Discussion Paper No. 795

THE EASTERLIN PARADOX AND ANOTHER ANATOMY OF INCOME COMPARISONS: EVIDENCE FROM HYPOTHETICAL CHOICE EXPERIMENTS

Katsunori Yamada
Masayuki Sato

September 2010

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

The Easterlin Paradox and Another Anatomy of Income Comparisons: Evidence from Hypothetical Choice Experiments*

Katsunori Yamada † Masayuki Sato ‡

Abstract

This paper provides evidence from Internet-based, large-scale survey data of hypothetical choice experiment on the relative utility hypothesis. The methodology exploited here complements previous empirical results from happiness studies, incentivized choice experiment studies, and neuroscience studies in such a way that methodological problems among previous studies within these fields are resolved. We show that not only the intensity but also the distribution of relative utility are different across specific comparison benchmarks (internal reference group), and across types of reference groups people are facing in the experiments (external reference group). The relative utility effect among Japanese respondents, while shown to exist in the form of jealousy, is found to be not as strong as can validate the Easterlin paradox. Comparison benchmark with daily contacts is related to stronger jealousy. We also provide empirical evidence, which nuances that the reference group is chosen endogenously.

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

JEL classifications: C25; D03; D62

*The authors are grateful for helpful comments, including those on the previous version of the paper: from Takahiro Ito, Fumio Ohtake, seminar participants at Kyoto University and Osaka University, and conference participants at Singapore Management University. The usual disclaimer applies. Financial support provided from the Global COE program, "Human Behavior and Socioeconomic Dynamics" of Osaka University, and from JSPS research fund KAKEN (21830057), is appreciated.

†*Corresponding author:* Institute of Social and Economic Research, Osaka University, kyamada@econ.osaka-u.ac.jp, Mihogaoka 6-1, Ibaraki 567-0047, Japan.

‡Field Science Education and Research Center, Kyoto University, msat@kais.kyoto-u.ac.jp, Kitashirakawa Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan.

1 Introduction

Traditional economic theories focus on the role of absolute income/consumption, whereas behavioral evidence suggests that social comparisons influence well-being and decisions (Fliessbach et al. 2007, p1305). Whether social comparison affects individual utility is of key importance for understanding human behavior in any social context, and in understanding asset pricings, which are the consequences of human interactions in markets. This study provides new evidence of social comparison from a hypothetical choice experiment, especially focusing on the distribution of social comparison effect. In doing so, we take account of differences in comparison benchmarks, which subjects perceive in mind (internal reference group), and in comparison targets, which subjects face in experiments (external reference group).

In the literature of happiness study, a seminal paper of Easterlin (1974) argued using individual-level surveys between 1946 and 1970 that although income per capita rose steadily in the United States during this period, average reported happiness showed no increasing trend and even declined between 1960 and 1970. Easterlin (1974, 1995, 2001) explained this paradox by alluding to Duesenberry's relative utility theory (Duesenberry 1949), which suggests that people often compare themselves to a reference group, and so they care not only about their own absolute consumption levels, but about how much they consume relative to that benchmark.

The Easterlin paradox prompted the literature of *economics of happiness* to investigate further the impact of absolute and relative income on subjective well-being measures. The majority of these studies, headed by van de Stadt et al. (1985) and Clark and Oswald (1996), found a negative relationship between satisfaction and various comparison income variables. The view that *increasing the income of all increases the happiness of no one* has been shared among studies such as McBride (2001), Luttmer (2005), and Clark et al. (2008). On the other hand, there are studies observing a relative-income effect that is either relatively smaller in absolute value or statistically insignificant such as Sloane and Williams (2000), Ferrer-i Carbonell (2005), Blanchflower and Oswald (2004), and Brown et al. (2008). Recently, Stevenson and Wolfers (2008) thoroughly re-examined data sets previously exploited in happiness studies, and provided a conclusion that data indicated a clear role for absolute income and a more limited role for relative income comparisons in determining happiness. These mixed results from happiness studies clearly call for further evidence on the relative utility hypothesis.

A crucial matter in empirical studies of relative utility is how to define and collect information about an appropriate reference person (or groups) for each subject; the possibilities, including the average (representative) person in society - colleagues, friends, family, neighbors, and so forth - are endless. Majority of previous studies along the line of happiness study assumed that the reference groups were defined by some social averages. However, this proxy of reference income may be problematic, since the social average may not be a salient reference group for subjects. Some papers tried to overcome this drawback by considering other plausible reference groups, and examples of such studies include: reference group of work places (Cappelli and Sherer 1988; Brown et al. 2008; Clark et al. 2009b), neighbors (Luttmer 2005; Clark et al. 2009a; Knight et al. 2009; Senik 2009) and friends and family members (Senik 2009). Senik (2009) considered various reference groups at the same time, and showed that local comparisons to one's parents,

former colleagues or high school mates are more powerful than self-ranking in the social ladder.

A recent prominent progress from the happiness literature is that Clark and Senik (2010) investigated on the issue of individual specific comparison benchmark, using data of *who compares to whom?*. Their finding was 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 *friend* tend to compare more than those whose comparison benchmark group is *work colleague* do.

Despite the recent development Clark and Senik (2010), studies along the line of happiness study approach encompassed other shared problems, which call for new tests on the Easterlin paradox with other research methodologies. One problem is that proxies of reference income used in empirical tests of happiness studies were imposed on subjects by an econometrician, since usually information regarding *direct* and *cardinal* measures of reference income was missing.[1] Another more minor issue among happiness studies is the use of subjective well-being (SWB) information as a proxy of utility in happiness regressions. There has been debates about the nature of SWB, since researchers have not been sure whether it represents cardinal or ordinal utility (van Praag 1991, Ferrer-i Carbonell and Frijters 2004). Also, SWB measures of respondents will be influenced by many aspects of their life, including individual characteristics and mood on the day of the survey (Kahneman et al. 2006).[2] Hence, it is beneficial if we re-examine findings from happiness studies on the relative utility hypothesis with empirical tests in which direct measures of reference income of valid comparison benchmarks are available, and in which we do not need to rely on SWB measures as a proxy of utility.

Possible complementary research agenda for happiness study on the relative utility hypothesis is actually threefold. The first agenda is incentivized choice experiments (Fehr and Schmidt 1999, Fehr and Schmidt 2006). Laboratory experiments have found evidence of the asymmetric inequality aversion, which reflects concerns for relative payoff. The second approach is neurosciences. The recent developments in neuroscience have clarified the brain mechanism of social preferences, which can be considered to be a hard-scientific micro-foundation for the relative utility hypothesis (Fliessbach et al. 2007, Fehr and Camerer 2007, Tricomi et al. 2010). With these methodologies, researchers can obtain more direct evidence of relative utility than the one gathered in happiness studies, since reference groups in these laboratory experiments are well defined (typically, coupled persons in the experiments), and the levels of income/reward to comparison benchmark persons are explicitly shown to the subjects. Nevertheless, these methodologies encompass a drawback: namely, that it may be difficult to generalize findings from multiple person and incentivized choice experiments, and ones from functional MRI studies with relatively small real rewards, to a social setting such as the Easterlin paradox. Also, the

[1] As far as we notice, the only exception in the literature of happiness study that has information of *direct and cardinal* measure of reference income is de la Garza et al. (2009). Knight et al. (2009) and Senik (2009) exploited information on perceptions of relative position in respondents' villages or in friends/family members. However, their proxies of relative comparisons were ordinal, so that interpretations of the magnitudes of coefficients for comparison effects were not straightforward in examining the Easterlin paradox.

[2] Some researchers such as Hollander (2001), Kahneman and Krueger (2006) argue that a consensus on the validity of SWB as a proxy of utility has been attained.

subjects who participated in most of these experiments were college students, being standard practice in neurosciences, as well as behavioral and experimental economics. Finally, it is not clear if coupled persons in laboratory experiments are salient reference persons in subjects' lives. These are sources of concern when we want to interpret the results from these research methodologies as being socially representative.

However, the third alternative methodology of *hypothetical choice experiments* is able to overcome these issues, and is a nice complement for the other techniques. A first merit of hypothetical choice experiments is that we can collect a lot of information, which could amount to a socially representative data set that is absent in incentivized choice experiment studies and neuroscience studies. Another merit of hypothetical choice experiments in testing the relative utility hypothesis is that information about comparison reference groups and direct measures of reference income is available to subjects, which is likewise present in incentivized choice experiments and neuroscience studies, but absent in happiness studies. Similar to Clark and Senik (2010), we gathered information about each subject's relevant comparison benchmark, and we examined if the intensity of relative utility changes in combination with specific comparison benchmark that respondents perceive, and with types of reference groups that people are facing. Finally, in hypothetical choice experiments, happiness measures are not necessary in estimating the intensity and sign of the relative utility effects, and hence, we can avoid suspicion on validities of empirical results arising from use of SWB measures.[3] Despite these merits, we recognize that our preferred methodology of hypothetical choice experiments also contains problems that can bias estimates of relative utility effects. We will discuss the issue in detail later.

