Relative Deprivation, Personal Income Satisfaction, and Average Well–Being under Different Income Distributions*

Christian Seidl[a], Stefan Traub[a], Andrea Morone[b]

October, 2002

[a]Dept. of Economics, University of Kiel, Germany
[b]ESSE, University of Bari, Italy

Abstract: This paper uses the data gained from an income categorization experiment for five shapes of income distributions to investigate background context effects, relative deprivation, range–frequency theory to explain background context effects, individual income satisfaction versus aggregate well–being, and the dual patterns of income categorization and limen setting. It is shown that background context effects exist and are reflected in relative deprivation. Not all precepts of range–frequency theory can be evidenced. Moreover, we demonstrate a welfare paradox which concerns a contradiction between individual income satisfaction and aggregate well–being. Finally, income categorization and limen setting harbor no response–mode effects, but exhibit conformity.

Keywords: Relative Deprivation; Income Distributions; Income Satisfaction; Context Effects.

JEL Number: D31, D63, C91

*Corresponding author: Christian Seidl, Institut für Volkswirtschaftslehre, Abteilung für Finanzwissenschaft und Sozialpolitik, Christian–Albrechts–Universität zu Kiel, Olshausenstraße 40, D–24098 Kiel, Germany. Tel.: ++49/431 880 3315, Fax: ++49/431 880 4621, e–mail: cseidl@economics.uni–kiel.de.

1 Introduction

Suppose you have a finite support of incomes and 14 equally spaced incomes which cover this whole support. If you ask subjects to assign these 14 incomes to seven categories ranging from "very bad" to "excellent", you will expect them to assign precisely two neighboring incomes in increasing order to the respective categories. Yet, this picture changes dramatically when these equally spaced income stimuli are embedded in sets of adventitious or background income stimuli which serve to create different income distributions. The background context causes subjects to rate the same income stimulus higher if there are only few higher incomes in the respective income distribution and lower if there are many incomes ahead of the considered income in the respective income distribution. Thus, income categorization and, *a fortiori*, income satisfaction, depend on the background context.

When subjects are asked to categorize incomes, they seem to step into the shoes of the income recipients and categorize the respective incomes with respect to *relative deprivation.* Although context dependence of categorization was widely investigated in psychology, it has, to the best of our knowledge, never been systematically studied with respect to the satisfaction with and the categorization of incomes. This is perhaps due to the prevalence of positively–skewed income distributions in virtually all societies. However, it is tempting to examine the effects of relative deprivation of other shapes of income distributions and compare the results. For a given aggregate income in an economy this implies that different patterns of income distributions engender different welfare effects. The present paper canvasses context effects of five different shapes of income distributions.

This paper is organized as follows: Section 2 informs succinctly on research on context dependence, Section 3 presents a short survey on range–

frequency theory, Section 4 describes the experiment, Section 5 discusses its results, and Section 6 concludes. The instructions and the stimulus material of the experiment have been relegated to the Appendix.

2 Context Dependence: A Succinct Appraisal

Parducci (1968, p. 84) observed that acts of wrongdoing are rated more leniently in a context of rather nasty behavior than in a context of mild misbehavior. Experimental research by Birnbaum (1973, 1974b), too, evidenced that subjects tend to judge persons by their worst bad deed.

Birnbaum et al. (1971) presented subjects lines of different length. They found that the effects of any particular line upon the judgment of average length varied inversely with the length of the other lines within the same set. Birnbaum (1974a) investigated subjects' perceptions of the magnitude of numerals. He observed that the categorization of 45 to 47 numerals ranging from 108 to 992 depended decisively on the shape of their distributional arrangement. Birnbaum (1992) found that certainty equivalents of binary lotteries are rated higher when associated with negatively–skewed than with positively–skewed distributions of proposals.

Parducci (1982) observed that subjects' categorization of squares of different size depended decisively on the skewness of the distribution according to which the differently sized squares were presented. In a similar experiment, Mellers and Birnbaum (1982) found that the members of identical sets of squares were assigned to higher categories of darkness (expressed as the number of dots contained in a square) when their presentation was embedded in a positively–skewed dot distribution of other squares than when it was embedded in a negatively–skewed dot distribution.

Notice that these findings, although related to, go beyond mere anchoring effects[1] and simple context effects[2]. They establish a relationship between the shape of the distribution of the presented stimuli and subjects' judgments on a categorial scale. Strong contextual effects exist for category ratings. Parducci (1982) has characterized such effects as a constituent of human behavior:

> I would have little interest in subjects' expressions of value experiences if these did not change with context. A particular income that might have seemed magnificent at an early stage in one's career would seem totally inadequate at a later stage. If a response scale did not reflect this change, it would miss the all important decline in experienced value. (p. 90)

Closer inspection shows that categorization of stimuli depends not only on the shape of the distribution of the stimuli but also on their range. It differs also for closed sets of categories and open–ended categories.

While Luce and Galanter (1963, p. 268) had deplored the lack of a sophisticated theory of category judgments which defines a scale of sensation that is invariant under experimental manipulations, Parducci and associates, upon having noticed that subjects' evaluation and categorization of objects depended on their background context, set to work to develop such a theory, to wit, *range–frequency theory*. It was developed from Parducci's (1965) *li-*

[1] Anchoring has been studied by Hunt and Volkmann (1937), Rogers (1941), McGarvey (1942/43), Helson (1947). For more recent work compare, for example, Tversky 1974, p. 154; Tversky and Kahneman 1974, p. 1128; Quattrone et al. 1984; Northcraft and Neale 1987; Green et al. 1995; Jacowitz and Kahneman 1995.

[2] See, for example, Glaeser and Sacerdote (2000) for sentences for crimes which increase if the victim is considered as "more valuable" or if the offender is considered as "less valuable".

men model and has proved to reveal important insights into category rating (Parducci 1968, 1974, 1982; Parducci et al. 1960; Parducci and Perret 1971; Birnbaum 1973, 1974a, 1974b, 1992; Birnbaum et al. 1971; Mellers 1982, 1986; Mellers and Birnbaum 1982). This theory takes into account that the *distribution of stimuli* in which they are embedded matters for the evaluation and categorization of the very same objects.

3 Range-Frequency Theory: A Short Survey

Range–frequency theory captures the dependence of category assignments on the distribution and the range of stimuli. It comprises equations for the range value, the frequency value, for judgment, and for the category assignment (Parducci 1982, pp. 94–5).

The *range value* R_i of stimulus S_i depends on the value of this stimulus and on stimulus range:

$$R_i := \frac{S_i - \min_j\{S_j\}}{\max_j\{S_j\} - \min_j\{S_j\}}. \tag{1}$$

The *frequency value* F_i of stimulus S_i depends on the rank of this stimulus, r_i, and on the ranks of the largest and smallest stimulus values, N and 1:

$$F_i := \frac{r_i - 1}{N - 1}. \tag{2}$$

The *judgment* of stimulus i, J_i, is modelled as a weighted mean of the range value and the frequency value:

$$J_i := wR_i + (1-w)F_i, \quad 0 \leq w \leq 1. \tag{3}$$

The *category assignment* of stimulus i, C_i, is then the simple transformation:

$$C_i := bJ_i + a, \tag{4}$$

where b denotes the range of possible categories and a the rank assigned to the lowest category, in most cases: 1. Thus, category assignment assumes that categories are equally spaced; adjoining categories differ precisely by 1.

For $w = 0$ only the frequency value matters, that is, the same number of stimuli is assigned to each category in increasing order. For instance, if there were seven categories, then the seventh one of lowest–ranked stimuli would be assigned to the lowest category, and so on.

For $w = 1$ only the range value matters, that is, the range of the stimuli is equally split. Stimuli are assigned to categories according to the limens of the equally–wide intervals of the range of the stimuli. This means that, if S_i is placed in another context with the same minimum value but a higher maximum value of the stimuli, then S_i tends to fall back in judgment and categorization. On the other hand, if the minimum of the stimuli decreases while their maximum remains unchanged, S_i tends to advance in judgment and categorization.

