with 135 earning $40K or less. There were 191 who worked 38-42 hours per week, with median and mean incomes of $57K and $73.8K USD.

## 2.3 Results

Some participants with high incomes rated all of the hypothetical salaries of the study as "1", whereas others with lower incomes rated all of the salaries as "7"; such data are not diagnostic among theories of contextual effects and would be considered "unusual" in a study with psychophysical stimuli. There were 104 (of 325) participants who either gave the same response to all salaries, who preferred a middle-level salary to both the highest or lowest, or who showed another unusual pattern; these unusual data were analyzed separately and are described in the section after next; the unusual data are included in the section following the next, which analyzed judgments in relation to incomes. Excluding the unusual data, there were 221 remaining participants who formed the "main" groups of 100 and 121 in C1 and

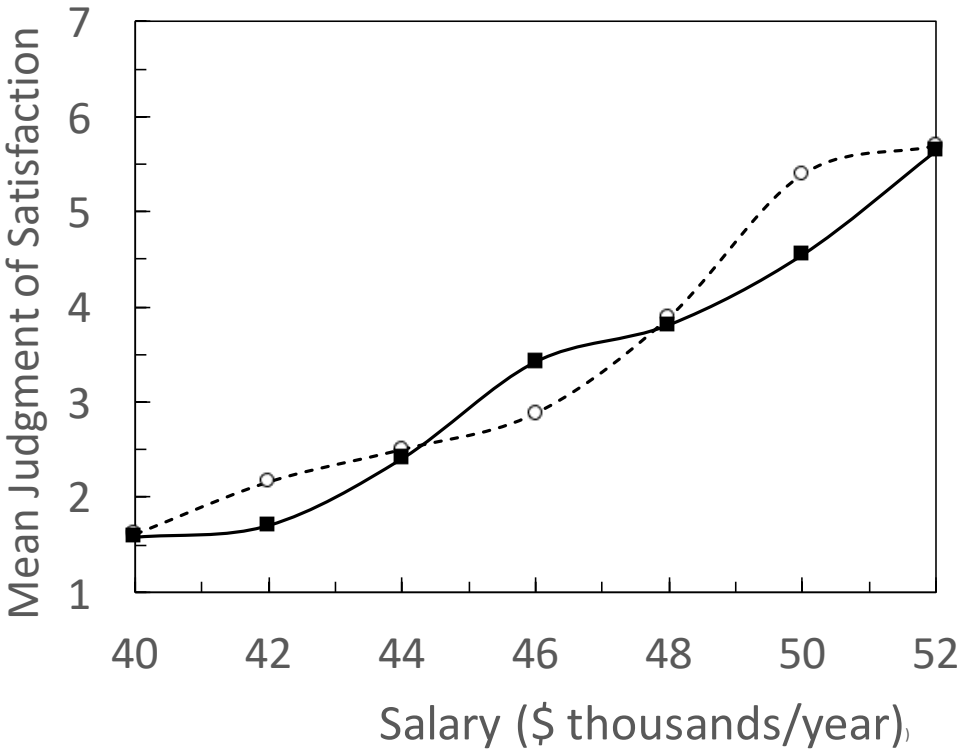


Figure 2: Mean judgments of satisfaction for the main groups of participants in the two conditions of Experiment 1, plotted as a function of Salary. Condition C1 is shown with open circles and dashed curve; Condition C2 is shown with filled squares and solid curves.

C2, respectively, whose results are described in the next section.

### 2.3.1 Experimental Context Effects

Figure 2 shows mean judgments of salary satisfaction for the main groups of participants as a function of salary, with a separate curve for each experimental context condition, for the seven levels of salary common to both conditions. Recall that of the 22 stimuli in Condition C1 (unfilled circles in Figure 3), there were five extra stimuli between \$40K and \$42K, and ten extra between \$46K and \$50K; whereas in Condition C2 (filled squares), there were

ten extra between $42K and $46K and five between $50K and $52K. Consistent with the frequency principle of RF theory or the ranking principle of DbS, the empirical curves are steeper in regions that have a greater density of stimuli. The empirical curves cross twice, near $44K and $48K, corresponding to the predicted crossovers of the simplified RF theory in Figure 1. Standard errors of the means in Figure 2 range from 0.11 to 0.16, roughly the size of the markers in the figure.

These results show significant effects of the ranking of the stimuli. The differences in mean judgments (C1 − C2) are significant ($p < 0.01$) for Salaries of $42K, $46K, and $50K, $t(219) = 2.65, -2.49$, and 4.10, respectively, with signs consistent with RF predictions in Figure 1.

Note that the mean judgment of $46K in Condition C2 is higher than that in Condition C1. A Salary of $46K is 14th (from the bottom) in C2 and only 9th in C1. However, the means of salaries presented are $46.4K in C2 and only $45.7K in C1. If people judged salaries in comparison with the mean, as in AL, CR, ID, or EN theories, they would give equal or lower responses to $46K in C2 than C1. Instead, the results show that ratings are significantly higher in C2 where the relative rank is higher (despite the higher mean salary), contradicting the predictions of those four theories, but consistent with RF or DbS theories.

The double crossover in Figure 2 contradicts the EN theory that judgments are a function of mean and endpoints and independent of rank. Nor is such a double crossover compatible with any fixed function of mean and standard deviation, as in AL, CR or ID. Instead, ratings depend on the cumulative frequency distribution (i.e., ranking), consistent with RF and DbS theories.

### 2.3.2 Analysis of Unusual Data

There were 104 sets of "unusual" data; most of these (57 people) gave the same response to all of the salaries listed, including 34 who rated all salaries as "1" and 14 who rated all

as "7". Some of those who assigned all "1" wrote comments that one could not live on such low salaries, and others who gave all "7" wrote that all of these same salaries were unbelievably high. From the perspective of AL or RF theories, such responses indicate that participants brought in very different prior contexts that overwhelmed the context provided by the stimuli used in the experiment. Some comments, however, expressed another reason one might respond all "1": some wrote that they would be unhappy to work where equally deserving people were paid unequally. Participants were not asked to evaluate "fairness" but salary equity (Birnbaum, 1983; Mellers, 1982, 1986) and salary satisfaction are no doubt related.

There were 34 people who had data patterns in which all salaries except the highest were evaluated as "1" and the highest was given another rating. The most common (14 people) was to assign "2" to the highest salary. Such patterns could occur in RF or DbS theories from a prior distribution in which the lowest salaries of the experiment were rare and below all experience in the prior context. This data pattern might also be compatible with the idea, expressed in a couple of comments, that it would be intolerable to be paid anything less than the highest amount the employer was willing to pay for the same work.

There were 13 people who gave higher ratings to salaries in the middle of the range than to the highest or lowest salaries. Presumably, these people would be unhappy to be the one receiving the highest salary when workers are not paid equally, as if they might become targets of jealousy or suspected of having done something improper to receive special treatment.

