ANOTHER AVENUE FOR ANATOMY OF INCOME COMPARISONS: EVIDENCE FROM HYPOTHETICAL CHOICE EXPERIMENTS

Katsunori Yamada
Masayuki Sato

Another Avenue for Anatomy of Income Comparisons: Evidence from Hypothetical Choice Experiments[*]

Katsunori Yamada[†]

Institute of Social and Economic Research, Osaka University

6-1 Mihogaoka, Ibaraki 567-0047, Japan.

MOVE, Universitat Autonoma de Barcelona.

UAB Edifici B (s/n) 08193 Cerdanyola del Valles, Barcelona, Spain.

Masayuki Sato

Graduate School of Human Development and Environment, Kobe University.

3-11 Tsurukabuto, Nada-ku, Kobe 657-8501, JAPAN

Abstract

We propose a new avenue for studying income comparisons effects, namely hypothetical discrete choice experiments in which respondents are presented with alternative combinations of hypothetical monthly income amounts, both for themselves and certain reference persons. With this experimental method we can avoid the problems associated with researcher-imposed reference persons' incomes that are found in most of the happiness studies testing comparison effects. This approach

[†]*Corresponding author:* Parmanent Address: 6-1 Mihogaoka, Ibaraki 567-0047, Japan. Tel/Fax: +81-6-6879-8566. Email: kyamada@iser.osaka-u.ac.jp

allows investigation of the differences in comparison effects across types of reference groups as well as respondents' individual characteristics, including specific comparison benchmarks, which are the main open questions in the literature. Some results from our original, large-scale, Internet-based survey are provided.

Keywords: Relative utility; Hypothetical choice experiment; Reference group; Comparison benchmark

JEL classifications: C9; D1; D3

1 Introduction

Traditional economic theories focus on the role of absolute income or consumption, whereas behavioral evidence suggests that social comparisons influence well-being and decisions (Fliessbach et al. 2007, p. 1305). Whether social comparisons do in fact affect individual utility is critically important in understanding human behavior in any social context, and in evaluating the outcomes of economic behavior.

This study provides a new way to measure income comparison effects using hypothetical discrete choice experiments. A critical component in empirical studies of income comparison and relative utility is defining the appropriate reference person or group for each subject. Our approach estimates the income comparison parameters in the utility function through the *decision utility approach*, which was originally put forth by Kahneman et al. (1997) in the field of environmental valuation. Hence, this method is different from the standard *experienced utility approach* used in happiness studies, where respondents' subjective reports of well-being and proxies for reference income are used to estimate the relative utility effects. By using the decision utility approach, we are able to avoid the use of researcher-imposed reference persons' incomes, which must be constructed by econometricians without knowledge of the subjects who provided data on happiness but who were not asked with whom they made a comparison or how much they think their rivals earn.[1] In our experimental approach, we impose alternative combinations of hypothetical monthly income amounts on subjects, both for themselves and certain reference persons. Using data on respondents' choices of preferred income scenario, we can estimate the utility function parameters that capture the intensity and sign of income comparisons with certain reference persons. As such, the methodological merits of our study include that (i) the subjects can see the characteristics of reference persons and how much these persons earn compared with themselves in clear situation choice tasks, and (ii) the subjects can choose discrete choices with a lower cognitive burden than when evaluating their lives with more general happiness scores.

Exploiting the features of the choice experiment, we can investigate differences in comparison effects by reference person type through changing the characteristics of reference persons in hypothetical income scenarios. Another important issue that can be addressed with our method is determining how individual characteristics such as age, gender, educational attainments, and respondents' specific comparison benchmarks affect attitudes in

---

[1]Critiques of using researcher-imposed reference persons' incomes to estimate income comparison effects in happiness regressions were detailed in Manski (1993) and Sloane and Williams (2000).

income comparisons. Among these variables, the comparison benchmarks are of particular interest for researchers in the field, because these were recently studied by Clark and Senik (2010) via the experienced utility approach. We provide empirical results on these issues by using our original, large-scale, Internet-based survey of Japanese subjects. Our data set is socially representative in terms of age and gender distribution, which is an important virtue of our sample, as many studies often rely on potentially biased student samples. Problems of sample bias also affect experimental studies that provide incentives for a small number of subjects such as Fehr and Schmidt (1999) and Fehr and Schmidt (2006).[2] Our Internet-based survey overcomes this issue by accessing an enormous number and wide variety of registered subjects through a reliable subcontracted research company.

While the Internet offers immediate access to a large and diverse subject pool and research opportunities at cheaper costs, there are also caveats and potential pitfalls that pertain to Internet-based experiments. Typical critiques include the inability to monitor the motivations and understanding of participants and uncertainty about the precise identity of the experimental subjects, among others. On these issues, Horton et al. (2011) provided clear evidence that online experiments on social preferences through Amazon Mechanical Turk (MTurk) replicate previous experimental results from physical laboratories, encouraging the use of Internet-based surveys.[3] The fact that our experiment was hypothetical may be another source of concern, but neuroscience studies such as Bickel et al. (2009) and Kang et al. (2011) show that incentivized and hypothetical experiments do not generate significant differences in human neural activity. In addition, Amir et al. (2012) found that experimental participants in MTurk showed no differences in responses between incentivized experiments and non-incentivized experiments in the public goods game and in the trust game.

The remainder of the paper is organized as follows. Section 2 reviews the empirical literature on income comparisons and relative utility and places this study in historical context. In Section 3 we explain the experimental design of hypothetical discrete choices for income comparisons and the methods employed in our original Internet-based survey. The questionnaire and data construction for empirical analyses are also explained in the section. Section 4 outlines the estimation method following Train (2009). Section 5

---

[2]Recent experimental studies on social preferences by Falk et al. (2011) and Exadaktylos et al. (2012) report that only slight student bias can be observed, if any, and argue that experimental results from student samples are useful even in designing policy for the whole population.

[3]See also Paolacci et al. (2010), Suri and Watts (2011), Rand (2011), and Amir et al. (2012) for the potential pitfalls of Internet-based experiments and surveys and rebuttals to those critiques in the literature of social preferences.

presents the results for our benchmark task, in which the reference person is defined as "the social average" as in the series of studies by Richard Easterlin. Section 6 provides some additional results when different types of reference persons are presented in hypothetical income scenarios. Section 7 discusses potential biases related to experimental studies and concludes the paper.

2   Related studies on income comparisons

In the literature of happiness studies on income comparisons, reference persons are defined according to the researchers' choices, and the list of reference groups examined in previous studies is almost endless: an average (representative) person in society (Easterlin 1974, Easterlin 1995, Easterlin 2001, Stevenson and Wolfers 2008), someone like you (Clark and Oswald 1996, Ferrer-i Carbonell 2005), colleagues (Cappelli and Sherer 1988, Brown et al. 2008, Clark et al. 2009b), friends (Senik 2009), family (Senik 2009), neighbors (Luttmer 2005, Clark et al. 2009a, Knight et al. 2009, Senik 2009), and so forth. The accumulated evidence on the various cases is plentiful, and a prominent area of progress in the happiness literature recently is reported by Clark and Senik (2010), who investigated the effects of individual-specific comparison benchmarks on comparison attitudes using data on *who compares to whom.* They found that the intensity of social comparison changes in combination with specific groups that people ascribe to their comparison benchmarks. For example, those who think that their reference group is *friends* tend to make more comparisons than those whose think the comparison benchmark group is *work colleagues.*[4]

Despite all these evidence from field data on happiness scores, an alternative experimental approach for testing the relative utility hypothesis is necessary (Falk and Heckman 2009). One reason is that the reference income proxies used in empirical tests in previous happiness studies were imposed on subjects by an econometrician because information on both the *direct* and *cardinal* measures of reference income typically was missing. To our knowledge, the only exception in the literature that has information on both the *direct and cardinal* measures of reference income is de la Garza et al. (2010). Knight et al. (2009) and Senik (2009) used information on the perceptions of relative position in the respondents' villages or among friends and family members. However, their proxies of relative comparisons were ordinal, so interpretations of the magnitudes of coefficients for

---

[4]See also Mayraz et al. (2009).

comparison effects were not straightforward. Another reason is that the use of information on subjective well-being is sometimes a cause for criticism by economists in other fields, even though the view that subjective well-being information is valid has been well established (Hollander 2001, Ferrer-i Carbonell and Frijters 2004, Kahneman and Krueger 2006, Oswald and Wu 2010).

Aiming to overcome these issues, we present an alternative approach to hypothetical discrete choice experiments. In our experiments, similar to Clark and Senik (2010), we investigate how people change their comparison behavior (intensity and sign of social comparison) on the basis of their demographics, including comparison benchmarks. Moreover, similar to Senik (2009), heterogeneity of comparison effects driven by differences in reference groups can be examined by changing the definition of reference persons in hypothetical situation choice tasks. This method provides a much easier way of clarifying such heterogeneity compared with researchers who collect information in surveys on direct and cardinal measure of reference income for each reference group, together with happiness scores.

Solnick and Hemenway (1998), Johansson-Stenman et al. (2002), Alpizar et al. (2005), Carlsson et al. (2007), and Andersson (2008) investigated the intensity of social comparisons by addressing the methodologies of hypothetical choice experiments. One issue in Johansson-Stenman et al. (2002), Alpizar et al. (2005), and Andersson (2008) is that the choice format was designed in such a way that respondents made iterative choices to arrive at the point of indifference. This strategy is known to result in starting point bias (Carson 1991). Solnick and Hemenway (1998) and Carlsson et al. (2007) did not use iterative choices. Instead, each respondent only made a single choice between two alternatives related to relative income. With these strategies, we cannot apply a mixed logit framework to estimate the distribution of a parameter of relative utility. Also, the *degree of positionality* inferred by these previous studies, except for Solnick and Hemenway (1998), contained measurement error because the assigned value for the degree of positionality was given arbitrarily. To our knowledge, Carlsson et al. (2009) is the only study that conducted hypothetical and discrete choice experiment on income comparisons with repeated choice questions. They considered changes in the intensity of relative utility across different caste classes in India, but they did not provide results from factorial design analyses with different reference persons being tested in the same experiment. In Carlsson et al. (2009), sample representativeness is also an issue, as they conducted in-person surveys of

498 college students.[5]

Finally, studies on social preferences that differ from income comparison studies are mentioned. Using game theoretical frameworks such as the dictator game, the ultimatum game, and the public goods provision game, researchers had subjects interact in their experiments and examined the implications on reciprocity, trust and fairness. These studies include, for example, Andreoni and Bernheim (2009), Paolacci et al. (2010), Andreoni and Rao (2011), Suri and Watts (2011), Rand (2011), Horton et al. (2011), and Amir et al. (2012). Unlike these studies, studies on income comparisons, including happiness studies and our study of discrete choice experiments, there are no strategic interactions among subjects. This feature is actually important when we would like to estimate parameters of utility functions that can be used for macroeconomics analyses, e.g., Abel (1990), Gali (1994), Futagami and Shibata (1998), Liu and Turnovsky (2005), and Garcia-Penalosa and Turnovsky (2008), as in macroeconomics the number of agents is infinite, which is different from game theoretical situations.

3 Experimental design, questionnaire, and data collection

3.1 Hypothetical discrete choice question: Social average task

First we explain the experimental paradigm in the survey, using the example of a benchmark experiment called the social average task. The methods described for the social average task are representative of the methods for all the tasks we conducted in our study. We discuss the results of the other two tasks in Section 6 and provide detailed explanations of their experimental settings and empirical results in the Appendix because of space constraints.

The Easterlin paradox, which suggests that "increasing the income of society as a whole will not increase the well-being of anyone," has been discussed in the literature on *experienced utility* that considers how national average income acts as a driving force of relative utility effects. The social average task provides a useful alternative method for investigating the validity of the paradox and relative utility effects in general. The merits

---

[5]In terms of the representativeness of the sample, the respondents in Solnick and Hemenway (1998), Johansson-Stenman et al. (2002), and Alpizar et al. (2005) included only students and the respondents in Andersson (2008) were only people in academia. In contrast, the respondents in Carlsson et al. (2007) were from a socially representative survey.

of our method are that subjects recognize that they are competing with the Japanese social average when making choices and that the economic situations are explicitly shown to them.