Range–frequency theory considers categorization to be a weighted average of these two components. Thus, it posits that categorization is linear both in stimulus value (range component) and in stimulus rank (frequency component). Arranging stimuli on the abscissa and categories on the ordinate should produce a nonlinear graph if $w > 0$ and if the distribution of stimuli is not uniform. Nonlinearity of this graph is caused solely by the assumption of linearity of categorization in stimulus rank (frequency component). Psychologists sometimes estimated $w = 0.45$ (Parducci et al. 1960, p. 74) or $w = 0.475$ (Birnbaum 1974a, p. 92), sometimes they just adopted $w = 0.5$ (for example, Parducci and Perrett 1971, p. 429).

Tests of range–frequency theory use sundry distributions of stimuli, for example, uniform, symmetrical unimodal, symmetrical bimodal, positively–

skewed and negatively–skewed distributions. The respective distributions are generated either by appropriate spacing and/or appropriate frequency of stimuli (see, for example, Parducci 1965, 1974; Parducci and Perrett 1971), or by embedding a set of (usually equally spaced) stimuli into a superset of adventitious stimuli (see, for example, Mellers 1982, 1986; Mellers and Birnbaum 1982; Parducci 1982) which shape the intended distribution.

For all distributions of stimuli the judgment function becomes steeper where the stimuli are more densely packed. Thus, if the subsets of equally spaced stimuli (which are common to all distributions) are arranged on the abscissa and the mean categorial value on the ordinate, symmetric unimodal distributions produce an S–shaped curve, bimodal distributions produce an ogival–shaped curve, positively–skewed distributions produce a concave curve, and negatively–skewed distributions a convex curve, where the curve of positively–skewed distributions lies above the curve of negatively–skewed distributions. The distance between curves is greater the less categories are admitted. Moreover, subjects tend to exhaust the available categories. If the set of stimuli is truncated, all categories are nevertheless occupied, although relatively more tenuously.

Range–frequency theory has been successfully employed by Mellers (1982, 1986) for the investigation of equity judgments such as equitable salaries or equitable taxation as functions of merit. Mellers winnowed out the "Aristotelian" subjects, that is, those, whose responses conformed with proportionality. For the rest, she placed merit ratings on the abscissa and mean salaries on the ordinate, and received precisely the pattern described in the preceding paragraph (Mellers 1982, pp. 259–261; 1986, pp. 82–86).

In an attempt to rescue his linear equity model (Harris 1976, 1980), Harris (1993) transformed Mellers' merit stimuli to yearly salaries, used these as stimuli, and observed a linear relationship between his stimuli and the equitable salaries. However, when using Mellers' merit design proper as stimuli, he found Mellers' results confirmed. Thus, he concluded that stimulus dimension, too, matters for subjects' behavioral conformity with range–frequency theory. Note, however, that Harris' treatment contains an element of *equitable redistribution* of a given salary structure, which is different from a primordial assignment of salaries according to merit.

4 The Experiment

4.1 Aims and Scope

This paper pursues four aims. Firstly, we examine whether *background context* matters. In other words, we canvass how the categorization of the same set of stimuli systematically depends on the background context. Indeed, categorization of incomes using different distributions of stimuli has never been studied thoroughly. Mellers and Harris examined the judgment of equitable salaries, not income categorization based on different income distributions.

Secondly, we investigate *relative deprivation* by way of income categorization.[3] When subjects categorize incomes, they cannot wholly avoid stepping into the shoes of the income recipients whose incomes they are asked to judge. Thus, they feel relative deprivation of an income position if many incomes

[3]Relative deprivation was introduced by Stouffer et al. (1949), and further elaborated by Runciman (1966). Similar ideas were developed by philosopher Temkin (1986, 1993). Temkin suggests that inequality aversion results from the complaints of income recipients in the low income echelons akin to relative deprivation. In an experimental investigation of the Temkin theory, Devooght (2002) found particular support for the dependence of complaints on the weighted sum of the gaps of incomes in excess of mean income and of mean income.

are encountered which are ahead of this income (likewise, they may feel "relative elation" if the particular income figures among the higher income strata within the income distribution).

Thirdly, we investigate whether *range–frequency theory* is a valid description of the categorization of incomes. Moreover, we focus on the proper weights of the range and frequency components, an issue which has been understudied in earlier research. After deriving the weights of the range and frequency components, we will investigate which income distribution generates most happiness both in terms of personal income satisfaction and aggregate well–being.

Finally, the present study investigates also the reverse side of income categorization, to wit, the *production of the limens of income categories.* In particular, we check whether the structure of the limens matches income categorization.[4] If the limens of income categories depend on the distribution of the presented stimuli, then utility functions of income estimated from such data cannot but reflect the respective pattern.[5]

[4]It seems that only Birnbaum (1974a) had paid attention to the reverse side of income categorization. Instead of asking subjects for the limens of income categories, he asked his subjects for their judgments of their *typical numbers* for each category.

[5]For instance, the Leyden school has ventured to estimate utility functions or individual welfare functions of income from data of limens of income categories. Cf., e.g., van Praag (1968, 1971), van Praag and Kapteyn (1973), van Herwaarden et al. (1977), Kapteyn and van Herwaarden (1980), van Herwaarden and Kapteyn (1981). For a criticism of the Leyden approach cf. Seidl (1994). The present paper offers another explanation of the lognormal hypothesis of the Leyden utility function of income, to wit, that it is a reflection of income categorization stemming from everyday experience with positively–skewed income distributions. In a seminal study, Birnbaum (1974a) reconciled range–frequency theory with the existence of a psychophysical function which is indeed *invariant* with respect to background context effects. In the realm of income, this function is but a utility function of income. In this view, the lognormal utility function emerges as a

4.2 The Experimental Design

The experiment was computerized and told subjects a cover story of the income distribution on a planet called Utopia, inhabited by small green individuals with the UFO as the local currency (see the Appendix). This extraterrestrial story was employed to distort as much as possible any connotation with the extant positively–skewed income distributions and, thereby, provide an unbiased test of context dependence of categorization. For this purpose, we chose a support of 100 and 1,000 UFOs for all income distributions and used Italian subjects who were at the time of the experiment accustomed to a completely different dimension of currency units.

Insert Table 1 about here.

For our experiment, we used five distributions, which were truncated to secure the required finite support: uniform, normal, bimodal (mixture of two normal distributions), positively–skewed (lognormal), and negatively–skewed (negative lognormal). To generate the experimental design, we used the parameters stated in the second and third columns of Table 1. In a first step, we computed the respective truncated distribution functions, divided their range (the unit interval) by 43 and computed the projection of these equally spaced values on the support (the 100–1,000 interval), which produced the mathematical bases of our stimuli.

In order to be able to compare a subset of identical stimuli across the five

manifestation of a unique utility function of income which owes its particular shape to the positively–skewed appearance of empirical income distributions.

experimental designs, we formed a sequence of 14 equally spaced values,[6] which were embedded in 28 adventitious income values which provided the background context of the respective experimental distributions. To accomplish that, we replaced the nearest values in the mathematical bases of the distributions by the values of the equally spaced subsets of stimuli, which formed our experimental design. The mathematical bases and the experimental values of our stimuli are presented in Table A1 in the Appendix. The subsets of identical stimuli across all experimental stimulus distributions are shown in boldface. The right side of Table 1 provides mean, standard deviation, skewness, and kurtosis of the experimental stimulus distributions.

To check whether our manipulation to create the experimental stimulus distributions changed the character of the mathematical distributions, we applied a Wilcoxon signed–ranks test. Table 2 shows us that the hypothesis of identity cannot be rejected.

Insert Table 2 about here.

4.3 Procedure

As a warm–up introduction, subjects were first shown 25 values taken from the mathematical distributions. Then the 42 values of the experimental design were presented to the subjects in a random order, first as a synopsis, and then one at a time. Subjects were asked to assign them to one of the categories *excellent, good, sufficient, barely sufficient, insufficient, bad, very*

[6]We started at 135 UFOs, and formed the sequence using a distance of 64 (in two cases 63) UFOs.

bad. After that, all stimuli were again shown together with the subject's categorization. Subjects were asked to confirm or change their categorization assignment. Thereafter, subjects were asked to provide limens of the seven income categories.