Although participants were randomly assigned to conditions, it was the case that among those working full time, there were 10 more in C1 than C2 who had salaries less than $55 thousand and 9 fewer in C1 who had incomes greater than that value. Possibly related to this difference, there were 64 and 40 people in Conditions C1 and C2 who displayed one of the unusual data patterns respectively, an unanticipated significant difference, Yates'


Figure 3: Mean judgments of satisfaction as a function of salary, for participants who worked full time, with separate curves for each level of reported income (Inc). Data are averaged over Conditions C1 and C2. Mean judgments by those who reported incomes below \$40K per year (Inc $<$ \$40K) are shown as open circles. Mean judgments by individuals who had full-time incomes from \$40K to \$52K, between \$52K and \$85K, and above \$85K per year are shown as filled squares, open triangles, and filled diamonds, respectively. The curves show predicted values calculated from RF theory with the assumption that the effective context can be approximated by a beta distribution.

$\chi^2(1) = 6.87$, $p < 0.01$.

### 2.3.3 Residual Context Effects

The residual context refers to the distribution of prior experiences that a participant brings to the experiment and which is not under experimental control. The effective context is (in

theory) a combination of the residual context and the immediate context provided by the stimuli and background of the experiment. Among factors that are likely correlated with a person's residual context in a study of satisfaction with salaries would be the individual's income.

To examine the relationships between income and judgments, we divided data for the 191 participants who reported working full time (38-42 hours per week) into four groups according to self-reported income. This analysis includes both main and unusual data and combines across experimental contexts. There were 48, 36, 48, and 59 individuals who had incomes less than \$40K, \$40K to \$52K, between \$52K and \$85K, and \$85K and above, respectively.

Figure 3 shows mean judgments of satisfaction for these income groups as a function of salary: unfilled circles show judgments for those with lowest incomes; filled squares are for incomes from \$40K to \$52K; unfilled triangles and filled diamonds show results for those with two highest ranges of income. Figure 3 shows that people earning more than \$52K rate salaries from \$40K to \$52K lower than do those who earn \$52K or less.[9]

The mean judgments in Figure 3 were fitted using a variant of the simplified RF theory, modified by the assumption that the average effective context is distributed as a beta distribution with endpoints and shape parameters that depend on a group's income level. It was assumed that $u(x) = x$, $w = 0.5$, and that the rating scale was uniform and equally spaced from 1 to 7. The data were fit to the equation:

$$P_g(x) = 6[w\frac{(x - y_{0g})}{(y_{mg} - y_{0g})} + (1 - w)B(x, \alpha_g, \beta_g, y_{0g}, y_{mg})] + 1 \tag{9}$$

[9]Incomes derived from part-time or temporary work seem less relevant to a person's context for judging satisfaction with full-time salaries. For example, a Computer Science major who is working 10 hours/week as an assistant on campus may have a context based more on the salaries of friends who have taken computer science jobs than based on the wages of a part-time assistant. Nevertheless, we found similar, but smaller magnitude relationships to Figure 3 for part-timers: part-timers earning less judged a given salary as more satisfying on average than those earning more.

where $P_g(x)$ is the predicted mean judgment of salary $x$ by income Group $g$; $B()$ is the cumulative Beta distribution; $\alpha_g$ and $\beta_g$ are the estimated shape parameters for Beta distribution in Group $g$; $y_{0g}$ and $y_{mg}$ are the estimated minimum and maximum in the effective context for Group $g$; that is, these are the stimuli that would have been judged 1 and 7, respectively.[10]

For groups with lowest to highest incomes, respectively, least-squares estimated minima were \$26.94, \$35.31, \$39.89, and \$39.35 thousand; estimated maxima were \$58.28, \$56.77, \$67.32, and \$69.15 thousand, respectively. The estimated shape parameters for the Beta distribution were $(\alpha, \beta) = (5.99, 3.72), (6.91, 3.67), (4.20, 5.53)$, and $(4.25, 5.18)$, respectively. These are single-peaked distributions that shift to the right as income increases, as one might expect. Summed over all four curves, the sum of squared deviations was 0.124. Figure 3 shows that the predictions (curves) provide a reasonable approximation to average judgments (markers).

In this curve fitting, the estimated "effective" minima and maxima are now estimated parameters (instead of the actual minima and maxima controlled by the experimenter), and so they can fall outside the actual range of the stimuli used in the experiment. Their estimation depends crucially on the assumed beta distribution used to extrapolate to their values. Therefore, although this fitting method gives a good reproduction to these data and we think that these estimated parameters could be used to predict new results on the same range for the same income groups, we suggest caution in extrapolating its predictions outside the range of salaries actually used in the study. Nevertheless, we think it might be informative to compare estimates of the effective context using this method against other procedures for eliciting participants' contexts directly.

[10]The beta distribution is a fairly flexible distribution on a fixed interval that can take on a variety of shapes, depending on just two shape parameters, $\alpha$ and $\beta$.

## 2.4 Discussion of Experiment 1

The data for the main group show that ratings as functions of salary can cross twice for contexts that differ in their frequency distributions. The results show that people do not simply evaluate salaries relative to the mean, as one might expect from the perspective of AL theory. Nor do the data agree with the theory that judgments are a fixed function of the mean and standard deviation or mean and endpoints of the distribution, as in CR, ID, or EN theories. Instead, the double crossover shows that ratings reflect the ranking of the stimuli as predicted by the frequency principle of RF theory (Figure 1) and DbS.

The data for the main group are reasonably compatible with previous judgments of the magnitude of numbers with similar cubic distributions (Birnbaum, 1974), which were also are well-described by RF theory. However, the overall data also show three systematic differences between the data and the predictions of simplified RF model: First, many people showed patterns that would have been unusual in psychophysical studies. Some of these unusual patterns might be compatible with RF theory, assuming that people bring individual, residual contexts for salaries into the lab, which for these participants overwhelm the experimental manipulations. However, some people may also judge satisfaction as related to concepts of fairness and equity.

Second, whereas predictions in Figure 1 range from 1 to 7, mean judgments in Figure 2 range from 1.6 to 5.6. Besides regression one might expect with error-filled empirical data, the reduced range of responses is consistent with the theory that people in the main group are reserving more extreme responses for more extreme salaries, presumably experienced in their prior contexts. Consistent with this idea, those who reported higher full time incomes are inferred by RF theory to have higher endpoints in their effective contexts.

Third, the ratings in Figures 2 and 3 show a positively accelerated trend relative to objective salary levels. If it is assumed that the context-free utility function for salaries, $u(x)$, is negatively accelerated (as is often supposed) or even linear, RF theory would interpret this

positive acceleration to imply that the salaries used in the present study fell in the left tail of the effective contexts for many of the participants. Indeed, the majority of participants who reported working full time reported higher incomes than $52, the highest salary used in this study.

In sum, Experiment 1 shows that manipulation of the frequency has significant effects that refute the implications of AL, CR, ID, and EN theories. Those theories assume that ranking has no effect on the judgments beyond what is inferred from mean, standard deviations, or endpoints. Experiment 1 also shows the importance of individual differences in prior contexts that participants bring to the study. In Experiment 2, we manipulate the endpoints to evaluate and compare the theories' implications for this manipulation.

# 3 Experiment 2: Range Effects

Without additional modifications (such as those in ID theory), DbS implies no effect of the endpoints, holding rank constant. RF theory in contrast, implies that each endpoint affects judgments of all salaries.