Before the subjects began responding to repeated choice questions, they were shown an instruction screen displaying the following:

> *The following figures show your hypothetical monthly income (before tax). Also displayed in the same figure is Japan's overall average monthly income (before tax). Suppose that these are current situations of your monthly income (before tax) and Japan's overall average monthly income (before tax).*

In the subsequent screens, we asked respondents hypothetical discrete choice questions while showing them various figures for different alternatives after the question as shown below.[6]

> *Comparing situation 1 and situation 2 shown in the figures, which is more preferable to you? Suppose that the price levels in the two situations are the same. Please choose from the following options.*
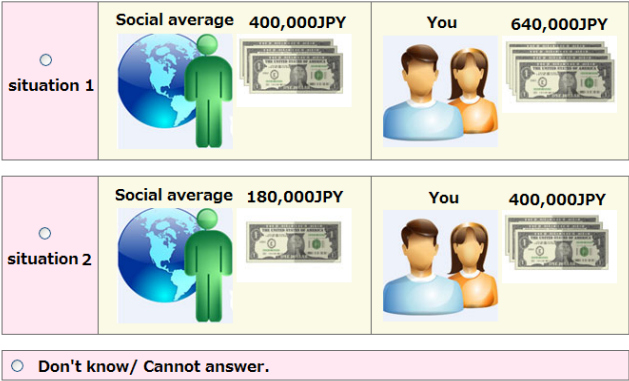
In the figure, each situation is defined by two *attributes*, one's own monthly pre-tax income and the monthly pre-tax income of the reference group. The choice scenario also provided the option "Don't know / Cannot answer."[7] Section 4 explains how we estimate

---

[6]In the survey information in the figures was presented in Japanese. The images for monthly income differ in terms of number of banknotes shown according to the attribute levels. Subjects repeated five questions and they were not allowed to go back to a previous question once they had made a choice. This survey format was also used in the other two tasks: the Leyden task and the "who-compares-to-whom task.

[7]We provided this no-choice option because of the suggestion by Arrow et al. (1993) and Haaijer et al. (2001), who pointed out the importance of including a no-choice option in hypothetical choice experiments. We then removed observations in which the no-choice option was selected from our regressions. An alternative way of coping with these observations is to interpret them as showing indifference between

parameters for a utility function using observations of discrete choices on preferred income scenarios. Next, we explain how we constructed our choice questions.

Considering the monthly income distribution in Japan, the levels of the attributes are set using the following distribution: 180,000 JPY, 240,000 JPY, 400,000 JPY, 640,000 JPY, and 900,000 JPY. The two attributes (own income and reference income) and five possible income levels for each attribute provide 25 potential variations in the income situation scenario. In the literature, these scenarios are called *alternatives*.

Researchers have to make their own choices about which alternatives to use in survey questions and which ones to discard. Following Louviere et al. (2000), we conducted *orthogonal planning* in choosing the alternatives to be used in choice questions. This method effectively pairs multi-dimensional and multiple-level attributes in alternatives, and provides an experimental plan with the greatest amount of information using the least number of observations. Further, employing orthogonal planning, we can avoid multicollinearity problems in the regressions of the random utility model explained in Section 4, because the independent variables in the regressions become orthogonal. We used SPSS Conjoint (ver. 15.0) for orthogonal design of alternatives in this study.

Next, we constructed *choice sets* consisting of pairs of alternatives and the no-choice option. By the requirement of orthogonal design, we generated two different alternative vectors, each of which consists of 25 pairs of own income levels and reference income levels.[8] Finally, choice sets, with a no-choice option, are created by pairing two alternatives, one of which is taken from an alternative vector and the other of which is chosen from the other alternative vector. The pairing strategy is at the discretion of the researchers, but all the variations must be exploited and same alternative cannot be used twice. Because the orthogonality in the alternative matrix is maintained for each row permutation, we can arbitrarily pair alternatives to meet the requirement.

As documented in Huber and Zwerina (1996) and Viscusi et al. (2008), it is ideal if the choice design can be paired so as to balance the utility of each alternative. One difficulty in choice experiments of relative utility, however, is that an increase (decrease) in a reference

---

the two situations, rather than a failure to understand the survey question. Unfortunately, we have no information about the true reason why the no-choice option was chosen. Hence, following the literature, we use the results of the first choice out of the five questions for a robustness check and found that the results presented are robust (not shown here due to space constraints, but available upon request).

[8]As such, in the social average task, 25 alternatives out of 25 potential variations *had to be used* to meet the requirement of orthogonal design. In the case of the Leyden task, 25 out of 1,000 potential alternatives were chosen to make an alternative vector, while in the "who-compares-to-whom task, 25 out of 125 potential variations were selected via orthogonal design. These alternative vectors were generated from different random seeds.

group's income does not necessarily mean that there is a decrease (increase) in one's own utility level; as such, we did not exclude the possibility of altruistic preference. Given these constraints, our best strategy for pairing alternatives is as follows.

Suppose we have the scenario $S = (x, y)$, where x denotes the level of one's own income and y is others income. Then, qualitatively, candidates of paired scenarios consist of the following 8 variations: $(x, y+)$, $(x, y-)$, $(x+, y)$, $(x+, y+)$, $(x+, y-)$, $(x-, y+)$, $(x-, y)$, and $(x-, y-)$, where $x+$ means some value greater than x and $x-$ means some value smaller than x. Since we do not exclude the possibility of altruism a priori, there are no a priori dominant choices for S from these eight alternatives. We then made pairs such that these eight situations appear as evenly as possible. Using the procedures discussed here, we were able to efficiently obtain parameter estimates. The following table of income levels (in 10,000 JPY) shows the set of questions we used in the survey. Each respondent answered five randomly assigned questions out of the 25 total questions.

	Alternative 1		Alternative 2			Alternative 1		Alternative 2	
Q	Own income	Ref. income	Own income	Ref. income	Q	Own income	Ref. income	Own income	Ref. income
1	64	40	40	18	14	64	90	90	90
2	90	90	40	90	15	18	64	24	40
3	90	64	64	90	16	24	90	18	64
4	90	18	90	40	17	24	18	64	24
5	90	40	90	64	18	18	90	40	40
6	90	24	90	18	19	24	24	64	18
7	24	40	64	64	20	18	24	24	90
8	64	64	90	24	21	18	40	18	24
9	40	40	64	40	22	40	64	24	64
10	64	24	18	18	23	40	18	18	90
11	40	24	24	4	24	18	18	18	40
12	64	18	40	24	25	24	64	24	18
13	40	90	40	64					

3.2 Logistics of the survey and questionnaire

Our data set was created using an original, Internet-based survey.[9] A Japanese consumer monitoring company, Nikkei Research Inc., conducted the survey under the direction of the authors. As clearly documented in Horton et al. (2011), "[t]he validity of economics experiments depends heavily upon trust, particularly subjects' trust that the promulgated rules will be followed and that all stated facts about payment, $\cdots$, are true." In Japanese society, the Nihon Keizai Shimbun (NIKKEI) Group is widely viewed as a trustworthy and neutral media outlet. Nikkei Research, Inc., as a part of NIKKEI group, has established a high reputation among researchers and consumers. For example, in order to provide highly reliable research data, its registered subjects are subject to monthly screenings.

---

[9]We conducted three preliminary tests before the main test. We then took differences in reference groups into account in designing the questions used in the choice questions.

The company keeps information up to date and excludes double registrations. Incentives for respondents are provided by cash voucher, rather than by points; point incentives can lead to bias, as particular respondents with points tend to answer.

Nikkei Research, Inc., and the authors reached an agreement that the number of subjects should be over 10,000, given the volume of the research fund. Then, in consideration of the unweighted average of response rates for seven similar academic choice experiment surveys conducted by the same company in 2008 and 2009, the company sent an invitation email for the survey to 60,482 subjects (out of more than 160,000 total registered subjects in the Nikkei Database). Subjects between the ages of 20 and 65 were selected using stratified random sampling so that the cohort profile of our sample mirrored the Japanese census statistics of age and gender distribution. Because subjects are not required to declare their educational attainments during the registration process, we did not use such information in our stratified sampling. In the email, we specified that the survey is being conducted for research purposes and followed the disclosure requirements for research involving human beings provided with incentives as set forth by the ethics committee. We specified our payment rules in the invitation email and stated that the incentive would be on a lottery basis. We informed subjects that 800 winners among those who completed the survey would be paid 500 JPY per person. The anonymity of subjects was completely secured. If subjects wished to participate in the survey, they were instructed to follow the link in the email that directed them to our stand-alone survey website, written in html, that was launched in February 2010. The survey was open for one week (Feb. 18 to 25, 2010), and 14,370 subjects completed the survey.[10] This approach is in contrast to one-day research, which many other research companies typically adopt for cost reasons. With one-day research, subjects are chosen on a first-come first-served basis, causing potential bias as the resulting sample may well not be representative. Sampling over a one-week period likely reduces this bias. The structure of the questionnaire is as follows.

3.2.1 Part 1: Introductory questions

At the beginning of the survey, respondents were asked to choose one of five possible categories on their level of satisfaction about *income*. Category 1 corresponds to "Not at all satisfied," while category 5 denotes "Extremely satisfied". The second question

---

[10]This response rate of 23.8% (14,370/60,482) is smaller than might have been desired. The decision of subjects to participate in the survey was driven by unobservable characteristics that likely differ between participants and non-participants. If the unobservable characteristics are independent of the income comparison effects, then sample selection will not bias the results.

related to social comparison and was phrased as "How much are you concerned, anxious, or jealous about the amounts of income received by other people?" The respondents were asked to choose from five response options, where category 1 corresponded to "Not at all" and category 5 denoted "Very concerned." The third question concerned the respondents' definition of their reference group. They were asked to choose one category, from those applicable to them, as their reference group, with the choices being: (i) family, (ii) neighbors, (iii) friends, (iv) colleagues, (v) do not care, and (vi) others. From these last two questions, we can observe "who compares to whom?" and "how much?," which were investigated as the framework of the happiness study of European countries by Clark and Senik (2010). The following table shows the distribution of the reference groups chosen by the respondents.

	Family	Neighbors	Friends	Colleagues	Do not compare	Others
Observations	483	578	4279	2024	2592	247
%	4.73	5.67	41.94	19.84	25.40	2.42

We can see that the most often cited reference group is *friends*, followed by *work colleagues*. These rankings are the opposite of those in European countries, as documented in Clark and Senik (2010), but it is interesting that in both Europe and Japan these two groups are the two most important reference groups. *Family* and *neighbors* play minor roles as reference groups. These findings were used in setting up our hypothetical choice experiment in the "who-compares-to-whom" task. In our data set, one-fourth of subjects answered they do not have comparison groups.

3.2.2 Part 2: Hypothetical discrete choice questions

1. Social average task (randomly assigned 5 questions)

2. Leyden task (randomly assigned 5 questions)

3. Who-compares-to-whom task (randomly assigned 5 questions)

3.2.3 Part 3: Demographic questions

The last part of the survey consisted of questions about individual characteristics, including age, gender, educational background, employment, marital status, type of residence, residence area, and annual pre-tax personal income in 2009.

3.3 Survey strategy and data construction

One clear deficiency of such web-based surveys is that researchers cannot actively monitor and encourage subjects to participate in the survey. In particular, when subjects do not have a clear understanding of questions, they will try to complete questions as quickly as possible by making up answers without contemplation. Hence, there is a trade-off between (1) better understanding of the questionnaire by subjects, which reduces the cognitive experimenter demand effect (EDE), and (2) the benefits of conducting the survey "behind the veil of ignorance," which reduces the social EDE (Zizzo 2010). In our case, given that the survey was on the Internet and that specificity was required to meet the study's academic purpose and the requirements of the ethics committee, we leaned toward the former (better understanding by subjects) and explicitly stated that the survey was a "Survey on socio-economic attitudes by Osaka University" in both the invitation email and the top page of the survey website.[11]

A good way to monitor subjects' willingness to participate in the survey is to look at the elapsed time for completing the survey.[12] If the elapsed time is extremely short for a subject, it is obvious that he completed the survey without contemplation, and it is plausible that he just wanted to join the lottery for the research reward. The average elapsed time to finish the survey was 9 min 9 s for our survey, with the median value of 6 min 5 s.[13] For subjects, the easiest way to finish the survey is to provide the same answers for conjoint questions documented above. We eliminated those who provided the same number for all five questions in any tasks (2,218) after we confirmed that those who provided the same answers on five consecutive questions in a task tended to finish

---

[11]In our survey, the questionnaire started with questions about income satisfaction and comparison attitudes. Because these introductory questions are followed by the hypothetical choice tasks, the question order might make individuals conscious of making pecuniary comparisons. If this is the case, this bias will also be related to social EDE. That said, the fact that the subjects were reminded about social comparisons does not necessarily lead to over- or under-estimates of the true effect. On the one hand, after the instructions subjects may be motivated to "beat" the reference persons in the hypothetical choices, which will over-estimate the true effects of social comparison. On the other hand, one can just as easily think of mechanisms shifting the results in the other direction. Namely, many people dislike thinking of themselves as status-seeking and they therefore underestimate the degree to which they state that they care about social comparisons. This resembles the purchase of moral satisfaction in Kahneman and Knetsch (1992). See the concluding section for more discussion on potential biases in experimental studies.

