

EMPIRICAL ARTICLE

Range-frequency models of within-subjects contextual effects: Salary satisfaction

Michael H. Birnbaum¹ and Julien Rouvère²

¹Psychology, California State University, Fullerton, CA, USA and ²Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA

Corresponding author: Michael H. Birnbaum; Email: mbirnbaum@fullerton.edu

Received: 24 August 2024; **Revised:** 24 February 2025; **Accepted:** 26 February 2025

Keywords: salary satisfaction; range–frequency (RF) theory; contextual effects; within- and between-subjects designs; accepting job offers

Abstract

In four studies, participants judged satisfaction with hypothetical salaries, given the salaries of others doing the same work. Unlike previous research, contexts (distributions of others' salaries) were manipulated within- rather than between-subjects. These studies enabled tests of an extension of range–frequency (RF) theory that assumes that judgments are a compromise between RF predictions based on between- and within-trial contexts. This extension to within-subjects designs correctly predicted the cases in which people assign higher satisfaction ratings to lower salaries. The manipulation of the context within-subjects confirmed phenomena previously observed in between-subjects research. However, a violation of this within-subjects RF model was also observed: When one's salary is lowest compared to others within the same firm, satisfaction varies inversely with the highest salary paid to another at the same firm. Apparently, judgments of satisfaction also depend on inequity. This finding was not observed in previous between-subjects research; indeed, salary and inequity are perfectly confounded for the participant in such a design. We theorize that satisfaction is not merely a judgment of where one's salary falls relative to other salaries, but also depends on how much one is underpaid relative to the distribution of underpayments. A revision of the within-subjects RF model (incorporating the distribution of inequities) gave a good description of judgments of salary satisfaction and of the likelihood to accept a job offer.

1. Introduction

When people are asked to rate how satisfied they are with their salaries, it has been observed that satisfaction is not simply a function of monetary salary but also depends on the salaries of others who are doing the same work (Boyce et al., 2010; Brown et al., 2008; Card et al., 2012; Tripp and Brown, 2016).

1.1. Models of salary satisfaction in context

Putnam-Farr and Morewedge (2021) reported a series of studies to determine which social comparisons affect satisfaction with one's salary. They proposed an 'ensemble' theory (EN), 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'. This theory conflicts with rank-affected theories such as decision by sampling (DbS), as in (Stewart et al., 2006) or (Boyce et al., 2010), or as in Parducci's (1965, 1968, 1995)'s range–frequency (RF) theory.

Wort et al., (2022) proposed an inferred distribution (ID) model in which people infer a normal distribution from the mean and endpoints of the distribution and use the rank in the ID to mediate their judgments of satisfaction.

Birnbaum and Rouvère (2023) noted that contextual effects for salary satisfaction behave like contextual effects observed in other psychophysical and social judgments (Birnbaum, 1974; Hayes and Wedell, 2023; Helson, 1964; Mellers, 1986; Parducci, 1968, 1995; Stevenson, 2019), which appear most compatible with RF theory. Birnbaum and Rouvère designed and conducted critical tests that would test RF theory against EN, ID, and DbS models, as well as those of Adaptation-Level (AL) Theory (Helson, 1947, 1964)¹ and Correlation-Regression (CR) Theory (Johnson and Mullen, 1969). Birnbaum and Rouvère (2023) used cubic distributions like those used by Birnbaum (1974) to produce curves relating ratings of satisfaction to salaries that should cross twice, as predicted by RF theory (and consistent with DbS), but which violate EN, AL, CR, or ID theories. Those four theories could be ruled out by this 'double crossing' effect, which was observed in the results of (Birnbaum and Rouvère, 2023, Experiment 1).²

To test DbS, (Birnbaum and Rouvère, 2023, Experiment 2) independently manipulated the minimum and maximum salaries, keeping the ranks constant. It was found that ratings of stimuli holding the same ranks are linearly related to each other for different ranges, with the curves having different heights and slopes as predicted by RF (and consistent with CR theories), but not by DbS. The EN theory implies systematically nonlinear relationships between judgments of stimuli, for which no evidence was observed. In sum, 5 of the 6 theories were rejected by one or both of the critical tests of Birnbaum and Rouvère (2023), leaving only RF theory as a viable description of the effects of manipulating contexts between subjects.

1.2. Within- and between-subjects research

Within-subjects and between-subjects research often give different or even opposite results (Birnbaum, 1982). For example, judgments of the fault or blame attributed to rape victims show that between-subjects, a divorcee is rated less at fault than married or virgin victims (Birnbaum, 1982; Jones and Aronson, 1973). However, within-subjects, the opposite is obtained (Birnbaum, 1982). Between-subjects, the number 9 is rated 'bigger' than the number 221 (Birnbaum, 1999); but within-subjects, 221 is correctly judged larger than 9. Between-subjects, people appear to neglect the base rate when forming Bayesian inferences (Bar-Hillel, 1980; Hammerton, 1973; Kahneman and Tversky, 1973); however, participants utilize the base rate when it is manipulated within-subjects (Birnbaum and Mellers, 1983), even though they do not conform exactly to Bayes theorem. These seemingly contrary results can be reconciled by a theory of contextual effects.

Birnbaum (1982, 1999) argued that in between-subjects research, context and stimulus are confounded because a stimulus evokes its own context. Between-subjects, 9 evokes a context of smaller numbers than is evoked by the number 221, and in the context evoked by 9, 9 is a bigger number than it seems in the context evoked by 221. Between-subjects, the virgin victim is compared to the context of virgins, and the divorcee victim is compared to divorcees, rather than to all women. Within-subjects,

¹ Anchoring and Adjustment (Tversky and Kahneman, 1974) is a special case of AL theory, so it is also violated by evidence that refutes AL theory.

² Birnbaum and Rouvère (2023) found that mean ratings of satisfaction as a function of salary were significantly higher in Context 1 (than in Context 2) for \$42K, significantly lower for \$46K, and significantly higher again for \$50K; for values of \$44K and \$48K, the mean values were virtually identical; therefore, the curves crossed twice, as predicted by RF theory for the cubic density contexts used.

when the participant is exposed to both types of victims before making judgments, there is a common context for the two judgments.

In Bayesian inference, the highly scattered distributions of responses in between-subjects studies suggest that subgroups of participants have different understandings of the problems (Bar-Hillel, 1980; Birnbaum and Mellers, 1983), among which is the possibility that the source's diagnostic ratio already incorporates the base rate (Birnbaum, 1983a).³

When base rate is held fixed for any participant, it is a study of a variable that is not varied; the participant can easily assume that the source has incorporated the base rate in his or her signal detection criterion. In a within-subjects study, it is possible to explain to participants that the hit to false alarm ratio of the source is independent of the base rate. In such studies, people utilize the base rate (Birnbaum and Mellers, 1983).

In the studies of salary satisfaction reviewed above, like most (but not all) studies of contextual effects, contexts were manipulated between-subjects. That is, each distribution of salaries was presented to a different group of subjects, and participants rated satisfaction with each of the possible salaries within the context. One reason to vary contexts between-subjects is the concern that when the same person experiences multiple contexts, contexts might combine and cancel each other.

1.3. Hierarchical contexts

It seems that people can understand and utilize multiple, hierarchically related contexts when there is at least one identifying variable that distinguishes these contexts and their relations. People correctly recognize that a *large mouse* is smaller than a *small elephant* if they know the three contexts of size for mice, elephants, and mammals. (Parducci et al., 1976) found that if people are asked to judge the sizes of circles and squares in the same study, the skew of the distribution of one did not affect judgments of the other, as if contexts were maintained independently for squares and circles. However, when asked to judge the size while ignoring the shape, the responses behaved as if contexts were combined.

For judgments of salary satisfaction, by analogy, it seems plausible that people can distinguish different situations (workplaces), where salaries and salary distributions (contexts) may differ. If so, it should be possible to manipulate salary contexts between-trials and within-subjects, in order to test theories of how contexts combine and to test whether conclusions about theories of salary satisfaction inferred from between-subjects research are also compatible with within-subjects research.

1.4. Purposes of present research

This research will address the following 4 specific questions:

(1) In between-subjects research, it is found that a lower dollar salary having a higher rank in context is rated on average more satisfying than the rating given by a different group of people who judge a higher dollar figure having lower rank in another context. Will this frequency (rank) effect also be observed when the same person experiences both contexts? That is, will the same person say they would be happier with a lower salary, if most others working in the same place are earning less?

(2) Between-subjects, it has been observed that a lower salary can be rated higher than a higher salary if its position relative to the endpoints is higher. Would this range effect also hold within-subjects: That is, would the same person say that they would be happier with a lower salary if the highest and lowest salaries were manipulated (holding rank fixed) to make the lower salary higher relative to the endpoints of the range?

(3) If people say they would be more satisfied with a lower salary in one context than a higher one in another context, would they also say they would be more likely to accept a job offer at the lower salary?

³As noted in Birnbaum (1983a), the so-called 'correct' solution presented by Kahneman and Tversky (1973) is not actually correct, because it is not a proper Bayesian analysis of the problem, because it assumes that the diagnostic ratio is independent of the base rate, which is not a realistic assumption.

Specifically, will results observed for ratings of salary satisfaction also be observed for judgments of the likelihood to accept job offers?

(4) Can we represent the judgment of salary satisfaction within-subjects by means of an extension of RF theory in which the within- and between-trials contexts combine by an average of the cumulative distributions?

The next section reviews the RF model in the form applicable to between-subjects studies (Birnbaum and Rouvère, 2023; Parducci, 1965, 1968, 1995), and the section following the next presents an extension of RF theory to handle the case of within-subjects studies such as the ones we report in this article.

1.5. RF model (between-subjects contexts)

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 ; $F_k(x_{0k}) = 0$ and $F_k(x_{mk}) = 1$.

The RF theory posits that judgments are a compromise between two principles of judgment: the range principle, which expresses the relative value of $u(x)$ to the endpoints of the distribution, and the frequency principle, which relates $u(x)$ to the cumulative probability (relative rank).

1.5.1. The range principle

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

$$H_k(x) = \frac{u(x) - u(x_{0k})}{u(x_{mk}) - u(x_{0k})}, \quad (1)$$

where $u(x)$ is the utility of salary x , independent of context (Birnbaum, 1974; Parducci, 2011). $H_k(x)$ will range from 0 to 1, as x ranges from x_{0k} to x_{mk} .

1.5.2. 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 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:

$$F_k(x) = \frac{r_{xk} - 1}{n - 1}. \quad (2)$$

The frequency value also ranges from 0 to 1.⁴

1.5.3. RF compromise

The RF compromise is an average between the range and frequency values.

$$RF_k(x) = wF_k(x) + (1 - w)H_k(x), \quad (3)$$

where w is the weight of the frequency principle. Hayes and Wedell (2023) noted that w is typically about 0.5 in empirical studies, though it can vary (Wedell and Parducci, 1988).

1.5.4. Response scale

The transformation from the subjective range frequency value, RF , to the overt response, R , depends on the subjective values of the response scale, the spacing and frequency distribution of example responses, the number of categories in a rating scale, and the response mechanism (Birnbaum, 1982; Mellers and Birnbaum, 1982; Parducci, 1982; Wedell and Parducci, 1988). Suppose the rating scale is uniformly

⁴If u is a strictly increasing, monotonic function, the cumulative frequency of $u(x)$ is the same as that of x .

20. ☒

Your Salary = \$50,000
Others = \$40,000; \$72,000
Not at all Happy ☐ ☐ ☐ ☐ ☐ ☐ ☐ Extremely Happy

21. ☒

Your Salary = \$46,000
Others = \$40,000; \$41,000; \$42,000; \$43,000; \$44,000; \$45,000; \$52,000
Not at all Happy ☐ ☐ ☐ ☐ ☐ ☐ ☐ Extremely Happy

Figure 1. Format for display of two trials of a salary satisfaction study.

distributed and equally-spaced, with minimum and maximum of R_0 and R_m , respectively. With these simplifying assumptions:

$$R_k(x) = (R_m - R_0)RF_k(x) + R_0, \quad (4)$$

where $R_k(x)$ is the predicted rating of salary x in Context k .