Figure 4 shows predictions of the simplified RF model for the design of Experiment 2, which used 4 between-subjects contexts in which both lower and upper endpoints were varied. The lowest salary was either $26K or $40K and the highest salary was either $52K or $70K. There were 13 values ranging from $42K to $50K that were common to all four range contexts and which held the same ranks in all contexts.

The simplified RF predictions in Figure 4 ignore background and residual contexts, assume that $s = u(x) = x$, $w = 0.5$, and that the rating scale is linear. Circles (filled or unfilled) connected by dashed lines show predicted judgments for the common values when the maximum salary was $52K; Squares connected by solid lines show predicted judgments for maximum of $70K. Unfilled and filled markers indicate predictions when minimum salary

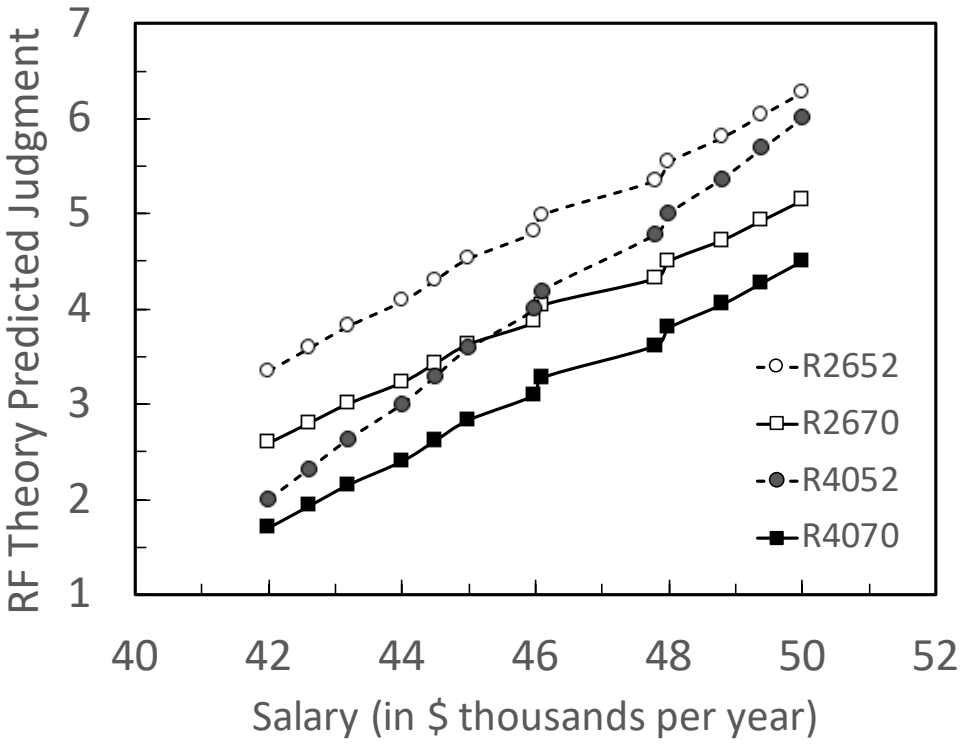


Figure 4: Predictions of simplified RF theory for manipulation of the lower and upper endpoints, for the 13 salaries common to all four range conditions. Conditions are labeled by the lower and upper endpoints of their ranges; for example, R2670 had lowest and highest salaries of \$26K and \$70K, respectively.

was \$26K or \$40K, respectively.

The two curves in Figure 4 with filled markers show the effect of varying the upper endpoint, holding the minimum salary at \$40K. The two curves with unfilled markers show the predicted effect of the upper endpoint when minimal salary was \$26K. Note that these pairs of curves diverge to the right, meaning that the predicted effect of changing the upper endpoint (the vertical gap between the curves) will be greater for salaries above the mean than for those below the mean. This implication of RF theory is distinct from the prediction

of EN theory, which implies that there should be no effect of the upper endpoint for judgments of salaries below the mean.

The two dashed curves connecting circles show the effect of varying the lower endpoint, holding the upper endpoint fixed at \$52K. The two solid curves connecting squares show the same effect when maximum salary is \$70K. Note that these pairs of curves converge to the right, meaning that the predicted effect of changing the lower endpoint is greater for salaries below the mean than above.

Although the predictions in Figure 4 are for a simplified RF model in which $s = u(x) = x$, Birnbaum (1974, p. 95) showed that for any $u(x)$ function, ratings of stimuli holding the same ranks in contexts differing in endpoints should be linearly related across contexts. Birnbaum (1974) noted that previous tests of the range principle in RF theory had not held the ranks constant; as far as we are aware, this study is the first pure test of this linearity implication of RF theory when endpoints are varied with ranks held fixed.

In contrast with RF theory, EN theory implies that ratings will not be linearly related between contexts over the entire range, because the upper endpoint should affect only judgments above the mean and the lower endpoint should affect only judgments below the mean. For Condition R2652 in Figure 4, when the endpoints are \$26K and \$52K (Context 1), assuming $s = u(x) = x$, the mean is \$44.72K, so Equation 4 implies, for $x <$ \$44.72, $e_1 = (x - 44.72)/(44.72 - 26)$ and for $x >$ \$44.72, $e_1 = (x - 44.72)/(52 - 44.72)$. Context 2 (R4070) has endpoints of \$40 and \$70; in this context, the mean is \$47.57, so Equation 4 implies for $x <$ \$47.57, $e_2 = (x - 47.57)/(47.57 - 40)$ and for $x >$ \$47.57, $e_2 = (x - 47.57)/(70 - 47.57)$. It follows that for $x <$ \$44.72, $e_2 = 2.47e_1 - 0.38$ and for $x >$ \$47.57, $e_2 = 0.32e_1 - 0.13$. Note that the slopes for these two sub-segments of the range differ by a factor of almost eight to one, so EN implies that judgments in Context 2 (R4070) should be concave downwards relative to Context 1 (R2652).

The theories of CR and ID allow slopes and heights of the curves to depend on the means

and standard deviations, which are affected by manipulation of the endpoints in this design. These theories can thus accommodate, at least qualitatively, effects of these manipulations. Assuming $s = u(x) = x$ and using objective means and standard deviations, the predictions of CR and ID are similar to those of RF in Figure 4, except these theories imply that the curve for R4052, with $\mu = 46.01$ and $\sigma = 3.67$, should cross all three of the other curves and have the lowest response for the three salaries below \$44K and the highest response for the three salaries above \$48. In addition, the ID theory implies that judgments in R2670 should be nonlinearly related to those in R4052, with an S-shape induced by the cumulative normal applied across two differing ranges.

The theory of DbS (Stewart, et al., 2006; Boyce, Brown, & Moore, 2010) implies endpoints of the stimuli in the experiment should have no effect on judgments of those stimuli that maintain the same ranks. AL theory allows main effects due to changes in the means but it implies no interactive effects of the endpoints, so the slopes cannot change and the curves cannot cross.