Similar to the present analysis, Solnick and Hemenway (1998); Johansson-Stenman et al. (2002); Alpizar et al. (2005) and Carlsson et al. (2007) investigated the intensity of social comparisons by estimating the *degrees of positionality* using data of hypothetical choice experiments.[4] These studies, however, considered that reference incomes were merely the national average income levels. Another issue among them is that the format of choice was designed in such a way that respondents made iterative choices to get them to a point of indifference. This choice experiment strategy is known to encompass severe biases, as pointed out by Carson (1991) and Arrow et al. (1993), who suggested use of discrete choice experiment of the sort conducted in this study for less severe bias and easier understanding of the questionnaire. Also, degree of positionality inferred by the previous studies encompassed measurement errors, since assigned value of degree of positionality for each point of indifference was given arbitrarily. The results in these studies concluded that relative income, on average, is about as important as absolute income at the margin (p.53, Carlsson et al. 2009), meaning that the Easterlin paradox is valid.

We have found just one paper of Carlsson et al. (2009) in which a hypothetical and discrete choice experiment on relative utility was conducted. Carlsson et al. (2009) considered changes in intensity of relative utility across different reference groups. With face-to-face surveys from 498 college students, they analyzed social comparisons in the caste system in India. They focused on two caste classes, and the importance of relative

[3]One last merit of hypothetical choice experiments over happiness regressions would be that when we quantify the intensity and sign of relative utility, we can fix all non-essential variables other than attributes in hypothetical choice experiments.

[4]Respondents in Solnick and Hemenway (1998); Johansson-Stenman et al. (2002), and Alpizar et al. (2005) were students when Carlsson et al. (2007) has given the social representative survey.

income within and between castes was examined. Because caste membership is hereditary and cannot be changed in India, these reference groups are considered to be exogenous. On the other hand, Clark et al. (2008) argued that despite limited empirical evidence in economics, reference groups will be chosen endogenously, possibly as groups of people's close contacts.[5] This suggestion motivates studies in which we consider reference groups in a more general setting, since societies are not usually as clearly divided as they are in India, where the caste system culture still remains. It is obvious that a good starting point in thinking about this issue is to consider reference groups of colleagues, friends, family, and neighbors - all of which were the focus of recent happiness studies of Senik (2009) and Clark and Senik (2010).

All in all, it is beneficial that we conduct a hypothetical choice experiment study of relative utility in which different types of reference groups and subject-specific comparison benchmarks are considered, and information on direct measures of reference income level is explicitly given to subjects. Our large survey with subjects of various socio-economic characteristics can be nationally representative, which is beneficial to investigate the Easterlin paradox. In this study, similarly to Clark and Senik (2010), we investigate how people change their comparison behavior (intensity and sign of social comparison) in combination with specific comparison benchmarks. Similar to Senik (2009), the difference of intensity of relative utility across different types of *external* reference groups is also examined. We also provide empirical evidence, which suggests that reference groups are chosen endogenously using a nested logit model framework. Our results from hypothetical choice experiments complement previous findings on relative utility from other research methodologies. Detailed explanations of our choice tasks are documented in Section 3.2.

The paper is organized as follows. Section 2 introduces our data, explains how they were collected and draws attention to possible biases. Section 3 presents the construction of the questionnaire. Section 4 introduces empirical framework to test the relative utility hypothesis, and then shows the results. Section 5 presents some concluding remarks and suggests directions for further research.

2 Data

2.1 Research company

Our data set is taken from an original Web-based survey conducted in March 2010.[6] A Japanese consumer monitoring company, Nikkei Research Inc., conducted the survey; as part of the Nihon Keizai Shimbun Group (NIKKEI), it is known to be trustworthy for academic research purposes among Japanese research companies. In order to provide highly reliable research data, its registered subjects are subject to monthly screenings,

[5] An exceptional economics study, which provided evidence of endogenous determination of reference group is Falk and Knell (2004). They argued that reference standards are devices to serve motives of self-improvement and self-enhancement. They empirically showed that reference standards increase in individuals' abilities, and that people thus tend to compare themselves to similar others for the purpose of self-improvement.

[6] The use of the Internet for social surveys is new and is becoming increasingly active. Examples of recent studies with web-based surveys include Viscusi et al. (2008), Ida and Goto (2009), Olsen (2009), and Bech et al. (2010).

which keep information up to date and exclude double registrations. The total number of registered subjects in the Nikkei Database is around 160,000. Incentives for respondents are provided by cash voucher, rather than by points; point incentive is biased, as particular respondents with points tend to answer. The research period is one week in our experiment, and only those subjects who can access the Internet during the weekend can join the survey. This is a sharp contrast with one-day research, which other research companies usually adapt for cost advantage, since with one-day research, subjects are chosen on a first-come basis, causing a bias. All these, and other techniques, allow data to be collected in such a way that it is as valid as possible.

2.2 Collection of data

In the survey, 60,482 Japanese subjects between the ages of 20 and 65 were chosen with stratified random sampling from the Nikkei Database, such that the cohort profile of our sample mirrored the Japanese census statistics. 14,370 subjects completed the survey. This response rate of 23.8 percent looks rather low; however, it should be noted that the un-weighted average of the response rates for similar academic conjoint surveys conducted by the same company in 2008 and 2009 (7 projects) is 28.8 percent. The low response rate is a source of concern, as is this deficiency of web-based surveys that is absent in face-to-face survey studies such as Solnick and Hemenway (1998) and Johansson-Stenman et al. (2002). Nonetheless, we proceed with data from a web-based survey with an apparent merit that we can collect a large number of samples that cover varieties of individual characteristics from throughout Japan. As is documented below, we find that our data set captures significant features of the Japanese people, which is difficult to pursue with small sample, face-to-face surveys.[7]

Because of the low response rate, in table 1, we present a comparison of the sample characteristics to overall statistics for the Japanese population. In Columns (1) and (2), we present descriptive statistics for our sample. Column (1) is for the whole sample and Column (2) is for our study sample, the construction of which is documented below. Column (3) illustrates the feature of Nikkei Database, and the last column shows national statistics.

Since the stratified random sampling was designed to mirror the population cohort profile of Japanese census statistics, the age and gender structures of our sample look quite comparable to national statistics. Women account for slightly more than 50 percent of the entire sample in our data set and in the national data set. Regarding marital status, female subjects those who divorced or separated seem under-sampled. This discrepancy from national statistics happened because the national data accounts for everyone who is over 15 in making the statistics. Since the average length of life for Japanese women is as high as around 87 (with that of men around 78), women tend to be separated at a final phase of their lives, which raised the rate of female divorce/separation in national statistics. On the other hand, in our data set, we do not include those who are older than 70, so that the rates of divorce/separation for males and females are lower than national statistics.

[7]As documented in Ida and Kuroda (2006), broadband Internet services are pervasive in Japan compared to other countries, thus the sample selection problem from a web-based survey will not be serious.

With respect to educational record, in our male samples, just one percent of the sample completed middle school only, 18 percent completed high school, 10 percent pursued more advanced studies in some college, and the remaining 70 percent of the sample holds college or grad degrees. This high-education biased distribution holds for female samples. This is an obvious over-sampling of better-educated participants, probably reflecting the digital divide of lower-educated people. As with lower-educated people, students and the unemployed are also under-sampled, while the discrepancies from national statistics for them do not look so severe. The reasons are not very clear, but we guess that these are also reflections of some digital divide among them. Finally, information on residence areas is compared. There is a clear over-sampling from the Kanto region, which includes Tokyo. Also, people from the Kansai region, which contains Osaka, are slightly over-sampled. Similiar to the case of educational record, those who are living in urban cities seem to have easier access to the Internet, which will be the reason for this bias.