1.6. RF theory for contexts varied within-subjects

In the present experiments, people are asked to rate how satisfied they would be to earn specified salaries in different contexts (different hypothetical workplaces), where the distribution of salaries earned by others varies from trial to trial within-subjects. Two example trials are shown in Figure 1.

When context is manipulated within-subjects, there are at least two contexts to consider (Birnbbaum et al., 1971): The Within-Trial (WT) context consists of the stimuli presented on a given trial; and the Between-Trial (BT) context consists of all of the stimuli presented in all trials of the entire study. The BT context is fixed in our study; we use B to designate it; the index k is used to refer to the WT context, which is now varied within-subjects.⁵ To extend RF theory to within-subjects manipulations of context, we theorize that the effective context on each trial is an aggregation of the WT and BT contexts. We represent the combination of contexts via a ‘vertical average’ of the cumulative distributions. This theory is analogous to the theory in Birnbbaum and Rouvère (2023) for the combination of a person’s prior (‘residual’) context (related to income level) with the context manipulated (between-subjects) in the experiment.

From these assumptions, it follows that the theoretical response to stimulus x in Context k is an average of the predictions of RF theory applied to the WT context (stimuli presented in Context k) and to the BT context (all stimuli presented in the study).

$$R_{Bk}(x) = w_B R_B(x) + (1 - w_B) R_k(x), \quad (5)$$

where $R_{Bk}(x)$ is the RF predicted rating of Salary x in an experiment with BT context of B and with Context k within the trial; $R_B(x)$ is the RF predicted judgment based on the context of the entire experiment; $R_k(x)$ is the RF predicted judgment based only on context of stimuli presented on trial k ; and w_B is the weight of the background context relative to the within-trial context. Thus, the predicted

⁵Birnbbaum et al., (1971) used the terms ‘between-sets’ and ‘within-sets’ to refer to the BT and WT contexts, since each trial consisted of a set of stimuli; however, because the term ‘set’ is used in other ways in Psychology, the current terminology seems preferable. In Birnbbaum and Stegner (1979), WT contexts have been termed ‘configurations’, and changes in response represented by changes in theoretical parameters related to stimulus configurations are termed ‘configural’ effects.

value is a compromise between the responses that would be predicted from RF theory based on the WT and the BT contexts.⁶

Note that if $w_B = 1$, then responses would be a function of salary, and people would always assign greater satisfaction to higher salaries within-subjects, so contexts varying from trial to trial would have no effect. If $w_B = 0$, then within- and between-subjects experiments would produce the same effects of context, as if the response to each new trial was the same as the response by a new group of participants who experienced only that trial.

1.7. Within-subjects RF (WSRF) theory predictions

To illustrate how this WSRF theory works, we calculated predictions using a simplified RF theory in which $u(x) = x$, $w = 0.5$ and $w_B = 0.5$, and the effects of the residual context (Birnbaum and Rouvère, 2023) are ignored. These ‘prior’ predictions (based on prior parameters rather than parameters estimated from data) are illustrated in figures for the experimental sub-designs, which illustrate not only the predictions of this model but also show how the experimental designs provide tests among rival theories of contextual effects.

To illustrate calculations for this WSRF model, suppose there is a within-subjects experiment with just two items, both of which the person has already read before making ratings. Suppose these are the two items in Figure 1: (1) How satisfied would you be with a salary of \$50,000 when there are two others who get \$40,000 and \$72,000? (2) How satisfied would you be with \$46,000 if the others receive \$40,000, \$41,000, \$42,000, \$43,000, \$44,000, \$45,000, and \$52,000? (To save space, hereafter, yearly salaries are stated in thousands of USD per year, so \$50,000 will be denoted \$50K, where K represents thousands, or simply as 50, when the meaning is clear.)

The range value of \$50K in the WT context of Item 1 is $H_1(50) = (50 - 40)/(72 - 40) = 0.3125$; the frequency value, $F_1(50) = 0.5$, because \$50K is the median, so the RF compromise is $RF_1 = 0.5 * 0.3125 + 0.5 * 0.5 = 0.4062$. On a 7-point response scale, the predicted judgment would be $R_1(50) = 1 + 6 * 0.4062 = 3.44$ (rounded), which is below 4, the midpoint of the scale of satisfaction. However, \$50K looks better in the BT context, which consists of the 11 salary values combined across the two trials of this example. The BT range value, $H_B(50) = (50 - 40)/(72 - 40) = 0.3125$, and because 50 is ninth from the bottom among 11, $F_B(50) = (9 - 1)/(11 - 1) = 0.8$, so $RF_B(50) = 0.5562$; therefore $R_B(50) = 4.34$. The predicted judgment of WSRF is then the average of these two, 3.89.

In the second item, \$46K has a WT range value of $H_2(46) = (46 - 40)/(52 - 40) = 0.5$; because \$46K ranks 7th out of 8 in the second trial, $F_2(46) = (7 - 1)/(8 - 1) = 0.8571$; averaging range and frequency values, $RF_2(46) = 0.6786$; on the rating scale, $R_2(46) = 5.07$. Relative to the BT context, $H_B(46) = (46 - 40)/(72 - 40) = 0.1875$ and because 46 ranks 8th among the 11 values, $F_B(46) = (8 - 1)/(11 - 1) = 0.7$; averaging range and frequency values, $RF_B(46) = 0.4438$; mapped to the rating scale, $R_B = 3.66$; averaging 5.07 (WT) and 3.66 (BT) values, the predicted response of WSRF is 4.37.

In this case, the model implies that people will be more satisfied to receive \$46K (4.37) than to receive \$50K (3.89), if others around them (the context) are paid even less. When $0 \leq w_B < 1$ this model can imply that people can assign higher satisfaction to a lower salary, depending on the context.

Figure 2 plots the predictions of the simplified WSRF theory for a sub-design of Experiment 3 of the present article in which Your Salary is either \$42K, \$46K, or \$50K, and 3 Others have salaries of either (\$40K, \$43K, \$52K) or (\$40K, \$49K, \$52K). In these two cases, the endpoints of the distributions are fixed (so range is fixed), and the rank of \$46K changes from second to third (of four) between the two contexts. Accordingly, the prediction of satisfaction for \$46K is 0.5 category higher (filled circle) in the context with \$43K than the one with \$49K (unfilled square). The predicted judgments of \$42K and \$50K, which did not change ranks, do not differ between these contexts. This sub-design tests RF and

⁶This equation follows from averaging the WT and BT distributions; via Equation (3) it is also equivalent to averaging the WT and BT range-frequency values, and then transforming to responses via Equation (4).

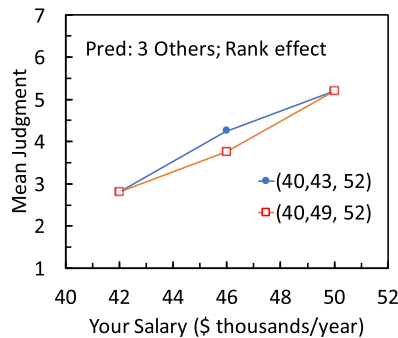


Figure 2. Predicted judgments of satisfaction when 3 other people are doing the same job, plotted as a function of Your Salary, with a separate curve for each distribution of the Others' Salaries. Both contexts have the same endpoints (\$40K and \$52K); the ranks of \$42K or \$50K are the same in both contexts; however, the rank of \$46K changes from second to third (of four) between the two contexts.

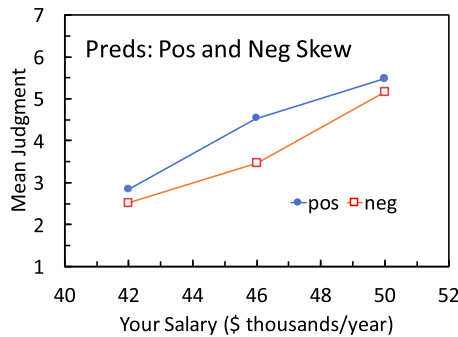


Figure 3. Predicted judgments of satisfaction when 7 other people are doing the same job, plotted as a function of Your Salary, with a separate curve for cases where salaries of Others' were either positively ('pos') or negatively ('neg') skewed on the same endpoints (\$40K to \$52K), with five values from \$41K to \$45K or from \$47K to \$51K, respectively.

DbS theories against AL, CR, and ID theories, which imply different relationships between the curves in Figure 2 from the concave downward form implied by RF theory (Birnbbaum and Rouvère, 2023).

A similar, but stronger manipulation of the frequency term can be achieved when there is a larger number of other people in the context. Figure 3 shows predictions for a sub-design in which there were 7 others. In the positively skewed distribution (filled circles), the Others' salaries were (\$40K, \$41K, \$42K, \$43K, \$44K, \$45K, \$52K) and in the negatively skewed distribution (unfilled squares), they were (\$40K, \$47K, \$48K, \$49K, \$50K, \$51K, \$52K). In these distributions, \$46K drops in rank from 7th (of eight) to second, implying a decrease in response of more than one category; in contrast, \$42K and \$50K drop by only 1.5 ranks, implying smaller decreases. The wider gap between the curves in the middle and the smaller gaps at the end values (combined with no gaps at the ends in Figure 2) are compatible with RF theory or DbS, but not with the other theories.

Figure 4 shows the predictions of the WSRF model for a sub-design of Experiment 3 in which Your Salary was always the middle of three people doing the same work. The lower and upper endpoints were independently manipulated to alter range, holding rank fixed. This sub-design was a $3 \times 2 \times 2$ factorial sub-design of Your Salary (\$42K, \$46K, or \$50K) by Lower Endpoint (\$20K or \$40K) by Upper Endpoint (\$52K or \$72K). Predictions for the narrowest range (\$40K to \$52K, unfilled squares connected by dashed line) produces the steepest slope as a function of Your Salary. The widest range produces the lowest slope (triangles). When both endpoints are low (\$20K and \$52K, filled circles),

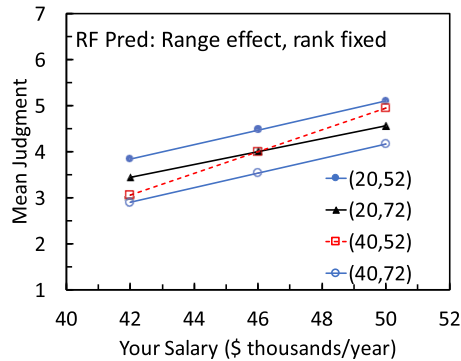


Figure 4. Predicted judgments of satisfaction when two other people are doing the same job, plotted as a function of Your Salary, with a separate curve for each pair of Others' salaries. In this design, Your Salary is always middle in rank, and the lower and upper endpoints have been independently manipulated.

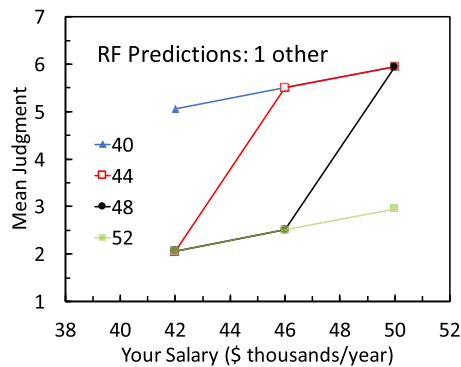


Figure 5. Predicted judgments of satisfaction when one other person is doing the same job, plotted as a function of Your Salary, with a separate curve for each level of the Other's salary.

the judgments are highest and slopes intermediate; and when both endpoints are high (\$40K and \$72K, unfilled circles), judgments are lowest.

Note that in Figure 4, \$46K receives a higher rating in the context of (\$20K, \$52K) than does \$50K in the context of (\$40K, \$72K); similarly, \$42K is rated higher (when both endpoints are low) than \$46K is rated (when endpoints are high). This sub-design provides a test of RF theory against DbS, which does not imply effects of endpoints with ranks fixed. Further, EN theory implies that the curves in Figure 4 should not be linearly related to each other (Birnbaum and Rouvère, 2023).