## 3.1 Method

The task, materials, instructions, and rating scale were similar to those of Experiment 1: Participants rated how satisfied they would be with a salary, given a list of 19 people who were doing the same job and evaluated as equally experienced, qualified and productive. This study was not preregistered. Complete instructions and materials are available via the following URL: https://konstanzworkshop.neocities.org/CSUF22/index.htm. From this page, participants clicked a link that randomly assigned them to one of four conditions, including for example, the condition at the following link: https://konstanzworkshop.neocities.org/Salary22/salary_r2652.htm

### 3.1.1 Design

The design was a between-subjects, $2 \times 2 \times 13$, Lowest Salary by Highest Salary by Common Salary, factorial design, with subjects nested in the $2 \times 2 = 4$ Range conditions of Lowest by Highest Salary. The 2 levels of Lowest Salary were \$26K or \$40K; the 2 levels of Highest Salary were \$52K or \$70K.

There were 13 Salaries common to all four Range conditions which held the same ranks in all conditions: \$42K, \$42.6K, \$43.2K, \$44K, \$44.5K, \$45K, \$46K, \$46.1K, \$47.8K, \$48K, \$48.8K, \$49.4K, and \$50K.

There were six additional contextual stimuli to establish ranges that differed for each condition added to the 13 common levels, making a total of 19 salaries per condition. The four Range conditions are named by the lowest and highest salaries:

Condition R2652 had contextual levels of \$26K, \$32K, \$40K, ..., \$50.5K, \$51.7K, and \$52K.

Condition R2670: \$26K, \$32K, \$40K, ..., \$52K, \$62K, or \$70K.

Condition R4052: \$40K, \$41K, \$41.5K, ..., \$50.5K, \$51.7K, and \$52K.

Condition R4070: \$40K, \$41K, \$41.5K, ..., \$52K, \$62K, and \$70K. Note that the 13 common salaries, indicated by "...," are nested in each range and held the same ranks in all conditions.

### 3.1.2 Procedure

Participants were instructed to imagine themselves as a company employee. They read a list of salaries of 19 people doing the same work who are equally experienced, qualified and productive. Participants were then instructed to rate how dissatisfied or satisfied they would be if they received each of those salaries after learning what others are paid for doing the same work. Ratings were made on a 7-points scale from 1 = Not at all happy to 7 = Extremely happy. The task consisted of a warm-up of 4 trials that included the condition's

endpoints, followed by the experimental block of 19 trials.

As in Experiment 1, participants were requested to indicate gender, age, level of education, nationality, hours per week worked for pay, and yearly income in thousands of USD. A box was provided for comments.

### 3.1.3 Participants

Participants were 561 students at California State University, Fullerton, who served as one option toward an assignment in lower division psychology and 46 who had been recruited from Reddit, as in Experiment 1. There were 107 participants whose data patterns were unusual (see Experiment 1), including 20 of 46 recruited from Reddit. As in Experiment 1, the unusual data were analyzed separately, leaving 500 in the main group. Of the 500 in the main group 137, 126, 118, and 119 were in conditions R2652, R2670, R4052, and R4070, respectively. The median age was 19 years; 154 identified as male (31%), 337 female, and 9 did not indicate gender. Only 30 of the 561 students (5%) reported working full time.

Upon acceptance of this paper, data will be available in anonymous form from the archive at the following url: http://psych.fullerton.edu/mbirnbaum/archive.htm

## 3.2 Results and Discussion of Experiment 2

Figure 5 shows mean judgments of satisfaction for the 13 salaries common common to all conditions, with a separate curve for each condition, for the main group of participants. Condition R2652 is shown as unfilled circles connected by dashed lines. This condition has the lowest minimum and maximum salaries ($26K and $52K), and as predicted by RF theory, it has the highest judgments. The lowest curve (filled squares) is for condition R4070, which has the highest minimum and maximum salaries. The condition with the smallest range (R4052, with filled circles connected by dashed curves) has the steepest slope, and the condition with the greatest range (R2670, shown as unfilled squares connected by solid line)

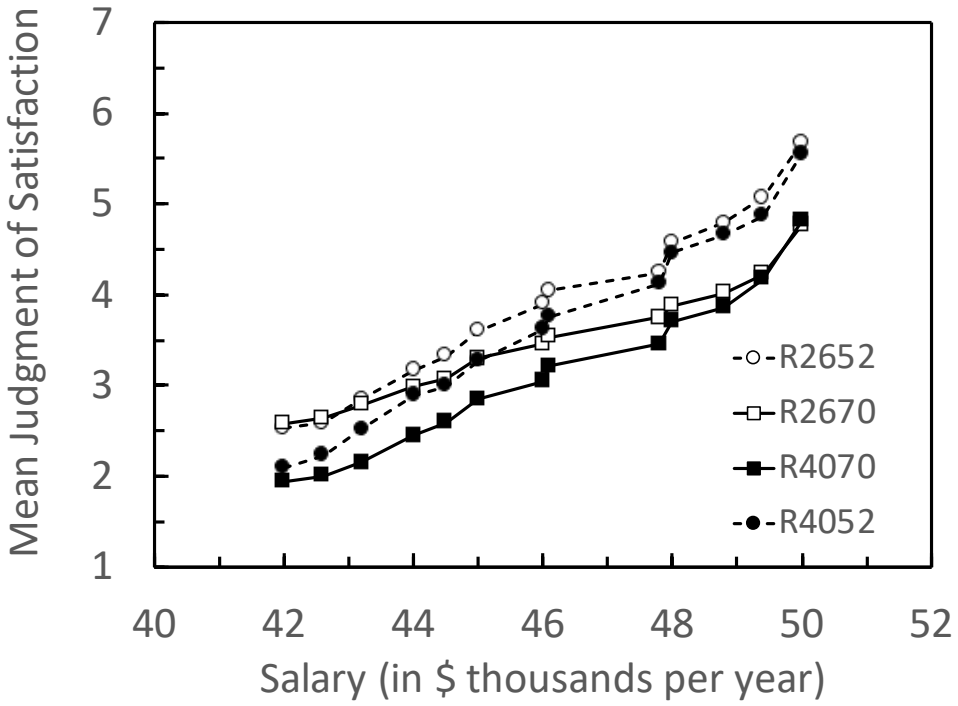


Figure 5: Mean judgments as a function of salary with a separate curve for each condition of lower and upper endpoints, for the 13 salaries common to all conditions, which held the same ranks in all contexts.

has the smallest slope. The relative heights and slopes of the curves are compatible with the predictions of the simplified RF theory in Figure 4. The standard errors of the means in Figure 5 range from 0.09 to 0.12, so the markers in Figure 5 are slightly larger than a standard error in each case.