Hence, in our data set, bias arose with respect to educational background and residence area. At this stage, the direction of bias for relative utility effect is actually not very clear. People may acquire higher ethics through pursuits of higher education and they may become altruistic toward others, as argued in Johansson-Stenman et al. (2002). At the same time, those who have higher academic backgrounds tend to earn more, which in turn, can end up with raising pecuniary jealousy in the search for higher status. Japanese people living in urban cities are nowadays said to be indifferent to their neighbors, and this indifferent attitude will make social comparison less meaningful. On the other hand, they have higher frequency of encountering others, which may raise the feeling of rivalry. Having said this, if evidence from Europe can be applicable to our Japanese data set, Clark and Senik (2010) showed that those who have Internet access, city dwellers, and those with higher educational backgrounds are more comparison conscious. Hence, our data set may capture the *upper bound* of the intensity of relative utility.[8]

All in all, this study does not intend to provide an accurate depiction of the representative Japanese people, since our sample is limited to registered subjects of Nikkei Database, and since the direction of possible biases in estimating relative utility effects is not very clear, as described above. This means that the results below may not be generalizable to the entire Japanese population; for instance, our dataset does not cover those who finished middle school only. Nonetheless, due to the large size of our sample, the breadth of all of 47 prefectures across Japan, and a variety of socio-economic back grounds, we believe that this dataset does capture significant features of the Japanese people.[9]

[8]See Johansson-Stenman et al. (2002) for discussions of other potential biases, including purchase of *moral satisfaction* (Kahneman and Knetsch 1992), associated with hypothetical choice experiments of relative utility.

[9]Kuriyama (2008) argued that biases from hypothetical choice experiments may not be so severe, since for subjects, it is difficult to strategize within a short time-span of surveys, and since subjects usually recognize that their strategized answers do not have much impact on the main results of the survey when the number of observations is large.

2.3 Construction of study sample

One clear deficiency for the web-based survey is that researchers cannot eagerly monitor and encourage subjects to participate in the survey. Especially, when subjects do not have a clear understanding of the meanings of questions or backgrounds of the survey, they will try to complete questions as quickly as possible by making up answers without contemplation. Our research company tries to minimize this motivation by making the research fee pay on a lottery basis: not all of the respondents can earn a reward so that those who join only for reward purposes are encouraged not to take part in the survey.[10] This scheme reduces response rate, while it will increase the validity of data when collected. Furthermore, at the beginning of the survey, we stated that the survey is designed for revealing comparison consciousness with academic purposes. This strategy is different from *behind the veil of ignorance* scheme employed by studies such as Johansson-Stenman et al. (2002) and Alpizar et al. (2005), who conducted face-to-face surveys with college students. We realize that our preferred strategy causes bias in pushing subjects toward being comparison conscious, but later we will show that even with this bias and one from the sample selection documented above, the comparison effect in Japan is not so strong as to validate the famous Easterlin paradox.

A good way to monitor subjects' willingness to participate in the survey is to look at elapsed time in completing the survey. If elapsed time is too 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 investigation time to finish the survey was 9 minutes 9 seconds in this survey, with the median value of 6 minutes 5 seconds. In figure 1, we show the cumulative distribution function (CDF) for our whole sample as sample (1).[11] For subjects, the easiest way to finish the survey is to provide the same answers for conjoint questions documented below. In this survey, we conduct three types of choice tasks, and subjects are required to answer five consecutive questions for each task. We eliminated those who provided the same number for all five questions in any tasks (2,218). The CDF of elapsed time for the remaining respondents are reported in figure 1 as sample (2). The figure clearly shows that those who provided the same answers through five consecutive questions in a task tended to finish the survey very quickly, probably without contemplation. We also discard the information of subjects whose elapsed time is shorter than four minutes (968), given feedbacks from a within company pilot test by NIKKEI. So far, we are left with 10,988 respondents.

Next, the observations dropped either missed information for some of the variables used in the empirical analysis below (219), or contained an inconsistency of retirement before the age of 55 (1). Finally, we deleted samples who report their personal annual pre-tax income in 2009 to be higher than twelve million Japanese Yen (565).[12] To ensure that this

[10] At the beginning of the questions, respondents were told that in total, 800 subjects will win the lottery and the reward is JPY500 for each person.

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

[12] The cut-off point, JPY12 million, is higher than the sum of the average of personal pre-tax annual income and 3 standard deviations of the income distribution. There will be two major reasons for the high frequency of high-income level subjects. One reason is that subjects are biased to those with higher education, Internet access, and city dwellers. It is natural that those individuals acquire more income than those who are not. The other reason is that they tried to "cheat" with respect to their income levels.

cut with income variable was not due to 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 this drop in the number of observations in our working sample does not seem to be due to a sample selection problem. At this final stage, we are left with 10,203 respondents. The above cuts, which amounted in total to 30 percent of our whole sample, were done because we would like to proceed with high-quality data. It is, however, probably fortunate for us that the main results documented below remain qualitatively unchanged when we use the whole sample as our study sample.

3 Data description

The survey consisted of three sections. The first section was a warm-up in which respondents answered one question of satisfaction about their levels of income, and two questions of social comparisons. Because these warm-up questions are followed by the hypothetical choice tasks in the second part of the survey, it is likely that the question order makes individuals more comparison-conscious. We proceeded with this ordering, since researchers cannot encourage subjects on site for web-based surveys, and if subjects had been kept in the dark about the study's topic, they would have lost concentration in answering conjoint questions, making the data set less valid.[13] Hence, again, this study will capture the upper bound of the intensity of relative utility effects. The last part of the survey consisted of questions about individual characteristics, including age, sex, educational background, job, marital status, types of residence, residence area, and annual pre-tax personal income in 2009. Part of descriptive statistics is presented in table 1.

3.1 Warm-up questions

In the first warm-up question, respondents are asked to choose one of five possible categories to provide information on their own levels of satisfaction about *income*. Category 1 corresponds to "least satisfied," while category 5 denotes "most satisfied". The second question was related to social comparison and was phrased as "How much do you care about others' income levels?". The respondents were asked to choose from five response options, where 1 corresponded to "do not care at all," and 5 denoted "care very much". The second question concerned the respondent's definition of their reference group. They were asked to choose one category, from those applicable to them, as their reference group among (i) family, (ii) neighbors, (iii) friends, (iv) colleagues, (v) do not care, and (vi) others. From these last two questions, we can see "who compares to whom?" and "how much?", which was investigated in depth in a framework of happiness study for European countries by Clark and Senik (2010).

Table 2 shows the distribution of the reference groups chosen by the respondents. We can see that the most often cited reference group is *friend*, followed by *work col-*

As it is documented in the text, we found that deletion of those high-income subjects is not related to sample selection issue. With the exchange rate in March 2009, 12 million Japanese Yen is around 130,000 US dollars.

[13]Face-to-face surveys such as Johansson-Stenman et al. (2002) and Alpizar et al. (2005) followed the behind the veil of ignorance strategy.

league. These rankings are opposite to those in Europe countries, as documented in Clark and Senik (2010), but it is interesting that both in Europe and in Japan, those two groups are the most important reference groups. Reference groups of *family* and *neighbor* play minor roles, and these findings are exploited in setting our hypothetical choice experiment in Section 3.2.3.[14] In our data set, one-fourth of subjects answer that they do not have comparison groups. This figure is lower than the one for European countries, as found by Clark and Senik (2010).

Table 3 shows the cross-tabulation of the distribution of reference groups and the intensity of social comparison. The table suggests that those who do not think that comparison is important tend not to have comparison groups, which is natural. Those who declare that their reference groups are family, neighbors, or colleagues tend to think that comparison is more important than respondents whose reference groups are friends do. This intriguing pattern of intensity of social comparison among those who think that their reference groups are friends is further discussed in Section 3.2.3, where we compare the reference groups of friend and work colleague.

3.2 Conjoint questions

Our research strategy was to elicit respondents' preference toward social comparison through hypothetical choice experiments (stated preference method). Among some of the previous theoretical studies on preference externalities, such as Futagami and Shibata (1998), it was assumed that people care about relative amount of wealth, that is, a stock variable. However, in our survey, in order to avoid the complexities involved in inter-temporal economic decision-making, we concentrate on the intra-temporal aspect of social comparison. Specifically, we consider pre-tax monthly income levels as our attributes. In choice tasks, the respondents make choices among presented different situations. Situations are described by their own income, income of reference group, and the characteristics of the reference group.

As documented in Section 1, we would like to investigate how people change their comparison behavior (intensity and sign of social comparison) across different reference groups. Like Clark and Senik (2010), we also see if the intensity and the sign of relative comparison effects are affected by reference groups that subjects have in mind, which is done by exploiting "who compares to whom" information obtained in one of the warm-up questions. Finally, we also would like to investigate if reference groups are determined endogenously, probably in accordance with the frequency of daily contacts. This is done in a nested logit model framework. In order to pursue these issues, we conducted the following three hypothetical choice tasks in the survey.