Figure 5 shows predictions of the WSRF model for a sub-design with one other person. Your Salary, has three values: $x = \$42\text{K}$, $\$44\text{K}$, or $\$50\text{K}$, the Other's Salary had 4 levels: \$40K, \$44K, \$48K, or \$52K. These trials are of the form, (x, y) , where you are paid x and the Other is paid y . Note that if each trial were presented to a separate group of participants (between-subjects), there would be just two ranks possible: either you are paid the most or the least. In such a case, if $x > y$, then both the range and frequency values of x would be 1, and both the range and frequency values of y would be zero, so the RF predicted judgments within-trial of x would be 7 or 1, if $x > y$ or $x < y$, respectively.

Figure 5 shows predicted values of WSRF for this factorial design as a function of Your Salary, x , with a separate curve for each level of the Other's, y . The separation between the curves shows the effect of the Other's salary. The top and bottom curves show the effect of x when Your Salary is highest or lowest, respectively. Note that if $w_B = 0$, the two outer curves would be horizontal (zero slope), and

if $w_B = 1$, then the curves would coincide in a single curve. Indeed, in Figures 2, 3, and 4, if $w_B = 1$, the curves collapse to a single function of Your Salary. Recall that $w_B = 0.5$ in these calculations.

There are two other points to note about the predictions in Figure 5: First, the model with its prior parameters implies 5 cases where a person would be more satisfied to receive a lower salary, if the other person is paid less. For example, it implies that a person would be more satisfied with a salary \$42K when the Other is paid \$40K than with a salary of \$50K when the Other is paid \$52K. (Spoiler alert: These implications are indeed observed in our experiments.) Second, note that if a person is paid the most or the least, the model implies no effect of the Other's salary; e.g., as in Figure 5. (Spoiler alert: As found in results of all four experiments, this implication of the WSRF model is violated in empirical data. A revision of the model to account for the violation is presented in the Discussion.)

2. Experiments 1–3

2.1. Methods of Experiments 1–3

In three experiments, participants rated how satisfied they would be with hypothetical salaries in the contexts of salaries of the other people who were doing the same job and who were equally experienced, qualified, and productive. Participants first served in a between-subjects study, as described in Birnbaum and Rouvère (2023), with a single context, followed by one of the following within-subjects experiments. The three experiments were similar but used different designs, different participants, and slightly different procedures.

2.1.1. Instructions

Subjects were instructed to ‘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 . . . people who are doing the same work and have been evaluated as equally experienced, qualified, and productive’.

Instructions stated, ‘On each trial, you will see the salary that you are paid, and the salaries of others who are doing the same work’. They were told, ‘Assume that the others doing the same work are equally experienced, qualified, and productive as you are’.

The task was to rate how satisfied, how happy or unhappy, they would be with a given salary considering the salaries paid to the others, using a 7-point scale from 1 = Not at all happy to 7 = Extremely happy.

2.1.2. Design of Experiment 1

In Experiment 1, there were either one or two others who were doing the same job. The design was a within-subjects 4×12 , Your Salary by Other's (or Others') salaries, factorial design. The 4 levels of Your Salary were \$40K, \$44K, \$48K, or \$52K. The 12 levels included four with one Other's salary: \$30K, \$42K, \$50K, or \$62K; and eight cases with two Others' salaries: \$30K and \$32K; \$30K and \$42K; \$30K and \$50K; \$30K and \$62K; \$42K and \$46K; \$42K and \$50K; \$42K and \$62K; or \$60K and \$62K. Overall, there were 16 trials having two salaries and 32 trials with three salaries, making a total of 128 salaries presented in the 48 experimental trials.

Complete instructions and materials for Experiment 1 are available at the following URL: https://konstanzworkshop.neocities.org/CSUF22/salary_wic_01.htm

2.1.3. Design of Experiment 2

In Experiments 2 and 3 there were from one to seven others who were doing the same work. The design of Experiment 2 was a within-subjects, 3×12 , Your Salary by Others' salaries, factorial design in which the three levels of Your Salary were \$42K, \$46K, or \$50K; the 12 levels of Others' salaries were (one other:)(\$40K), (\$44K), (\$48K), (\$52K); (two others:)(\$26K, \$52K), (\$40K, \$70K); (seven others: positively skewed) (\$40K, \$41K, \$42K, \$43K, \$44K, \$45K, \$52K); (seven others: negatively skewed) (\$40K, \$47K, \$48K, \$49K, \$50K, \$51K, \$52K); (seven others: endpoints manipulated) (\$26K, \$42K,

\$44K, \$46K, \$48K, \$50K, \$52K), (\$26K, \$42K, \$44K, \$46K, \$48K, \$50K, \$70K), (\$40K, \$42K, \$44K, \$46K, \$48K, \$50K, \$52K), (\$40K, \$42K, \$44K, \$46K, \$48K, \$50K, \$70K). Overall, there were 12 trials with two salaries, 6 trials with three salaries, and 18 trials with 8 salaries, so participants saw 186 salaries in the 36 trials.

The instructions and randomized trials are available at the following URL: https://konstanzworkshop.neocities.org/CSUF22/salary_wick_01.htm

2.1.4. Design of Experiment 3

Experiment 3 used a within-subjects, 3×12 , Your Salary by Others' salaries, factorial design in which the three levels of Your Salary were \$42K, \$46K, or \$50K, and the 12 levels Others' salaries were (one other:) (\$40K), (\$44K), (\$48K), (\$52K); (two others:) (\$20K, \$52K), (\$20K, \$70K), (\$40K, \$52K), (\$40K, \$70K); (three others:) (\$40K, \$43K, \$52K), (\$40K, \$49K, \$52K); (seven others: positively skewed [POS(7)]) (\$40K, \$41K, \$42K, \$43K, \$44K, \$45K, \$52K); and (seven others: negatively skewed [NEG(7)]) (\$40K, \$47K, \$48K, \$49K, \$50K, \$51K, \$52K).

Each experiment's design can also be viewed as the union of several factorial sub-designs. For example, Experiment 3 can be viewed as the union of four sub-designs: A 3 by 4, Your Salary by Other's salary (two people); a 3 by 2 by 2, Your Salary by Lowest Salary by Highest Salary (3 people); a 3 by 2, Your Salary by Rank of Your Salary (with endpoints fixed in sets of 4 people); a 3 by 2, Your Salary by Rank of Your Salary (with endpoints fixed in sets of 8 people in positively or negatively skewed distributions). Participants experienced a total of 132 salaries in the 36 experimental trials.

The instructions and one set of randomized trials are displayed at the following URL: https://konstanzworkshop.neocities.org/CSUF22/salary_wick_01a

2.1.5. Procedure

In all 3 studies, there were warm-up blocks of representative trials to familiarize the participants with the range of salaries (9 trials in Experiment 1 and 5 in Experiments 2 and 3), followed by a randomly ordered block of either 48 experimental trials (Experiment 1) or 36 trials (Experiments 2 and 3).

In Experiment 1, when the experimental trials were completed, participants were instructed to perform three other, unrelated judgment tasks, after which they were to complete a second repetition of this salary satisfaction task, which also included the 9 warm up trials and the same 48 experimental trials. Because participants were free to work at their own paces, not all participants completed the second repetition of the task in the time allotted.

In Experiments 2 and 3, participants were instructed to click a button upon completing the task, which led to another random ordering of the same task, (including the warm-up block), which was repeated for the duration of the 30 minute study. Participants completed different numbers of repetitions, from 2 to 6 repetitions.

In each of the experiments, there were additional questions that requested age, gender, education, and nationality. A box was provided for comments.

2.1.6. Subjects

In all three experiments, subjects were undergraduates at California State University, Fullerton, who participated as one option to earn partial credit toward an assignment in lower division Psychology.

In Experiment 1, subjects were 270 students (75% female), 210 of whom completed two repetitions of the task. Overall, there were 480 judgments of each of the 48 experimental trials. Age ranged from 18–38, with 75% 20 years or younger. For those who completed both repetitions and who had defined reliability, the median correlation between responses in the two repetitions was 0.74.

In Experiment 2, there were 95 students (68% female), 92 of whom completed at least 3 repetitions of the task; 70 of these completed 5 or 6 repetitions. There were a total of 439 judgments per item. Age ranged from 18 to 29, with 82% aged 20 or younger. Median correlation between responses in successive repetitions was 0.80.

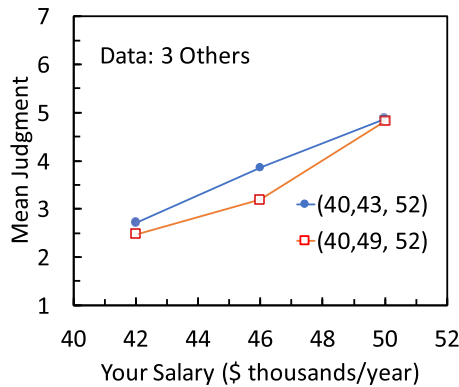


Figure 6. Mean judgments of satisfaction as a function of Your Salary, with a separate curve for each set of Others' salaries. Compare with predictions in Figure 2.

In Experiment 3, there were 215 students (70% female), all of whom completed at least two repetitions; 211, 180, and 157 of them completed 3, 4, or 5 repetitions, respectively. In sum, there were 978 judgments for each of the 36 items. Median correlation between successive repetitions was 0.79.

2.2. Results and discussion of Experiments 1–3

All three experiments gave similar results. Because Experiment 3 confirmed and extended findings of Experiments 1 and 2, this section presents Experiment 3 in detail, which covers the main findings of the three studies. Detailed results of Experiments 1 and 2 are presented in the Appendix.

Figure 6 shows mean judgments of satisfaction corresponding to the predictions in Figure 2 for the sub-design of Experiment 3 in which there were 3 others, with salaries of either (\$40K, \$43K, \$52K) or (\$40K, \$49K, \$52K). Data are plotted as a function of Your Salary, with separate curves for the Others' salaries. Each mean is the average of 978 judgments.

Consistent with predictions of the WSRF model in Figure 2, the mean rating of \$46K is higher when its rank is higher in context than when its rank is lower (mean difference = 0.66, median = 1).⁷

Figure 7 shows mean judgments with positively ('pos') and negatively ('neg') skewed distributions (Experiment 3). As predicted by WSRF model (Figure 3), the effect of these positively and negatively skewed distributions was greatest for \$46K, where the difference in rank between the two distributions was the greatest and smaller for \$42K and \$50K, where the rank difference was smaller. This pattern was observed for subjects who completed 2, 3, 4, or 5 repetitions, and within every repetition for those who completed all 5 repetitions. The difference was also greater at \$42K than at \$50K, which is not predicted by the WSRF model. This pattern was also observed with the same design in Experiment 2 (Appendix).⁸

⁷Data were analyzed separately for each repetition; in addition, data were analyzed separately for those 157 subjects who completed all 5 repetitions of the task (and for each repetition within that group). These analyses produced 10 mean differences and 10 (partly redundant) *t*-tests for the effect of context on the judgment of \$46K. The mean differences ranged from 0.47 to 0.84 across repetitions and subject groups, with no apparent relation to differences among groups or repetitions, and with *t* = 5.91 to 9.39, all significant. Although people judged \$42K to be slightly higher on average in this context (mean difference = 0.24, statistically significant in 8 of 10 *t*-tests), contrary to predictions that there should be no difference in its judgment, the median difference was 0 and only 33% gave higher judgments to \$42K when the context of Others' salaries was (\$40K, \$43K, \$52K) instead of (\$40K, \$49K, \$52K).

⁸The mean differences in Experiment 3 were 0.46, 1.26, 0.30 averaged over all judgments by all Ss. The 30 *t*-tests for these three tests of context, for each repetition by all Ss, and by only those 157 who completed all five repetitions, were all positive and significant, ranging from *t* = 2.33–13.84.