The differences between predictions in Figure 4 and obtained mean judgments in Figure 5 are similar to differences observed in Experiment 1 between Figures 1 and 2: First, all curves show lower slopes and smaller vertical gaps between the curves than do the predictions. Second, there is a positive acceleration to the right, as found in Experiment 1. Nevertheless,

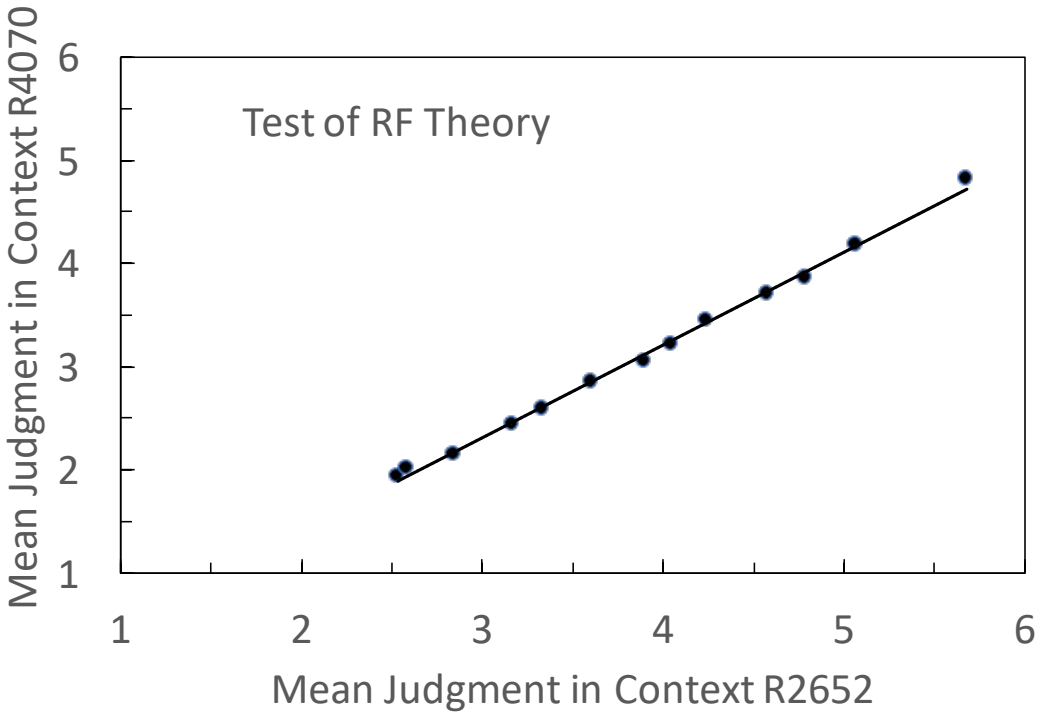


Figure 6: Mean judgments in the context with endpoints of \$40K and \$70K are plotted against mean judgments in the context with endpoints of \$26K and \$52K, with a separate marker for each stimulus common to the two contexts. RF theory implies that the curve should be linear, whereas EN theory implies that the curve should be concave downwards, with a slope for the lower five points more than seven times as steep as the slope for the upper five points.

652 the major trends agree with those predicted by RF theory.

653 Although EN theory allows that endpoints affect the judgments, it does not correctly
654 describe these results. According to that theory, each endpoint should only affect judgments
655 of salaries that are on the same side of the mean as the endpoint. However, the two curves
656 in Figure 5 with filled symbols (R4052 and R4070, which have the same lower endpoint,
657 \$40K, and different upper endpoints) show that the entire curve for R4052 is above that of

R4070, even for stimuli below the mean, and that the gap between the curves increases to the right, as in Figure 4. Similarly, the two curves with unfilled symbols (R2652 and R2670, with lower endpoint of $26K and different upper endpoints) also show similar divergence to the right without any discontinuity across the mean.

The two curves with circles (R2652 and R4052, with upper endpoint of $52K) converge to the right and show no change as they cross the mean, as do the two curves with squares (R2670 and R4070), which share upper endpoint of $70K. Thus, the effect of an endpoint does not seem to be limited to stimuli on the same side of the mean, as implied by EN theory, but instead each endpoint affects the entire curve, as implied by RF theory.

Figure 6 plots the judgments from Context R4070 against those from R2652 with a separate marker for each of the 13 common stimuli. RF theory implies that judgments of the same stimuli holding the same ranks in contexts differing in endpoints should be linearly related to each other (Birnbaum, 1974), whereas EN theory implies that the judgments should not be linear across the whole range. The line in Figure 6 is the least-squares regression line, showing that the mean judgments (markers) fall close to linearity. EN theory implies that this curve should have been the greatest departure from linearity and should have been concave downwards, with the lowest five points having a slope more than seven times greater than the slope for the highest five points. Similar graphs (not shown) for the data between other pairs of contexts also appeared linear, compatible with RF theory, showing no evidence of nonlinearity implied by EN theory.

Because endpoints affect the standard deviation of a distribution, the changes in slope in Figure 5 are qualitatively compatible with CR and ID theories. However, the curve for R4052 in Figure 5 does not cross the other three curves, contrary to predictions of these theories if objective values of the means and standard deviations are used to calculate predictions. This curve (R4052) also showed no evidence of the slight S-shape predicted by ID theory when plotted against R2670. These quantitative discrepancies of CR and ID might be remedied by

Table 2: Compatibility of the Results with Theories of Contextual Effects

| Abbrev | Theory | Double Cross | Endpoints |
|---|---|---|---|
| AL | Adaptation Level | No | No |
| CR | Correlation-Regression | No | Yes |
| ID | Inferred Distribution | No | Yes |
| EN | Ensemble | No | partial |
| DbS | Decision by Sampling | Yes | No |
| RF | Range-Frequency | Yes | Yes |

fitting other functions for $u(x)$ and by allowing subjective evaluations of means and standard deviations.

Because the ranks of the stimuli are the same in all four contexts, DbS does not provide any explicit explanation for the changes in slope in Figure 5 due to changes in the endpoints. The changes in slope (including crossover of R4052 and R2670) in Figure 5 are not consistent with AL theory, which implies that the curves should have been parallel.

# 4 Discussion

Experiment 1 found that judgments of salary satisfaction can show a double crossover when the stimuli are spaced to form cubic distributions. This finding shows that participants respond to more than just the mean, standard deviation, and endpoints of the distribution but instead show that differences in response are proportional to differences in rank. Experiment 2 found that that ratings of salary satisfaction do not depend entirely on ranks but also depend on the minimum and maximum salaries in the experimental context.

Table 2 summarizes the implications of the results for the six theories of contextual effects considered here. Each "Yes" or "No" in the column under "Double Cross" indicates a theory

that can or cannot account for the double crossover observed in Experiment 1 (Figure 2). Only DbS and RF theories account for this result from Experiment 1.

Similarly, theories that can or cannot account for effects of endpoints in Experiment 2 (Figure 5) are noted with "Yes" or "No" in the column labeled "Endpoints". The term "partial" for EN in this column indicates that although EN implies effects of endpoints, it is only partially consistent with the results because it implies that the endpoint affects only judgments of salaries on the same side of the mean, whereas the data show that each endpoint affects judgments of all salaries. There was no evidence of discontinuities or changes in slope at the mean implied by EN theory. The results have the main properties of the predictions of the simplified RF theory, used to calculate predictions in Figures 1 and 4. The only theory in Table 2 qualitatively compatible with the results of both experiments is RF theory.

## 4.1 Estimating the Effective Context

If RF theory is assumed, and if we can assume the shape of the $u(x)$ function or estimate it from an independent method such as judgments of "differences" (Birnbaum, 1982; Rose & Birnbaum, 1975), RF theory can be used to estimate the effective context using the method of Equation 9.[11] The effective context is assumed to reflect a combination of experimental, background, and residual (or prior) contexts. Because the Web (Reddit) recruits in Experiment 1 had a wide range of income levels, we were able to estimate the effective contexts in for groups differing in income. Those who have higher full-time incomes rate salaries lower than do those with lower incomes. It was possible to fit the mean judgments by groups of people with different incomes (Figure 3) using RF theory with the assumption that effective contexts can be approximated as beta distributions with different endpoints and shape

[11] Quotation marks are used to distinguish instructions to judge "differences" and "ratios" or judgments obtained with such instructions from mathematical models used to represent such data or theoretical statements about mathematical differences and ratios. For example, when people are instructed to judge "ratios" of subjective magnitude, they might actually evaluate subjective differences in sensation.

parameters for groups who earn different incomes.