3.2.1 Representative agent task

The first task is a benchmark task and we call it *representative agent task*. The specification here is that the reference group is given by "some representative person in society (Japan)," a definition which is often used in previous empirical and theoretical studies.

[14] We conducted a pre-test and found the same pattern of the distribution of reference groups. We then designed the questions used in conjoint analyses, which take into account differences in reference groups. For detail, see the conjoint questions explained below.

Despite its simplicity, this task will provide a useful benchmark in investigating if the Easterlin paradox is valid when a reference group is explicitly given to subjects in the questionnaire. This task is also useful for theoretical macroeconomists. Previous theoretical studies were usually conducted without estimating important parameters of relative utility. And these put forward various propositions in accordance with the parameters assumed. Parameters estimated in this study will be of help for the reality of their policy implications.

In each choice situation, the respondents make a choice among three options: (i) situation 1, (ii) situation 2, and (iii) do not know/ cannot answer. We provided the no-choice options because of a suggestion by Arrow et al. (1993) and Haaijer et al. (2001), who pointed out the importance of including the no-choice option in hypothetical choice experiments. We then removed observations in which the no-choice option (3) was selected from our regressions. An alternative way of coping with those observations is to interpret them as showing *indifference* between the two situations, rather than a failure to understand the survey question. Unfortunately, we have no information of the true reason why no-option was chosen. Hence, following the literature, we see the results of the first choice in each choice task for a robustness check. It was shown that results presented below are robust (not shown here due to space constraints, but available upon request). The above procedure applies to empirical tests on the following tasks, as well.

At the beginning of this task, respondents are told to "In the following figures, pairs of your monthly pre-tax income, and that of reference persons are presented. Imagine that each set of your income profile and the other's income profile regulates the society's socio-economic situations." The reference group in this task is described by the Japanese social average. To ease the understandings of respondents, we provided them with figure images.[15] Situations are described by two attributes of one's own monthly pre-tax income and the monthly pre-tax income of the reference group. It was also repeatedly stressed that price levels were the same in these two situations, and that the respondents are asked to choose the situation that would be best for their own interests.[16]

Respondents repeat five situation choices, which are defined along these two dimensions. After three pretests, we determined the levels of these attributes and the survey included the following situation variations: JPY 180 thousands, JPY 240 thousands, JPY 400 thousands, JPY 640 thousands, and JPY 900 thousands for each dimension. Since we have two attributes and five levels for each attribute, there are 25 potential variations

[15]It is obvious that figure images helped respondents understand the meaning of the questionnaire of social comparison. After pretests, in which everything was explained in sentences, we found that cognitive burdens associated with understanding the questionnaires of choice experiments were not negligible. This could be confirmed from estimated standard deviations in logit regressions, which reflected the variances of choice patterns; these variances partly reflect less-clear understanding of questionnaires by the subjects. We then decided to provide figure images in the final test, which halved the average elapsed time to complete the survey, and reduced the standard deviations of logit estimations.

[16]In Johansson-Stenman et al. (2002), Alpizar et al. (2005), and Carlsson et al. (2007), respondents are asked to consider the well-being of offspring, rather than their own. This framing was used in order to help the respondents liberate themselves from their current circumstances. In this study, we asked about respondents' own interests, since we would like to know the *current* situations within Japanese society. Parameters thus estimated will be more useful for drawing current policy implications. Also, we can control individual fixed effects, as we asked respondents to make repeated, five situation choices for each task, thereby interpreting the data as panel.

in an alternative. In choice experiments, two alternatives are paired to form choice sets, and the pairing strategy is the researcher's choice to determine. As is documented in Viscusi et al. (2008), the choice design should be paired so as to balance the utility of each alternative.[17] In the choice experiment of relative utility, however, it does not necessarily mean that an increase in a reference group's income leads to decline in one's own utility level; in such, we did not exclude the possibility of altruistic preference. Given that, our best strategy for pairing alternatives is as follows:

Suppose we have a situation of $S = (x, y)$ where x denotes the level of own income and y is other's income. Then, qualitatively, candidate sets of paired situations consist of the following 8 situations; $(x, y+)$, $(x, y-)$, $(x+, y)$, $(x+, y+)$, $(x+, y-)$, $(x-, y+)$, $(x-, y)$, and $(x-, y-)$ where $x+$ means some value greater than x while $x-$ describes a value smaller than x, for example. Since we do not exclude the possibility of altruism a priori, there are no a priori dominant choices over S in these eight alternatives. We then made pairs such that these eight situations appear as evenly as possible. We exploited all of 25 potential variations in the survey. Since a respondent answered five questions for this task, there are five different question sets for it. They are assigned uniform-randomly to respondents. With all these devices, we could efficiently obtain parameter estimates.

3.2.2 Leyden task

The second task is called *Leyden task*, which is designed to see how individual characteristics of reference persons affect social comparisons. As a first step of such an examination, we employ a reference group characterized by age, gender, and educational attainment. This characterization is borrowed from the Ledyen school definition of reference group.[18] Hence, in this task, reference group is not solely the social average. We aim to investigate how intensity (and the sign) of social comparison changes when people face different characterizations of rivals.

As in the *representative agent task*, the respondents make a choice among three options: (i) situation 1, (ii) situation 2, and (iii) do not know/ cannot answer. At the beginning of this task, respondents are again told to "Imagine that each set on your own monthly pre-tax income and the other's monthly pre-tax income regulates society's socio-economic situations." They are also told that, "This time, your rival is not just the social average. It can be, for example, women who are 28 years old, with a college degree, or men who are 48 years old, with high school degrees." The situations are described by two attributes of one's own monthly pre-tax income, and monthly pre-tax income of the reference group, and the rival's individual characteristics: age, gender, and educational back ground.[19] It was also repeatedly stressed that price levels were the same in these two situations, and respondents were asked to choose the situation that would be best for their own interests.

Respondents repeated five situation choices, which were defined along these five di-

[17] See also Huber and Zwerina (1996).

[18] The Ledyen school is a group of researchers who initiated the use of subjective well-being as a proxy of utility. See van Praag and Frijters (1999) for a survey of contributions by the Leyden school. The school also referred to relative utility and its preferred specification of reference group being a group of people of one's age, the same gender, and the same educational attainment.

[19] Blanchflower et al. (2009) employed the Leyden school type definition of comparison group, which is the person's age group for each gender in each nation.

mensions. After pretests, we determined the levels of these attributes as follows. First, income variables contain variations of JPY 180 thousands, JPY 240 thousands, JPY 400 thousands, JPY 640 thousands, and JPY 900 thousands, as before. For age, we included four levels: 22, 32, 45, and 58. These figures reflect different stages of workers' careers. Gender includes male and female; for educational attainment, we considered five variations of middle school, high school, technical school, under graduate, and graduate. Hence, potential variations of sets of attributes are 1,000, out of which 25 were chosen for our survey via orthogonal design.[20] We repeat this procedure two times, and paired alternatives to form 25 choice sets. Our pairing strategy for this task is the same as in the *representative agent task*, except that information on three of the socio-economic characteristics was not taken into account. Since each respondent answers five questions for this task, five question sets were made, and they were distributed in a random fashion to each subject.

3.2.3 Who compares to whom task

Last task is *who compares to whom task*, in which we provided two types of reference group of *friend* and *colleague* in choice experiments. We focused on these two groups after the finding from one of the warm-up questions that these two are dominant among Japanese respondents (and also among European people, as was shown in Clark and Senik 2010). Then, we have three attributes of one's own income, reference income of friends, and reference income of colleagues in this task. While we can elicit intensity and signs of relative utility for friend and colleagues with two-situation-choice framework, as in the previous tasks, the framework of the choice task has five options: (i) situation 1, (ii) situation 2, (iii) situation 3, (iv) situation 4, and (v) do not know/ cannot answer. As before, the levels of these attributes have five variations of JPY 180 thousands, JPY 240 thousands, JPY 400 thousands, JPY 640 thousands, and JPY 900 thousands for each dimension. Since we have three attributes in this task, there are 125 potential variations in alternative. Orthogonal design was used to pick up 25 out of the 125. We repeated this procedure four times to form four-situation choice task used in the survey. We paired these four situations to form a choice set such that we can exploit a tree structure of the conjoint task. This is documented below. Since a respondent answers five questions for this task, there are five different question sets for it. They are assigned uniform-randomly to respondents.