[12]Rubinstein (2007) conducted an Internet-based survey experiment and recorded response time for some questions. He found significant differences in response time across types of questions, and suggested that choices made on the basis of an emotional response require less response time than choices that require the use of cognitive reasoning. In our case, we have information on total response time to complete the entire survey, while response times for individual questions are not available.

[13]Observations with no time records (90) and elapsed time longer than 60 minutes (106) are excluded from our study sample.

the survey very quickly, most likely without contemplation.[14] We also discarded the information of subjects whose elapsed time is shorter than 4 min (968) on the basis of feedback from an internal company pilot test by NIKKEI. So far, we are left with 10,988 respondents.

Next, observations were dropped if they were either missing information for some of the variables used in the empirical analysis below (219), or contained an inconsistency in the data, such as retirement before the age of 55 (1). Finally, we excluded observations of respondents who report their personal annual pre-tax income in 2009 to be higher than 12 million JPY (565).[15] To ensure that this cut-off for the income variable was not a result of sample selection, we compared the observations in the two groups along different dimensions including age, education, marital status, and residence area. We are happy to report that the number of observations excluded from our working sample does not seem to be a result of any sample selection problems, and that the main results documented below remain qualitatively unchanged when we use the whole sample as our study sample. At the final stage, we were left with 10,203 respondents.

The descriptive statistics of our data are shown in Table 1. Because the stratified random sample was designed to mirror the population cohort profile of Japanese census statistics, the age and gender structures of our sample appear quite similar to national statistics. There is however considerable under-representation of women who are divorced, separated or widowed. This difference from national statistics comes about because the latter include everyone aged over 15. As the average length of life for Japanese women is around 87 (with that of men being around 78), women tend to be widowed towards the end of their lives, which is reflected in the rate of female divorce/separation/widowhood in national statistics. However, as our sample only includes those who are aged up to 65, the rates of divorce/separation/widowhood for both men and women are lower than those in national statistics.

With respect to educational attainment, in our samples of males, just 1% of the sample completed middle school only, 18% completed high school only, 10 percent completed some

---

[14]Rand (2011) reported that at least 80% of experiment participants in MTurk were not merely making random selections on survey questions, which resembles the figure in our case.

[15]The cut-off point, 12 million JPY, is higher than the sum of the average of personal pre-tax annual income and three standard deviations of the income distribution. There are two major reasons for the high frequency of high-income level subjects. One reason is that survey participation is biased toward persons with higher education, Internet access, and urban residence. It is natural that these individuals have higher income than others without these characteristics. The other reason is that they tried to cheat by inflating their income levels. According to the exchange rate in March 2009, 12 million JPY is around 130,000 USD.

of college, and the remaining 70% held college or post-graduate degrees. This bias toward higher education also holds for the female samples. This is an over-sampling of more highly educated participants. Related to this issue, the average income levels from our sample are greater than that in the national statistics. The survey requests that subjects indicate their own income level from a list of 11 categories, where category 1 denotes annual wages of less than 2 million JPY and category 11 corresponds to an annual income level of more than 50 million JPY. When we measure individual income levels as the mid-point in each of the 9 intermediate categories, and use ad hoc values of 1.5 and 55 million JPY for the two extreme categories, respectively, we obtained that the average annual income of our whole sample was 5.69 million JPY for males and 2.93 million JPY for females.

The differences from the national statistics for students and the unemployed do not look severe. Information on residence location is compared. There is an over-sampling from the Kanto region, which includes Tokyo. Also, people from the Kansai region, which contains Osaka, are slightly over-sampled. Overall, we find that our data set captures significant features of Japanese society, except for the distribution of educational attainment. It is difficult to obtain a representative sample with small face-to-face surveys.

4   Random utility model and empirical method

In this section we introduce the econometric foundation on how subjects' choice data can be used to estimate their utility functions. We start by describing a discrete choice model with a general utility function. To analyze decisions in hypothetical choice experiments, we use a random utility model framework. The model deals with data on repeated choices over available alternatives. It is assumed that subjects choose an alternative since they obtain higher utility out of the alternative than from the other available alternatives. When there are two alternatives available (A and B, for example), and if they chose A rather than B, then the choice data is recorded as 1 for alternative A and 0 for alternative B, along with the levels of the explanatory variables (attributes) in alternatives A and B, respectively. These pieces of information comprise the observation for regression analyses.

Now more specifically, there are N subjects and they answer $T(\geq 1)$ repeated choice questions. The utility of subject n when s/he chooses alternative i at question $t \in T$, U_{itn}, consists of observable components in experiments V_{itn} and unobservable components ϵ_{itn} so that utility can be viewed as $U_{itn} = V_{itn} + \epsilon_{itn}$. Utility from observable components are assumed to be linear combinations of each attribute as $V_{itn} = \sum_{k=1}^{K} \beta_k X_{ik}$, where

$k = 1, \cdots, K(K \geq 2)$ represents the variety of attributes, X_k denotes the levels of kth attributes, and β_k measures marginal utility of each attribute. In the following analysis, the vector of $\beta \equiv (\beta_1, \cdots, \beta_K)$ that maximizes the log likelihood function of observed choice patterns by subject is the estimator of conditional or mixed logit model regressions. Following McFadden (1974), ϵ_{itn} is distributed following independent and identical distribution of extreme value type 1 (IIDEV1) with variance σ^2.

The logit formula of choice probability P_{itn} that subject n chooses alternative i from the set of alternatives S_t (choice set) in question $t \in T$ can be written as

$$P_{itn} = prob(U_{itn} > U_{jtn}, \forall j \neq i \in S_t) = prob(\epsilon_{jtn} - \epsilon_{itn} < V_{itn} - V_{jtn}, \forall j \neq i \in S_t).$$

McFadden (1974) showed that $P_{itn} = \exp(\lambda V_{itn})/\sum_{j \in S} \exp(\lambda V_{jtn})$, where $\lambda = \pi/\sqrt{6}\sigma$ is the scale parameter.

Finally, a dummy variable d_{itn} is defined, taking a value of 1 if subject n choose alternative i for question $t \in T$, and 0 otherwise. Together with the logit formula of choice probability P_{itn}, the log likelihood function of repeated choices observed in experiments can be written as

$$LL(\beta) = \sum_n \sum_t \sum_{i \in S_t} d_{itn} \ln P_{itn}.$$

In the conditional logit model, the parameters of utility function, β, can be obtained with the first-order condition of $\partial LL(\beta)/\partial \beta = 0$ (McFadden 1974). To be more specific, when we estimate the model assuming that the independence of irrelevant alternatives (IIA) holds, we obtain a conditional logit model where all of N subjects share the same set of parameter in β. Alternatively, when we allow for distributions of some parameters in β across subjects, we obtain the mixed logit model. In the latter case, while we assume that the error term is independently and identically distributed as in the conditional logit model, non-IIA situations are allowed. In the case of the mixed logit model, we can obtain the distribution of parameters $f(\beta)$ as follows. Following Train (2009), we specify that $f(\beta)$ is either a normal or a log normal distribution function with parameters set as θ. The choice probability function P_{itn}^{ML} for the mixed logit model can be written as

$$P_{itn}^{ML} = \int P_{itn}(\beta) f(\beta|\theta) d\beta,$$

where P_{itn} is the logit choice probability in the conditional logit model given β. θ can be

obtained via simulation which maximizes the simulated log likelihood function[16]

$$SLL(\theta) = \sum_n \sum_t \sum_{i \in S_t} d_{itn} \ln P_{itn}^{ML}.$$

Next, we specify the shape of the utility function for our own purposes. Here we present the specific theoretical framework for the social average task only because of space constraints. Individuals derive utility not only from their own income $X_1 = y$ but also from the social average income $X_2 = \bar{y}$. From textbook assumptions, we suppose that subjects value attribute y positively. On the other hand, the social average income $\bar{y}$ can be valued positively (altruism) or negatively (jealousy). Following Johansson-Stenman et al. (2002), Dupor and Liu (2003), Liu and Turnovsky (2005), we consider the constant relative risk aversion-type utility function as

$$(1) \qquad V = \frac{(y\bar{y}^\gamma)^{1-\rho}}{(1-\rho)},$$

where $\rho > 0$. If $\rho = 1$, it reduces to the log felicity function. The parameter γ regulates the intensity and sign of relative utility and is the central topic of this study. If $\gamma < 0$, the individual has jealousy. If $\gamma > 0$, the individual has an altruistic preference, whereas if $\gamma = 0$, there is no relative utility.

Again, let i denote the alternative and n denote the subject. We take the logarithms of both sides in Equation 1 to obtain

$$(2) \qquad \ln V_{ni} = (1-\rho) \ln y_{ni} + (1-\rho)\gamma \ln \bar{y}_{ni} - \ln(1-\rho).$$

With an error term ϵ_n, the probability P_{in} that respondent n prefers alternative i to alternative j is given by

$$P_{in} = Prob\big((1-\rho)\ln y_{in} + (1-\rho)\gamma \ln \bar{y}_{in} - \ln(1-\rho) + \epsilon_{in}$$
$$> (1-\rho)\ln y_{jn} + (1-\rho)\gamma \ln \bar{y}_{jn} - \ln(1-\rho) + \epsilon_{jn}\big), \quad \text{for all} \quad j \neq i.$$

Using maximum-likelihood estimation we obtain coefficients for $\ln y$ as $\beta_1 = 1 - \rho$ and $\ln \bar{y}$ as $\beta_2 = (1-\rho)\gamma$. β_1 and β_2 are regarded as marginal utility in the random utility model framework. It is noteworthy here that estimated β_1 and β_2 are divided by the scale parameter λ, which is unknown to researchers (Train 2009, p41). This means that

---

[16]See Section 6 of Train (2009) for details.

we cannot obtain true magnitudes of all the parameters in Equation 2. However, we can obtain true estimates of our interested variable γ by dividing β_2 with β_1, thus canceling λ out.[17]

5   Benchmark results from the social average task

Our benchmark results from social average task are shown here. Results in this section will be informative to theoretical macroeconomists because previous theoretical studies on relative utility effects such as Abel (1990), Gali (1994), Liu and Turnovsky (2005), and Garcia-Penalosa and Turnovsky (2008) were conducted without estimating important parameters in the utility function. As such, they put forward various propositions in accordance with the parameters assumed and do not necessarily reflect "reality."

Table 2 provides results for the whole sample. The first column shows the results from the conditional logit model. We see that a person's own income affects utility positively and significantly, as is expected. Next, from the coefficient of the reference income term, it is shown that relative utility exists among Japanese respondents, and that, on average, the effect appears in the form of *jealousy*. These two coefficients provide the true magnitude for the parameter of relative utility γ, by dividing the second by the first. From the estimates in column (1) of Table 2, we obtain that $\gamma = -0.458$. From a decision utility framework, this result stands in the middle of two extremes in experienced utility frameworks: the fully relative utility function of Easterlin (1995) and the solely absolute utility function of Stevenson and Wolfers (2008).[18] It is noteworthy that a recent finding in the happiness studies literature by de la Garza et al. (2010) reached the same conclusion that money buys happiness *to some extent* in Japan, through the use of direct and cardinal measures of reference income. Also note that our result is perfectly in line with the other studies based on hypothetical choices in different societies such as Solnick and Hemenway

[17]Ida and Goto (2009) compared estimated parameters in a logit model framework by assuming that $\sigma = 1$ through all regressions of subsamples. However, researchers cannot compare estimates from different subsamples without taking differences in σ into consideration. As stated in the body, the coefficients that are estimated indicate the effect of each *observed* variable relative to the variance of the *unobserved* factors. It is useful to recognize that the likelihood ratio test designed for confirming statistically significant difference of parameters among subgroups is not suitable here. For example, a larger σ in a subsample leads to smaller coefficients in its regression, even when the observed factors in two subsamples have the *same* effect on utility. A heteroscedastic logit model can be used in investigating the difference of σ between the subgroups.