The experiment was administered from April 24, 2001, to May 5, 2001, at the Laboratorio Informatico, Department of Economics, University of Bari, Italy. 250 subjects participated in this experiment, 50 for each of the five distributions. Subjects were only admitted to a single participation. Each subject received a lump–sum reimbursement of 15,000 Italian Lire (about 7.5 EURO. Subjects spent between 6 and 43 minutes to complete the experiment (mean: 16.1 minutes, standard deviation: 6.2).

5 Results

Comparing subjects' primary and revised category assignments we found them to be not significantly different. This allowed us to use only the revised assignments of categories for our analyses.

5.1 Background Context Matters

Table 3 contains the mean (μ) and median (M) assignments of the 14 common stimuli to the seven categories, coded from 1 (very bad) to 7 (excellent). The table shows that the subjects actually exhausted the categories irrespective of the distribution of stimuli because the categories coincide for the tails (135 UFOs and 965 UFOs, respectively).

Insert Table 3 about here.

For our experiment, testing on background–context effects is equivalent to testing on whether the five sets of observations have the same underlying distribution for a given stimulus income. That is, the null hypothesis for stimulus i, $i = \{135, 199, \ldots, 965\}$, is given by

$$\mathcal{H}_0^i \ : \ F_{un}^i(z) = F_{no}^i(z) = F_{bi}^i(z) = F_{ps}^i(z) = F_{ns}^i(z) \ \forall \ z \in \mathbb{R} \ ,$$

where un=uniform, no=normal, bi=bimodal, ps=positively–skewed, ns=negatively–skewed. Since neither normality nor cardinality of the observations hold, we use the (non–parametric) Kruskal–Wallis (KW) test in order to test on background–context effects. The results (χ^2 values and significance levels p) of this test for each of the 14 common stimuli are given in the last two columns of the table.

For the interior common stimuli (191–901 UFOs), we observe considerable background context effects: The respective Kruskal–Wallis tests are significant at the 1% level, except for the 454 UFOs stimulus which is significant at the 10% level. That is, for 12 of 14 tests performed, we have to reject the null hypothesis that the five different sets of observations came from the same distribution. Assuming stochastic independence of the 14 observations (per subject) under the null hypothesis,[7] a supplementary binomial test would strongly reject the null hypothesis that this results from pure chance ($p = .006$).

This subsection demonstrates that background context matters for income categorization. Our test was global in the sense that it did not allow pairwise comparisons. In the next subsection, we are concerned with a directional hypothesis.

[7] This assumption is, of course, not unproblematic.

5.2 Relative Deprivation

The figures in Table 3 show a clear tendency: The positively–skewed distribution by and large exhibits the highest mean assignments, followed by the bimodal distribution, and the uniform and the normal distributions. Under the negatively–skewed distribution, subjects' categorization of incomes turns out worst. Take, for example, the 773 UFOs category: The difference between the positively–skewed and the negatively–skewed distributions amounts to no less than 0.84 categories.

Consequently, we hypothesize that identical income stimuli are perceived to belong to higher evaluation categories if the background context shifts more income mass to lower income brackets or, the other way round, if the background context exhibits more income mass concentrated among higher income strata, then the evaluation of identical income stimuli is downgraded (*relative deprivation*). In order to test on relative deprivation, we compute Spearman's rank correlations between the subjects' categorizations of a stimulus and the number of incomes larger than that stimulus as a measure of relative deprivation. Note that, if alternative measures of relative deprivation such as the sum of incomes exceeding the stimulus etc. are applied, results do not change qualitatively.

Insert Table 4 about here.

Table 4 contains the relevant data and the results of the test. The null hypothesis is

$$\mathcal{H}_0^k \ : \ X^k \text{ and } Y^k \text{ are independent },$$

where X^k and Y^k denote the distributions of the categorizations of stimulus k and the corresponding number of incomes larger than the stimulus (see Table 4), respectively. As can be taken from Table 4, the null hypothesis of independence is rejected for 11 of 14 tests at the 5% significance level and for 12 of 14 test at the 10% significance level. Furthermore, all significant correlations exhibit the right, negative, sign. Again, a binomial test would strongly confirm that this does not result from pure chance ($p = .006$).

Hence, we conclude that relative deprivation is an important factor in the evaluation of incomes: The more incomes exist which exceed the income to be evaluated, that is, the greater the relative deprivation associated with this income is, the worse is this income's categorization for the respective background.

5.3 Range–Frequency Theory

In order to test the empirical performance of range–frequency theory in the context of income categorization, we generalize the judgment equation (3) to:

$$J_i^k = \alpha^k + w_R^k R_i^k + w_F^k F_i^k + u_i^k \ , \tag{5}$$

where the judgment of stimulus i under income distribution k is given by solving the category–assignment equation (4) for J_i^k:

$$J_i^k = \frac{1}{6}(C_i^k - 1) \ . \tag{6}$$

α^k denotes an intercept term, w_R^k and w_F^k denote the weights of the range and the frequency components, respectively, and u_i^k is an error term.

Using equation (5), we can test three postulates of range–frequency theory: The first postulate requires the intercept term α^k to equal zero (*neu-*

trality), that is, we will test

$$\mathcal{H}_0^k \ : \ \alpha^k = 0 \quad \text{vs.} \quad \mathcal{H}_1^k \ : \ \alpha^k \neq 0 \ .$$

The second postulate requires the weights to add up to 1 (*additivity*):

$$\mathcal{H}_0^k \ : \ w_F^k + w_R^k = 1 \quad \text{vs.} \quad \mathcal{H}_1^k \ : \ w_F^k + w_R^k \neq 1 \ ,$$

and the third postulate demands that the weights must be *nonnegative* (if additivity holds this means that they must not exceed one):

$$\mathcal{H}_0^k \ : \ w_F^k, \ w_R^k \geq 0 \quad \text{vs.} \quad \mathcal{H}_1^k \ : \ w_F^k \leq 0, \quad \text{or} \quad w_R^k \leq 0 \ .$$

Furthermore, we can also test on whether different distributions of income stimuli generate different sets of weights (*background context*), that is, we test

$$\mathcal{H}_0^k \ : \ w_R^k = w_R^\ell, \ w_F^k = w_F^\ell \quad \text{vs.} \quad \mathcal{H}_1^k \ : \ w_R^k \neq w_R^\ell, \ w_F^k \neq w_F^\ell \quad \text{for} \quad k \neq \ell.$$

In order to test on additivity, we estimate a restricted equation

$$J_i^k - F_i^k = \alpha^k + w_R^k (R_i^k - F_i^k) + u_i^k \tag{7}$$

in addition to (5) and compute the respective F tests. Background–context dependence is tested by means of a pooled sample and dummy variables. Note that we did not run any regressions for the uniform distribution since range and frequency values coincide [the numerator of the range equation (1) becomes exactly F_i times the denominator of (1)].

Table 5 contains the estimates of the weights using OLS. For every distribution of income stimuli, the table compares the restricted (above) with the unrestricted regression (below). The model summary shows a much better fit of the unrestricted model. Hence, the F test (last column) strongly rejects the null hypothesis of additivity, that is, the restriction $w_F = (1 - w_R)$ does not hold. In all 4 cases, the sum of the estimated coefficients for R_i and F_i slightly exceeds 1.[8] Hence, we focus our attention on the unrestricted model in the following.

Insert Table 5 about here.

With the exception of the positively–skewed distribution, the intercept terms are insignificant as maintained by the neutrality hypothesis. The intercept term of the positively–skewed distribution exhibits a negative sign. This means that a positively–skewed distribution of stimuli biases subjects' categorizations of incomes downwards. That is, although relative deprivation is lowest and, thus, income categorizations are highest under the positively–skewed income distribution, a (relatively small) premium is attached to the judgment function of the positively–skewed income distribution. This result is possibly due to an endowment effect [Tversky and Griffin (1991, p. 117] caused by the relatively low mean income of the positively–skewed income distribution.

Except for the normal distribution, the estimated weights of the range and the frequency components are inside the unit interval, that is, the non-negativity hypothesis cannot be rejected. The weight of the range component amounts to about 0.8, that is, distinctly more weight is given to the range component than to the frequency component. With regard to the normal distribution, we observe a weight of the range component larger than 1. Computing the t value for the null hypothesis $w_R - 1 = 0$ shows, however,

[8]This contradicts, a result obtained by Parducci et al. (1960, p. 75).

that w_R does not differ significantly from 1. On the other hand, the frequency component does not matter at all for the categorization of incomes.