## 4.2 Representing Contextual Distributions

In DbS and in ID theories, memory and inference processes are assumed to create what we call here the effective context. In DbS, it is assumed that people sample from instances stored in memory to construct a ranking that determines the evaluation of each stimulus, and in ID, a ranking is induced by inference of a normal distribution from memories of the mean and endpoints of the context. The problems for these two models is that each of them makes a simplifying assumption that is contradicted by the data of one of the experiments. Instead of assuming that people retain only a ranking (that does not reflect the endpoints) or infer a normal distribution from mean and endpoints (which oversimplifies the ranking), RF theory holds that the effective context retains both a metric scale of the stimuli relative to the endpoints and a relative frequency representation.

The idea of EN theory is that people represent distributions by an ensemble of estimates of statistics of the distribution and that they do not retain details about its shape that are not preserved by those summary statistics. This EN theory is based on findings that people can estimate the mean and endpoints of values that they have experienced. However, because people can estimate certain statistics of a distribution does not rule out the idea that they retain other information about the distribution that is not retained in those statistics.

Mellers, Richards, and Birnbaum (1992) asked people to estimate probability distributions of how much they would like people described by adjectives. Similarly, Ronayne and Brown (2017) elicited distributions of options available in a market for multiattribute goods. From these studies and others, it seems that people are capable of reporting distribution information directly, and it does not appear necessary to assume that people only retain information about a limited set of statistics. It would be interesting to compare estimated effective distributions (e.g., using the techniques of Equation 9) with those that might be

elicited by such direct methods.

## 4.3 Combining Distributions

How do prior contexts and experimental contexts combine to produce the effective context? In Mellers, et al. (1992), participants were asked to imagine hypothetical people described by single adjectives or by adjective combinations and to estimate the probabilities that the people would have various degrees of likeableness. The question addressed was, how does the distribution of a combination of adjectives relate to the separate distributions of individual adjectives that were combined to describe a person? Three different models of how distributions combine were evaluated in that study.

A similar technique to that in Mellers, et al. (1992) might be employed to investigate models of how experimental and prior contexts combine to produce the effective context. Participants in different randomly assigned conditions might be asked to estimate salaries that would be judged to be rated as 1, 2, 3, etc. either before or after being exposed to experimental contexts such as used in this study. One might estimate the effective context using a uniform experimental distribution, for example, followed by presentation of a skewed experimental distribution, and measure the effective context again, in order to ascertain how the effective context responds to a changing distribution of stimuli.

## 4.4 Using RF theory to estimate psychophysical function

In Birnbaum's (1974) version of RF theory, the range function of RF theory is interpreted as a context-free psychophysical function. By manipulating the frequency distribution while holding endpoints fixed, one can estimate this psychophysical function from the data and test if this estimate is indeed independent of context.

The estimated psychophysical function for numbers from Birnbaum (1974) agreed with

estimates from the subtractive theory of judgments of "ratios" and "differences" of numbers, presented as pairs in a factorial design (Rose & Birnbaum, 1975), who fit the model,

$$D(x, y) = J[u(x) - u(y)] \tag{10}$$

where $D(x, y)$ is the predicted judgment of "difference" between stimuli $x$ and $y$; $u(x)$ is the psychophysical function of $x$; $J$ is a strictly increasing monotonic function that can be estimated from the data to reproduce the rank order of judgments of "differences." If $x$ and $y$ are spaced properly, one can define a scale in which the intervals of $u(x)$ are constrained; in the limit, $u(x)$ forms an interval scale (Krantz, Luce, Suppes, & Tversky, 1971).

The function, $u(x)$, estimated from the subtractive model of "differences" (Equation 9) in Rose and Birnbaum (1975) was found to be linearly related to the other estimated $u(x)$ function, estimated from RF theory applied to judgments in Birnbaum (1974). These two estimates were also in fair agreement with psychophysical functions estimated using other techniques (Rule & Curtis, 1973; Schneider, Parker, Ostrosky, Stein, & Kanow (1974).

In the present studies, we did not estimate $u(x)$ from the data; instead, we assumed for simplicity that $u(x) = x$ for the (relatively small) range of salaries used here. Given the experimental designs used here, and given the large individual differences in prior contexts (as evidenced in Figure 3), we did not consider our study to be sufficient to isolate and identify the psychophysical function separate from the effective context. For that purpose, it would have been useful to have obtained an independent estimate of the $u(x)$ function for the same individuals by another technique such as "difference" judgments.

## 4.5 DbS and Psychophysics

A thesis of DbS (Stewart, et al., 2006) is that people do not represent subjective values of stimuli on a ratio or interval scale, but only on an ordinal scale in which stimuli can

be ranked but not evaluated for higher metric properties such as ratios or differences. A problem for this thesis is that it fails to account for findings that are consistent with the use of two operations on a common scale. The agreement between psychophysical scales obtained from matrices of data involving different tasks and models using only the ordinal information in the data is called "scale convergence" in this literature; and there is a body of evidence showing scale convergence (Birnbaum, 1982; Birnbaum & Sutton, 1992).

In particular, judgments of "ratios of differences" and "differences of differences" show two different, appropriately interrelated rank orders that agree with algebraic ratios and differences on a common scale of intervals (Birnbaum, 1982; Birnbaum, Anderson, & Hynan, 1989; Hagerty & Birnbaum, 1978; Veit, 1978). These studies observed the appropriate ordinal constraints indicating that it is possible to construct a ratio scale of intervals. In other words, evidence is consistent with the proposition that people can compare magnitudes by a metric process rather than merely an ordinal one.

One might theorize that when comparing stimuli, people sample a distribution of stimuli, rank them, and then can compute the differences in ranks between them and can judge both ratios and differences of intervals in rank. But this complex interpretation seems to contradict the original assumption that people can only rank stimuli and do not judge quantitative relationships among them. It could be tested by randomly assigning babies to different environments in which stimuli are presented with different frequency distributions, which should result in different estimated psychophysical functions from "difference" judgments. It seems unlikely that this long-term developmental study will be done in the near future, but short-term studies have found evidence consistent with the proposition that psychophysical functions estimated from judgments of "differences" may be independent of context, as described in the next section.

## 4.6 Loci of Contextual Effects

Birnbaum (1982) theorized that contextual effects might operate at the level of the psychophysical function or at the level of the judgment function– the transformation between integrated impressions and overt responses–or both. Mellers and Birnbaum (1982) tested these theories with judgments of single stimuli presented in different distributions and with judgments of "differences" between pairs of stimuli spaced in the same contexts. They found that judgments of "differences" between pairs of stimuli are not monotonically related to differences in judgment between the stimuli. They concluded that when stimuli are presented for single judgments, responses depend on contexts produced by spacing of the stimuli as would be expected from RF theory; however, when the same stimuli in the same spacings are presented in pairs for "difference" judgments, the rank order of "difference" judgments appears to be independent of stimulus spacing. Thus, contextual effects in these studies could be attributed to the judgment function that relates responses to subjective values. Mellers and Birnbaum (1982) thus concluded that when comparing stimuli within the same modality, contextual effects operate at the level of the response function, and the estimated psychophysical functions were apparently independent of how the stimuli were spaced. The rank order of "difference" judgments did not differ systematically between contexts, even though the rank order of response differences did differ between contexts. See also Mellers and Birnbaum (1983).