[18]According to Figure 3 in Easterlin (1995), the Easterlin paradox evidently held for Japan in the period from 1958 to 1987. One reason why the comparison intensity we estimated falls short of the level validating the paradox reason is that the social comparison effect is just one of many explanations of the Easterlin paradox. Habit formation, for example, explains the paradox as well (van de Stadt et al. 1985).

(1998), Johansson-Stenman et al. (2002), Alpizar et al. (2005) and Carlsson et al. (2007).

5.1 Representativeness of the result

As was documented in section 3.3, our data set under-sampled those who completed middle school only. A simple estimation of Mincer equation with our data showed that educational attainment was positively correlated with income levels. As was shown by Ravallion and Lokshin (2010), it is well known in the literature that people with lower income tend to become less jealous, and we found the same result in our data set, as is documented below in Section 5.2. Hence, the representativeness of our results is an issue, and it is likely that we *over-estimated* the relative utility effects because of the oversampling of persons with higher educational attainment. In column (2) of Table 2, we show the result when we adjust the sampling weights using a post-stratification method.[19] The estimated γ after the post-stratification adjustment suggests, against our expectation, that the original γ in Column (1) was a slight *under-estimation* of the negative relative utility effect compared with the adjusted, representative data set in terms of gender, age, and educational attainment.

The reason for this result is as follows. First, the coefficient of the own income term in Column (2) is smaller in comparison with that in Column (1). Comparing coefficients of own income term from columns (1) and (2), we see that subjects with higher educational attainment and higher income levels enjoy higher marginal utility from own income.[20] Second, after the adjustment, the coefficient of the reference income term in Column (2) is smaller in absolute value than the one in Column (1). The direction of the change is hence in line with the presumption that people with higher educational attainment and higher income levels tend to become more jealous. In our data set, the first effect dominates the second effect to provide γ of stronger jealousy after adjusting for the over-sampling of persons with higher educational attainment.

Given the results, an accurate depiction of the representative relative utility effects in

---

[19]We computed post-stratification adjustments to survey sampling weights. The sampling weights in gender and educational attainments were adjusted such that the sum of the weights equals the control total for each stratum.

[20]Because marginal utility from consumption becomes smaller as your consumption levels increases in neoclassical economics theory, at first sight this seems odd. This observation, however, can be justified when we allow for heterogeneity of a parameter in the utility function between the poor and the rich. When a shift parameter of the utility function is greater for the rich than the poor, the marginal utility of consumption at a certain level of consumption becomes higher for the rich. The heterogeneity of the utility function, reversely, could explain why some become rich while the other stay poor, even when the other demographic conditions are the same for all the subjects.

Japanese may not be necessary since, as described above, samples with lower educational attainment are under-sampled. Nonetheless, because of the large size of our sample, the breadth of coverage across Japan's 47 prefectures, and the wide variety of job types from public servants to students and the unemployed, we believe that our data set does capture significant features of the relative utility effects in Japan. Hereafter, we show results using the unweighted sample for brevity.

5.2   Heterogeneity of preference parameters

In the last columns of table 2, we show the result from the mixed logit model in which normal distributions of parameters across subjects are allowed.[21] We find that own income affects utility positively, whereas the reference income has a negative impact on utility on average. We also find a similar ratio in the values estimated for the own income term to that of reference income term in both the conditional logit model and the mixed logit model, which validates the robustness of previous findings from the conditional logit model.

It is interesting to note that the standard deviation terms estimated in the mixed logit model are both significant at the 1% level. Behavioral economics has provided evidence that demographic differences lead to substantial differences in preference parameters, such as the time discount rate and the level of risk aversion.[22]

Following Viscusi et al. (2008), we identify the effects of individual characteristics on preference parameters by controlling for interaction terms of attributes in the surveys and demographic variables in conditional logit models.[23] In doing so, we consider two organic factors, six acquired individual characteristics, and three subjective variables as potential sources of parameter heterogeneity. The two organic variables are age and gender. We consider annual income level, educational attainment, urban residence, marital status,

---

[21]The STATA module for mixed logit model estimation is provided by Hole (2007).

[22]Small et al. (2005) applied the framework of a mixed logit model to investigate the distribution of commuters' preferences for speedy and reliable highway travel, finding that there was substantial heterogeneity in motorists' values of travel time and reliability. Hole (2008) investigated the preferences of patients about general practitioner appointments using standard logit, mixed logit, and latent class logit models. He showed that there was significant preference heterogeneity for all the attributes in the experiment. Viscusi et al. (2008) showed that eco-conscious individuals have a lower rate of time discounting than those who are not eco-friendly. Ida and Goto (2009) showed that smokers have with a higher value of time discounting and a lower value of risk aversion than nonsmokers.

[23]The introduction of interaction terms into conditional logit frameworks is acceptable as long as we confine our attention to the sign and significance of the interaction terms, as is clearly explained on page 22 in Train (2009). See Ai and Norton (2003) for interpretations of the *marginal effects* of dummy interaction terms in logit models. As long as one can interpret the coefficients as marginal utilities, as we do in a random utility model framework, Ai and Norton's point is not relevant.

unemployment status, and student dummies for the six individual characteristics. The three subjective dummy variables include a "do not compare" dummy, a "very happy" dummy, and a "very comparison conscious" dummy.

We report the results without a detailed table to save space (the full results are available upon request). We find that people tend to become more jealous if they are rich, female, highly educated, or married. Interestingly, urban residents do not have stronger comparison attitudes when compared to those who do not live in major cities. It is also interesting that age does not affect comparison intensity. In terms of the marginal utility of own income, as previously introduced, those with higher income and higher educational attainment tend to obtain higher utility from a certain amount of income. Regarding the subjective variables, we find that those who report that they do not compare have weaker comparison attitudes, which we discuss in more depth in the following section, and that the more they care about comparisons, the stronger their jealousy becomes. These findings are as expected. Feelings of being happier do not affect comparison intensity. Our findings are robust against changes in the threshold level for the comparison conscious group, the happy group, high income group, and elder group.

Thus, we can confirm that heterogeneity plays a role in determining the intensity of social comparison, just as previous behavioral economics studies have found in other fields.

5.3 Analysis with comparison benchmark information

A recent caveat from the happiness study of Clark and Senik (2010) is that comparison attitudes can differ depending on the reference group that people ascribe as their comparison benchmark. In our data set, similar to Clark and Senik (2010), information on specific and relevant reference groups for each subject is available. It is interesting to see how people change the intensity of comparison on the basis of their comparison benchmarks.

Since the target reference group is based on a general concept, the differences in relative utility intensity derived in this task reflect basic differences in the intensity of relative utility across subgroups. We divide our study sample into subgroups of individually relevant reference groups and compare the obtained *true* magnitudes of γ across the subgroups.

The variances of the estimated γ by subgroup for the comparison benchmarks are obtained by using the Delta method to examine the statistical significance of the differences. Since γ takes the form $\gamma = r/s$, where r and s are stochastic variables, the variance of γ

is obtained as follows:

$$Var(\gamma) = \begin{pmatrix} \frac{\partial \gamma}{\partial r} & \frac{\partial \gamma}{\partial s} \end{pmatrix} \begin{pmatrix} Var(r) & Cov(r,s) \\ Cov(r,s) & Var(s) \end{pmatrix} \begin{pmatrix} \frac{\partial \gamma}{\partial r} \\ \frac{\partial \gamma}{\partial s} \end{pmatrix} = \frac{1}{\beta_s^2}Var(r) + \frac{\beta_r^2}{\beta_s^4}Var(\beta_s) - \frac{2\beta_r}{\beta_s^3}Cov(r,s),$$

where for r and s, β_r and β_s are the averages and $Var(r)$ and $Var(s)$ are the variances, respectively. $Cov(r,s)$ is the covariance of r and s.

Columns (1) through (5) of Table 3 show the coefficients and estimated γ across subgroups for the comparison benchmarks. We exclude the subgroup of "others." An interesting estimate of γ appears in column (2), where the comparison benchmark is neighbors. People who tend to compare themselves with neighbors are the most jealous in Japan. The value of γ for the "neighbors" subgroup is significantly different from the other subgroups ($p < 0.01$).

The finding that those who compare themselves with neighbors have the strongest intensity of relative utility requires further attention. The reference group of neighbors is characterized by close contact. Hence, the result seems natural if we accept that people endogenously choose their reference groups from groups of close contacts, as Clark et al. (2008) argued. However, it is interesting to see a weaker intensity of relative utility for those whose comparison benchmark is work colleagues, another type of close contact, shown in column (4). The difference of estimated γ between the two subgroups of comparison benchmark is significant ($p < 0.01$). Regarding this finding, it may make sense when considering the tunnel effect proposed by Hirschman and Rothschild (1973). He argued that an increase in work colleagues' income could be interpreted as a positive signal regarding likely future outcomes. However, the effect seems not so strong as to provide *positive* relative utility, as was found by Senik (2004) using a Russian data set.[24]

It is important to note that the "Do not compare" group in column (5) has a significantly smaller estimated γ than the other groups ($p < 0.01$).[25] However, the fact that social comparison effects are observed among those who explicitly state that they do not compare, and that differences in the estimated gamma from other subgroups appears marginal, despite being significant, draws our attentions.

The benefits of our experimental approach are that we showed subjects clearly illustrated income comparison scenarios with information on their own income levels and

---

[24]Card et al. (2012) compared the positive effects of the tunnel effects and negative effects of relative utility in a social experiment setting, and showed that the negative effects are dominant in the United States.

[25]In the previous section we picked up the same effect when we interacted a "Do not compare" dummy variable with reference income terms.

reference income levels, and that we elicited information on comparison benchmarks. This procedure resolves uncertainty in existing happiness studies that do not elicit comparison benchmarks as to whether the negative coefficients for the relative income proxies in happiness regressions are in fact capturing social comparisons. Hence, a natural expectation for the results from our experimental approach is that we find no social comparisons effects among those who say they do not compare. It is also noteworthy that we did not get the result because of a biased construction of the experiment: from the construction of choice sets as explained in Section 3.1, we can obtain positive, negative, or no relative utility effects depending on patterns of subjects' choices.

Table 4 shows the results when we divide the observations of the "Do not compare group into subgroups of survey-elicited intensity of jealousy (from 1 to 5). The number of subjects who declared the maximum intensity of jealousy (5) was too few to provide a relevant result, as shown in Column (5) of the table. As it can be seen, the estimated values of γ across subgroups of survey elicited intensity of jealousy make some sense. In subjects who declared weaker jealousy in the questionnaire, their choice patterns in the choice experiment provided weaker intensity of income comparisons of γ. For those who declared the minimum intensity of jealousy (1), the estimated γ was -0.342. This figure is a 25% reduction from the comparison effect for the whole sample, but again, is significantly different from zero. Our interpretation of the result, considering that the result is not found because of a biased construction of the experiment favoring negative comparisons, is that ultimately humans make comparisons, even though they themselves declare that they do not. In support of this argument, Fliessbach et al. (2007), Takahashi et al. (2009), and Tricomi et al. (2010) revealed a neurological basis for making comparisons in human brains. These studies imply that we inherently cannot escape from making comparisons. We even suggest that our study has provided stronger support for the existence of negative relative utility effects than research that relies on evaluation and rating data can provide: it is easy for subjects to "cheat" in questionnaires to say that they do not make comparisons, or to rate their intensity of jealousy very low, even when they indeed are very jealous. It is not, however, very easy for them to expect what their choices in choice experiments will indicate about their jealousy without knowing the technical aspects of discrete choice experiments.

6 Discussion of results from extended tasks

A salient feature of the hypothetical discrete choice experiment on income comparisons is that we can investigate differences in such effects across various types of reference persons in well-controlled experimental conditions. Here we introduce the main results from other such applications. The detailed construction of the experiments and empirical results are given in the Appendix.

6.1 Leyden task

In the Leyden task, reference persons are characterized by specific demographic variables of age, gender, and educational attainment, which we adapted from the Leyden School definition of reference group (van Praag and Frijters 1999). We then would like to see how the intensity and sign of relative utility change in accordance with the characteristics of reference persons.