We addressed this expanded framework, since thereby we would like to exploit a *tree structure* of the choice options. Our purpose is that we investigate if people perceive two different reference groups as actually different. More specifically, we would like to exclude a possibility that people define their comparison benchmark to be merely *others*, and that characterization amongst *others* are not important.

In this task, two attributes of situation 1 and situation 2 are characterized by the same level of income for *colleagues*, while one's own income and income levels of *friend* are randomly chosen. Regarding attributes of situation 3 and situation 4, the income level of *friend* is fixed constant while one's own income and income levels of *colleague* are randomly chosen. We call the nest of situation 1 and situation 2 *C-fixed*, while the

[20] We used SPSS version 15.0 for orthogonal design.

second nest of situation 3 and situation 4 is called *F-fixed*. For respondents who consider that *only* reference income of friends matters, the F-fixed nest exhibits similarity of the choice options in the nest. Also, for respondents who consider that *only* reference income of work colleague matters, the C-fixed nest shows 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 friend and that of work colleague - put differently, if they think that both reference groups of friend and work colleague are just taken as *others* - the tree structure of the choice options turn irrelevant.

4 Empirical Results

4.1 Random utility model

Here we describe our theoretical foundation of the concept of conjoint analysis, but only for the case of *representative agent task* due to space constraints. We can easily apply it to the other two tasks.

Individuals derive utility, not only from their own income y but also from by income of the social average $\bar{y}$. Since each individual cannot choose the levels of income by the social average, they are considered as regulating *utility externality*. In this case, the utility function is generally given by

$$U = U(y, \bar{y}), \tag{1}$$

where U represents the utility function. From textbook assumptions, we suppose that subjects value attribute y positively. On the other hand, the social average income can be valued positively (altruism) or negatively (jealousy). Following Johansson-Stenman et al. (2002), Dupor and Liu (2003), Liu and Turnovsky (2005), we address the constant relative risk aversion (CRRA) type utility function as

$$U = \frac{(y\bar{y}^{\gamma})^{1-\rho}}{(1-\rho)}, \tag{2}$$

where $\rho > 0$. If $\rho = 1$, it reduces to log felicity function. The parameter γ is the central topic of this study and regulates intensity and sign of relative utility. If $\gamma < 0$ the individual has jealousy and $\gamma \leq -1$, it means that the Easterlin paradox is valid, since only relative income matters. If $\gamma > 0$, the individual has altruistic preference, whilst if $\gamma = 0$, there is no relative utility.

To analyze the conjoint decisions, we use a random utility model framework. Let i denote alternative i, and n denote respondent n. Then, in order to bring the theoretical framework to the data, we take the logarithms of both sides in Equation 2 to obtain

$$\ln U_{ni} = (1-\rho)\ln y_{ni} + (1-\rho)\gamma \ln \bar{y}_{ni} - \ln(1-\rho). \tag{3}$$

The above expression can be used for estimating parameters via a random utility model together with an error term ϵ_n. The probability p_{ni} that respondent n prefers alternative i to alternative j is given by

$$\begin{aligned} p_{ni} = Prob\big(&(1-\rho)\ln y_{ni} + (1-\rho)\gamma \ln \bar{y}_{ni} - \ln(1-\rho) + \epsilon_{ni} \\ &> (1-\rho)\ln y_{nj} + (1-\rho)\gamma \ln \bar{y}_{nj} - \ln(1-\rho) + \epsilon_{nj}\big), \quad \text{for all} \quad j \neq i. \end{aligned} \tag{4}$$

When ϵ_n is distributed following independent and identical distribution of extreme value type 1 (IIDEV1), we obtain a conditional logit model (McFadden 1974). Logit models estimate the parameters in equation 3 using the maximized likelihood method, with data from choice experiments.

Here, it should be remembered that estimated variables are divided by the scale parameter σ, which is unknown to researchers (Train 2003, p41). Namely,

> [e]ach of the coefficients is scaled by $1/\sigma$. The parameter σ is called scale parameter, because it scales the coefficients to reflect the variance of the unobserved portion of utility. (Train 2003)

We cannot obtain true magnitudes of all the parameters in equation 3, but we can obtain true estimates of γ by dividing the coefficient of $\bar{y}$ with that of y thus canceling σ out. We innocuously assume that $\rho = 1$ so that we consider log utility functions.[21]

4.2 Representative agent task

4.2.1 Result of test on the Easterlin paradox

In order to examine the validity of the Eastelrin paradox in Japan from the data of hypothetical choice experiment, we provide results from the conditional logit model. In the conditional logit model, it is assumed that the IIA assumption holds: random components of each alternative are not correlated. We also assume that the random components within each subject are not correlated. On the other hand, in the mixed logit model that we address later, these assumptions are relaxed and the distributions of parameters across subjects are allowed.[22] Table 4 summarizes the results for the test of the Easterlin paradox.

In column (1) of table 4, we see that self income affects utility positively and significantly, as is expected. The 95 percent confidential interval of the estimate for self income term lies in the positive domain. That self income affects utility positively will validate our structure of the survey because the results from the hypothetical choice experiments conform to textbook economic assumptions. Next, it is shown that relative utility does exist among Japanese respondents, and that the effect appears in the form of *jealousy* on average from the negative and significant estimate of reference income term in the conditional logit model. These two estimates from the conditional logit model provide the *true* magnitude for the parameter γ of relative utility, by dividing the second by the first. From the estimates in column (1), we obtain that $\gamma = -0.458$. This estimate indicates that the intensity of negative relative utility effect among Japanese people is not

[21] Ida and Goto (2009) tried to compare estimated parameters in a logit model framework by assuming that $\sigma = 1$ through all regressions of sub-samples. However, researchers cannot compare estimates from different sub-samples 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 an LR test designed for confirming statistically significant difference of parameters among sub-groups is not suitable here. For example, a larger σ in a sub-sample leads to smaller coefficients in its regression, even when the observed factors in two sub-samples have the *same* effect on utility. A heteroscedastic logit model can be used in investigating the difference of σ between the sub-groups.

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

as strong as that which validates the Easterlin paradox. To paraphrase Easterlin (1995), increasing the income of all increases the happiness of all to some extent.[23] Hence, it seems that our result from hypothetical choice experiment stands in the middle of two extremes: fully relative utility function of Easterlin (1995) and only absolute utility function of Stevenson and Wolfers (2008). It should be noted here that from the above discussion on possible direction of bias of estimated relative utility effect, we are picking up the *upper bound* of the effect (in the form of jealousy). We argue that our robust finding here is that *money buys happiness* since γ will not exceed one in the absolute value, even after we take the bias into account.

4.2.2 Who compares most to the social average?

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

In the representative agent task, the reference group is fixed as the social average. Further, the differences of relative utility effects derived in this task will reflect *basic* difference in the intensity of relative utility across sub-groups, since the focused reference group has a general concept. The simplest examination here is that we divide our study sample into sub-groups of individually relevant reference groups, estimating with the conditional logit framework and comparing obtained *true* magnitudes of relative utility parameters across the sub-groups. Columns (2) - (7) in table 4 provide the results.

Column (2) of table 4 shows the estimate of γ for those who have in mind the reference group of family. While the number of subjects in this category is smaller than the other sub-groups, except for the one for "others" in column (7), estimates from the conditional logit model are shown to be significant at the one percent level, and inferred γ falls just around the one from whole sample in column (1). An interesting estimate of γ appears in column (3), in which the in-mind reference group is neighbors. The intensity of relative utility is the strongest among the sub-groups: people who tend to compare to neighbors are the most jealous in Japan, while there are not so many of those people. The most cited relevant in-mind reference group, friends, provides the second strongest jealousy among reference groups, as is documented in column (4). On the other hand, the second most cited reference group of work-colleagues in column (5) does not seem to provide strong relative utility for Japanese subjects. The intensity of relative utility is smaller than the average, which is surprising when we consider severe intra-firm career competition in Japanese firms. Column (6) shows the result for those who do not compare. It is shown that they actually compare and they have jealousy just as the other respondents. On the other hand, the intensity of comparison is the weakest among sub-groups, as is expected. The last finding will validate our data set and the structure of our choice experiments.