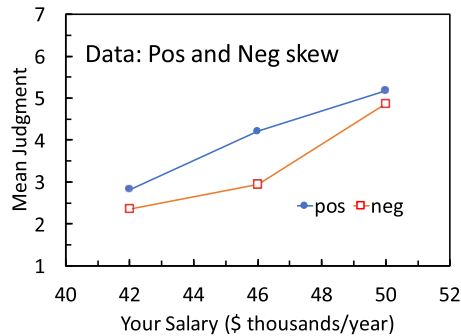


Figure 7. Mean judgments of satisfaction as a function of Your Salary, with a separate curve for each context induced by 7 Others, whose salaries were either positively ('pos') or negatively ('neg') skewed. Compare with Figure 3.

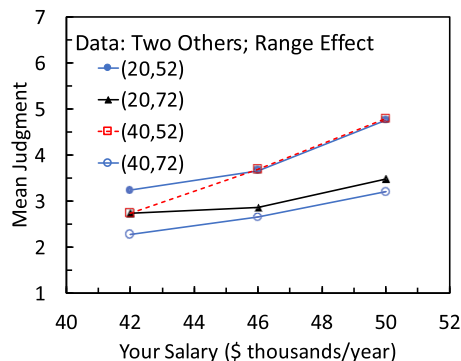


Figure 8. Mean judgments of satisfaction as a function of Your Salary, with separate curves for each level of Other's salaries. Rank of Your Salary is fixed and endpoints (ranges) are manipulated. Compare with Figure 4.

The theory of AL, including the 'anchoring and adjustment' interpretation of AL theory (Helson, 1947), as well as CR, and ID theories cannot account for the concave downward, nonlinear relationship between the higher relative to the lower curves in Figures 6 and 7 (positively to negatively skewed distributions). Both RF theory and DbS correctly predict this relationship; that the vertical gap should be greatest for \$46K in both figures, and that the vertical gaps should be greater in Figure 7 than in Figure 6.

Figure 8 shows mean judgments for the sub-design of Experiment 3 with exactly two others, in which rank of Your Salary is fixed (always the median), and the endpoints are varied. The highest curve (filled circles) is observed when the endpoints are relatively low (\$20K, \$52K), and the lowest curve (unfilled circles) is observed when the endpoints are higher (\$40K, \$72K)⁹; the steepest curve (dashed line) is observed when the range is least (\$40K, \$52K), and the lowest slope (triangles) occurs when the range is greatest (\$20K, \$72K). The relative heights and slopes of the curves agree with the predictions in Figure 4, except that the dashed curve appears relatively higher in the data, compared to the predictions.

The DbS theory does not predict effects of variation of the endpoints when rank is fixed. Because rank is fixed for all markers in Figure 8, that model cannot explain why the curves do not coincide, nor even why the slopes are not zero. The AL theory could explain the height changes but not the slope changes in Figure 8. The EN theory implies that the highest curve (20, 52) should be concave

⁹This result agrees with results for a similar design in Experiment 2 (Appendix), which used slightly different values.



Figure 9. Mean judgments of satisfaction as a function of Other's salary, with a separate curve for each level of Your Salary. Compare with predictions in Figure 5.

downwards relative to the lowest curve (40, 72), which does not appear in the data. These effects of endpoints with rank fixed agree with previous findings observed in between-subjects designs (Birnbbaum and Rouvere, 2023).

Consistent with predictions of the WSRF model in Figure 4, the mean rating of satisfaction in Figure 8 with \$46K is higher, when Others' salaries are (\$20K, \$52K), than is the mean satisfaction with \$50K, when Others' salaries are (\$40K, \$72K). The mean difference was 0.45 ($t = 8.02$, 5.48, 0.10, 5.39, and 5.34 for Repetitions 1 through 5, respectively, all significant, except in Repetition 3). Similarly, \$42K is rated on average 0.59 higher when Others' salaries are (\$20K, \$52K) than \$46K is rated when Others' salaries are (\$40K, \$72K), with t ranging from 4.56 to 9.97. Similar results were found when analyzed only for the 157 participants who completed all 5 Repetitions. Experiment 2 found similar results using slightly different salary levels (Appendix).

Figure 9 shows mean judgments from the sub-design in which there was only one other co-worker, as a function of Your Salary (x), with a separate curve for each level of the Other's salary, y .

The prior WSRF model predicted 5 cases in Figure 9 (see Figure 5) where lower salaries should be rated more satisfying in context than higher salaries in other contexts. In all 5 cases, the observed mean differences were as implied and significant in every Repetition.¹⁰ The smallest mean difference of these cases was 0.99, for the most extreme salary difference of $x = \$42K$ when $y = \$40K$ compared to $x = \$50K$ when $y = \$52K$. In this case, $t(214) = 8.45$, with 147 of the 215 Ss (68%) showing a higher mean judgment for the lower salary in the more favorable context. In all five cases, the differences were slightly larger in the last Repetition than in the first.

Consistent with predictions of WSRF model in Figure 5, the empirical curves in Figure 9 nearly coincide when Your Salary is greatest ($x > y$). However, when Your Salary is least ($x < y$), mean judgments of satisfaction decrease as the Other's salary (y) increases, contrary to the implication that Other's salary should not matter when Your Salary is lowest. The mean satisfaction to receive \$42K when the Other gets \$44K is higher than that to receive \$46K when the Other gets \$52K, contrary to implication of WSRF. Experiments 1 and 2 also observed this same type of deviation (Appendix). This violation of the WSRF model is taken up in the Discussion, where a revision of the model is proposed to describe it.

Although Experiments 1–3 show that people say they would be happier with lower salaries, if their salaries are higher than others, would people accept job offers that (in context) make them happier with lower salaries? People might instead consider higher salary more important than greater satisfaction when deciding to accept a job. Experiment 4 tested whether people say they would be more likely to accept job offers at lower salaries that were rated in Experiment 3 as more satisfying.

¹⁰Each of the five differences was tested in each Repetition, resulting in 25 t -tests, ranging from 5.84 to 19.58. Restricting the data to only the 157 Ss who completed 5 Repetitions, the 25 partially redundant t values ranged from 5.09 to 16.31.

3. Experiment 4: Salary satisfaction and accepting job offers

Experiment 4 investigated the relationship between judgments of salary satisfaction and judgments of the likelihood to accept job offers. This study also explored a reviewer's suggestion that instructions intended to focus attention on the WT or BT contexts might impact the results.

3.1. Method

As in Experiments 1–3, people judged satisfaction with hypothetical salaries, given the salaries of others doing the same work, and the same people also judged how likely they would be to accept job offers given the same situations. In addition, there were two between-subjects conditions of instructions, which were intended to focus attention on either the WT or BT contexts.

3.1.1. Instructions

The overall design of instructional manipulations was a 2 by 2, Task (Salary Satisfaction or Job Acceptance) by Instructional Focus (attention focused on WT or BT context), with subjects nested in Focus and crossed with Task.

Full instructions, stimuli, and materials for Salary Satisfaction with BT or WT and Job Acceptance with BT and WT focus, can be found at the following URLs, respectively:

https://konstanzworkshop.neocities.org/CSUF24/salary_W4_01.htm

https://konstanzworkshop.neocities.org/CSUF24/salary_W4_I_01.htm

https://konstanzworkshop.neocities.org/CSUF24/salary_C4_01.htm

https://konstanzworkshop.neocities.org/CSUF24/salary_C4_I_01.htm

3.1.2. Task instructions

The salary satisfaction task instructions were similar to those in Experiments 1–3. Instructions read (in part) as follows: ‘Your task is to rate how satisfied or dissatisfied, how happy or unhappy, you would be with your salary, . . . now that you know what other people are getting who are doing the same work . . . Please make your ratings . . . on the 7 point scale . . . from: Not at all Happy to Extremely Happy’.

The job acceptance task instructions read (in part) as follows: ‘Your task is to rate how likely you would be to accept each job offer, now that you know what other people are getting who are doing the same work at the company. Please make your ratings . . . along the 7 point scale . . . ; Very Unlikely to Accept the Job Offer . . . to . . . Very Likely to Accept the Job Offer’ . . .

‘Assume that the others doing the same work are equally experienced, qualified, and productive as you are’.

‘You should assume that all other aspects of the companies are the same; that . . . they have equal resources, that raises will be the same . . . , and that the gaps in salaries between people . . . will remain the same Please do not imagine that things will change over time or imagine other differences . . . aside from what is given in each scenario’.

3.1.3. Focus instructions

Each participant in Experiment 4 was randomly assigned to one of two conditions of instructions that differed with respect to the focus to be placed on the BT as opposed to the WT contexts.

In the Salary Satisfaction, BT focus Condition, instructions read (in part), ‘. . . read over the trials to get an overall picture of the different situations, before you start. You can think of each trial as a scenario describing a different company and a different salary you might receive there. When judging your salary satisfaction in each scenario, we want you to think of all the other scenarios. That is, your ratings should be higher if you think you would be happier with your salary in that company than you would be in the other companies of this study’.

In the WT focus Condition, satisfaction instructions stated, ‘When judging your salary satisfaction in each scenario, we want you to imagine that this is the only scenario you would experience . . . , we only

want you to think of how satisfied you would be with the salary in the context of the others' salaries in that specific scenario; do not consider any outside options you might have such as those in the other trials'.

In the job acceptance task, instructions for BT and WT conditions were similar to those for the satisfaction task, with wording suited to accepting a job. For example, the WT condition was expressed as follows: 'When judging your likelihood of accepting an offer in each scenario, we only want you to think of the information for that specific scenario. Judge how likely you would be to accept the offer in the context of the others' salaries in that scenario; do not consider the other scenarios or options from outside the experiment. Please rate how likely you would be to accept each job offer taken in isolation of all other trials. That is, judge each case on its own, without thinking of the other trials in this study or other considerations'.

3.1.4. Stimuli and design

Within each combination of instructional manipulations, the stimuli and designs were the same as in Experiment 3, producing 36 scenarios of Your Salary and Others' salaries.

3.1.5. Procedure

Each subject performed five sub-tasks: In the first sub-task, they read overall instructions about both salary satisfaction and job acceptance, and they made seven warm-up judgments of salary satisfaction and of accepting job offers. They also made a direct choice between two job offers with the same salary in the context of different salaries paid to others. An open-ended question asked why they made the decisions they did. This warm-up sub-task can be viewed at the following URL: https://konstanzworkshop.neocities.org/CSUF24/salary_choice.htm

When subjects clicked a button to complete this sub-task, they were directed to a page that randomly assigned them to one of two conditions of instructions, intended to make either WT or BT contexts more salient, which were maintained for the next four sub-tasks: salary satisfaction, accepting job offers, second repetition of salary satisfaction, and second repetition of accepting job offers.

The first presentation of each sub-task included full, detailed instructions for that task, five representative warm-up trials, and the 36 experimental trials of each sub-task. The second repetition of each task included only a two sentence summary of the task to be done next (satisfaction or job acceptance) and a reminder specifying that only the WT or BT context was to be considered in forming the judgments.

3.1.6. Subjects

There were 160 undergraduates who completed the experiment as one option to gain credit toward an assignment in lower division psychology. Of these, 79 received the instructions to consider only the WT context, and the other 81 received the BT instructions asking them to compare each scenario with all others in the study. A few additional people who started the experiment were excluded because they failed to complete all sub-parts of the task, or because they gave the same response in all cases. Of the 160 who completed everything, 126 were female (79%), and 81% were 20 years of age or younger.

The median correlation coefficients within-person between repetitions, which were separated by one other sub-task, for ratings of salary satisfaction and job acceptance were 0.73 and 0.79, respectively.

3.2. Results and discussion of Experiment 4

The mean ratings of salary satisfaction are shown in Figure 10. Left and right columns show ratings of salary satisfaction with instructions focused on the BT or WT contexts, respectively. The rows of panels, from lowest to highest, show the mean judgments plotted as in Figures 6–9. Each marker in Figure 10 represents the mean over either 158 or 162 judgments (2 repetitions by 79 or 81 subjects, for WT and BT instructions, respectively).