We found that reference groups with higher educational attainment tend to be the target of stronger jealousy. We also find that if the reference person is older than the subject, feelings of pecuniary emulation are mitigated. The altruistic attitudes toward elderly persons, or admiration of them, are interesting since both higher age and higher education are associated with higher income on average. As we confirmed in the social average task, reference persons with higher income levels draw stronger jealousy. These intriguing relative utility effects for the elderly may be a good reflection of Japanese culture. Regarding the effect of gender, it was found that males are the target of stronger jealousy from both males and females. To sum up, from the Leyden task, we can say that comparison attitudes change on the basis of the features of reference persons. These findings suggest that consideration of social averages as the salient reference group is not sufficient when examining the relative utility effects in the whole society. Instead, researchers should pay attention to the features of the true reference groups of subjects, since they greatly affect the outcome of empirical investigations of relative utility.

6.2 Who-compares-to-whom task

In this task, we simultaneously consider two types of reference groups, friends and work colleagues, in choice questions. The selection of these reference groups comes from the result of our preliminary tests, which showed that these two groups were the most cited by respondents.

In addition to the simultaneous treatment of two external reference groups, in this study we also have information on subjects' specific comparison benchmark, and we examine the following issues: (i) if friends and colleagues are recognized as different type of reference groups, and if so, (ii) how different they are; and (iii) if comparison attitudes toward these two groups vary by specific comparison benchmarks. The answers are as follows.

- (i) The result suggests that friends and colleagues are recognized as different types of reference groups and subjects frequently change their comparison benchmarks from one to the other, rather than stick to one, in accordance with situations they face. This result hints at the fact that reference groups are chosen endogenously.[26]

- (ii) The average comparison effects toward the friend group and the work colleague group are very different in magnitude. The intensity of jealousy toward work colleague group is more than 10 times stronger than that toward the friend group.

- (iii) A noteworthy finding is the large standard deviation of the income comparisons effect toward "friends," compared with the small mean effect. We found that a mean estimate of comparison effects toward friends reflected differing attitudes in the subjects toward their friends. In fact, around 30% of subjects feel altruism toward friends. Another intriguing pattern is found when we compare the estimates of standard deviations for the relative utility effects toward work colleagues by two subgroups of subjects with friends and those with work colleagues as the comparison benchmark. On one hand, the standard deviation is found to be significant (though it is not particularly large) in the subgroup with a comparison benchmark of friends. On the other hand, for the subgroup with a comparison benchmark as work colleagues, the standard deviation is found to be *insignificant*, implying that the negative relative utility effect against work colleagues spikes near the group average. These outcomes suggest that reference groups of friends and work colleagues are different, not only in terms of the average intensities of the relative utility effects, but also in the sense that their distributions differ by subsample groups of specific comparison benchmarks. We argue that this outcome reflects that *friends* encompasses many aspects of life, such as benevolent rivals, persons of understanding, and so forth, and that work colleagues tend to be regarded as rivals.

---

[26]See Train et al. (1987) and Herriges and Kling (1996) for technical discussion on the nested logit model that derived this implication.

7 Concluding remarks

In this concluding section, we discuss biases that are normally associated with *stated preference methods.*

As Hausman (1993) and Carson et al. (2001) pointed out, there are some potential biases in stated choice methods. Bateman et al. (2002) categorized these biases into three broad categories: (i) incentives to misrepresent responses, (ii) implied value cues, and (iii) scenario misspecification. The first category relates to false answers to the survey questions. This bias arises when the questions and scenario settings are not well designed. The second and the third biases result from respondents' misunderstandings of the survey questions, namely cognitive EDE as coined by Zizzo (2010). To avoid these potential biases, researcher should carefully design choice tasks through pilot surveys, repeated preliminary tests and close investigation of the preliminary results. In the present study, we conducted pilot surveys and three preliminary tests to fine tune our questionnaire. The most substantial change in our main test from the preliminary test was the introduction of visual images in the choice situation tasks. By introducing visual images, the cognitive burden on subjects was decreased, with the average elapsed time for finishing the survey reduced by half from the initial preliminary test without images. We also see that standard errors in the logit model estimations were also reduced compared to the estimations obtained from preliminary test data. By comparing our results with those from the preliminary tests, we also find that the order of questions and selection of questions, other than the choice questions in the present study, do not seem to seriously affect the main results of our study.

Dolan and Kahneman (2008) critically summarized biases associated with stated preference methods including hypothetical discrete choice experiments. They then advocated happiness (or experienced utility) research for situations where researchers would like to infer the market values of non-market goods. Note, however, that footnote 4 in Dolan and Kahneman (2008) holds that "[their] critique is focused on the use of measures of decision utility to elicit values of this kind, rather than their usefulness in other contexts, such as predicting behavior." The purpose of this paper is to elicit the sign and the intensity of comparison effects which affect human behavior.[27]

---

[27]The other drawback inherent in stated choice methods that is often mentioned is the artificial nature of the questions and incentive incompatibility for subjects in making choices. Regarding the issue, Lusk and Schroeder (2004) showed that stated choice methods provided similar results for marginal effects compared with the results in non-hypothetical settings. They held that careful design of the survey is the key issue in avoiding this bias, a requirement that we argue that we have satisfied through the use of

In stated choice method studies of Johansson-Stenman et al. (2002), Alpizar et al. (2005), Carlsson et al. (2007), and Andersson (2008), respondents were asked to consider the well-being of their offspring, rather than their own well-being. This framing was used in order to help the respondents liberate themselves from their current circumstances, disentangling their actual consumption from the hypothetical consumption choices in the survey. In the present study, we instead asked about the respondents' own interests. This design choice was made because we would like to know the current situation within Japanese society. The biases associated with ignoring the previous strategy are not expected to be especially severe because we can control for individual fixed effects, as we asked respondents to make five repeated choices in each task, unlike in the previous studies. Our strategy here was also motivated by Dolan and Kahneman (2008), who took a critical view on having subjects make hypothetical choices on the basis of future expectations and past memory.

We suggest a future research agenda as follows. The merit of the hypothetical choice experiment framework under a random utility model is that we do not rely on information of subjective well-being to obtain the true parameters of the (decision) utility function for the relative utility. Because subjective well-being information is usually strongly influenced by country fixed effects and by social norms, hypothetical choice experiment frameworks will be useful alternative avenues in conducting international comparisons of the relative utility effects.

References

ABEL, A. B., "Asset Prices under Habit Formation and Catching Up with the Joneses," *American Economic Review* 80 (May 1990), 38–42.

AI, C. AND E. C. NORTON, "Interaction terms in logit and probit models," *Economics Letters* 80 (July 2003), 123–129.

ALPIZAR, F., F. CARLSSON AND O. JOHANSSON-STENMAN, "How much do we care about absolute versus relative income and consumption?," *Journal of Economic Behavior & Organization* 56 (March 2005), 405–421.

AMIR, O., G. RAND AND Y. GAL, "Economic Games on the Internet: The Effect of $1 Stakes.," *Plos One* 7 (February 2012), e31461.

---

multiple preliminary tests. See also Falk and Heckman (2009).

ANDERSSON, F. W., "Is concern for relative consumption a function of relative consumption," *The Journal of Socio-Economics* 37 (February 2008), 353–364.

ANDREONI, J. AND B. D. BERNHEIM, "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects," *Econometrica* 77 (September 2009), 1607–1636.

ANDREONI, J. AND J. M. RAO, "The power of asking: How communication affects selfishness, empathy, and altruism," *Journal of Public Economics* 95 (August 2011), 513–520.

ARROW, K., R. SOLOW, P. R. PORTNEY, E. E. LEAMER, R. RADNER AND H. SCHUMAN, "Report of the NOAA Panel on Contingent Valuation," Technical Report, National Oceanic and Atmospheric Administration, 1993.

BATEMAN, I., R. CARSON, B. DAY, M. HANEMANN, N. HANLEY, T. HETT, M. JONES-LEE, G. LOOMES, S. MOURATO, E. OZDEMIROGLU, D. PEARCE, R. SUGDEN AND J. SWANSON, *Economic Valuation with Stated Preference Techniques* (Edward Elgar, 2002).

BICKEL, W. K., J. A. PITCOCK, R. YI AND E. J. C. ANGTUACO, "Congruence of BOLD Response across Intertemporal Choice Conditions: Fictive and Real Money Gains and Losses," *The Journal of Neuroscience* 29 (2009), 8839–8846.

BROWN, G. D. A., J. GARDNER, A. J. OSWALD AND J. QIAN, "Does Wage Rank Affect Employees ' Well-being?," *Industrial Relations* 47 (July 2008), 355–389.

CAPPELLI, P. AND P. D. SHERER, "Satisfaction, Market Wages, & Labor Relations: An Airline Study," *Industrial Relations* 27 (Winter 1988), 56–73.

CARD, D., A. MAS, E. MORETTI AND E. SAEZ, "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction," *American Economic Review* 102 (October 2012), 2981–3003.

CARLSSON, F., G. GUPTA AND O. JOHANSSON-STENMAN, "Keeping up with the Vaishyas? Caste and relative standing in India," *Oxford Economic Papers* 61 (January 2009), 52–73.

CARLSSON, F., O. JOHANSSON-STENMAN AND P. MARTINSSON, "Do You Enjoy Having More than Others? Survey Evidence of Positional Goods," *Economica* 74 (November 2007), 586–598.

CARSON, R., "Constructed Markets," in J. B. Braden and C. D. Kolstad, eds., *Measuring the Demand for Environmental Quality, Volume 198* (North Holland: Elsevier Science Publisher, 1991), 122–162.

CARSON, R., N. FLORES AND N. MEADE, "Contingent Valuation: Controversies and Evidence," *Environmental & Resource Economics* 19 (June 2001), 173–210.

CLARK, A. AND C. SENIK, "Who compares to whom? The anatomy of income comparisons in Europe," *Economic Journal* 120 (2010), 573–594.

CLARK, A. E., P. FRIJTERS AND M. A. SHIELDS, "Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles," *Journal of Economic Literature* 46 (March 2008), 95–144.

CLARK, A. E., N. KRISTENSEN AND N. WESTERGARD-NIELSEN, "Economic Satisfaction and Income Rank in Small Neighbourhoods," *Journal of the European Economic Association* 7 (04-05 2009a), 519–527.

———, "Job Satisfaction and Co-worker Wages: Status or Signal?," *Economic Journal* 119 (March 2009b), 430–447.

CLARK, A. E. AND A. J. OSWALD, "Satisfaction and comparison income," *Journal of Public Economics* 61 (September 1996), 359–381.

DE LA GARZA, A., G. MASTROBUONI, A. SANNABE AND K. YAMADA, "The Relative Utility Hypothesis With and Without Self-reported Reference Wages," Iser discussion paper, Institute of Social and Economic Research, Osaka University, 2010.

DOLAN, P. AND D. KAHNEMAN, "Interpretations of Utility and Their Implications for The Valuation of Health," *Economic Journal* 118 (January 2008), 215–234.

DUPOR, B. AND W.-F. LIU, "Jealousy and Equilibrium Overconsumption," *American Economic Review* 93 (March 2003), 423–428.

EASTERLIN, R. A., "Does Economic Growth Improve the Human Lot? Some Empirical Evidence," *In: David, P.A., Reder, M.W. (Eds.), Nations and Households in Economic Growth: Essays in Honour of Moses Abramowitz* (1974).

———, "Will raising the incomes of all increase the happiness of all?," *Journal of Economic Behavior & Organization* 27 (June 1995), 35–47.

———, "Income and Happiness: Towards an Unified Theory," *Economic Journal* 111 (July 2001), 465–84.

Exadaktylos, F., A. M. Espin and P. Branas-Garza, "Experimental Subjects are Not Different," Technical Report, 2012.

Falk, A. and J. J. Heckman, "Lab Experiments Are a Major Source of Knowledge in the Social Sciences," *Science* 326 (October 2009), 535–538.

Falk, A., S. Meier and C. Zehnder, "Did we Overestimate the Role of Social Preferences? The Case of Self-Selected Student Samples," *Journal of European Economic Association* Forthcoming (2011).

Fehr, E. and K. M. Schmidt, "A Theory Of Fairness, Competition, And Cooperation," *The Quarterly Journal of Economics* 114 (August 1999), 817–868.

———, "The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories," in S.-C. Kolm and J. M. Ythier, eds., *Handbook on the Economics of Giving, Reciprocity and Altruism, Vol. 1* (Amsterdam: Elsevier, 2006), 615–691.