However, Mellers and Birnbaum (1982) also tested cross modality comparisons in which stimuli from two different modalities were compared; in this case, they concluded that contextual effects operate before stimuli are compared. They theorized that in order to compare the darkness of a dot pattern with the size of a circle, for example, people compare darkness to other levels of darkness and compare size of the circle to other circles, and then compare the two relative positions to each other.

## 4.7 Happiness

According to AL theory, one cannot escape a "hedonic treadmill" because the sum of deviations about the mean is zero (Parducci, 1968, 1995; Edwards, 2018). If one has a good experience, it raises the mean, which lowers judgments of experiences that were once pleasurable. Mark Twain (1898) wrote, "Every man is a suffering-machine and a happiness-machine combined. The two functions work together harmoniously, with a fine and delicate precision, on the give-and-take principle. For every happiness turned out in the one department the other stands ready to modify it with a sorrow or a pain ... Sometimes for an hour's happiness a man's machinery makes him pay years of misery. "

In contrast with the hedonic treadmill implied by AL, RF theory (Parducci, 1968, 1995; 2011) provides a solution to escape the treadmill, because in RF theory, the neutral point is between the midpoint (range) and the median (frequency). According to RF theory, "Happiness is a negatively skewed distribution," because in such a distribution, most experiences will fall above this neutral value (Wedell & Parducci, 1988). Consistent with this theory, Parducci (2011) and Tripp and Brown (2016) found that the average rating of satisfaction with payments in a negatively skewed distribution was indeed higher than the mean rating of satisfaction in a positively skewed distribution with the same mean payment.

A counter-intuitive implication of RF theory is that if one has an opportunity for a rare and wonderful experience that can be enjoyed but once in life, one should avoid it, lest it extend one's range upwards and thereby lower the hedonic experiences of everyday life. Instead, one should strive for a life in which the best, if modest, experiences are available consistently and the worst experiences, which are unavoidable, occur only rarely (Parducci, 1968, 1995).

According to our results, people would be happier with lower salaries if they are paid more than their co-workers compared to a situation in which they would receive higher salaries but receive less than others doing the same work. These conclusions are based on judgments

obtained between-subjects who experience different contexts. What would a person do when asked to choose between these two job offers: (1) a higher salary in the context of co-workers who are paid even more versus (2) a lower salary that is the highest among the co-workers? This choice problem converts the issue from comparing people who are in different isolated contexts to one in which both contexts are available within the same person.

## 4.8 Within and Between-Ss contexts

It has been shown that the results of between-subjects studies do not always agree with findings of within-subjects studies. For example, when people are randomly assigned to conditions, the number 9 can be judged to be a "bigger" number than 221 when they are rated by different groups of people but not when both numbers are judged by the same people (Birnbaum, 1982, 1999). There are other situations in which both between- and within-subjects experiments give similar results (Birnbaum, 2008). It seems of interest to determine if salary satisfaction is an area where people can imagine how they would feel in different contexts to make reasonable choices for their own happiness.

In many studies and in Experiments 1 and 2, context has been manipulated between subjects to avoid the possibility that contexts might combine and their effects thereby cancel. Nevertheless, this salary satisfaction paradigm is one in which it seems that participants can imagine different scenarios and evaluate how happy they would be in those scenarios to receive hypothetical salaries in different distributions. We are currently evaluating simple cases within-Ss.

# References

Birnbaum, M. H. (1974). Using contextual effects to derive psychophysical scales. *Perception & Psychophysics, 15*(1), 89-96. https://doi.org/10.3758/bf03205834

Birnbaum, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 401-485). Hillsdale, N.J.: Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203780947

Birnbaum, M. H. (1983). Perceived equity of salary policies. *Journal of Applied Psychology, 68*(1), 49-59. https://doi.org/10.1037/0021-9010.68.1.49

Birnbaum, M. H. (1999). How to show that $9 > 221$: Collect judgments in a between-subjects design. *Psychological Methods, 4*(3), 243-249. https://doi.org/10.1037/1082-989x.4.3.243

Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review, 115*, 463-501. https://doi.org/10.1037/0033-295x.115.2.463

Birnbaum, M. H., Anderson, C. J., & Hynan, L. G. (1989). Two operations for "ratios" and "differences" of distances on the mental map. *Journal of Experimental Psychology: Human Perception and Performance, 15*(4), 785–796. https://doi.org/10.1037/0096-1523.15.4.785

Birnbaum, M. H., & Sutton, S. E. (1992). Scale convergence and utility measurement. *Organizational Behavior and Human Decision Processes, 52*(2), 183-215. https://doi.org/10.1016/0749-5978(92)90035-6

Boyce, C.J., Brown, G.D.A., & Moore, S.C. (2010). Money and happiness: Rank of income, not income, affects life satisfaction. *Psychological Science, 21*(4), 471-475. https://doi.org/10.1177/0956797610362671

Brown, G.D.A., Gardner, J., Oswald, A.J., & Qian, J. (2008). Does wage rank affect employees' well-being? *Industrial Relations, 47*(3), 355-389. https://doi.org/10.1111/j.1468-232X.2008.00525.x

Card, D., Mas, A., Moretti, E., & Saez, E. (2012). Inequality at work: The effect of peer salaries on job satisfaction. *The American Economic Review, 102*(6), 2981-3003. https://www.jstor.org/stable/41724678

Edwards, J. (2018). Harry Helson's adaptation-level theory, happiness treadmills, and behavioral economics. *Journal of the History of Economic Thought, 40*(1), 1 - 22. https://doi.org/10.1017/S1053837216001140

Hagerty, M., & Birnbaum, M.H. (1978). Nonmetric tests of ratio vs. subtractive theories of stimulus comparison. *Perception & Psychophysics, 24*, 121–129 (1978). https://doi.org/10.3758/BF03199538

Hayes, W. M., & Wedell, D. H. (2022). Reinforcement learning in and out of context: The effects of attentional focus. *Journal of Experimental Psychology: Learning, Memory, and Cognition, xxx.* Advance online publication. https://doi.org/10.1037/xlm0001145

Hayes, W. M., & Wedell, D. H. (2023). Testing models of context-dependent outcome encoding in reinforcement learning. *Cognition, 230,*

Helson, H. (1947). Adaptation-Level as frame of reference for prediction of psychophysical data. *American Journal of Psychology, 60*(1), 1-29. https://doi.org/10.2307/1417326

Helson, H. (1964). *Adaptation-Level theory.* Oxford, England: Harper & Row. https://psycnet.apa.org/record/1964-35039-000

Johnson, D. M., & Mullally, C. R. (1969). Correlation-and-regression model for category judgments. *Psychological Review, 76*(2), 205–215. https://doi.org/10.1037/h0027227

Krantz, D. H. , Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement.* New York: Academic Press. https://doi.org/10.1016/B978-0-12-425401-5.50012-X

Mellers, B. A. (1982). Equity judgment: A revision of Aristotelian views. *Journal of Experimental Psychology: General, 111*(2), 242–270. https://doi.org/10.1037/0096-3445.111.2.242

Mellers, B. A. (1986). "Fair" allocations of salaries and taxes. *Journal of Experimental Psychology: Human Perception and Performance, 12*(1), 80–91. https://doi.org/10.1037/0096-1523.12.1.80

Mellers, B. A., & Birnbaum, M. H. (1982). Loci of contextual effects in judgment.