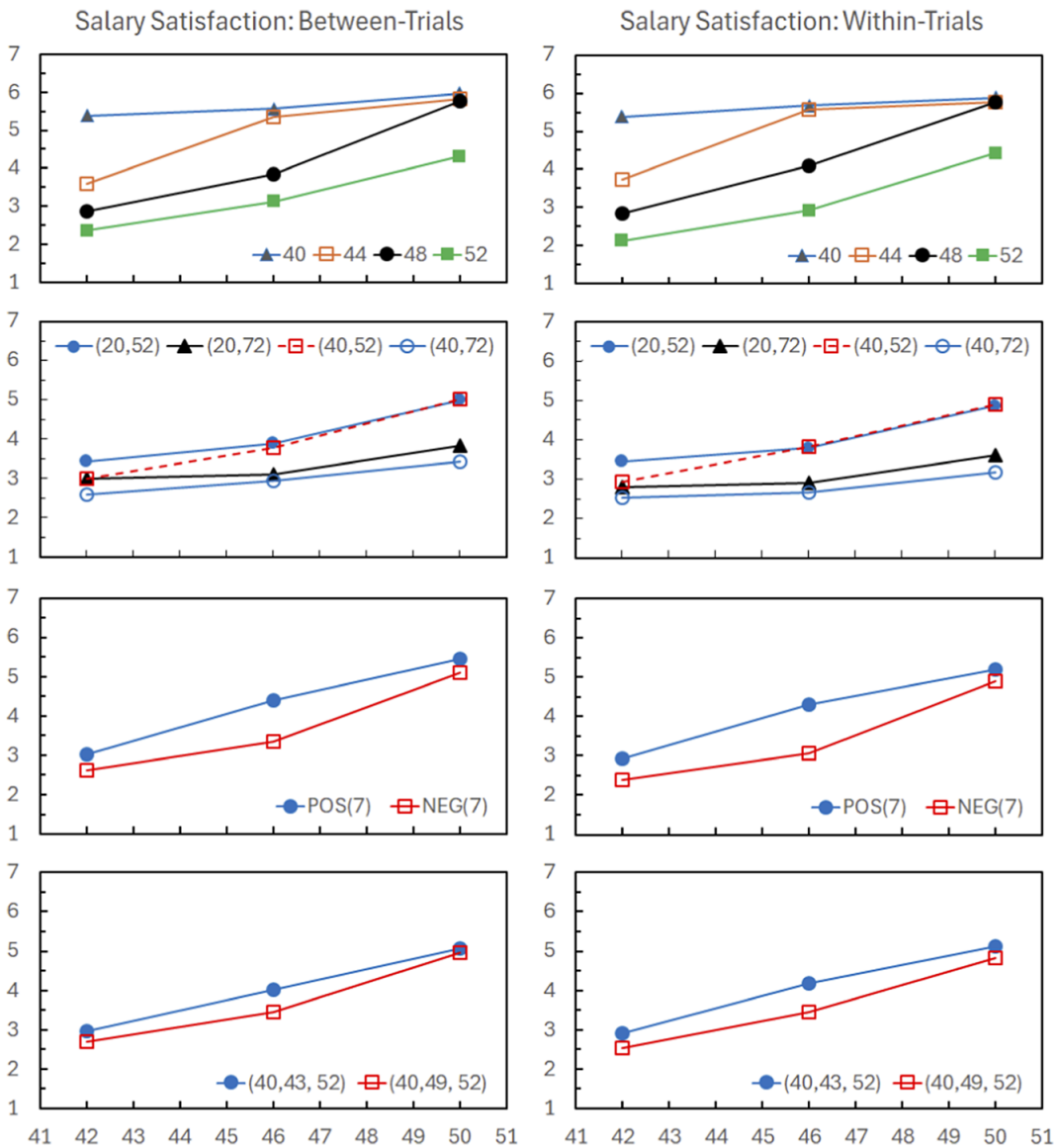


Figure 10. Ordinate shows mean judgments of satisfaction; abscissa represents *Your Salary*. Separate curves in each panel represent *Others' salaries*. Rows of panels from lowest to highest correspond to Figures 6–9. Left and right columns show results when instructions emphasized *Between-* and *Within-Trials* contexts, respectively.

The correlation coefficient between the mean judgments of satisfaction in Experiment 3 and those of Experiment 4 (averaged over focus instructions) is 0.999. Although high correlations can mask important differences, comparison of Figure 10 with Figures 6–9 shows that the new data in all panels of Figure 10 reinforce the findings and patterns observed in Figures 6–9 of Experiment 3.

By comparing columns in Figure 10 (focus on BT or WT contexts), it appears that the focus instructions had minimal effects. If people ignored the WT context (as instructed in the BT condition), the curves in each panel of Figure 10 (which represent different WT contexts) would coincide. Similarly, if people ignored the BT context (as instructed in the WT condition), then the curves in

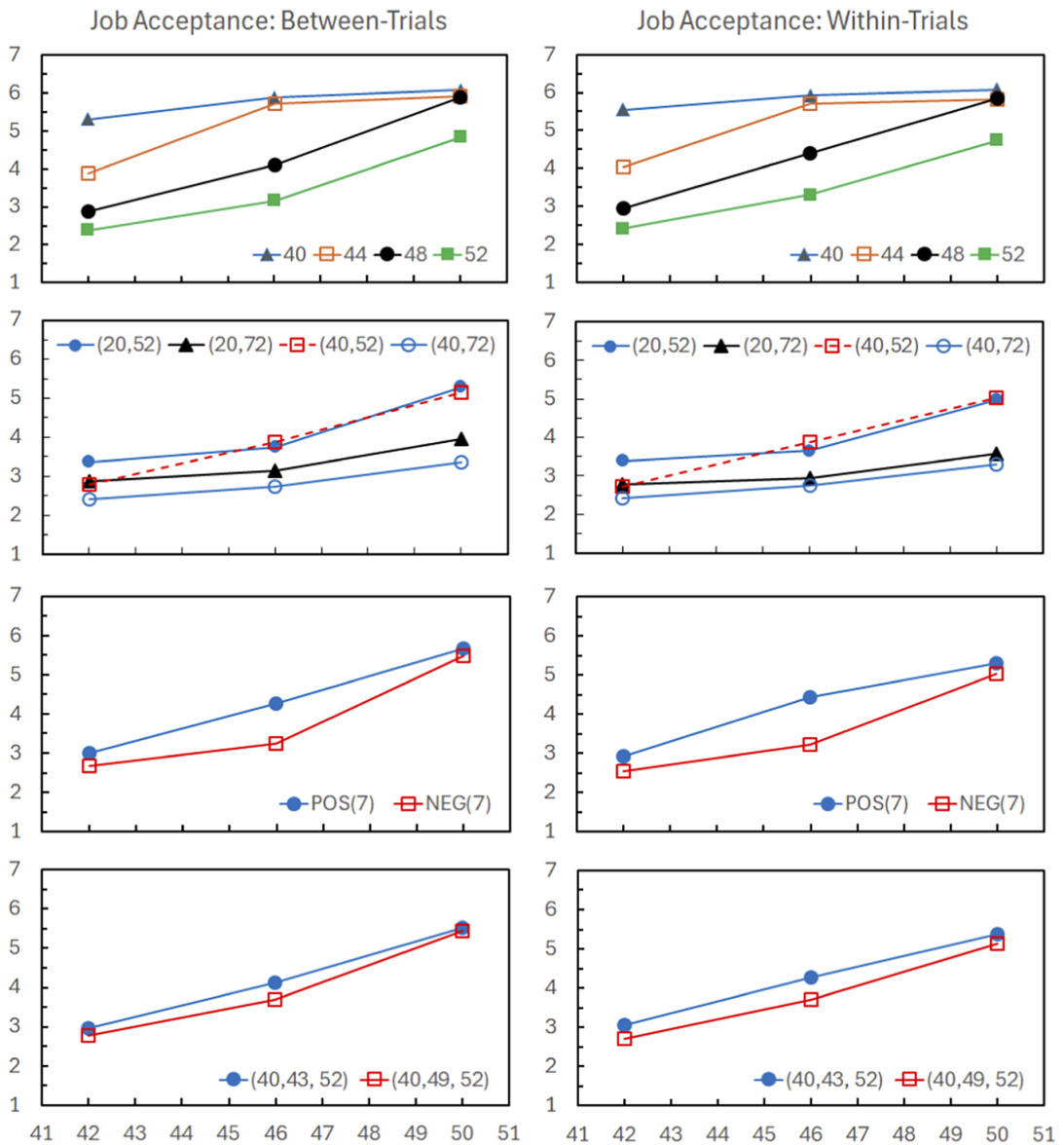


Figure 11. Ordinate shows mean judgments of Likelihood to accept a job offer, plotted as in Figure 10; abscissa represents Your Salary Offer. Left and right columns show results when instructions emphasized Between- and Within-Trials contexts, respectively.

the upper-most panels would be horizontal. However, the empirical curves show only tiny differences between WT and BT focus conditions.

Figure 11 shows job acceptance ratings, plotted as in Figure 10. Ratings of job acceptance are surprisingly similar to the satisfaction ratings. There is a slightly greater effect of Your Salary in the job acceptance ratings, but this increased effect is small (about 0.2) and does not produce any material change in the results. For example, in the upper panels of Figure 10 the mean rating of satisfaction for a salary of \$42K when the Other's salary is \$40K is 5.38, and the mean rating for \$50K when the Other's is \$52K is only 4.37. For job acceptance (Figure 11), corresponding ratings are 5.42 and 4.79. Although

the difference for job acceptance may be a little less than for satisfaction, the direction of the difference was not reversed, as might have been expected if salary was far more important than what Others earn for job acceptance.

The correlation coefficient between mean ratings of salary satisfaction and of job acceptance, averaged over focus conditions is 0.994. The correlations between ratings in BT and WT focus conditions are 0.993 and 0.991 for salary satisfaction and job acceptance judgments, respectively. The correlation coefficients between the two repetitions of satisfaction-BT, satisfaction-WT, job acceptance-BT, and job acceptance-WT are 0.983, 0.982, 0.993, and 0.994, respectively; these values might be considered measures of reliability. Given the similarity of the figures and considering correction for attenuation, it seems hard to argue that these four measures deviate much from the idea that they might be four measures of the same underlying construct.

In sum, it appears that Experiment 4 found little evidence of any sizable effect of either instructional manipulation; instead, Experiment 4 provides reinforcement of the findings of Experiment 3 in two new groups of participants using slightly different materials and procedures.

Although Experiment 4 found only small effects of instructions, finding small effects does not guarantee there is no effect, so it is possible that with some other procedures or participants, instructional manipulations might have greater effects. For example, it is possible that our procedures primed people to focus on salary satisfaction which carried over to evaluating job offers. It is also possible that young college students differ from older people who have differing experience. Perhaps those with more ‘life experience’, working for a salary and paying bills, would give higher ratings to job offers with higher salaries, even when context makes those salaries less satisfying.

Suppose judgments of satisfaction are not simply judgments of relative size of one’s salary, but involve inequity as well, as proposed in the next section. Perhaps if the instructions were to judge the *size* of each salary relative to either the WT or BT contexts, there might have been a more substantial effect of these instructions to focus on one context or the other. In this thought experiment, such a manipulation should have affected the parameter w_B , assuming that the WSRF model was descriptive of size judgments, even if not fully descriptive of satisfaction.

4. General discussion

All four experiments gave results consistent with the main conclusions. Within-subjects manipulation of contexts for salary satisfaction resulted in fairly large contextual effects; these changes were large enough to produce many cases in which a lower salary in context is rated significantly more satisfying than a higher salary in a different context.

With respect to the first two questions posed in the Introduction (Section 1.4), the results show that within-subjects, people assign significantly higher satisfaction ratings to lower salaries when the rank of the salary relative to others’ salaries is higher; further, they assign higher satisfaction to lower salaries when the position of the salary relative to the endpoints of the context is higher.