The finding that those who compare neighbors have the strongest intensity of relative utility draws some attention. The feature of reference group as neighbors will be

[23] To paraphrase Blanchflower and Oswald (2004), money buys happiness.

characterized by close contacts. 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 weak intensity of relative utility for those who compare themselves to another close contacts reference group, work colleague, in column (6). Regarding this finding, it may make sense to appeal to the tunnel effect conjecture by Hirschman and Rothschild (1973). They argued that the increase of work colleagues' income could be interpreted as a positive signal regarding likely future outcomes. The difference of relative utility effects between the two close contacts reference groups of neighbor and work colleague in columns (3) and (6) may be explained by the tunnel effect story, while the effect is not so strong as providing *positive* relative utility, as was found by Senik (2004) for a Russian data set.

4.2.3 Heterogeneity of preference parameters

In column (8) of table 4, we show the results from a mixed logit model, in which distributions of parameters are allowed. In other words, in the mixed logit model, we can observe heterogeneity of preference among respondents. Although we assume that the error term is independently and identically distributed, as in the conditional logit model, the mixed logit model allows the discrete choice model to apply in the non-IIA situation. Estimates in the table generally show that our previous results, obtained with the conditional logit model, are robust. Attributes for self income affect utility positively, whereas the reference income term is picked up as a negative. It is also suggested that the ratios of estimate of self income term to that of reference income term from the conditional logit model and from the mixed logit model are very similar, validating the robustness of previous findings from the conditional logit model.

Column (9) in table 4 also suggests that the standard deviations estimated in the mixed logit model may require some attention. In particular, the standard deviations for self income and reference income terms are both significant at the one percent level. This outcome implies that heterogeneity in preference parameters is an important issue. Especially, the associated standard deviation of reference income term is large compared to the mean. This implies that there is large heterogeneity in the parameter for relative utility.

In the literature of behavioral economics, it is well-recognized that demographic differences lead to substantial differences in preference parameters, such as the time discount rate and the rate of risk aversion.[24] Also, it is natural to expect that sign and intensity of relative utility depend on individual characteristics, just as it does on in-mind reference groups.

Following Viscusi et al. (2008), we pick up the effects of individual characteristics on preference parameters by controlling for interaction terms of attributes in the surveys

[24]Ida and Goto (2009) showed that smokers are endowed with a higher value of time discounting and a lower value of risk aversion than nonsmokers. Viscusi et al. (2008) showed that eco-conscious individuals have a lower rate of time discounting than those who are not eco-friendly. 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 preferences of patients over general practitioner appointments using standard logit, mixed and latent class logit models. He showed that there was significant preference heterogeneity for all the attributes in the experiment.

and demographic variables in conditional logit models. In this paper, we consider two organic factors, six acquired individual characteristics, and three subjective variables as potential sources of parameter heterogeneity. Related to the organic variables are the two variables of age and gender. We consider annual income level, educational attainment, city dwellings, marital status, unemployment status, and student dummies for individual characteristics. Three subjective dummy variables include no in-mind reference group dummy, very happy dummy, and very comparison conscious dummy. Table 5 reports the results with various dummy interactions.[25] Column (2) of table 5 is the benchmark case, and columns (3) - (6) are presented for robustness checks. We provide column (1), result of no dummy interactions regression, for readers' reference.

From column (2), we can see that people tend to become more jealous if they are rich, female, highly educated, or married. Interestingly, city dwellers do not have stronger comparison attitudes when compared to those who do not live in capital cities. It is also interesting that age does not affect comparison intensity. Regarding subjective variables, it is shown that those who do not have an in-mind reference group have weaker comparison attitudes, and that the more they care about comparisons, the stronger their jealousy becomes. These findings are as expected, while the feelings of being happier do not affect comparison intensity. These findings are robust against changes in threshold level for the Comparison Conscious group, the Happy group, High Income group, and Elder group, as is shown in columns (3) - (6).[26] Thus, we can confirm that heterogeneity plays a role in determining the parameters, just as previous behavioral economics studies have found in other fields.

4.2.4 Welfare implications of estimated relative utility effects

The Commission on the Measurement of Economic Performance and Social Progress (CMEPSP), initiated by the President of the French Republic, Nicholas Sarkozy, documented dissatisfaction with the present state of statistical information about the economy and the society, especially about GDP measures. The report from the commission (Stiglitz et al. 2009) considered what additional information might be required for the production of more relevant indicators of social progress than GDP. The report argued that they should capture distributional effect and relative utility effect among others. From our data set, we can infer the relevance of introducing such additional information into aggregative measures of economic activity.

First we consider a simple social welfare measure, which can be obtained from information of $\tilde{y}$, the average income of all subjects in the study sample, for our benchmark. This $\tilde{y}$

[25]Note that introducing interaction terms into conditional logit frameworks is fine as long as we confine our attentions on the sign and significance of the interaction terms. See Ai and Norton (2003) for interpretations of marginal effects of dummy interaction terms in logit models.

[26]Column (2) is the baseline case, where threshold levels of Comparison Conscious, Happy group, High Income, and Elder are given as category 5 (highest), category 4 (second highest), category 5, and the age of 53 (75 percent of study sample), respectively. Column (3) is the case in which threshold changes in subjective variables of Comparison Conscious and Happy into categories 4 (second highest) and 3 (middle), respectively, are allowed. Column (4) is the case where threshold change in High Income to category 4 is allowed. Column (5) is the case where threshold changes in Elder category to the age of 43 (50 percent of study sample) is tested. Finally, Column (6) is the case where threshold changes in Comparison Conscious, Happy, High Income, and Elder documented above are allowed simultaneously.

will be a proxy of a standard per capita GDP measure. In our data set, $\tilde{y} = 3,674,017.446$ in Japanese Yen. Now we assume that utility function is given by log form, and per capita welfare is obtained as

$$Welfare_1 = \log \tilde{y} = 15.117.$$

Following the spirit of Stiglitz et al. (2009), we next consider the effect of introducing income distribution into the benchmark measure of social welfare. In this case, the relevant indicator is the average welfare level of the subjects, which is obtained as

$$Welfare_2 = \frac{\sum_i \log y_i}{10203} = 14.917,$$

where i denotes subject index and y_i is individual annual income level from the data set. The reduction of per capita welfare level of 1.33 percent from $Welfare_1$ to $Welfare_2$ is due to the rightly skewed income distribution. However, against our presumption and significance documented in Stiglitz et al. (2009), the magnitude of introducing income distribution effect looks rather minor. Thirdly, we take into account of relative utility effects estimated from the above conditional logit model. Now, the reference group is the social average, so that we can use the value of $\tilde{y}$. In this case, the per capita welfare, which is adjusted for relative utility effect, is obtained as

$$Welfare_3 = \log(\tilde{y}\tilde{y}^{-0.458}) = 8.193.$$

As it can be seen from $Welfare_3$, the introduction of relative utility effect provided a major reduction of 44 percent in the per capita welfare, compared to the one from the standard and previously most cited per capita welfare measure of $Welfare_1$. It is noteworthy here that the effect is much stronger in considering relative utility effect than the one obtained when considering distributional effects. Finally, we also allow variations of relative utility effects. Here, we need information of individual parameters of relative utility γ_i, which can be obtained using Bayesian Reverse Formula after mixed logit model estimations (Train 2003).[27] Again, the reference group is the social average, and the per capita welfare which is adjusted for *individual* relative utility intensity, is obtained as

$$Welfare_4 = \frac{\sum_i \log(y_i \tilde{y}^{\gamma_i})}{10203} = 5.610,$$

where γ_i denotes individual parameter of relative utility.[28] Although around 5 percent of subjects have positive relative utility, altruism, toward the reference group of the social average, the per capita welfare level, which considers variations in relative utility

[27] We found that the individual-level estimates of γ_i are positive for approximately 5% of subjects, indicating that they are altruistic toward the social average. This finding serves to support the strategies adopted in theoretical studies on preference externality, such as Liu and Turnovsky (2005), which dealt with both jealousy and altruism (admiration). A caveat here is that economists may be better served to think about to what extent jealousy or admiration is pervasive in a macro-economy, rather than to think about an economy where relative utility is characterized only by a negative or a positive one. In that sense, the research agenda of Garcia-Penalosa and Turnovsky (2008), which considers the heterogeneity of preference parameters, is more promising.

[28] We followed the procedure of Revelt and Train (1998) in estimating individual parameters of relative utility: we fixed the effect of own income as constant across individuals. See footnote 8 of Revelt and Train (1998) for details on this point.

effect and income distribution, reduces a further 31 percent from $Welfare_3$ because of socially dominating jealousy effects. In total, the welfare reduction from the $Welfare_1$ to $Welfare_4$ is as much as 70 percent, outweighing the effect of introducing merely income distribution effect. To sum up, between income distribution and relative utility, the latter plays a more important role in terms of magnitude when we consider improvements of GDP measures, as is suggested by Stiglitz et al. (2009). Here, we consider merely cross sectional *level* effects of incorporating additional information to GDP, and we stay away from the dynamic aspect of it because of data limitations. This is a topic of future research.