*Journal of Experimental Psychology: Human Perception and Performance, 8*(4), 582- 601. https://doi.org/10.1037/0096-1523.8.4.582

Mellers, B. A., & Birnbaum, M. H. (1983). Contextual effects in social judgment. *Journal of Experimental Social Psychology, 19*(2), 157–171. https://doi.org/10.1016/0022-1031(83)90035-5

Mellers, B. A., Richards, V., & Birnbaum, M. H. (1992). Distributional theories of impression formation. *Organizational Behavior and Human Decision Processes, 51(*3), 313-343. https://doi.org/10.1016/0749-5978(92)90016-z

Otto, A. R., & Vassena, E. (2021). It's all relative: Reward-induced cognitive control modulation depends on context. *Journal of Experimental Psychology: General, 150*(2), 306–313. https://doi.org/10.1037/xge0000842

Parducci, A. (1965). Category judgment: A range-frequency model.*Psychological Review, 72*(6), 407–418. https://doi.org/10.1037/h0022602

Parducci, A. (1968). The relativism of absolute judgments. *Scientific American, 219(*6), 84–90. https://doi.org/10.1038/scientificamerican1268-84

Parducci, A. (1982). Category ratings: Still more contextual effects! In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 89-105). Hillsdale, N.J.: Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203780947

Parducci, A. (1995). *Happiness, Pleasure, and Judgment: The Contextual Theory and its Applications.* Mahwah, NJ: Erlbaum.

Parducci, A. (2011). Utility versus pleasure: the grand paradox. *Frontiers in Psychology, 15* https://doi.org/10.3389/fpsyg.2011.00296

Parducci, A., & Perrett, L. F. (1971). Category rating scales: Effects of relative spacing and frequency of stimulus values. *Journal of Experimental Psychology, 89*(2), 427–452. https://doi.org/10.1037/h0031258

Putnam-Farr, E., & Morewedge, C. K. (2021). Which social comparisons influence hap-

piness with unequal pay? *Journal of Experimental Psychology: General, 150*(3), 570-582. https://doi.org/10.1037/xge0000965

Rethlingshafer, D., & Hinckley, E. D. (1963). Influence of judges' characteristics upon the adaptation-level. *American Journal of Psychology, 76*(1), 116-119. https://doi.org/10.2307/1420007

Ronayne, D., & Brown, G. D. A. (2017). Multi-attribute decision by sampling: An account of the attraction, compromise and similarity effects. *Journal of Mathematical Psychology, 81*, 11–27. https://doi.org/10.1016/j.jmp.2017.08.005

Rose, B. J., & Birnbaum, M. H. (1975). Judgments of differences and ratios of numerals. *Perception & Psychophysics, 18*(3), 194–200. https://doi.org/10.3758/BF03205967

Rule, S. J., & Curtis, D. W. (1973). Conjoint scaling of subjective number and weight. *Journal of Experimental Psychology, 97*(3), 305–309. https://doi.org/10.1037/h0034095

Schneider, B., Parker, S., Ostrosky, D., Stein, D., & Kanow, G. (1974). A scale for the psychological magnitude of number. *Perception & Psychophysics, 16*, 43-46. https://doi.org/10.3758/BF03203247

Slovic, P. (1995). The construction of preference. *American Psychologist, 50*(5), 364–371. https://doi.org/10.1037/0003-066X.50.5.364

Stevenson, M. K. (1992). The impact of temporal context and risk on the judged value of future outcomes. *Organizational Behavior and Human Decision Processes 52*(3),455-491. http://DOI:10.1016/0749-5978(92)90029-7

Stevenson, M. K. (2019). Temporal discounting and context: Discounting weights for gains and losses presented in isolation and in combination. *Decision, 6*(3), 261–276. https://doi.org/10.1037/dec0000099

Stewart, N., Chater, N., & Brown, G.D.A. (2006). Decision by sampling. *Cognitive Psychology, 53*, 1-26. https://doi.org/10.1016/j.cogpsych.2005.10.003

Tripp, J., & Brown, G. D. A. (2016). Being paid relatively well most of the time:

Negatively skewed payments are more satisfying. *Memory & Cognition, 44*(6), 966–973. https://doi.org/10.3758/s13421-016-0604-0

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. `Science, 185(4157)`, 1124–1131. http://www.jstor.org/stable/1738360

Twain, M. (1898). The mysterious stranger. *The project Gutenberg eBook of The Mysterious Stranger, and other stories, by Mark Twain.* https://gutenberg.org/cache/epub/3186/pg3186-images.html

Veit, C. T. (1978). Ratio and subtractive processes in psychological judgment. *Journal of Experimental Psychology: General, 107*(1), 81-107.https://doi.org/10.1037/0096-3445.107.1.81

Wedell, D.H., Hayes, W.M. & Kim, J. (2020). Context effects on reproduced magnitudes from short-term and long-term memory. *Attention, Perception & Psychophysics, 82*, 1710–1726. https://doi.org/10.3758/s13414-019-01932-z

Wedell, D. H., & Parducci, A. (1988). The category effect in social judgment: Experimental ratings of happiness. *Journal of Personality and Social Psychology, 55*(3), 341–356. https://doi.org/10.1037/0022-3514.55.3.341

Wollschlaeger, L. M., & Diederich, A. (2020). Similarity, attraction, and compromise effects: Original findings, recent empirical observations, and computational cognitive process models. *The American Journal of Psychology, 133*(1), 1–30. https://doi.org/10.5406/amerjpsyc.133.1.0001

Wort, F., Walasek, L., & Brown, G. D. A. (2022). Rank-based alternatives to mean-based ensemble models of satisfaction with earnings: Comment on Putnam-Farr and Morewedge (2020). *Journal of Experimental Psychology: General, 151*(11), 2963–2967. https://doi.org/10.1037/xge0001237

Yearsley, J. M., Pothos, E. M., Barque-Duran, A., Trueblood, J. S., & Hampton, J. A. (2022). Context effects in similarity judgments. *Journal of Experimental Psychology:*

*General, 151*(3), 711–717. https://doi.org/10.1037/xge0001097

| Abbrev | Theory | Author(s) |
|---|---|---|
| AL | Adaptation Level | Helson (1947, 1965) |
| CR | Correlation-Regression | Johnson & Mullally (1969) |
| ID | Inferred Distribution | Wort, Walasek, & Brown (2022) |
| EN | Ensemble | Putnam-Farr & Morewedge (2021) |
| DbS | Decision by Sampling | Stewart, Chater, & Brown (2006) |
| RF | Range-Frequency | Parducci (1965, 1995) |