It is noteworthy that context can cause the same person in the same experiment to give a higher rating of satisfaction to a substantially lower salary. This finding seems more impressive than when such findings are observed between-subjects, as in Birnbaum and Rouvère (2023). Recall that between-subjects, one group of participants says 9 is a ‘bigger’ number than another group rates 221, but within-subjects, 221 is judged bigger (Birnbaum, 1999). The within-subjects results of Experiments 1–4, however, show that the same people assign a higher rating to a lower salary (if it exceeds salaries of their co-workers) than they give to a higher salary (when the co-workers are paid more than they are). Perhaps even more striking is that the same people judge that they would be more likely to accept a lower salary offer, if others at the same company are paid less. Whether people would actually turn down higher salary offers is an open empirical question.

The manipulations of rank holding range fixed produced results that, as in previous research using between-subjects variations of context, rule out rival theories of contextual effects. The results in Figures 6 and 7 (replicated in Figure 10), which show nonlinear effects of salary due to manipulations

of the salary's rank, are not consistent with AL, CR, or ID theories. Effects of rank are compatible with WSRF (Figures 2 and 3) and DbS theories. These results with within-subjects manipulations of context are similar to those reported previously in between-subjects research on contextual effects (e.g., Birnbaum and Rouvère, 2023; Parducci, 1965, 1995; Wedell and Parducci, 1988).

The manipulations of range holding rank fixed, (Figures 4 and 8), are correctly predicted by the WSRF model, but are not consistent with implications of DbS or AL theories. EN theory implies that the curves in Figure 8 and the third panels from the bottom of Figure 10 should not be linearly related to each other (Birnbaum and Rouvère, 2023), a prediction that did not materialize either in previous between-subjects research or in the present within-subjects studies.

With respect to the third question of Section 1.4, the data (Figures 10 and 11) show that judgments of the likelihood of accepting job offers show the same main properties as judgments of satisfaction. The similarity of job acceptance ratings to satisfaction may be due to the procedure in which the same people made both judgments, or perhaps it may be limited to undergraduate subjects, who might choose jobs by anticipated happiness rather than by salary. It seems plausible that people with more life experience might place higher priority on other considerations, such as the benefits of higher salary to one's spouse or children, over their own personal satisfaction when evaluating job offers.

The fourth question of Section 1.4 asked whether an extension of RF theory, WSRF, proposed to reconcile within- and between-subjects research, would give a good description of the effects of manipulating context within-subjects. Here, the answer is more nuanced; the model does a fairly good job of predicting the effects of manipulating rank of one's salary with endpoints fixed or manipulating endpoints (and range) with rank fixed. This model had considerable success in these predictions even with prior parameters. However, a systematic violation of WSRF was observed: salary satisfaction decreased with the gap between one's salary and the highest salary of others, contrary to the implication that when one's salary is lowest, satisfaction should be independent of what others are paid. The implications and analysis of this discrepancy are the foci of the next sections, which discuss implications of the discrepancy and present a revision the WSRF model to better describe the data.

4.1. Discrepancy from WSRF model

The WSRF model implies that when one is paid the least, the amount paid to the highest paid person should have no effect, but this property was violated in all four experiments.

The violations require us to rethink how we conceptualize salary satisfaction. As in other theories of salary satisfaction (Putnam-Farr and Morewedge, 2021; Tripp and Brown, 2016; Wort et al., 2022), the WSRF model holds that judging satisfaction is analogous to judging the magnitude of one's salary compared to the salaries of others. However, we now think that the discrepancies from the simple WSRF model are due to perceived inequities produced by differences in salary between people who are equally qualified and productive at the same job.

The WSRF model is based on how a salary relates to two distributions: the distribution of salaries in the entire study (BT) and the distribution of salaries within a single trial (WT). However, there is a third distribution to consider, which is the distribution of differences in salary (inequities) between one's salary and the highest salary paid to a person of comparable merit (Birnbaum, 1983b; Birnbaum and Hynan, 1986).

This distribution of inequities is not distinguishable to a participant when context is manipulated in a between-subjects design, because the distribution of inequities (differences from the highest salary) maps exactly to the distribution of salaries; that is, the highest salary in a between-subjects condition is also the smallest difference from highest salary, and the lowest salary is also the greatest difference from the highest salary. Thus, in the between-subjects design, the range value in RF is isomorphic to the inequity (difference from the highest salary); that is, x is isomorphic to $H_k(x)$ and to $x_{mk} - x$ when x_{mk} and x_{0k} are fixed for a person. However, when contexts are manipulated between trials and within-subjects, inequity and the range term become disentangled.

According to this argument, the violations observed in Experiments 1–4 might only have been detected using a within-subjects design, because in a between-subjects design, the endpoints are fixed for any participant. However, in a within-subjects design x_{mk} varies, so a participant can experience relative inequity as a separate variable from the salary's position in the range.¹¹

Theoretically, we can propose that satisfaction depends not only on the relative position of one's salary in the distribution of salaries, but also on the salary inequity, relative to the distribution of inequities.

The next section develops a more general model in which relative inequity is incorporated into the WSRF model.

4.2. Revised WSRF model with inequities

Define the inequity as the difference between one's salary and the top salary in Context k : $d_k(x) = u(x) - u(x_{mk})$. This inequity variable is non-positive, because the value is zero when one is paid the maximum and otherwise negative. This non-positive, inverse scale of inequity is positively related to satisfaction.

As in the WSRF model (Equation (5)), let k index the WT context and let B represent the BT context of salaries. Let D index the within-subjects and between-trials distribution of these inequities. For each of the three distributions, k , B , and D , we can define a range and frequency component. For contexts k and B , the range and frequency terms are as stated earlier, in Equation 5.

For D , the distribution of inequities, the range and frequency terms are denoted $H_D(d_k)$ and $F_D(d_k)$, which represent the position of the inequity relative to the greatest and least inequities, and the rank of the inequity, respectively. Each of these RF components varies from 0, when the person is most underpaid, to 1 when the person is least underpaid; i.e., when $x = x_{mk}$, $d_k(x) = 0$.

Salary satisfaction is then represented as a weighted average of three RF compromises, representing the WT, BT, and inequity contexts.

The WT frequency and range components, which varied from trial to trial, are again denoted $F_k(x)$ and $H_k(x)$, as in Equations (2) and (3).

In fitting models to data, the approximation $u(x) = x$ was retained for this range of salaries.

For the BT calculations, it was assumed that $w = 0.5$. Given the BT distribution of salaries presented in Experiment 3, the RF compromise values, $RF_B(x)$, for Your Salary of \$42K, \$46K, and \$50K were calculated to be 0.35, 0.50, and 0.65, respectively.

The range and frequency components of the inequity are denoted $F_D(d_k)$ and $H_D(d_k)$, respectively, and are calculated from RF theory applied to the distribution of inequities.

The following general model was used to fit the data; this model includes original RF theory, the WSRF model, and the new revised model as special cases. It can be written as follows:

$$R_{BDk}(x) = a[w_1 RF_B(x) + w_2 F_k(x) + w_3 H_k(x) + w_4 F_D(d_k) + w_5 H_D(d_k)] + b, \quad (6)$$

where $R_{BDk}(x)$ is the predicted judgment of satisfaction with salary x in an experiment with overall distribution between-trials B , with contexts varying from trial to trial indexed by k ; and with distribution of differences from the top salary of D ; a and b are parameters that map the RF values onto the response scale; and the w_i are weights that are constrained to be non-negative and sum to 1.

Original RF theory, applied to a between-subjects manipulation of contexts indexed by k , follows from this equation, if $w_1 = w_4 = w_5 = 0$.

The 'prior' WSRF model, where k (context) varies between trials and within-subjects, used to calculate predictions in Figures 2–5, follows from Equation (6) if $a = 6$, $b = 1$, $w_1 = 0.5$, $w_2 = w_3 = 0.25$ ($w_2 + w_3 = 0.5$), and $w_4 = w_5 = 0$. That prior model has a correlation of 0.883 with the

¹¹ In the within-subjects design, the highest salary, x_{mk} , can vary from trial to trial, independent of the minimum, in which case, $x_{mk} - x$ is not perfectly correlated with $(x - x_{0k}) / (x_{mk} - x_{0k})$ for the participant.

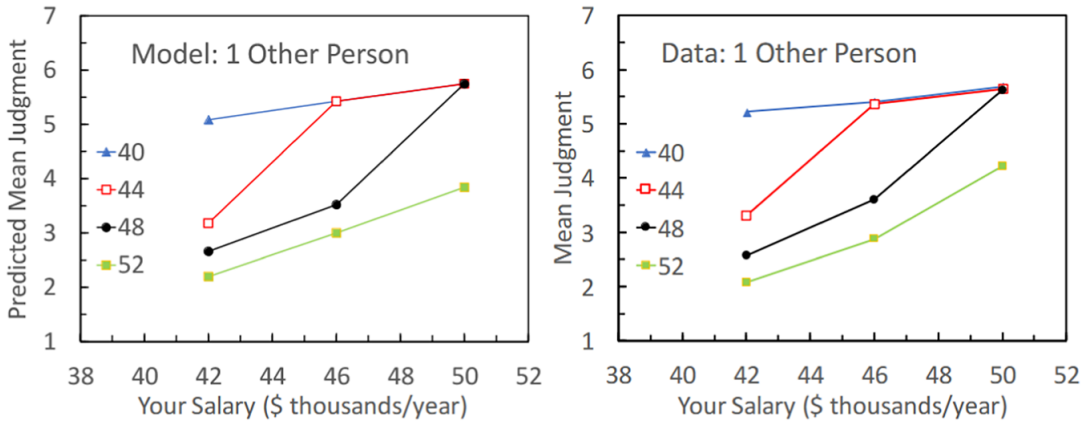


Figure 12. Left panel: Predicted judgments of Revised WSRF Model as a function of Other's salary, with a separate curve for each level of Your Salary. Corresponding data from Experiment 3 are shown in panel on the right.

mean judgments of Experiment 3 (Figures 6–9), with a sum of squared deviations between calculated predictions and observed mean judgments of 11.48.

If we solve for least-squares parameters for the WSRF model (with $w_4 = w_5$ fixed to zero) in Experiment 3, the estimates are: $a = 5.35$, $b = 1.09$, $w_1 = 0.58$, $w_2 = 0.15$, and $w_3 = 0.27$. In this case, the sum of squared errors is only slightly reduced from 11.48 to 9.10, with $r = 0.893$ between predictions of WSRF and data.

When all 5 weights are estimated in the revised model and fit to Experiment 3, however, the sum of squared deviations drops substantially, from 9.10 to 0.86, with $r = 0.990$ between theory and data. For the revised model, the best-fit values of the parameters are $a = 5.83$, $b = 0.86$, and w_1 to w_5 are 0.38, 0.21, 0.04, 0.37, and 0, respectively.

Figure 12 shows the best-fit predictions of the revised model of Equation (6) corresponding to the data of Figure 9, where the violations of WSRF (Equation (5)) were observed. The unrevised model of Equation (5) implies that the curves (representing Others' salaries) should have coincided when Your Salary is lowest, independent of what the Other is paid. The revised model now provides a much better description of the data, since it correctly predicts that when one is paid least, satisfaction varies inversely with the Other's salary. The predicted mean judgments for other cases are presented in the Appendix and Supplementary Material.

This revised model (Equation (6)), with parameters estimated from the data of Experiment 3, had a cross-validation correlation of $r = 0.989$ with the mean satisfaction ratings in Experiment 4, averaged over focus instructions. Figures of predictions (as in Figures 10 and 11) show that the revised model reproduces the main features of the data of both experiments quite well, including the pattern in Figure 9 and in the upper panels of Figures 10 and 11 that showed systematic deviations from the original WSRF model of Equation (5).

Essentially, one parameter, w_4 , representing the inequities, reduced the deviations to less than one-tenth the previous value.¹²

This new model was also fit separately to all four sets of data from Experiment 4. Correlations of fit in the four cases ranged from 0.984 to 0.990. The estimated weights, shown in Table 1, are similar to those estimated in Experiment 3 with some slight differences. As in Experiment 3, the estimated $w_5 = 0$ in all cases (hence not shown in the table). The estimates of w_1 (weight of the BT context

¹²Because the study did not manipulate the distribution of differences, the range and frequency terms of the inequity term are highly correlated, so we cannot properly distinguish between contributions of the range and frequency components of the inequity term from this analysis.