4.3 Leyden task

In the Leyden task, we would like to see how intensity (and sign) of relative utility change when subjects face reference groups, which are characterized with demographic variables, not just the social average. We then are able to see how salient the reference group of the social average was. Following the spirit of Leyden school, we consider reference groups with individual characteristics of age, gender, and educational background.

Table 6 presents the results from the conditional logit model framework. In all columns, we pick up the effects of types of reference groups by interaction terms of reference income level and reference group characteristic dummy variables. We created those dummy variables as follows. Regarding gender, we made a *different sex* dummy variable, with information of subjects' own gender and that of reference group in the choice questions (omitted category is "same gender").[29] With respect to age, we created *higher age* and *same age* dummy variables (omitted category is "younger age") from information of subjects' own age and that of the reference group in the choice questions. Finally, from the information of subjects' own educational background and that of the reference group in the choice questions, we created dummy variables of *higher education* and *same education* (omitted category is "lower education"). These interaction terms are added into the conditional logit framework in the representative agent task, thereby examining how this additional information on reference group is salient.

In column (1) of Table 6, we show the result of a conditional logit model estimation where we do not control for subjects' own individual characteristics on their own income, reference income, and reference group type dummy interactions. The column shows that one's own income effect is positive and significant, as with the case of representative agent task. The partial effect of reference income is picked up as negative and significant, which is also the same for the case of representative agent task. An interesting finding here is that we can see the sign and magnitude of relative utility toward specific types of reference group. As the column shows, people feel *strong jealousy* if they are facing reference groups of different sexes (compared to reference groups of the same gender). It is also shown that reference groups of *higher education* tend to be a target of stronger jealousy.[30] Finally, in

[29] In the results we provide here, we do not control for asymmetric gender effects between men and women.

[30] This is the case when compared to the reference group of lower education. The reference group of same education is also associated with stronger jealousy, compared to the reference group of lower education, but the intensity of jealousy toward this group is found to be weaker than that of higher education.

Japan, if the reference person is elder, it is shown that the feeling of pecuniary emulation is mitigated. The effect of altruism toward an elderly person, or admiration of them, is comparable to the negative feeling against those with higher education. These intriguing patterns of relative utility effects for different types of relative utility may be a good reflection of Japanese cultures.

In columns (2) - (6) of Table 6, we show the result of a conditional logit model estimation, in which we control for subjects' own individual characteristics on one's own income, reference income, and reference group type dummy interactions. The creation of dummy variables for one's own individual characteristics is the same as the cases of representative agent task regressions in columns (2) to (6) of Table 5.[31] From the columns, we can see the previous result of column (1) is robust to the inclusion of subjects' own individual characteristics, with respect to higher age dummy and higher education dummy. Unfortunately, the significance of the effect of different sexes disappears after the inclusions of one's own individual characteristics. To further exploit this point, we divided our sample into sub-groups of males and females. Column (7) of Table 6 shows the result of male samples while Column (8) illustrates the females' case. In both regressions, interaction terms considered in the bench mark regression (column 2) are controlled. Those additional regressions for genders suggest that previous results on the effects of age and educational backgrounds of reference groups remain unchanged. Regarding the effect of gender of reference group on relative utility, it is suggested that males feel stronger jealousy toward people of the same gender than do females, while females feel weaker jealousy to people of the same gender, females, than do to males. Hence, it is suggested that males are the target of stronger jealousy.

All in all, from the Leyden task, we can say that the intensity of relative utility changes in combination with the features of reference groups. Among Japanese respondents, those with a higher educational background are a target of higher jealousy, and elderly people tend to draw altruism (or admiration). People with higher educational backgrounds and elderly people are both associated with higher income, while the relative utility effects toward these two groups go in opposite directions of each other, interestingly. From these findings, it is suggested that in examining a social setting such as the Easterlin paradox, it may not be sufficient if we consider the salient reference group to be social averages, and that features of true reference groups seriously affect the outcome of empirical investigations of the Easterlin paradox.

4.4 Who compares to whom task

In this task, we consider two types of reference groups: friends and colleagues at the same time, in choice questions. The choice of the reference groups come from the result of our warm-up question in pre-tests, that these two were the most cited ones among others. Also, these two external reference groups are brought into focus by Senik (2009). In addition to the simultaneous treatment of two external reference groups, in this study, we have information of subjects' specific comparison benchmark (internal reference group), and we would like to examine the following issues: (i) we investigate if friend and colleagues are respectively recognized as different type of reference groups, and if so, how different

[31] See for detail footnote 26.

they are; and (ii) we examine if perceptions toward these two groups are different across subject sub-groups of different specific comparison benchmarks.

In order to investigate these issues, we first employ a conditional logit model framework as is shown in Table 7. In the who compares to whom task, there are two external reference groups, and the true parameters of relative utility toward these two groups are measured by dividing the estimates of reference income term for friend, and that for work colleague, with the estimate of own income term, respectively.

The first column of Table 7 shows the result of conditional logit estimation for the whole study sample. Firstly, it is shown that one's own income effect is found to be positive and significant as before, which validates the framework of choice task in this study. Secondly, the relative utility effects toward friend group and colleague group are both estimated to be significantly negative, as with the case of reference group of social average. The difference in magnitudes of these two reference groups, however, draws attention. Looking at the true estimates of relative utility parameters, the intensity of jealousy toward work colleague group is more than ten times stronger than that of the friend group. Another interesting finding is that from columns (2) to (7), where estimation results of sub-sample of individual specific comparison benchmark are provided, relative utility effect toward the reference group of friend disappears in some cases. Especially, from column (4), for those who think that their reference group is friend, the relative utility effect is not significant in the conditional logit framework; while from column (5), for those whose reference group is work colleague, relative utility effect toward friend is significantly negative. At first glance, this outcome is a puzzle. Finally, we point out that the intensity of jealousy toward the reference group of work colleague by those who answered in the warm-up question that they do not compare is the weakest among sub-groups (column (6)). Together, with the same finding in the representative agent task, it will validate our data set.

In order to further investigate the issue of the weak result on the reference group of friend, we employ a mixed logit model framework. The result is shown in column (8) of Table 7. Firstly, we obtain very similar magnitudes of estimates on one's own income, reference income of friend, and reference income of colleague, with the case of conditional logit model in column (1) and with respect to the means of relative utility effect distributions. This finding is good for the robustness of the previous result. However, the interesting discovery from the mixed logit model framework is found in standard deviation terms: while all standard deviation terms for three attributes are significant, the relative magnitudes of them compared to their mean estimates are very different. A noteworthy finding is the large standard deviation compared to the mean of the external reference income term for the friend group. With this finding, we guess that the estimate of reference income of friend that is close to zero reflects not only weak average intensity of relative utility toward the group, but also various attitudes among subjects toward their friends. In order to obtain intuition on this point, we provide figure 2 where CDFs of 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.[32] The figure clearly shows that the distributions of these true magnitudes of relative utility effects exhibit different patterns

[32] γ_a is obtained in the representative agent task. We obtained individual parameters of relative utility using Bayesian Reverse Formula after mixed logit model estimations (Train 2003). See also footnote 28.

with each other. The CDF of γ_f shows that all the subjects in our study sample have negative relative utility, while the CDF of γ_w shows that more than 20 percent of subjects feel altruism towards friends. We also see from the CDF of γ_a that the distribution of γ_a has the largest variance, which may be a reflection of weak validity of reference group as the social average.