Ferrer-i Carbonell, A., "Income and well-being: an empirical analysis of the comparison income effect," *Journal of Public Economics* 89 (June 2005), 997–1019.

Ferrer-i Carbonell, A. and P. Frijters, "How Important is Methodology for the estimates of the determinants of Happiness?," *Economic Journal* 114 (July 2004), 641–659.

Fliessbach, K., B. Weber, P. Trautner, T. Dohmen, U. Sunde, C. E. Elger and A. Falk, "Social Comparison Affects Reward-Related Brain Activity in the Human Ventral Striatum," *Science* 318 (November 2007), 11305–308.

Futagami, K. and A. Shibata, "Keeping one step ahead of the Joneses: Status, the distribution of wealth, and long run growth," *Journal of Economic Behavior & Organization* 36 (July 1998), 109–126.

Gali, J., "Keeping Up with the Joneses: Consumption Externalities, Portfolio Choice, and Asset Prices," *Journal of Money, Credit and Banking* 26 (February 1994), 1–8.

Garcia-Penalosa, C. and S. Turnovsky, "Consumption externalities: a representative consumer model when agents are heterogeneous," *Economic Theory* 37 (December 2008), 439–467.

Haaijer, R., W. Kamakura and M. Wedel, "Mixed Logit With Repeated Choices: Households' Choices Of Appliance Efficiency Level," *International Journal of Market Research* 43 (2001), 93–106.

Hausman, J., *Contingent Valuation: A Critical Assessment* (North-Holland, 1993).

Herriges, J. A. and C. L. Kling, "Testing the consistency of nested logit models with utility maximization," *Economics Letters* 50 (January 1996), 33–39.

Hirschman, A. O. and M. Rothschild, "The Changing Tolerance for Income Inequality in the Course of Economic Development; with a Mathematical Appendix," *The Quarterly Journal of Economics* 87 (November 1973), 544–66.

Hole, A. R., "Fitting mixed logit models by using maximum simulated likelihood," *Stata Journal* 7 (2007), 388–401.

———, "Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment," *Journal of Health Economics* 27 (2008), 1078–1094.

Hollander, H., "On the validity of utility statements: standard theory versus Duesenberry's," *Journal of Economic Behavior & Organization* 45 (July 2001), 227–249.

Horton, J., D. Rand and R. Zeckhauser, "The online laboratory: conducting experiments in a real labor market," *Experimental Economics* 14 (September 2011), 399–425.

Huber, J. and K. Zwerina, "The importance of utility balance in efficient choice designs," *Journal of Marketing Research* 33 (1996), 307–317.

Ida, T. and R. Goto, "Simultaneous Measurement of Time and Risk Preferences: Stated Preference Discrete Choice Modeling Analysis Depending on Smoking Behavior," *International Economic Review* 50 (2009), 1169–1182.

Johansson-Stenman, O., F. Carlsson and D. Daruvala, "Measuring Future Grandparents" Preferences for Equality and Relative Standing," *Economic Journal* 112 (April 2002), 362–383.

Kahneman, D. and J. L. Knetsch, "Valuing public goods: The purchase of moral satisfaction," *Journal of Environmental Economics and Management* 22 (January 1992), 57–70.

Kahneman, D. and A. B. Krueger, "Developments in the Measurement of Subjective Well-Being," *Journal of Economic Perspectives* 20 (Winter 2006), 3–24.

Kahneman, D., P. P. Wakker and R. Sarin, "Back to Bentham? Explorations of Experienced Utility," *The Quarterly Journal of Economics* 112 (May 1997), 375–405.

Kang, M. J., A. Rangel, M. Camus and C. F. Camerer, "Hypothetical and Real Choice Differentially Activate Common Valuation Areas," *The Journal of Neuroscience* 31 (2011), 461–468.

Knight, J., L. Song and R. Gunatilaka, "Subjective well-being and its determinants in rural China," *China Economic Review* 20 (December 2009), 635–649.

Liu, W.-F. and S. Turnovsky, "Consumption externalities, production externalities, and long-run macroeconomic efficiency," *Journal of Public Economics* 89 (June 2005), 1097–1129.

Louviere, J. J., D. A. Hensher and J. D. Swait, *Stated Choice Methods: Analysis and Applications* (Cambridge University Press, 2000).

Lusk, J. L. and T. C. Schroeder, "Are Choice Experiments Incentive Compatible?; A Test with Quality Differentiated Beef Steaks," *American Journal of Agricultural Economics* 86 (May 2004), 467–482.

Luttmer, E. F. P., "Neighbors as Negatives: Relative Earnings and Well-Being," *The Quarterly Journal of Economics* 120 (August 2005), 963–1002.

Manski, C. F., "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies* 60 (July 1993), 531–42.

Mayraz, G., J. Schupp and G. G. Wagner, "Life Satisfaction and Relative Income: Perceptions and Evidence," CEP Discussion Papers dp0938, Centre for Economic Performance, LSE, 2009.

McFadden, D., *Conditional Logit Analysis of Qualitative Choice Behavior in P. Zarembka (ed.) Frontiers in Econometrics* (Academic Press, 1974).

Oswald, A. and S. Wu, "Objective Confirmation of Subjective Measures of Human Well-Being: Evidence from the U.S.A.," *Science* 327 (January 2010), 576–579.

Paolacci, G., J. Chandler and I. P. G., "Running experiments on amazon mechanical turk," *Judgment and Decision Making* 5 (2010), 411–419.

Rand, D. G., "The promise of mechanical turk: how online labor markets can help theorists run behavioral experiments," *Journal of Theoretical Biology* 299 (2011), 172–179.

Ravallion, M. and M. Lokshin, "Who cares about relative deprivation?," *Journal of Economic Behavior & Organization* 73 (February 2010), 171–185.

Rubinstein, A., "Instinctive and Cognitive Reasoning: A Study of Response Times," *Economic Journal* 117 (October 2007), 1243–1259.

Senik, C., "When information dominates comparison: Learning from Russian subjective panel data," *Journal of Public Economics* 88 (August 2004), 2099–2123.

———, "Direct evidence on income comparisons and their welfare effects," *Journal of Economic Behavior & Organization* 72 (October 2009), 408–424.

Sloane, P. J. and H. Williams, "Job Satisfaction, Comparison Earnings, and Gender," *LABOUR* 14 (September 2000), 473–502.

Small, K. A., C. Winston and J. Yan, "Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability," *Econometrica* 73 (07 2005), 1367–1382.

Solnick, S. J. and D. Hemenway, "Is more always better?: A survey on positional concerns," *Journal of Economic Behavior & Organization* 37 (November 1998), 373–383.

Stevenson, B. and J. Wolfers, "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox," (Spring 2008), 1–87.

Suri, S. and D. J. Watts, "A study of cooperation and contagion in web-based, networked public goods experiments," *Plos One* 6 (March 2011), e16836.

Takahashi, H., M. Kato, M. Matsuura, D. Mobbs, T. Suhara and Y. Okubo, "When Your Gain Is My Pain and Your Pain Is My Gain: Neural Correlates of Envy and Schadenfreude," *Science* 323 (2009), 937–939.

TRAIN, K., *Discrete Choice Methods with Simulation, 2nd edition* (Cambridge University Press, Cambridge, 2009).

TRAIN, K. E., D. L. MCFADDEN AND M. BEN-AKIVA, "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices," *RAND Journal of Economics* 18 (Spring 1987), 109–123.

TRICOMI, E., A. RANGEL, C. F. CAMERER AND J. P. O'DOHERTY, "Neural evidence for inequality-averse social preferences," *Nature* 463 (February 2010), 1089–1091.

VAN DE STADT, H., A. KAPTEYN AND S. VAN DE GEER, "The Relativity of Utility: Evidence from Panel Data," *The Review of Economics and Statistics* 67 (May 1985), 179–87.

VAN PRAAG, B. M. AND P. FRIJTERS, "The measurement of welfare and well-being; the Leyden approach," Paul frijters discussion papers, School of Economics and Finance, Queensland University of Technology, June 1999.

VISCUSI, W. K., J. HUBER AND J. BELL, "Estimating discount rates for environmental quality from utility-based choice experiments," *Journal of Risk and Uncertainty* 37 (2008), 199–220.

ZIZZO, D., "Experimenter demand effects in economic experiments," *Experimental Economics* 13 (March 2010), 75–98.

Table 1: Descriptive Statistics

	(1) Our Survey (whole sample)		(2) Our Survey (study sample)		(3) NIKKEI	(4) National Data (b)	
	Male	Female	Male	Female		Male	Female
Age category							
20s	18.20	22.18	18.06	21.51	13.73	19.48	18.75
30s	24.53	22.90	25.37	23.18	36.42	24.48	24.06
40s	20.53	24.47	19.20	24.80	30.07	21.78	21.73
50s	23.07	18.61	22.54	18.85	13.88	22.11	22.64
60s	13.68	11.83	14.83	11.66	5.89	12.15	12.82
Education (a)					N.A.		
Middle school	1.00	0.97	0.94	0.89		18.18	20.80
High school	19.62	26.19	21.20	25.89		41.60	43.39
Some college	10.47	31.89	11.14	32.12		11.36	24.54
College	68.91	40.95	66.73	41.10		28.33	10.67
Marital Status							
Single	32.69	26.5	33.23	25.75	29.99	32.00	23.40
Married	63.95	67.24	63.27	67.92	60.66	61.80	57.60
Divorced/widowed	3.36	6.26	3.50	6.34	9.35	6.20	19.00
Region							
Hokkaido		4.31		4.65	3.97		4.3
Tohoku		4.21		4.36	4.06		7.4
Kanto		45.32		44.23	46.94		32.9
Koshinetsu		3.95		4.19	3.67		6.7
Chubu		10.09		9.96	9.45		11.9
Kansai		20.23		20.55	19.73		16.3
Chugoku		3.92		3.92	3.82		6.0
Shikoku		1.84		1.88	1.90		3.1
Kyushu		6.14		6.28	6.45		11.4
Female [0.1]		52.57		55.59	56.64		51.27
Student [0,1]		3.60		3.41	N.A.		7.60
Annual income (b)	5.69	2.93	4.90	2.71	N.A.	4.87	1.85
Unemployment		4.05		4.01	N.A.		4.90

All figures except for annual income (in million JPY) are percentages for each category.
(a) Those who are currently students are excluded from the figure.
(b) Demographic characteristics are from the Population Estimates by the Statistics Bureau (Sep. 2009); education attainment data are from the Employment Status Survey (Table 3; 2007) by the Statistics Bureau; marital status data are from the Population Statistics of Japan (Table 6.21; 2008) by the National Institute of Population and Social Security Research; region data are from the Population Statistics of Japan (Table 9.5; 2008); income information is from the Employment Status Survey (2008); and unemployment data are from the Labour Force Survey (Feb. 2010) by the Statistics Bureau.

Table 2: Conditional Logit and Mixed Logit Estimates (Social Average Task)

	(1)	(2)	(3)	
Model	Conditional logit		Mixed logit	
Dep. Var: Utility			Mean	SD
Own income	0.048***	0.039***	0.097***	0.077***
	(0.001)	(0.002)	(0.002)	(0.002)
Reference income	-0.022***	-0.021***	-0.044***	0.081***
	(0.001)	(0.001)	(0.001)	(0.002)
Estimated γ	-0.458	-0.546		
Observations	48172	48172	48172	
Pseudo R-squared	0.249			

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$. Robust standard errors clustered by subject in parentheses.

Table 3: Conditional Logit Estimates across Comparison Benchmark Subgroups (Social Average Task)

	(1)	(2)	(3)	(4)	(5)
Comparison benchmark	family	neighbor	friend	colleague	Do not compare
Model			Conditional logit		
Dep. Var: Utility					
Own income	0.045***	0.056***	0.052***	0.050***	0.041***
	(0.003)	(0.003)	(0.001)	(0.002)	(0.001)
Reference income	-0.021***	-0.031***	-0.025***	-0.022***	-0.017***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Estimated γ	-0.467	-0.554	-0.481	-0.440	-0.415
Estimated variance of γ	0.049	0.018	0.064	0.073	0.026
γ same as?	colleague	friend	family	Do not compare	
T statistics	68.2	16.8	21.5	52.4	
Observations	2255	2739	20442	9581	11982
Pseudo R-squared	0.228	0.320	0.279	0.253	0.190

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$. Robust standard errors clustered by subject in parentheses. For estimates of γ, we report variances constructed via the Delta method.