Eventually, we ran an (unrestricted) pooled regression with the positively–skewed distribution as the benchmark case and dummies for the differential intercepts and slopes of the other distributions in order to test on background context. The adjusted R^2 of this regression is .935 ($F = 3645$, $p \leq .01$). As compared to the positively–skewed income distribution, the intercept terms of the normal, the bimodal, and the negatively–skewed distributions are significantly larger (the t values are between 2.090 and 2.863; $p \leq .05$) which confirms that the neutrality hypothesis is rejected only for the positively–skewed income distribution, whose mean income is lowest. Moreover, the pooled regression confirms that the range component is given a significantly greater and the frequency component a significantly smaller weight, respectively, under the normal distribution (the t values are 2.081 and -2.576, respectively; $p \leq .05$). That is, the shape of the normal distribution seems to induce subjects to categorize the stimulus incomes by range alone. The differences between the weights of the bimodal, the positively–skewed, and the negatively–skewed distributions are not significant [bimodal vs. positively skewed: $t = -.345$, $p = .723$ (range), $t = .203$, $p = .839$ (frequency); negatively vs. positively skewed: $t = -1.132$, $p = .258$ (range), $t = 1.181$, $p = .238$ (frequency)]. For these three income distributions, the structural part of income categorization in terms of the weights entering the judgment function is equal and independent of the shape of the income distribution to be judged. In other words, under the bimodal, the positively–skewed, and the negatively–skewed income distribution background context matters for the categorization of incomes but not for the judgment function itself.

Summarizing, we find, first, the neutrality hypothesis of range–frequency theory violated for the positively–skewed income distribution but not for the other income distributions. Second, additivity is violated for all income distributions considered. The component weights are slightly super–additive. The estimates demonstrate that, third, the weights are within the interval $[0, 1]$ and, fourth, not significantly different for the negatively–skewed, the positively–skewed, and the bimodal income distributions but, fifth, far off from values around $w = 0.5$ estimated (and sometimes merely assumed) by psychologists (for example, Parducci et al. 1960, p. 74; Parducci and Perrett 1971, p. 429; Birnbaum 1974a, p. 92). Instead, the weight of the frequency component is much smaller, being close to .2.

Two reasons can account for the low weight of the frequency component. Firstly, Harris' (1993) conjecture can have something in it. Using incomes instead of ratings could have moved subjects' behavior closer to the linear model. However, relying on real monetary values, Mellers (1986) observed pronounced curvatures of the judgment functions in her work on equitalbe taxes. Also Parducci et al. (1960) and Birnbaum (1974a) found distict curvatures of the judgment functions of experiments on a size categorization of numerals which ranged within the interval from 108 to 992 (similar to the support of the income distributions used for our experiment).

Secondly, recall that Mellers (1982, 1986) winnowed out the subjects with Aristotelian equity values (who endorsed proportionality for distributive justice). This comes up to the elimination of all subjects who behaved in conformity with the range component only. This had somewhat increased the influence of the frequency component.

5.4 Income Satisfaction versus Well–Being: A Paradox

Whereas psychologists construct the graphs of the judgment functions or the category assignment functions for the common stimuli only, using the mean category assignments as exhibited in Table 3, we construct the graphs of the judgment functions for all 42 stimulus values using the estimates of the unrestricted weights of the range and the frequency components. The respective graphs are shown in Figure 1.

Insert Figure 1 about here.

This figure confirms the message conveyed by the entries in Table 3: The graph of the judgment function of the positively–skewed income distribution exhibits a concave shape and dominates all other judgment functions up to incomes of about 800 UFOs. The graph of the judgment function of the negatively–skewed income distribution exhibits a convex shape and is dominated by all other judgment functions over the whole interval of stimulus incomes. For the judgment functions of the normal and the bimodal income distributions we observe linear and S–shaped graphs, respectively. The latter two intersect several times, and lie for most incomes between the graphs of the judgments functions of the positively–skewed and the negatively–skewed income distributions. For incomes above about 800 UFOs the graph of the judgment function of the bimodal income distributions dominates all other income distributions. Thus, a positively–skewed income distributions generates the highest income satisfaction for small and moderate incomes. Concerning the top incomes, the highest income satisfaction is conveyed by a bimodal income distribution. Under a negatively–skewed income distribution, personal income satisfaction turns out to be lowest. These observations are perfectly in line with our previous result that a positively–skewed income distribution generates less relative deprivation than a negatively–skewed one.

Notice that income satisfaction is inverse to the means of the distributions. Mean income is highest for the negatively–skewed distribution, yet income satisfaction is lowest. For the positively–skewed distribution, the mean income is lowest, yet income satisfaction is highest. The mean income of the other three distributions is not much different among them and lies in between, as does by and large income satisfaction. Does this imply that the positively–skewed income distributions, which prevail in the real world, are able to elicit the highest income satisfaction from a given aggregate income? Our experiment even suggests that the negatively–skewed distribution elicits the minimum individual income satisfaction from the maximum total income.

However, greater individual income satisfaction does not necessarily imply a higher level of well–being or social welfare within the respective society. Rather do we have to aggregate the individual welfare of the income recipients. Applying a Harsanyi-type social welfare function, we sum individual income satisfaction from below and divide the partial sums by the number of income recipients whose incomes do not exceed y_i. Formally, we have

$$\bar{W}(y_i) = \frac{1}{i} \sum_{j=1}^{i} J_j(y_j) \; . \tag{8}$$

Accordingly, the graph of $\bar{W}$ shows average social welfare for all income recipients disposing of an income of y_i or less. Figure 2 graphically depicts the average well–being of the society under different income distributions.

Insert Figure 2 about here.

Figure 2 shows that average well–being is highest under a bimodal income distribution for those income recipients who do not dispose of more than about 400 UFOs. If we consider also better incomes between 400 and 800 UFOs, then the Utopians are best off with a normal income distribution. Eventually, if we take into account the top earners as well, the negatively–skewed income distribution generates greatest average well–being.

Comparing the graphs of the judgment functions and average well–being of the positively–skewed and the negatively–skewed income distributions, we strike a paradoxical situation: Under a positively–skewed income distribution every single income recipient, even the top earners, experiences higher individual income satisfaction than under a negatively–skewed income distribution; yet, for each stratified subset of subjects, average well–being under a negatively–skewed income distribution exceeds average well–being under a positively–skewed income distribution.

This is akin to an observation made by Camacho–Cuena, Seidl, and Morone (2002). When subjects had to assess income distributions as a whole from under a veil of ignorance, they seem to pay attention to all possible incomes to which they may be attributed within an income distribution. This affects their ratings of income distributions: Even for income distributions with identical means, negatively–skewed distributions are rated distinctly higher than positively–skewed distributions, possibly because they offer the better chance to end up at a comparatively satisfactory income level.[9]

[9]For ample experimental evidence see Camacho–Cuena, Seidl, and Morone (2002), who observed also a preference reversal phenomenon [cf. Seidl (2002)] between the rating and the evaluation of income distributions. For the context of the present paper, categorization

5.5 Pattern of Limens

After subjects had categorized the 14 stimulus incomes, they were told that, for the purpose of future use in Utopia's statistical office, they should state limens for the seven income categories. Moreover, they were told that the limens should properly reflect the income distribution prevailing in Utopia. Note that subjects were not forced to express consistent behavior in the sense that the upper limen of of a category had to be equal to the lower limen of the following category.[10]

We observed only one category overlap[11], but several empty intervals between category limens.[12] What might have prompted subjects to behave in this way? They are too many to explain their behavior simply by error, even more so as these subjects made their responses without overlaps between limens, but empty intervals between them. Therefore, it seems as if these subjects took our question under the proviso of making entirely unambiguous statements such as: "An income between 405 and 506 UFOs is certainly insufficient. But for incomes between 331 and 404 UFOs I am not entirely

of incomes is more akin to rating than to evaluation. Beckman, Formby, Smith, and Zheng (2002) observed less opposition to Pareto-improving moves of income distributions when subjects make their judgments under a veil of ignorance. For known positions, opposition against extra income is highest for income gains of persons in a higher income echelon, less for persons in a lower income echelon, and least for own extra income. This finding matches with our results for individual income satisfaction in the present paper.