Table 1. Estimated weights in the revised WSRF model (Equation (6)).

Experiment 4	Parameter			
Condition	w_1	w_2	w_3	w_4
Satisfaction BT	0.40	0.19	0.05	0.36
Satisfaction WT	0.34	0.21	0.02	0.43
Job acceptance BT	0.45	0.13	0.05	0.37
Job acceptance WT	0.37	0.17	0.00	0.46

Note: Estimated $w_5 = 0$ in all cases. $u(x) = x$.

RF values) was always substantial, slightly higher when instructions emphasized the BT context, and slightly higher for job acceptance than for salary satisfaction. The estimates of the WT inequity effect, w_4 , was always substantial, and slightly higher when the instructions emphasized WT context.¹³

According to the revised model, people are less satisfied when they are paid less than the highest-paid, equally qualified person doing the same work, and the degree of dissatisfaction depends on the size of the inequity compared to the distribution of inequities. That is, a discrepancy of \$2K should be more injurious to satisfaction if most of the other people deviated by less than \$2K from the maximum and the same discrepancy would have less of an effect if many others were underpaid by even greater amounts. This implication has not yet been tested, because D was fixed in this study.

Using systextual design (Birnbaum, 1975, 2007), it would be possible to independently manipulate the distribution of differences, $d_D(x_k) = x_{mk} - x$, holding the marginal distributions of x and x_{mk} fixed. Such experimental manipulations would allow one to test this implication of the revised model.

Before conducting this research, we thought of salary satisfaction as a judgment of how one’s salary compares to the distribution of salaries in a person’s context. We now think that salary satisfaction introduces something extra, namely, ‘inequity’ that is represented by the additional ingredient to the WSRF model. The unrevised WSRF model may still be descriptive of psychophysical magnitude judgments, which lack the extra components of ‘inequity’, envy, or jealousy aspects of salary satisfaction. Suppose we had asked participants to judge how ‘big’ each salary was in relation to either the other salaries within each trial (WT), or in relation to all of the salaries in the entire study (BT). It seems plausible that such instructions might lead to data that fit the WSRF model better without the inequity revision. Thus, we theorize that the instruction to judge magnitudes—e.g., ‘size’ of salaries rather than satisfaction with salaries—might reduce or eliminate the effect of inequities.

4.3. Equity research

In the present research, people were asked to examine lists of salaries and to judge how satisfied they would be to receive each salary, assuming that all of the people were doing the same job and were of equal merit, productivity, experience, etc. In contrast, equity research deals with the question of how rewards are or should be assigned to people who are not equally deserving.

The study of equity explores the questions of what people deserve and the consequences that follow when people receive undeserved benefits or undeserved punishments (Adams, 1965). What is a ‘just’ (‘fair’) method for distributing salaries or raises to people according to merit? Birnbaum (1983b) noted that some systems used in the ‘real world’ for assigning raises according to merit actually increase inequities over time. He found that people who are asked to assign raises ‘fairly’ according to merit

¹³The estimated weight of the WT range term, w_3 , is larger than in Table 1 when the inequity term is forced to have zero weight ($w_4 = w_5 = 0$), as one would expect from the fact that both terms are correlated functions of x . Similarly, w_4 has a larger, positive weight when w_3 is fixed to zero, as one would expect since both are monotonic functions of the inequity and D is fixed in this study.

make assignments that reduce inequities, and that they judge the implications of certain so-called ‘merit’ systems to be ‘unfair’.

Mellers (1982, 1986) asked people to adopt a neutral, ‘fair’ point of view and to assign salaries to people from a fixed budget according to merit. She found that their judgments were affected by the frequency distribution of the merits. She also found that when the total amount to distribute was made small, those with the least merit received a larger percentage of the total, as if there was a minimum salary below which the (disinterested) judges were reluctant to go. She concluded that salaries are ‘fair’ when the RF position of the salary in the distribution of salaries matches the RF position of the merit in the distribution of merits, consistent with the RF theory of cross-modality matching (Birnbbaum, 1982; Mellers and Birnbbaum, 1982).

A separate line of research studied evaluations of how payments are divided between two co-workers from the viewpoint of one of the (interested) workers (Bazerman et al., 1992; Loewenstein et al., 1989; Messick and Sentis, 1985). For examples, suppose you worked 2 hours on a project and another person worked equally productively for 4 hours; how satisfied would you be with the boss’s decisions if the boss paid you one half what the other received?; an equal share?; or twice as much as the person who worked twice as many hours? This research concluded that people are very dissatisfied if they receive less than a person who worked less than they did, happy to receive more than another person who did equal work, but people are less satisfied when their payment exceeded that of another person who worked more. Thus, satisfaction with the boss’s division appears to be a mixture of self-interest and equity.

4.4. Choices versus ratings

It is not always the case that if A is judged more favorably than B people will choose A over B in a direct choice. For example, Lichtenstein and Slovic (1971) found cases where people judge the value of Gamble A to be higher than that of B and yet people choose B over A. Such findings are termed ‘reversals of preference’ because two procedures for obtaining a preference between A and B yield opposite conclusions.

Ratings of attractiveness of gambles have also been found to not agree with judgments of the values of gambles (Tversky et al., 1988). Furthermore, two ways of judging value of gambles, buying prices (willingness to pay) and selling prices (willingness to sell), produce different orderings from each other (Birnbbaum and Stegner, 1979; Birnbbaum and Sutton, 1992; Birnbbaum et al., 2016).

(Mellers et al., 1992) conducted diagnostic tests of two theories of preference reversals between ratings and estimates of value, expression theory (Goldstein and Einhorn, 1987) and contingent weight theory (Tversky et al., 1988), and were able to reject these two theories in favor of the theory that ratings and estimations of prices are governed by different operations for aggregating the information describing the gambles.

(Mellers et al., 1992) were able to show two different ways in which ratings and price estimates can be affected by context. First, both ratings and buying prices are higher when the distribution of expected values was positively skewed than when it was negatively skewed. Second, the operation used to combine probability and payoff in ratings could be shifted from additive to multiplicative by including contextual stimuli close to zero, which also induced changes in preference order.

The findings that ratings and price estimates are both affected by context might induce a theorist seeking simplicity to hope that direct choices would not be affected by context. However, it has been shown that the choice between A and B depends on the context provided by a third option, C. (Huber et al., 1982) created multiattribute alternatives, A and B, such that the majority prefers B to A. When a third stimulus, C, that is dominated by A but not by B is added to the choice set, the percentage who choose A over B increases, contrary to many early, context-free theories of choice.

Many theories have been proposed to account for such contextual effects in choice (Evangelidis et al., 2024), including that the subjective values of the attributes depend on their range and frequency distributions (Pettibone and Wedell, 2000; Ronayne and Brown, 2017). Direct choice is more complex

than judgment because, in addition to the marginal distributions of the individual attributes and the joint distribution of combinations, there are also the joint distributions of contrasts among items in the choice set. Only a few aspects of these many possible distributions have yet been manipulated independently, leaving many opportunities for investigators to explore how context affects choices, ratings, and estimations.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/jdm.2025.6>.

Data availability statement. The data of this article are included in the Supplementary Material to this article.

Acknowledgements. Article dedicated to the memory of Allen Parducci, who passed away in August, 2023. We thank Kimmo Eriksson for suggestions on an earlier draft.

Competing interests. The authors declare no competing interests.

References

- Adams, S. J. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 267–299). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60108-2](https://doi.org/10.1016/S0065-2601(08)60108-2)
- Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, 44(3), 211–233. [https://doi.org/10.1016/0001-6918\(80\)90046-3](https://doi.org/10.1016/0001-6918(80)90046-3)
- Bazerman, M. H., Loewenstein, G. F., & White, S. B. (1992). Reversals of preference in allocation decisions: Judging an alternative versus choosing among alternatives. *Administrative Science Quarterly*, 37(2), 220. <https://doi.org/10.2307/2393222>
- 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. (1975). Expectancy and judgment. In F. Restle, R. Shiffrin, N. J. Castellan, H. Lindman, & D. Pisoni (Eds.), *Cognitive theory* (Vol. 1, pp. 107–118). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203781548-7>
- Birnbaum, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 401–485). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203780947>
- Birnbaum, M. H. (1983a). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology*, 96(1), 85–94. <https://doi.org/10.2307/1422211>
- Birnbaum, M. H. (1983b). 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. (2007). Designing online experiments. In A. Joinson, K. McKenna, T. Postmes, & U.-D. Reips (Eds.), *Oxford handbook of internet psychology* (pp. 391–403). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199561803.013.0025>
- Birnbaum, M. H., & Hynan, L. G. (1986). Judgments of salary bias and test bias from statistical evidence. *Organizational Behavior and Human Decision Processes*, 37(2), 266–278. [https://doi.org/10.1016/0749-5978\(86\)90055-5](https://doi.org/10.1016/0749-5978(86)90055-5)
- Birnbaum, M. H., & Mellers, B. A. (1983). Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology*, 45(4), 792–804. <https://doi.org/10.1037/0022-3514.45.4.792>
- Birnbaum, M. H., Parducci, A., & Gifford, R. K. (1971). Contextual effects in information integration. *Journal of Experimental Psychology*, 88(2), 158–170. <https://doi.org/10.1037/h0030880>
- Birnbaum, M. H. & Rouvère, J. (2023). Contextual effects in salary satisfaction. *Judgment and Decision Making*, 18, E31. <https://doi.org/10.1017/jdm.2023.26>
- Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology*, 37(1), 48–74. <https://doi.org/10.1037/0022-3514.37.1.48>
- 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](https://doi.org/10.1016/0749-5978(92)90035-6)
- Birnbaum, M. H., Yeary, S., Luce, R. D., & Zhao, L. (2016). Empirical evaluation of four models for buying and selling prices of gambles. *Journal of Mathematical Psychology*, 75, 183–193. <https://doi.org/10.1016/j.jmp.2016.05.007>
- 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>

- Evangelidis, I., Bhatia, S., Levav, J., & Simonson, I. (2024). 50 years of context effects: merging the behavioral and quantitative perspectives. *Journal of Consumer Research*, 51(1), 19–28. <https://doi.org/10.1093/jcr/ucad028>
- Goldstein, W. M., & Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. *Psychological Review*, 94(2), 236–254. <https://doi.org/10.1037/0033-295X.94.2.236>
- Hammerton, M. (1973). A case of radical probability estimation. *Journal of Experimental Psychology*, 101(2), 252–254. <https://doi.org/10.1037/h0035224>
- Hayes, W. M., & Wedell, D. H. (2023). Testing models of context-dependent outcome encoding in reinforcement learning. *Cognition*, 230, 105280. <https://doi.org/10.1016/j.cognition.2022.105280>
- 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*. Harper & Row. <https://psycnet.apa.org/record/1964-35039-000>
- Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, 9(1), 90–98. <https://doi.org/10.1086/208899>
- Johnson, D. M., & Mullanly, C. R. (1969). Correlation-and-regression model for category judgments. *Psychological Review*, 76(2), 205–215. <https://doi.org/10.1037/h0027227>
- Jones, C., & Aronson, E. (1973). Attribution of fault to a rape victim as a function of respectability of the victim. *Journal of Personality and Social Psychology*, 26(3), 415–419. <https://doi.org/10.1037/h0034463>
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251. <https://doi.org/10.1037/h0034747>
- Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, 89(1), 46–55. <https://doi.org/10.1037/h0031207>
- Loewenstein, G. F., Thompson, L., & Bazerman, M. H. (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology*, 57(3), 426–441. <https://doi.org/10.1037/0022-3514.57.3.426>
- 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., Ordóñez, L., & Birnbaum, M. H. (1992). A change-of-process theory for contextual effects and preference reversals in risky decision making. *Organizational Behavior and Human Decision Processes*, 52(3), 331–369. [https://doi.org/10.1016/0749-5978\(92\)90025-3](https://doi.org/10.1016/0749-5978(92)90025-3)
- Messick, D. M., & Sentis, K. P. (1985). Estimating social and nonsocial utility functions from ordinal data. *European Journal of Social Psychology*, 15(4), 389–399. <https://doi.org/10.1002/ejsp.2420150403>
- 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). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203780947>
- Parducci, A. (1995). *Happiness, pleasure, and judgment: The contextual theory and its applications*. Erlbaum.
- Parducci, A. (2011). Utility versus pleasure: the grand paradox. *Frontiers in Psychology*, 15. <https://doi.org/10.3389/fpsyg.2011.00296>
- Parducci, A., Knobel, S., & Thomas, C. (1976). Independent contexts for category ratings: A range-frequency analysis. *Perception & Psychophysics*, 20, 360–366. <https://doi.org/10.3758/BF03199416>
- Pettibone, J. C., & Wedell, D. H. (2000). Examining models of nondominated decoy effects across judgment and choice. *Organizational Behavior and Human Decision Processes*, 81(2), 300–328. <https://doi.org/10.1006/obhd.1999.2880>
- Putnam-Farr, E., & Morewedge, C. K. (2021). Which social comparisons influence happiness with unequal pay? *Journal of Experimental Psychology: General*, 150(3), 570–582. <https://doi.org/10.1037/xge0000965>
- 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>
- 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>
- Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, 95(3), 371–384. <https://doi.org/10.1037/0033-295X.95.3.371>