Table 4: Conditional Logit Estimates in the "Do Not Compare" Subgroup by Intensity of Jealousy (Social Average Task)

	(1)	(2)	(3)	(4)	(5)
Intensity of Jealousy	1	2	3	4	5
Model			Conditional logit		
Dep. Var: Utility					
Own income	0.038***	0.041***	0.044***	0.034***	-0.003
	(0.003)	(0.002)	(0.002)	(0.011)	(0.018)
Reference income	-0.013***	-0.015***	-0.022***	-0.038***	-0.035
	(0.002)	(0.001)	(0.002)	(0.010)	(0.040)
Estimated γ	-0.342	-0.366	-0.500	-1.118	N.A.
Estimated variance of γ	0.004	0.001	0.002	0.147	
γ same as?	jealous=2	jealous=3	jealous=4		
T statistics	17.5	145.8	22.1		
Observations	2181	5935	3662	188	16
Pseudo R-squared	0.162	0.183	0.224	0.229	0.127

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$. Robust standard errors clustered by subject in parentheses. For estimates of γ, we report variances constructed via the Delta method.

A   Appendices

Appendix A.1   Leyden task

Appendix A.1.1   Construction of choice tasks

The general method of constructing the choice scenarios in the Leyden task is the same as in social average task, including orthogonal planning. However, in this task, the reference group is not simply the social average, but instead is characterized by the gender, age, and educational attainment of the reference person.

Hence, a total of five alternatives were defined in this task. After the preliminary tests, we determined the levels of these attributes as follows. First, as before, the income variables contain the following variations: 180,000 JPY, 240,000 JPY, 400,000 JPY, 640,000 JPY, and 900,000 JPY. For age, we included four levels – 22, 32, 45, and 58 years old   with the goal of reflecting different stages of workers' careers. Gender was male and female. For the levels of educational attainment, we included five variations: middle school, high school, technical school, undergraduate, and graduate.

One thousand potential variations in the combinations of these attributes exist. The computer algorithm for orthogonal planning in SPSS Conjoint provided 25 sets of alternatives out of 1,000 potential variations. We replicated this procedure to obtain two sets of alternative vectors. To pair the alternatives for this task, we used the same strategy as in the social average task, with the exception that information on the three attributes of socioeconomic characteristics were not taken into account. We also added the no-choice options as in the previous task. The following table shows the set of questions we used in the survey.

Q	Alternative 1					Alternative 2				
			Reference person's characteristics					Reference person's characteristics		
	Own income	Ref. income	Gender	Education	Age	Own income	Ref. income	Gender	Education	Age
1	18	90	female	under graduate	32	40	40	male	junior high	32
2	40	18	male	graduate	32	64	40	male	tech school	45
3	24	24	male	graduate	45	40	24	male	tech school	22
4	90	24	female	junior high	32	64	64	female	junior high	22
5	18	40	female	graduate	58	24	40	male	under graduate	22
6	40	40	male	under graduate	22	64	24	male	graduate	22
7	40	24	female	high school	22	90	24	male	high school	32
8	40	64	male	junior high	58	24	64	male	high school	45
9	64	40	female	junior high	45	24	24	female	junior high	58
10	90	90	male	high school	58	64	90	female	under graduate	32
11	24	64	female	under graduate	22	18	64	male	graduate	32
12	90	40	male	tech school	22	90	90	male	junior high	45
13	64	90	male	graduate	22	24	18	female	tech school	32
14	64	24	male	under graduate	58	24	90	male	graduate	22
15	18	24	male	tech school	22	40	64	male	under graduate	58
16	64	18	female	high school	22	90	18	male	under graduate	22
17	18	64	male	high school	45	18	40	female	high school	22
18	40	90	female	tech school	45	18	18	male	junior high	22
19	90	64	female	graduate	22	40	18	female	graduate	45
20	90	18	male	under graduate	45	40	90	female	high school	22
21	24	40	male	high school	32	18	90	male	tech school	58
22	64	64	male	tech school	32	64	18	male	high school	58
23	18	18	male	junior high	22	18	24	female	under graduate	45
24	24	90	male	junior high	22	90	40	female	graduate	58
25	24	18	female	tech school	58	90	64	female	tech school	22

Appendix A.1.2 Instructions in the survey

Before the subjects started the repeated choice questions, they were shown an instruction screen saying that:

> *The next figure shows your hypothetical monthly income (before tax). It also shows the monthly income (before tax) of a certain other person. As in the previous question [social average task], suppose that the current situation of your monthly income (before tax) and the other person's monthly income (before tax) are both as shown.*

> *This certain other person might, for example, be a 28-year-old woman with a university degree, or a 58-year-old man with a high school diploma. The characteristics of this other person vary in each question.*

In the subsequent screens, we asked respondents to answer the following question while showing them various figures for different alternatives after the question, as shown below.[28]

> *Comparing situation 1 and situation 2 shown in the figures, which is more preferable to you? Suppose that the price levels in the two situations are the*

---

[28]In the survey information in the figures was presented in Japanese. The images for monthly income differ in terms of number of banknotes shown according to the attribute levels. Also, images for reference person differ, depending on his or her characteristics. Subjects repeated five questions and they were not allowed to go back to previous questions once they had made their choices.

same.

As we documented above, we prepared in a total of 25 choice sets consisting of specific hypothetical amounts for a person's own pre-tax monthly income and for the reference person. Each respondent answered five randomly assigned questions out of the 25 total questions.

Appendix A.1.3   Empirical results

We identify the effects of the reference person's characteristics on marginal utility by adding interaction terms for reference income levels and reference person characteristic dummy variables. We created these dummy variables as follows. Regarding gender, we made a *different gender* dummy variable, with information of the subjects' own gender and that of the reference group in the choice scenario (0="same gender"). With respect to age, we created dummy variables for *higher age* and *younger age* (0="same age") using information on the subjects' own age and that of the reference group in the choice scenarios. Finally, using information on the subjects' own level of educational attainment and that of the reference group in the choice scenarios, we created dummy variables for *higher education* and *lower education* (0="same education"). These interaction terms are added into the conditional logit model estimation, thereby examining how this additional information on reference groups affects social comparison. Table 5 presents the results.

In column (1) of Table 5, we show the results of a conditional logit model estimation when we do not control for the effects of subjects' own individual characteristics on their own income, reference income, and reference group type dummy interactions. Column (2) of Table 5 shows the results when controlling for these effects.[29] The columns show that one's own income effect is positive and significant, as is the case in the previous task. The

---

[29]Our findings are robust against changes in the threshold level for the comparison-conscious group, the happy group, the high-income group, and the elder group. These results are available upon requests.

main effect of reference income identified is negative and significant.

As columns (1) and (2) show, reference groups with higher educational attainment tend to be the target of stronger jealousy. We also find that if the reference person is older than you are, feelings of pecuniary emulation are mitigated. The altruistic attitudes toward elderly persons, or admiration of them, are interesting since both higher age and higher education are associated with higher income on average. As it was confirmed in the social average task, reference persons with higher income levels draw stronger jealousy. These intriguing relative utility effects for the elderly may be a good reflection of Japanese culture.

While columns (1) and (2) do not provide strong evidence showing effects of reference persons' gender on comparisons, we actually have significant effects when we divide our sample into subgroups of male and females. Column (3) (male) and column (4) (female) suggest that males have stronger jealousy toward people of the same gender than they do toward females, whereas females have weaker jealousy toward people of the same gender than they do toward males. Hence, we conclude that males are the target of stronger jealousy in Japanese society. Columns (3) and (4) also suggest that the previous results on the effects of age and educational backgrounds of reference groups remain unchanged.

To sum up, from the Leyden task we can say that comparison attitudes change on the basis of the features of reference persons. These findings suggest that consideration of social averages as the salient reference group is not sufficient when examining relative utility effects. Instead, researchers should pay attention to the features of the true reference groups of subjects, since they can greatly affect the outcome of empirical investigations on relative utility.

Appendix A.2 Who-compares-to-whom task

Appendix A.2.1 Construction of choice tasks

In the descriptive statistics from the pre-test, we could see that the most often cited reference group is *friends*, followed by *work colleagues*. Now these two groups are treated as reference persons.[30] We thus use three attributes, one's own income, reference income of friends, and reference income of colleagues. While we can elicit the intensity and signs of relative utility for friends and colleagues using a two-situation-choice framework as in

---

[30]In the Japanese social context, the two reference groups (friends/classmates and work-related) may not be mutually exclusive. The nested-logit regressions, however, show that respondents distinguished these two reference groups clearly. We thank Charles Yuji Horioka for pointing out this potential flaw in the structure of the choice experiment.

the previous tasks, the framework of this choice task has five options: (i) situation 1, (ii) situation 2, (iii) situation 3, (iv) situation 4, and (v) do not know/cannot answer.

We created this expanded framework so as to use a *tree structure* for the choice options. Our purpose is to investigate whether people perceive two different reference groups as actually different. More specifically, we would like to exclude the possibility that people define their comparison benchmark to simply be others and that the exact characterizations of others are not important.

As before, the levels of the three attributes have five variations: 180,000 JPY, 240,000 JPY, 400,000 JPY, 640,000 JPY, and 900,000 JPY. Given that we have three attributes in this task, there are 125 potential variations of alternatives. Again, orthogonal design was used to pick up 25 out of the 125 variations to make a vector of alternatives. We repeated this procedure four times to form the four-situation choice task used in the survey. We paired these four situations to form a choice set such that we can make use of a tree structure in the hypothetical choice experiments. Two attributes for situation 1 and situation 2 are characterized by the same level of income for *colleagues*, while one's own income and income levels of *friends* are randomly chosen. Regarding the attributes of situation 3 and situation 4, the income level of *friends* is fixed, while one's own income and income levels of *colleagues* are randomly chosen. We call the nest of situation 1 and situation 2 *C-fixed*, while the second nest of situation 3 and situation 4 is called *F-fixed*. For respondents who consider that *only* the reference income of friends matters, the F-fixed nest exhibits the similarity of the choice options in the nest. Also, for respondents who consider that only reference income of work colleagues matters, the C-fixed nest shows the equivalence of the choice options in the nest. With this tree structure, if subjects think that there is no difference between the reference group of friends and that of work colleagues– in other words, if they think of both reference groups of friends and work colleagues as being simply "others"–the tree structure of the choice options becomes irrelevant. If this is the case, from the nested logit model estimation, we would obtain that *Inclusive Value (IV)* parameters related to respective nests are estimated to be significantly different from one. The following table shows the set of questions we used in the survey.

| | Alternative 1 | | | Alternative 2 | | | Alternative 3 | | | Alternative 4 | | |
| | | Ref. income | | | Ref. income | | | Ref. income | | | Ref. income | |
Q	Self	colleague	friend	Self	colleague	friend	Self	colleague	friend	Self	colleague	friend
1	90	40	64	18	40	90	90	64	40	64	18	40
2	90	64	24	64	64	40	64	24	90	18	90	90
3	64	64	64	18	64	24	90	40	18	24	24	18
4	40	90	64	40	90	40	40	64	90	64	40	90
5	18	24	64	24	24	90	24	64	18	18	18	18
6	64	18	90	90	18	64	64	90	18	40	40	18
7	64	40	18	40	40	64	90	18	24	40	64	24
8	40	40	40	64	40	18	40	90	24	24	44	24
9	24	64	40	90	64	94	18	40	90	24	18	90
10	18	40	24	90	40	40	24	44	64	64	24	64
11	24	18	64	64	18	24	40	24	18	64	64	18
12	64	24	40	90	24	18	90	28	64	24	90	64
13	64	90	24	18	90	64	64	40	24	18	24	24
14	90	90	90	24	90	18	64	64	64	40	18	64
15	24	90	18	90	90	24	64	18	40	90	24	40
16	40	64	18	24	64	64	24	18	90	90	68	90
17	40	24	90	64	24	64	18	18	18	90	90	18
18	24	40	90	24	40	24	18	24	40	40	90	40
19	18	64	90	40	64	18	40	40	40	24	68	40
20	18	18	18	40	18	90	24	24	24	64	90	24
21	90	18	40	18	18	18	90	90	90	40	24	90
22	40	18	24	24	18	40	24	90	40	18	40	40
23	24	24	24	18	24	40	18	90	64	18	64	64
24	18	90	40	64	90	90	18	64	24	90	18	24
25	90	24	18	40	24	24	40	18	64	90	40	64

Appendix A.2.2 Instructions in the survey

Before the subjects started the repeated choice questions, they were shown an instruction screen saying that:
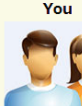
> The next figure shows your hypothetical monthly income (before tax). And in the same way as before, it pairs that amount with the monthly income (before tax) of certain other persons. Suppose that the current situations for these sets are as shown.