[10]This is in contrast to the surveys of the Leyden school, where subjects could indicate only one of these two figures.

[11]This one instance seems to be an error because this same subject exhibited empty intervals for the other categories.

[12]Among our 50 subjects per distribution, we observed 12 subjects with empty intervals for the uniform distribution, 10 for the symmetric distribution, 11 for the bimodal distribution, 9 for the positively–skewed distribution, and 14 for the negatively–skewed distribution.

sure whether they are still bad or already insufficient. Likewise, for incomes between 507 and 598 UFOs, I am not entirely sure whether they are still insufficient. Therefore, to be on the safe side, I make statements only for those areas for which I am entirely confident."[13]

Insert Table 6 about here.

Table 6 lists the means (μ) and medians (M) of the lower and upper limens of the seven income categories for the five income distributions. In analogy to Table 3, the test on background context effects with respect to income categorization, the Kruskal–Wallis test (see the last two columns of Table 6), shows that background context matters. For 9 of the 12 tests conducted, the null hypothesis of the 5 sets of observations coming from the same distributions has to be rejected ($p \leq .10$).

Comparing Table 6 with Table 3 and confining ourselves to the middle limens (from "bad" to "sufficient"), we see that the negatively–skewed distribution of stimuli, which exhibits the lowest category assignments and, therefore, income satisfaction, in Table 3, exhibits the highest limens in Table 6. It is followed rather indiscriminately by the uniform and the symmetric distributions which ranks third in Table 3, and then by the bimodal distribution, which occupies rank two in Table 3. The positively–skewed distribution of stimuli, which exhibits the highest category assignments in Table 2, shows

[13]Obviously Birnbaum (1974a) had anticipated such an attitude. Wisely, he asked his subjects only for their "typical numbers" for each category. Indeed, if in everyday life one asks subjects for sufficient incomes, one often gets a representative income level as an answer rather than an income interval.

the lowest limens in Table 6. Thus, the ordering of the limens corresponds by and large with the category assignments; subjects behaved consistently for both sides of the medal. This reflects again the influence of the background context on the perception of income limens for the calibration of categories. If the background context exhibits more income mass concentrated among the higher income brackets, subjects become more exacting, which shifts the limens of income categorization in the direction of higher incomes. However, if more income mass is concentrated among the lower income brackets, subjects become more humble as to income categorization, that is, categorial limens are shifted in the direction of lower incomes. Background context of stimuli matters also with respect to the perception of limens of income categorization.

This shows that limen setting reflects *relative deprivation:* Limens are higher the more incomes are ahead of the limen incomes. On the other hand, inspection of Tables 6 and A1 reveals that the *total income level*, too, matters. A modest endowment effect is, therefore, also at work. Higher income levels are capable of compensating for enduring more better–off income recipients, which constitutes the second influence on limen setting.

6 Conclusion

This paper uses the data gained from an income categorization experiment to investigate background context effects, relative deprivation, range–frequency theory to explain background context effects, individual income satisfaction versus aggregate well–being, and the dual patterns of income categorization and limen setting.

Five groups of 50 subjects were asked to assign 14 common income stimuli to seven income categories. These common stimuli were embedded in 28 adventitious stimuli to form five different income distributions, uniform, normal, bimodal, positively–skewed, and negatively–skewed. Each distribution was presented to a group of subjects.

Firstly, we found that background context matters. Using a Kruskal–Wallis test, we had to reject the hypothesis that the five different sets of observations of income categorization came from the same distribution. This means that the background of the 28 adventitious income stimuli had influenced income categorization.

Secondly, we investigated the direction of background context effects, which led us to discover that relative deprivation is at work to shape the pattern of background context. Spearman's rank correlations between income categorization and the number of incomes ahead of the respective stimuli shows a significantly positive relationship. Thus, identical income stimuli are perceived to belong to higher evaluation categories if the background context shifts more income mass to lower income brackets, or, the other way round, if the background context exhibits more income mass concentrated among higher income strata, then the evaluation of identical income stimuli is downgraded.

Thirdly, background context effects have been explained by means of range–frequency theory, which posits that the categorization of a stimulus is a weighted mean of this stimulus' range and frequency component. We found that neutrality is violated for the positively–skewed distribution, which reflects the working of a modest endowment effect. Furthermore, the weights are slightly super–additive and nonnegative. The frequency component is ruled out for the normal distribution. For the negatively–skewed, the positively–skewed, and the bimodal income distributions, the weight of the frequency component is about .2 and not significantly different for the negatively–skewed, the positively–skewed, and the bimodal income distributions. This result is remarkable, because for the negatively–skewed, the positively–skewed, and the bimodal income distributions background context matters for the categorization of incomes, but not for the judgment function itself.

Fourthly, we struck a paradox between individual income satisfaction and aggregate well–being. Whereas the judgment functions show that individual income satisfaction is highest for the positively skewed income distribution and lowest for the negatively skewed income distribution, a Harsanyi–type social welfare function demonstrates that average aggregate well–being is for all stratified subsets of subjects higher for the negatively skewed income distribution than for the positively skewed income distribution. This paradox results from the weighting of income satisfaction with the frequency of the involved subjects.

Finally, we found that limen setting of income categories provides a picture which is perfectly consistent with income categorization. This demonstrates that response–mode effects are absent for experiments on income categorization on the one hand, and limen setting on the other.

Acknowledgements

This research was supported by a grant from the European Commission under contract ERBFMRXCT 980248. We are indebted to Maria Vittoria Levati for having written the computer program of our experiment and for her suggestions concerning our experimental treatment, as well as to the participants of the Third Annual Meeting of LivinTaX in Madrid for an encouraging

discussion. The usual disclaimer applies. The data of this experiment are available from Stefan Traub, e–mail: traub@economics.uni-kiel.de.

7 References

Beckman, S. R., Formby, J. P., Smith W. J., Zheng, B. (2002), Envy, Malice and Pareto Efficiency: An Experimental Examination, *Social Choice and Welfare* 19, 349–67.

Birnbaum, M. H. (1973), Morality Judgement: Test of an Averaging Model with Differential Weights, *Journal of Experimental Psychology* 99, 395–9.

Birnbaum, M. H. (1974a), Using Contextual Effects to Derive Psychophysical Scales, *Perception & Psychophysics* 15, 89–96.

Birnbaum, M. H. (1974b), The Nonadditivity of Personality Impressions, *Journal of Experimental Psychology Monograph* 102, 543–61.

Birnbaum, M. H. (1992), Violations of Monotonicity and Contextual Effects in Choice–Based Certainty Equivalents, *Psychological Science* 3, 310–4.

Birnbaum, M. H., Parducci, A., Gifford, R. K. (1971), Contextual Effects in Information Integration, *Journal of Experimental Psychology* 88, 158–70.

Camacho–Cuena, E., Seidl, C., Morone, A. (2002), Income Distributions Versus Lotteries: Happiness, Response–Mode Effects, and Preference Reversals, Manuscript, Kiel: University of Kiel.

Devooght, K. (2002), Measuring Inequality by Counting ‘Complaints’: Theory and Empirics, Discussion Paper No. DARP 59, London: STICERD, London School of Economics.

Glaeser, E. L., Sacerdote, B. (2000), The Determinants of Punishment: Deterrence, Incapacitation and Vengeance, Discussion Paper No. 1894, Harvard Institute of Economic Research, Cambridge, MA.

Green, D., Jacowitz, K. E., Kahneman, D., McFadden, D. (1995), Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods, Dept. of Economics, University of California, Berkely.

Harris, R. J. (1976), Handling Negative Inputs: On the Plausible Equity Formulae, *Journal of Experimental Social Psychology* 12, 194–209.

Harris, R. J. (1980), Equity Judgements in Hypothetical, Four–Person Partnerships, *Journal of Experimental Social Psychology* 16, 96–115.

Harris, R. J. (1993), Two Insights Occasioned by Attempts to Pin Down the Equity Formula, in: B. A. Mellers and J. Baron (ed.), *Psychological Perspectives on Justice,* Cambridge University Press, Cambridge, 32–54.

Helson, H. (1947), Adaption–Level as Frame of Reference for Prediction of Psychophysical Data, *American Journal of Psychology* 60, 1–29.