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>

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>

Appendix: Results of Experiments 1 and 2

Experiment 1 was designed to investigate simple contexts consisting of only one or two other people doing the same job. The 12 by 4, Others’ salaries by Your Salary factorial design and predictions of the prior WSRF theory (Equation (5)) are shown in Table A.1, and the corresponding observed mean judgments are shown in Table A.2. Two important aspects to note in Table A.1 are (1) cases where higher ratings are predicted for lower salaries, given the prior parameters; and (2) cases where the theory implies that judgments should be equal. For example, when Your Salary is highest or lowest, there should be no effect of what others earn. Because this implication of WSRF theory is independent of the specification of $u(x)$, w , and w_B , it is a strong test of Equation (5).

Consider the first row and first column of Table A.1, where Your Salary is \$40K and the Other’s salary is \$30K. The WSRF predicted rating is 4.82, which exceeds 17 entries in columns to the right, where Your Salary would be higher but contexts are predicted to make those higher salaries less satisfying.¹⁴ The corresponding mean judgments are shown in Table A.2, where in all 17 cases, mean satisfaction for \$40K when the other gets \$30K (4.58) is in fact greater than the mean ratings for higher salaries, as predicted in Table A.1. Similarly, the mean rating of \$40K in Table A.2, when the others earn \$30K and \$32K (4.65), also exceeds the same 17 cases of higher salaries in less favorable contexts.

In cases where Your Salary is higher than salaries of all others, there appears a small effect of other salaries. For example, when Your Salary is \$52K (right-most column of Table 1), there are 8 cases where Your Salary is highest, and the prediction is 6.23. In these cases, the mean ratings in Table A.2 range from 4.71 to 5.10.

Table A.1. Prior WSRF model predictions for Experiment 1.

Others’ salaries	Your Salary			
	40	44	48	52
30	4.82	5.38	5.76	6.23
42	1.82	5.38	5.76	6.23
50	1.82	2.38	2.76	6.23
62	1.82	2.38	2.76	3.23
(30, 32)	4.82	5.38	5.76	6.23
(30, 42)	3.82	5.38	5.76	6.23
(30, 50)	3.32	3.93	4.86	6.23
(30, 62)	3.04	3.79	4.35	5.01
(42, 46)	1.82	3.88	5.76	6.23
(42, 50)	1.82	3.51	4.63	6.23
(42, 62)	1.82	3.28	3.96	4.73
(60, 62)	1.82	2.38	2.76	3.23

Note: WSRF predictions with $u(x) = x$; $w_B = w = 0.5$.
Salaries in \$ thousands/year.

¹⁴Note that when Your Salary is \$40 and the others receive \$30K and \$32K, the predicted rating is also 4.82.

Table A.2. Mean judgments from Experiment 1.

Others' salaries	Your Salary			
	40	44	48	52
30	4.58	4.56	4.81	5.01
42	3.60	4.77	4.81	4.89
50	2.68	3.00	3.72	5.10
62	2.14	1.97	2.14	2.65
(30, 32)	4.65	4.74	4.82	5.02
(30, 42)	3.73	4.49	4.64	5.01
(30, 50)	3.40	3.18	3.71	4.71
(30, 62)	2.71	2.83	2.88	3.49
(42, 46)	3.12	4.06	4.87	5.05
(42, 50)	2.76	3.47	3.93	4.81
(42, 62)	2.29	2.70	3.01	3.78
(60, 62)	1.78	1.89	1.97	2.70

Note: Salaries in \$ thousands/year.

However, when Your Salary is lowest, there is a large effect of Other's salary, contrary to WSRF. For example, when Your Salary is \$40K, and is lower than the Others' salaries, the WSRF model implies that Others' salaries should have no effect. There are 7 predictions of 1.82 in the first column of Table A.1; however, the corresponding mean judgments in Table A.2 decrease from 3.60 to 2.14 as the Other's salary increases from \$42K to \$62K, and it decreases from 3.12 to 1.78 as two Others' salaries increase from (\$42K, \$46K) to (\$60K, \$62K).

Also contrary to WSRF without revision, the mean rating is higher to receive \$40K when the Other's salary is \$42K (3.6) than when Your Salary is \$52K and the Other's salary is \$62K (2.65), $t = 9.76$ and 6.6 in Repetitions 1 and 2, respectively. If one is paid less than others, it is apparently more satisfying to be closer to the top salary than to be farther below it. Experiment 1 therefore observed systematic violations of the WSRF model.

Given these violations of the WSRF model, Experiment 2 was therefore designed to check the results, using different stimulus levels and different participants, and it also tested two other basic implications of WSRF theory: (1) the rank of Your Salary was manipulated holding the endpoints fixed and (2) the range of salaries was manipulated, holding rank of Your Salary fixed. Table A.3 shows the levels of the 12 by 3 design and the mean judgments of Experiment 2.

During data collection for Experiment 2, it was decided to revise part of the design for Experiment 3. The endpoint manipulation was not symmetric in Experiment 2: the lower endpoints were \$26K versus \$40K but the upper endpoints were \$52K versus \$70K. So Experiment 3 used \$20K–\$40K and \$52K–\$72K, making them equal. Furthermore, it was realized that when endpoints are manipulated in contexts with just two others, the ranks and values would not be confounded as they were in the sub-design with 7 others; that is, the rank of a salary of \$42K, \$46K, or \$50K will have the rank of second of three within any of the range conditions, whereas in the sets of 7 of Experiment 2, ranks increased with increasing value. Thus, the revised design (Experiment 3) was considered an improvement, so Experiment 3 replaced Experiment 2 until the end of the semester, when data collection ended.

Mean judgments for Experiments 2 and 3 are shown in Tables A.3 and A.4.

The sub-design of Experiment 2 with positively and negatively skewed distributions (with 7 others) yields results very similar to those in Experiment 3, with a slightly greater effect of context in Experiment 2 than 3.

Table A.3. Mean judgments from Experiment 2.

Others’ salaries	Your Salary		
	42	46	50
40	5.15	5.51	5.77
44	3.12	5.30	5.69
48	2.48	3.40	5.52
52	2.07	2.73	4.04
(26, 52)	3.55	3.87	4.88
(40, 70)	2.26	2.95	3.56
POS(7)	2.81	4.31	5.15
NEG(7)	2.28	2.89	4.81
26-52(7)	2.54	3.68	5.03
26-70(7)	2.49	3.35	4.49
40-52(7)	2.47	3.67	5.10
40-70(7)	2.33	3.36	4.32

Note: Salaries in \$ thousands/year.

Table A.4. Mean judgments from Experiment 3.

Others’ salaries	Your Salary		
	42	46	50
40	5.22	5.40	5.68
44	3.31	5.37	5.64
48	2.57	3.60	5.62
52	2.08	2.88	4.22
(20, 52)	3.24	3.66	4.76
(20, 72)	2.74	2.85	3.48
(40, 52)	2.73	3.69	4.79
(40, 72)	2.27	2.65	3.20
(40, 43, 52)	2.71	3.86	4.87
(40, 49, 52)	2.47	3.20	4.82
POS(7)	2.82	4.21	5.18
NEG(7)	2.35	2.95	4.87

Note: Salaries in \$ thousands/year.

The manipulation of endpoints in the sub-design with two others of Experiment 2 matches the corresponding portion of this sub-design of Experiment 3: the results are quite similar except Experiment 2 shows a slightly larger effect. The range manipulations in the sub-design with 7 others in Experiment 2 show smaller trends in the predicted directions, compared to the corresponding sub-design with 3 others in Experiment 3.

The sub-design with one Other’s salary was the same in both Experiments 2 and 3 and results are very similar. In both experiments, ratings show minimal effects of the Other’s salary when Your Salary is higher than the Other’s. However, both experiments also confirmed the violation of WSRF from Experiment 1 that when Your Salary is below the Other’s, mean judgments decrease systematically as the shortfall between Your Salary from the Other’s increases. For example, in Table A.3, when Your

Table A.5. Revised model predictions for Experiment 3.

Others' salaries	Your Salary		
	42	46	50
40	5.08	5.42	5.75
44	3.18	5.42	5.75
48	2.66	3.52	5.75
52	2.20	3.00	3.84
(20, 52)	2.96	3.79	4.67
(20, 72)	2.38	2.86	3.33
(40, 52)	2.84	3.72	4.64
(40, 72)	2.30	2.79	3.26
(40, 43, 52)	2.64	3.92	4.84
(40, 49, 52)	2.64	3.52	4.84
POS(7)	2.67	4.15	5.08
NEG(7)	2.41	3.29	4.81

Note: Salaries in \$ thousands/year.

Table A.6. Mean judgments of salary satisfaction and job acceptance (Experiment 4).

	Satisfaction			Job acceptance		
	Your Salary			Your Salary		
Others	42	46	50	42	46	50
40	5.38	5.61	5.93	5.42	5.90	6.08.
44	3.66	5.47	5.80	3.96	5.72	5.88.
48	2.86	3.97	5.77	2.92	4.24	5.87
52	2.25	3.03	4.37	2.40	3.23	4.79.
(20, 52)	3.44	3.84	4.93	3.38	3.70	5.14
(20, 72)	2.89	3.00	3.72	2.82	3.04	3.77
(40, 52)	2.96	3.79	4.97	2.75	3.89	5.09
(40, 72)	2.56	2.80	3.31	2.41	2.76	3.33
(40, 43, 52)	2.94	4.10	5.10	3.02	4.19	5.46
(40, 49, 52)	2.61	3.46	4.90	2.74	3.70	5.29
POS(7)	2.98	4.35	5.31	2.97	4.35	5.48
NEG(7)	2.50	3.21	5.01	2.61	3.23	5.26

Note: Data for 160 participants in Experiment 4.

Salary is \$42K, the mean rating is 3.12 if the Other's salary is \$44K, but only 2.07 if the Other receives \$52K; in Table A.4, the corresponding values from Experiment 3 are 3.31 and 2.08.

Table A.5 shows the predictions of the revised model (Section 4.2) for Experiment 3 (Table A.4). As can be seen in Table A.5, the revised model provides a fairly reasonable description of the empirical data.

Table A.6 shows the mean judgments of satisfaction and likelihood of job acceptance from Experiment 4. Experiment 4 reproduced all of the main features of Experiment 3 in each of the four conditions of different instructions.

In sum, the main findings of the four studies (with different participants, different levels of the stimuli, and slightly different procedures) are consistent; they are compatible with respect to both the apparent successes and failures of the WSRF model.