> This time for the question, imagine that the certain other persons as (1) a co-worker, (2) a friend.

In the subsequent screens, we asked respondents to answer the following question while showing the various figures for different alternatives, as shown below.
footnoteIn the survey everything in the figures was presented in Japanese. Subjects repeated five questions and they were not allowed to go back to previous questions once they had made their choices.

> Comparing situations 1 through 4 as shown in the figures, which would be the most preferable to you? Suppose that the price levels in the four situations are all the same.

Each respondent answered five randomly assigned questions out of the 25 total questions.

Appendix A.2.3   Empirical results

In the questionnaire for this task, we provided four choice options and a "Don't know/ Cannot answer" option, since we aim to use a tree structure for the choice options. The first two options are the *F-fixed* nest, while the third and the fourth options are the *W-fixed* nest in this task. Our purpose in making these nests is to test if people perceive the friend and colleague groups as independent from each other. If people define their rivals as being merely "others", then the characterization of reference persons is not important, nullifying the nested structure of the four options.

The results from the nested logit model are as follows (not shown in tabular form). First, the IV parameter for the *F-fixed* nest becomes 1.556, while that for the *W-fixed* nest becomes 1.627. Both of these results are significantly different from 1 at the 1% confidence level. These figures indicate that respondents perceive the two reference groups as different from each other. Secondly, an interesting finding here is that the estimated IV parameters exceed 1. According to Train et al. (1987), from a purely statistical perspective, the values of IV parameters indicate the relative substitutability within and among nests, and if they are greater than 1, it means that choice substitutability among nests are

7

more frequent.[31] In our choice setting, the outcome suggests that subjects frequently change their comparison benchmarks from one to the other, rather than stick to a single benchmark group, in accordance with situations they face. This result hints at the fact that reference groups are chosen endogenously.

To investigate the difference in relative utility effects toward friends and colleagues, we first employ a conditional logit model framework as is shown in Table 6. In this task, the true parameters of relative utility, γ_f and γ_w, are calculated by dividing the estimates of the reference income terms for friends and for work colleagues by the estimates of the own income term.

The first column of Table 6 shows the result of conditional logit estimation for the whole study sample. Firstly, it shows that the own income effect is found to be positive and significant, which validates the framework of the choice task in this study. Secondly, the relative utility effects toward the friend group and colleague group are both estimated to be significantly negative, as is the case when the reference group is the social average.

The difference in magnitudes for the terms of these two reference groups, however, warrants attention. Looking at the true estimates of relative utility parameters, the intensity of jealousy toward work colleague group is more than 10 times stronger than that of the friend group. Another interesting finding is that from columns (2) to (6), where estimation results of subgroups of individual-specific comparison benchmark are provided, the relative utility effect toward certain types of friends disappears in some cases.

Especially, in column (4), for those who state that their reference group is friends, the relative utility effect toward friends is not significantly different from zero; whereas in column (5), for those whose reference group is work colleagues, the relative utility effect toward friends is significantly negative.[32] At first glance, this outcome is puzzling.

A mixed logit model framework helps to understand the issue of the weak intensity of comparison attitude toward friends. Column (7) of Table 6 shows that we obtain very similar magnitudes for the mean effects of one's own income, the reference income of friends, and that of work colleagues, as is the case in the conditional logit model in column (1). The column, on the other hand, shows that the relative magnitudes of the

_____

[31]See Herriges and Kling (1996) for the relationship between the magnitude of IV parameters and the global necessary and sufficient condition of utility maximization behavior in a random utility model framework.

[32]We point out that the intensity of jealousy toward the reference group of work colleagues by those who answered that they do not compare is the weakest among subgroups (column 6). Together with the same finding in the social average task, this result validates our data set.

standard deviation terms compared to their mean estimates are very different from each other.

A noteworthy finding is the large standard deviation for the reference income level for friends compared to the mean. With this finding, we conclude that a mean estimate of reference income for friends that is close to zero reflects differing attitudes in the subjects toward their friends. In order to visualize the intuition of this point, we provide figure 1 in which the cumulative distribution functions (CDFs) of the true parameters of relative utility for the social average (γ_a), for the friend group (γ_f), and for the work colleague group (γ_w) are illustrated.[33] The figure clearly shows that the distributions of these true magnitudes of relative utility effects exhibit different patterns from each other. The CDF of γ_w shows that all the subjects in our study sample have negative relative utility toward work colleagues, whereas the CDF of γ_f shows that around 30%of subjects feel altruism toward friends. We also see from the CDF of γ_a that the distribution of γ_a has the largest variance.

It is also interesting to note the differences in the distribution of relative utility effects, in addition to those in the intensity of comparisons, by separately regressing subsamples of specific comparison benchmarks using the mixed logit model framework. Results are shown in Table 7.

The first noteworthy finding is that in each subgroup of specific comparison benchmarks, the means of the effect of one's own income provides quite similar magnitudes to each other. The means of relative utility effects for friends are found to be significantly negative, except for subgroups with neighbors as the comparison benchmark. For the subgroups with family as the comparison benchmark, the mean relative utility effect of friend's income is estimated to be significant, but it is only at the 10% confidence level. In all subgroups, the absolute values of the mean estimates of relative utility effects for friends are close to zero. Another noteworthy observation is that the estimated standard deviation terms of the reference income terms for friends are large compared to the means. Notice that those terms are significant for all subgroups, including the subgroup with work colleagues as the comparison benchmark.

The mean relative utility effects of work colleagues' income are found to be significantly negative for all subsample regressions. The magnitudes of the mean estimates compared to the own income terms are larger than they were in the case of using friends as the reference group.

---

[33]γ_a is obtained in the social average task. We obtained individual parameters of relative utility using the inverse Bayesian formula after the mixed logit model estimation (Train 2009).

An intriguing pattern is found when we compare the estimates of standard deviations for the relative utility effects toward work colleagues for the two subgroups of subjects with friends (column 3) and with work colleagues (column 4) as the comparison benchmark. On one hand, the standard deviation is found to be significant (though it is not particularly large) in the friend subgroup. On the other hand, for the work colleague subgroup, the standard deviation is found to be *insignificant*, implying that the negative relative utility effect against work colleagues spikes near the average point among those subsamples. This outcome suggests that reference groups of friends and work colleagues are different, not only in terms of the average intensities of the relative utility effects, but also in the sense that their distributions differ by subsample group for specific comparison benchmarks. We argue that this outcome reflects that *friends* encompass many aspects of life, such as benevolent rivals, persons of understanding, and so forth, and that work colleagues tend to be regarded as rivals.

Figure 1: Distributions of Relative Utility Parameters (Social Average, Friends, and Colleagues)
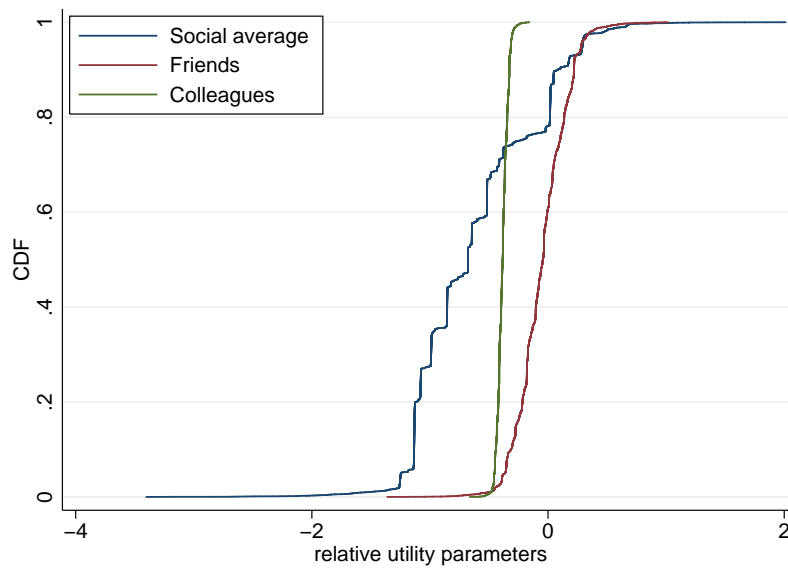
Table 5: Conditional Logit Estimates with Characterized Reference Groups (Leyden Task)

	(1)	(2)	(3)	(4)
Dep. Var: Utility		Conditional logit model		
Own income	0.028***	0.028***	0.029***	0.020***
	(0.001)	(0.001)	(0.002)	(0.001)
Reference income	-0.001	-0.001	-0.002	0.001
	(0.001)	(0.001)	(0.002)	(0.002)
Interactions:				
Reference person's demographic * reference income				
Different sex $*\bar{y}$	0.001	0.001	0.003***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
Higher age $*\bar{y}$	0.005***	0.005***	0.007***	0.007***
	(0.001)	(0.001)	(0.002)	(0.002)
Lower age $*\bar{y}$	-0.004***	-0.004***	-0.005***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Higher education $*\bar{y}$	-0.004***	-0.004***	-0.004***	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Lower education $*\bar{y}$	-0.000	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Interactions of own individual characterstics				
and income levels (a)	No	Yes	Yes	Yes
Observations	45554	45554	20328 (male)	25226 (female)
Pseudo R-squared	0.194	0.213	0.253	0.186

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$. Standard errors clustered by subject in parentheses. Omitted categories are "Same age $*\bar{y}$" and "Same education $*\bar{y}$."
(a) If Yes, individual characteristics are controlled with interaction terms for the own income term, and for reference income. The same set of individual characteristic variables controlled in the social average task regressions are taken into account.

Table 6: Conditional Logit and Mixed Logit Estimates (Who-Compares-to-Whom Task)

Comparison benchmarks	(1)	(2) family	(3) neighbor	(4) friend	(5) colleague	(6) Do not compare	(7) Mixed logit	
			Conditional logit				Mean	SD
Dep. Var: Utility								
Own income	0.028***	0.028***	0.031***	0.028***	0.030***	0.025***	0.035***	0.029***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)
Reference income (Friend)	-0.001***	-0.001	0.000	0.000	-0.003***	-0.002***	-0.002***	-0.011***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Reference income (Colleague)	-0.011***	-0.012***	-0.012***	-0.012***	-0.011***	-0.007***	-0.011***	0.006***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
Estimated γ_f	-0.036	0	0	0	-0.100	-0.080		
Estimated variance of γ_f		0.039	0.030	0.013	0.017	0.017		
Estimated γ_w	-0.393	-0.429	-0.387	-0.429	-0.367	-0.280		
Estimated variance of γ_w		0.047	0.036	0.015	0.019	0.019		
Observations	47180	2226	2664	20131	9387	11639	47180	
Pseudo R-squared	0.140	0.143	0.168	0.147	0.158	0.111		

$**p < 0.01, **p < 0.05, *p < 0.1$. Robust standard errors clustered by subject in parentheses. For estimates of γ, we report variances constructed via the Delta method.

Result omitted for the case of "other comparison benchmarks".

Table 7: Mixed Logit Estimates for Subgroups of Comparison Benchmarks (Who-Compares-to-Whom Task)

Comparison benchmark	(1) family		(2) neighbor		(3) friend		(4) colleague		(5) Do not compare	
					Mixed logit					
Dep. Var: Utility	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Own income	0.036***	0.031***	0.036***	0.024***	0.035***	0.028***	0.038***	0.028***	0.032***	0.030***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Reference income (Friend)	-0.002*	0.013***	-0.000	0.006*	-0.001**	0.012***	-0.004***	0.011***	-0.003***	0.008***
	(0.001)	(0.002)	(0.001)	(0.003)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
Reference income (Colleague)	-0.013***	0.010***	-0.012***	0.006**	-0.013***	0.007***	-0.011***	0.003	-0.007***	0.003
	(0.001)	(0.003)	(0.001)	(0.003)	(0.000)	(0.001)	(0.001)	(0.005)	(0.000)	(0.002)
Observations	2226		2664		20131		9387		11639	

$* * * p < 0.01, * * p < 0.05, * p < 0.1$. Robust standard errors in parentheses.
Result omitted for the case of "other comparison benchmarks".

14