Hunt, W. A., Volkmann, J. (1937), The Anchoring of an Affective Scale, *American Journal of Psychology* 49, 88–92.

Jacowitz, K. E., Kahneman, D. (1995), Measures of Anchoring in Estimation Tasks, *Personality and Social Psychology Bulletin* 21, 1161–6.

Kapteyn, A., van Herwaarden, F.G. (1980), Independent Welfare Functions and Optimal Income Distribution, *Journal of Public Economics* 14, 375–97.

Luce, R. D., Galanter, E. (1963), Psychophysical Scaling, in: R. D. Luce, R. R. Bush and E. Galanter (ed.), *Handbook of Mathematical Psychology, Vol. I,* Wiley, New York, 245–307.

McGarvey, H. R. (1942/43), Anchoring Effects in the Absolute Judgement of Verbal Materials, *Archive of Psychology* 39, 1–86.

Mellers, B. A. (1982), Equity Judgement: A Revision of Aristotelian Views, *Journal of Experimental Psychology: General* 111, 242–70.

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

Mellers, B. A., Birnbaum, M. H. (1982), Loci of Contextual Effects in Judgement, *Journal of Experimental Psychology: Human Perception and Performance* 8, 582–601.

Northcraft, G. B., Neale, M. A. (1987), Experts, Amateurs, and Real Estate: An Anchoring–and–Adjustment Perspective on Property Pricing Decisions, *Organizational Behavior and Human Decision Processes* 39, 84–97.

Parducci, A. (1965), Category Judgement: A Range–Frequency Model, *Psychological Review* 72, 407–18.

Parducci, A. (1968), The Relativism of Absolute Judgements, *Scientific American* 219/6, 84–90.

Parducci, A. (1974), Contextual Effects: A Range–Frequency Analysis, in: E. C. Carterette and P. Friedman (ed.), *Handbook of Perception, Vol. II: Psychological Judgement and Measurement*, Academic Press, New York–San Francisco–London, 127–41.

Parducci, A. (1982), Category Ratings: Still More Contextual Effects! in: B. Wegener (ed.), *Social Attitudes and Psychophysical Measurement*, Lawrence Erlbaum Associates, Publishers, Hillsdale, NY, 88–105.

Parducci, A., Calfee, R. C., Marshall, L. M., Davidson, L. P. (1960), Context Effects in Judgement: Adaptation Level as a Function of the Mean, Midpoint, and Median of the Stimuli, *Journal of Experimental Psychology* 60, 65–77.

Parducci, A., Perrett, L. F. (1971), Category Rating Scales: Effects of Relative Spacing and Frequency of Stimulus Values, *Journal of Experimental Psychology Monograph* 89, 427–52.

Quattrone, G. A., Lawrence, C. P., Finkel, S. E., Andrus, D. C. (1984), Explorations in Anchoring: The Effects of Prior Range, Anchor Extremity, and Suggestive Hints, Working Paper, Stanford University, Stanford, CA.

Rogers, S. (1941), The Anchoring of Absolute Judgements, *Archive of Psychology* 37.

Runciman, W. G. (1966), *Relative Deprivation and Social Justice*, Henley: Routledge and Kegan Paul.

Seidl, C. (1994), How Sensible is the Leyden Individual Welfare Function of Income? *European Economic Review* 38, 1633–59.

Seidl, C. (2002), Preference Reversal, *Journal of Economic Surveys*. [forthcoming]

Stouffer, S. A., Suckman, E. A., Devinney, L. C., Star, S. A., Williams, R. M. (1949), *The American Soldier: Adjustment During Army Life*, Vol. 1, Princeton University Press, Princeton.

Temkin, L. S. (1986), Inequality, *Philosophy and Public Affairs* 15, 99–121.

Temkin, L. S. (1993), *Inequality*, Oxford: Oxford University Press.

Tversky, A. (1974), Assessing Uncertainty, *Journal of the Royal Statistical Society* 36, 148–67.

Tversky, A., Griffin, D. (1991), Endowment and Contrast in Judgments of Well–Being, in: F. Strack, M. Argyle, and N. Schwarz (ed.), *Subjective Well–Being*, Pergamon Press, Oxford, 101–118.

Tversky, A., Kahneman, D. (1974), Judgement Under Uncertainty: Heuristics and Biased, *Science* 185, 1124-31.

van Herwaarden, F.G., Kapteyn, A., van Praag, B.M.S. (1977), Twelve Thousand Individual Welfare Functions, A Comparison of Six Samples in Belgium and the Netherlands, *European Economic Review* 9, 283–300.

van Herwaarden, F.G., Kapteyn, A. (1981), Empirical Comparisons of the Shape of Welfare Functions, *European Economic Review* 15, 261–68.

van Praag, B.M.S. (1968), *Individual Welfare Functions and Consumer Behavior*, North Holland, Amsterdam.

van Praag, B.M.S. (1971), The Welfare Functions of Income in Belgium: An Empirical Investigation, *European Economic Review* 2, 337–69.

van Praag, B.M.S., Kapteyn, A. (1973), Further Evidence on the Individual Welfare Functions of Income: An Empirical Investigation in the Netherlands, *European Economic Review* 4, 33–62.

Tables

Table 1 Descriptive Statistics of Experimental Distributions

	Parameters for generation		Moments of experimental distributions			
Distribution	μ	σ	Mean	Std.dev.	Skewn.	Kurt.
Uniform	(550)	(260)	550	258	.000	-1.200
Normal	550	230	549	198	.000	$-.544$
Bimodal[a]	325, 775	100	550	241	.000	-1.479
Positively–skewed[b]	6	1	408	230	.738	$-.349$
Negatively–skewed[c]	6	1	692	230	$-.735$	$-.348$

Table note. All distributions truncated at 100 on the left and at 1000 on the right.

[a]Mixture of two normal densities.

[b]Lognormal density.

[c]Lognormal density with $x^{-} = 1100 - x$.

Table 2 Test on Equality of Mathematical and Experimental Distributions

	Wilcoxon test	
Distribution	Z	p
Uniform	−.210	.834
Normal	−.280	.779
Bimodal	−.454	.649
Positively–skewed	−.314	.760
Negatively–skewed	−1.044	.296

Table note. Z statistic and significance level (p) of the Wilcoxon signed–ranks test. $n = 42$.

Table 3 Test on Background Context Effects — Kruskal–Wallis (KW) test

	Uniform		Normal		Bimodal		Pos.–sk.		Neg.–sk.		KW test	
Stim.	μ	M	μ	M	μ	M	μ	M	μ	M	χ^2	p
135	1.06	1	1.04	1	1.10	1	1.00	1	1.06	1	5.334	.255
199	1.38	1	1.18	1	1.36	1	1.48	1	1.14	1	13.399	.009
263	2.02	2	1.98	2	2.00	2	2.24	2	1.76	2	23.966	.000
327	2.42	2	2.40	2	2.56	3	2.78	3	2.28	2	21.353	.000
390	2.84	3	2.80	3	2.76	3	3.10	3	2.46	2	32.927	.000
454	3.48	4	3.26	3	3.44	3	3.58	4	3.24	3	8.624	.071
518	3.80	4	3.96	4	4.04	4	4.16	4	3.82	4	15.288	.004
582	4.12	4	4.16	4	4.16	4	4.42	4	3.86	4	27.036	.000
646	4.46	5	4.76	5	4.74	5	4.88	5	4.24	4	45.872	.000
710	5.24	5	5.10	5	5.40	5	5.46	5	4.78	5	31.661	.000
773	5.46	5.5	5.56	6	5.68	6	5.80	6	4.96	5	40.838	.000
837	6.04	6	6.02	6	6.32	6	6.18	6	5.92	6	18.646	.001
901	6.76	7	6.54	7	7.00	7	6.92	7	6.86	7	16.377	.003
965	6.96	7	6.98	7	7.00	7	7.00	7	6.98	7	4.065	.397

Table note. μ =mean, M =median. $n = 50 \times 5$, $df = 4$ for all tests.

Table 4 Test on Relative Deprivation

	# of incomes larger than the stimulus					Rank correlation[a]	
Stim.	un	no	bi	ps	ns	r_s	p
135	40	41	41	39	41	.100	.116
199	37	40	40	33	40	-.183	.004
263	34	38	36	27	39	-.277	.000
327	31	35	31	22	38	-.274	.000
390	28	32	26	18	36	-.298	.000
454	25	28	23	15	34	-.171	.007
518	22	23	21	12	32	-.188	.003
582	19	18	20	9	29	-.274	.000
646	16	13	18	7	26	-.359	.000
710	13	9	15	5	23	-.262	.000
773	10	5	10	3	18	-.353	.000
837	7	3	5	2	14	-.122	.055
901	4	1	1	1	8	-.183	.004
965	1	0	0	0	2	-.044	.491

Table note. un=uniform, no=normal, bi=bimodal, ps=positively–skewed, ns=negatively–skewed. $n = 250$ for all tests.

[a]Spearman's rank correlation between the number of incomes larger than the stimulus and the categorization of that stimulus.

Table 5 OLS Estimation of Weights

Coefficients			95% CI w_R	Model summary		Test on
α	w_R	w_F	95% CI w_F	F^a	R^2	additivity^b
			Normal distribution			
**.010	**.939	(.061)	[.853, 1.025]	460.986	.398	
.003	*.044*		—	.000		
-.001	**1.034	-.012	[.895, 1.173]	5814.075	.943	6664.298
.007	*.071*	*.061*	[−.131, .108]	.000		.000
			Bimodal distribution			
**.019	**.765	(.235)	[.612, .918]	96.151	.121	
.003	*.078*		—	.000		
.004	**.823	**.206	[.665, .981]	5357.242	.939	9346.656
.006	*.080*	*.078*	[.052, .359]	.000		.000
			Positively–skewed distribution			
-.001	**.850	(.150)	[.777, .923]	523.337	.428	
.006	*.037*		—	.000		
**-.028	**.855	**.187	[.783, .927]	5658.250	.942	6176.862
.009	*.037*	*.038*	[.112, .262]	.000		.000
			Negatively–skewed distribution			
*.015	**.783	(.217)	[.692, .874]	284.062	.289	
.008	*.046*		—	.000		
-.002	**.791	**.256	[.701, .882]	3717.255	.914	5065.407
.009	*.046*	*.047*	[.163, .348]	.000		.000

Table note. $n = 700$. *$p \leq .10$, **$p \leq .05$; tested against 0. Above: restricted model; below: unrestricted model. Standard errors in italics.

^{a}First row: F value, second row: significance level.

^{b}F value and significance level.

Table 6 Limens Under Different Income Distributions — Kruskal–Wallis (KW) Test

		Uniform		Normal		Bimodal		Pos.–sk.		Neg.–sk.		KW test	
Category	Limen	μ	M	μ	M	μ	M	μ	M	μ	M	χ^2	p
Very bad	Low	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)	—	—
	Up	200	199	204	200	190	200	188	195	205	200	4.230	.376
Bad	Low	219	200	213	200	201	200	196	200	212	200	3.535	.473
	Up	338	310	339	350	335	310	311	300	352	355	12.129	.016
Insufficient	Low	343	332	343	350	335	311	313	301	363	400	14.251	.007
	Up	468	450	468	473	452	450	430	427	472	500	13.031	.011
Barely sufficient	Low	473	456	470	473	457	455	431	428	482	500	14.552	.006
	Up	599	600	603	600	586	600	565	599	609	605	13.021	.011
Sufficient	Low	607	601	604	600	588	600	570	600	619	646	14.878	.005
	Up	726	750	733	750	704	700	700	700	750	770	21.998	.000
Good	Low	736	750	733	750	707	701	707	705	755	781	21.117	.000
	Up	859	864	856	880	841	850	853	864	862	895	5.834	.212
Excellent	Low	876	898	867	893	851	850	865	877	880	900	8.062	.089
	Up	(965)	(965)	(965)	(965)	(965)	(965)	(965)	(965)	(965)	(965)	—	—

Table note. μ =mean, M =median. $n = 50 \times 5$, $df = 4$ for all tests.

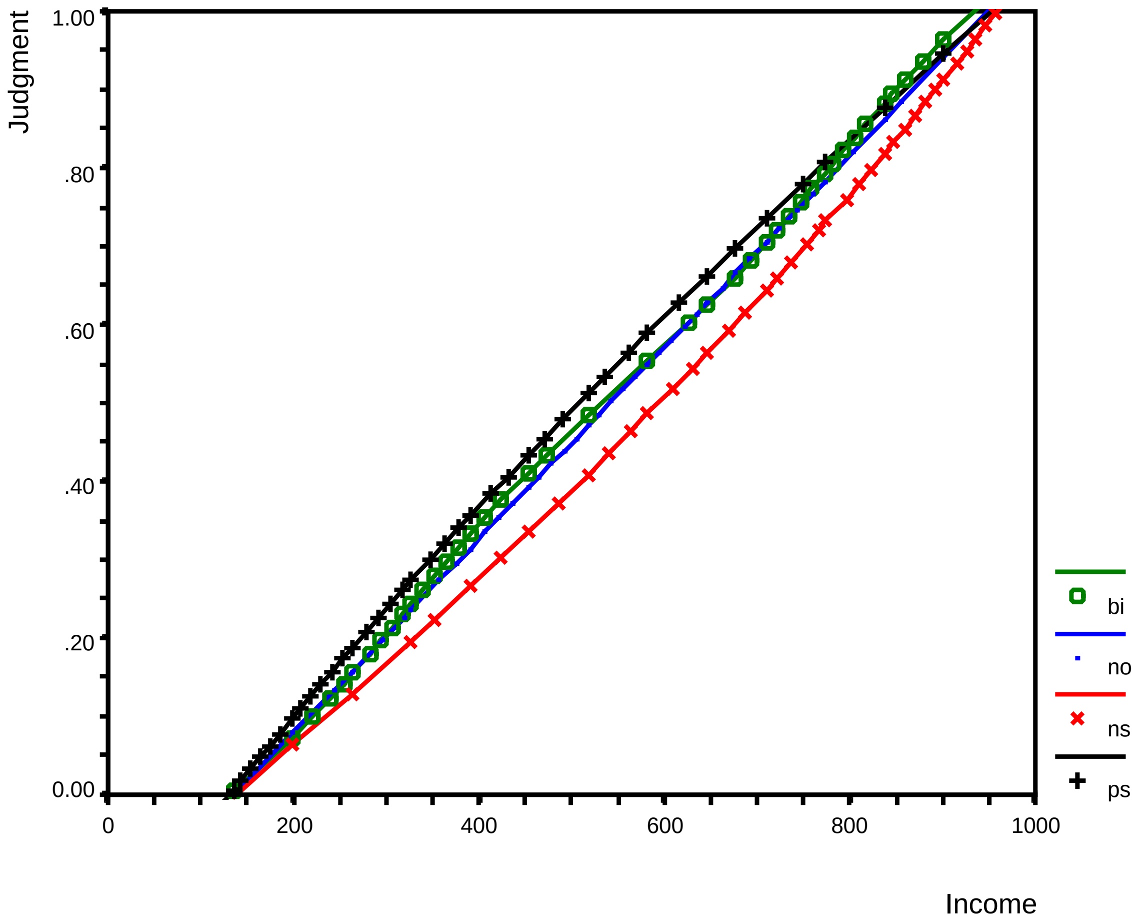

Figure 1 Graphs of Judgment Functions

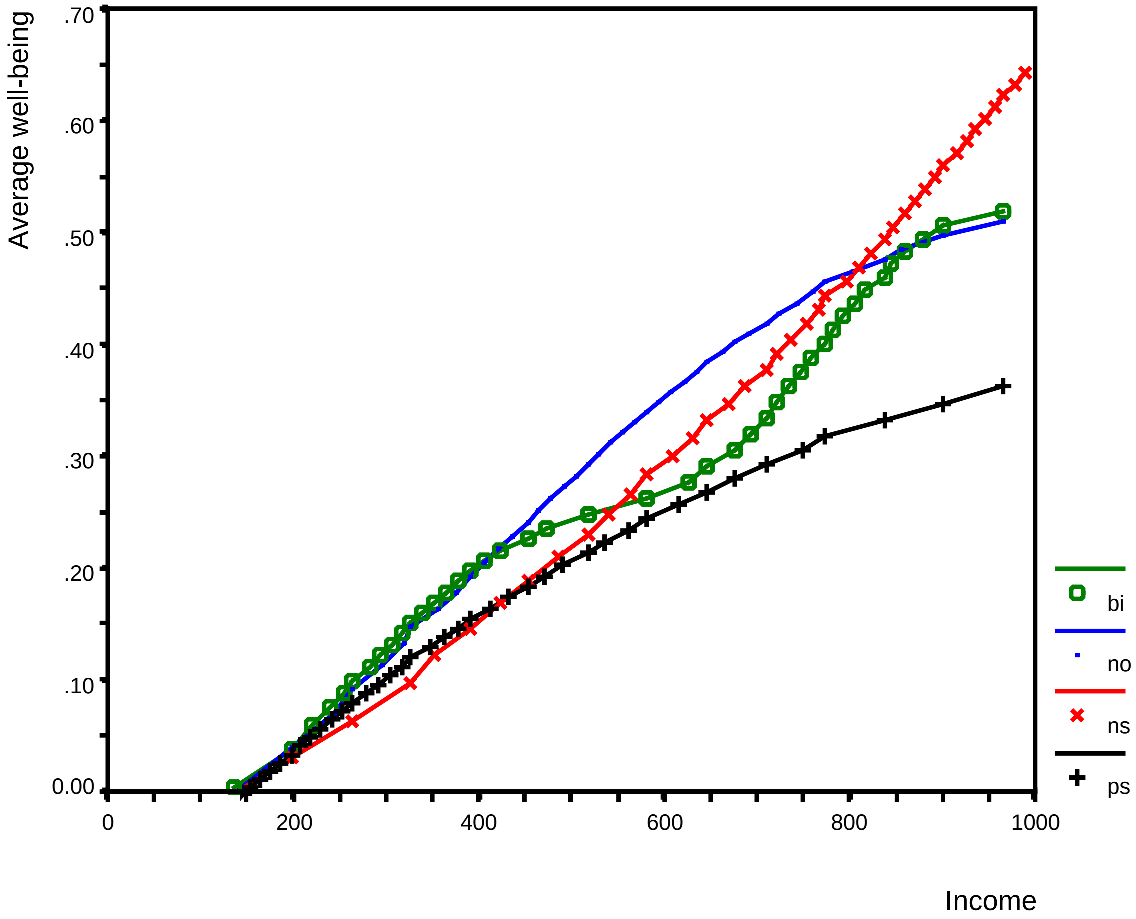

Figure 2 Average Well–Being Under Different Income Distributions

Appendix

Instructions and Stimulus Material

Income evaluation in Utopia[14]

Suppose you live in the future and participate in a space shuttle flight to the planet Utopia, which is inhabited by small green individuals. The local currency in Utopia is the UFO.

Suppose further that each small green individual bears on his or her chest a visible identification card, which (among other information) also shows his or her income. Utopia's constitution states that the lowest allowable income is 100 UFOs, while the upper income ceiling is 1000 UFOs: nobody must earn less than 100 UFOs, and nobody must earn more than 1000 UFOs. Consider that 100 UFOs is beyond the survival income level and that more income is always preferable.

After your landing on Utopia, you walk around in Utopia's capital, called Haley, and observe the income of several subjects.

> *Then* 25 *values taken from the true mathematical distribution of the respective group were shown in a random order to allow subjects to become acquainted with the experimental procedure.*

After your short trip through Haley, you meet Utopia's Prime Minister who had invited you to consult him with respect to an important issue: As you are an economist (a species completely unknown in Utopia), the Prime Minister asks you to make an evaluation of the incomes earned in Utopia. He wants you to categorize the incomes earned in Utopia into seven categories, viz.:

[14]The emphasized text illustrates what the subject was shown on the monitor.

1. Excellent
2. Good
3. Sufficient
4. Barely sufficient
5. Insufficient
6. Bad
7. Very bad

In order to perform this job properly, the Prime Minister presents to you a booklet containing a random sample of the incomes of 42 income recipients. You are assured that this sample is a perfect representation of the income distribution in Utopia.

In the following you can see the 42 entries of this booklet.

> *Now 42 values of the respective experimental distribution were shown in random order. First, the whole set of values was shown on the monitor and thereafter all entries were shown one at a time (in the very same order) and subjects were asked to assign them to one of the above categories. After all values had been assigned to categories, subjects were shown all values together with their categorization and could either confirm or change their categorization. Both the prior and posterior categorizations were recorded.*

The Prime Minister is quite happy with your categorization of incomes, which enables him to gain insights into the social stratification of Utopia. For future use of Utopia's statistical office, he asks you to state also limens for the seven income categories (notice that there is no inflation in Utopia).

For this purpose, he gives you a questionnaire and asks you to fill it in. Your limens should properly reflect the income distribution prevailing in Utopia.

A green individual's income is				
very bad	if it is	less than		UFOs
bad	if it is	between	and	UFOs
insufficient	if it is	between	and	UFOs
barely sufficient	if it is	between	and	UFOs
sufficient	if it is	between	and	UFOs
good	if it is	between	and	UFOs
excellent	if it is	higher than		UFOs

After this, your task is done. The Prime Minister thanks you and awards you the Utopian Order of the Garter in return for your services to his planet.

Table A1 Stimuli of the Experiment

	Mathematical					Experimental				
#	un	no	bi	ps	ns	un	no	bi	ps	ns
1	121	165	167	111	162	121	**135**	**135**	111	**135**
2	142	209	199	121	216	**135**	**199**	**199**	121	**199**
3	163	243	221	132	265	163	243	221	**135**	**263**
4	184	272	239	143	310	184	**263**	239	143	**327**
5	205	296	254	153	351	**199**	296	254	153	351
6	226	319	269	164	388	226	319	**263**	164	**390**
7	247	339	282	174	423	247	**327**	282	174	423
8	267	357	294	185	455	**263**	357	294	185	**454**
9	288	375	306	196	485	288	375	306	**199**	485
10	309	391	317	207	513	309	**390**	317	207	**518**
11	330	407	329	218	539	**327**	407	**327**	218	539
12	351	422	340	229	564	351	422	340	229	564
13	372	437	352	241	587	372	437	352	241	**582**
14	393	451	364	253	609	**390**	**454**	364	253	609
15	414	465	377	265	630	414	465	377	**263**	630
16	435	478	391	278	650	435	478	**390**	278	**646**
17	456	492	406	291	669	**454**	492	406	291	669
18	477	505	424	304	687	477	505	424	304	687
19	498	518	444	318	705	498	**518**	**454**	318	**710**
20	519	530	472	333	721	**518**	530	472	**327**	721
21	540	543	516	347	737	540	543	**518**	347	737

Table continues.

Continuation of Table A1

	Mathematical					Experimental				
#	un	no	bi	ps	ns	un	no	bi	ps	ns
22	560	556	583	363	753	560	556	**582**	363	753
23	581	569	627	379	767	**582**	569	627	379	767
24	602	581	655	395	782	602	**582**	**646**	**390**	**773**
25	623	594	675	413	796	623	594	675	413	796
26	644	607	693	431	809	**646**	607	693	431	809
27	665	621	708	450	822	665	621	**710**	**454**	822
28	686	634	722	470	835	686	634	722	470	**837**
29	707	648	735	491	847	**710**	**646**	735	491	847
30	728	662	747	513	859	728	662	747	**518**	859
31	749	677	759	536	871	749	677	759	536	871
32	770	692	770	561	882	**773**	692	**773**	561	882
33	791	708	782	587	893	791	**710**	782	**582**	893
34	793	724	793	615	904	812	724	793	615	**901**
35	833	742	805	645	915	**837**	742	805	**646**	915
36	853	760	817	677	926	853	760	817	677	926
37	874	780	830	712	936	874	**773**	**837**	**710**	936
38	895	803	845	749	947	**901**	803	845	749	947
39	916	827	860	790	957	916	**837**	860	**773**	957
40	937	856	878	835	968	937	856	878	**837**	**965**
41	958	890	900	884	979	**965**	**901**	**901**	**901**	979
42	979	934	932	938	989	979	**965**	**965**	**965**	989

Table note. un=uniform, no=normal, bi=bimodal, ps=positively–skewed, ns=negatively–skewed.