In column 8 of Table 7, we provide the result of mixed logit estimation for the whole study sample. As was the case of the conditional logit model framework, it is interesting to see the difference of distribution of relative utility effects, as well as the intensity of jealous by separately regressing sub-samples of specific comparison benchmarks. Results are shown in Table 8. The first noteworthy thing is that through all sub-groups 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 to the friend group are found to be significantly negative, except for sub-groups with in-mind reference group of neighbors who are found to feel *altruism* toward their friend, on average. For the sub-groups with in-mind reference group of family, the mean is estimated to be significant, but it is only at the ten percent level. In all the sub-groups, the absolute values of mean estimates of relative utility effects are small as before. A noteworthy thing here is that the estimated standard deviation terms for the reference income term of friend are large compared to the means. Notice that those terms are significant for all sub-groups, including one of the in-mind reference groups of work colleague. The means of relative utility effects to work colleague group are found to be significantly negative for all sub-sample regressions. The magnitudes of mean estimates compared to those of one's own income terms are larger than they were in the case of the reference group of friend. An intriguing pattern is found when we compare the estimates of standard deviations of relative utility effects to work colleague for two sub-groups of subjects with in-mind reference groups of friend (column (3)) and subjects with specific comparison benchmarks of a colleague (column (4)). On the one hand, the standard deviation is found to be significant (though it is not so large) in the former sub-group. On the other hand, for the latter sub-group, the standard deviation is found to be *in-significant*, implying that negative relative utility effect against work colleague spikes at the average point among those sub-samples. This outcome suggests that reference groups of friends and work colleagues are different, not only in the sense of the average intensities of relative utility effects, but also in the sense of their distributions across sub-sample groups of specific comparison benchmarks. We argue that this outcome reflects that 'friend' encompasses many aspects of life, such as good rivals, persons of understanding, and so forth, and that work colleagues tend to be regarded as rivals.[33]

4.4.1 Similarity of reference groups

In the questionnaire for this task, we provided four choice options, since we would like to exploit a *tree structure* of the choice options. Our purpose is that we test if people

[33]As pointed out before in section 4.2.2, regarding work colleague, the tunnel effects of signaling future prospects will have some consequences on the intensity and distribution of relative utility toward work colleagues, but the result shows that the effect does not seem very strong among Japanese subjects. Card et al. (2010) compared positive effects of the tunnel effects and negative effects of relative utility in a social experiment setting, and showed that the negative effects are dominat in the U.S..

perceive two of the reference groups as respectively independent. More specifically, we would like to exclude a possibility that people define their rivals being merely *others*, and that characterization, amongst others, is not important.

In the first two out of four options, we set the reference income levels of work colleague as constant, and we call this nest of choice options *C-fixed*. Similarly, in the last two out of four options, we set the reference income levels of friend as constant, and we call this nest of choice options *F-fixed*. For respondents who consider that *only* reference income of friends matters, the F-fixed nest exhibits similarity of the choice options in the nest. Also, for respondents who consider that *only* reference income of work colleague matters, the C-fixed nest shows equivalence of the choice options in the nest. With this tree structure, if subjects think that there is no difference between reference group of friend and that of work colleague - put differently, if they think that both reference groups of friend and work colleague are just taken as *others*, then the tree structure of the choice options turn irrelevant. If it is the case, from nested logit model estimation, we obtain that *Inclusive Value (IV)* parameters related to respective nests are estimated to be significantly different from one.[34]

The results from the nested logit model are as follows. First, IV parameter for the F-fixed nest turns 1.556, while that for the W-fixed nest becomes 1.627. Both of them are significantly different from one at the one percent level. These figures indicate that respondents perceive two reference groups as different to each other, as is expected. Secondly, an interesting finding here is that estimated IV parameters exceed 1. According to Train et al. (1987), from a purely statistical perspective, the value of IV parameters indicate relative substitutability within and among nests, and if they become greater than 1, it means that choice substitutability among nests are more frequent.[35] In our choice setting, the outcome implies that subjects regard both reference groups of friend and work colleague relevant and they *frequently* change their comparison benchmarks from one to the other, rather than stick to one, in accordance with situations they face. This will be an evidence which nuances that reference group is chosen endogenously.

5 Concluding remarks

This paper provides evidence from Internet-based, large-scale survey data of hypothetical choice experiment on the relative utility hypothesis. The methodology exploited here complements previous empirical results from happiness studies, incentivized choice experiment studies, and neuroscience studies in such a way that methodological problems among previous studies of these fields are resolved. We show that the intensity of negative relative utility among Japanese people is not as strong as that which validates the Easterlin paradox. To paraphrase Easterlin (1995), increasing the income of all increases the happiness of all *to some extent*. We clarify the way in which intensity (and sign) of relative

[34]Previous results of the conditional logit model showed that the magnitudes of relative utility towards the friend group and the work colleague group are different from each other. Especially, the average intensity of relative utility for friends was close to zero, suggesting that subjects plausibly distinguish the friend group and work colleague group.

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

utility is influenced by comparison benchmarks that subjects perceive, and by rivals to which they are comparing themselves. Stronger jealousy is associated with people who tend to compare themselves to reference groups of daily contacts. Comparison attitude toward friends is found to be diversified, even including altruism for a certain proportion of Japanese respondents, while comparison attitude toward colleagues is uniformly characterized by jealousy. We also provide empirical evidence, which suggests that a reference group is chosen endogenously.

Finally, we suggest a future research agenda. As documented above, a merit of the hypothetical choice experiment framework with a random utility model is that we do not rely on information of subjective well-being in obtaining *true* parameters of utility function with relative concerns. This is a great merit, especially when researchers would like to conduct international comparisons of relative utility, as usually subjective well-being information is strongly influenced by country fixed effects. Our future tasks will be testing the validity of the Easterlin paradox for various countries using hypothetical choice experiment frameworks.

References

Ai, C. and E. C. Norton (2003): "Interaction terms in logit and probit models," *Economics Letters*, 80, 123–129.

Alpizar, F., F. Carlsson, and O. Johansson-Stenman (2005): "How much do we care about absolute versus relative income and consumption?" *Journal of Economic Behavior & Organization*, 56, 405–421.

Arrow, K., R. Solow, P. R. Portney, E. E. Leamer, R. Radner, and H. Schuman (1993): "Report of the NOAA Panel on Contingent Valuation," Tech. rep., National Oceanic and Atmospheric Administration.

Bech, M., T. Kjaer, and J. lauridsen (2010): "Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment," *Health Economics*, Forthcoming.

Blanchflower, D. G. and A. J. Oswald (2004): "Well-being over time in Britain and the USA," *Journal of Public Economics*, 88, 1359–1386.

Blanchflower, D. G., A. J. Oswald, and B. Van Landeghem (2009): "Imitative Obesity and Relative Utility," *Journal of the European Economic Association*, 7, 528–538.

Brown, G. D. A., J. Gardner, A. J. Oswald, and J. Qian (2008): "Does Wage Rank Affect Employees 'Well-being?" *Industrial Relations*, 47, 355–389.

Cappelli, P. and P. D. Sherer (1988): "Satisfaction, Market Wages, & Labor Relations: An Airline Study," *Industrial Relations*, 27, 56–73.

Card, D., A. Mas, E. Moretti, and E. Saez (2010): "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction," Nber working papers, National Bureau of Economic Research, Inc.

Carlsson, F., G. Gupta, and O. Johansson-Stenman (2009): "Keeping up with the Vaishyas? Caste and relative standing in India," *Oxford Economic Papers*, 61, 52–73.

Carlsson, F., O. Johansson-Stenman, and P. Martinsson (2007): "Do You Enjoy Having More than Others? Survey Evidence of Positional Goods," *Economica*, 74, 586–598.

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

Clark, A. and C. Senik (2010): "Who compares to whom? The anatomy of income comparisons in Europe," *Economic Journal*, Forthcoming.

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

Clark, A. E., N. Kristensen, and N. Westergard-Nielsen (2009a): "Economic Satisfaction and Income Rank in Small Neighbourhoods," *Journal of the European Economic Association*, 7, 519–527.

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

Clark, A. E. and A. J. Oswald (1996): "Satisfaction and comparison income," *Journal of Public Economics*, 61, 359–381.

de la Garza, A., G. Mastrobuoni, A. Sannabe, and K. Yamada (2009): "The Relative Income Hypothesis With and Without Self-reported Reference Wages," *mimeo*.

Duesenberry, J. S. (1949): *Income, Saving and the Theory of Consumer Behavior*, Harvard University Press.

Dupor, B. and W.-F. Liu (2003): "Jealousy and Equilibrium Overconsumption," *American Economic Review*, 93, 423–428.

Easterlin, R. A. (1974): "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*.

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

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

Falk, A. and M. Knell (2004): "Choosing the Joneses: Endogenous Goals and Reference Standards," *Scandinavian Journal of Economics*, 106, 417–435.

Fehr, E. and C. F. Camerer (2007): "Social neuroeconomics: the neural circuitry of social preferences," *Trends in Cognitive Sciences*, 11, 419–427.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Ida, T. and T. Kuroda (2006): "Discrete Choice Analysis of Demand for Broadband in Japan," *Journal of Regulatory Economics*, 29, 5–22.

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

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

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

Kahneman, D., A. B. Krueger, D. Schkade, N. Schwarz, and A. A. Stone (2006): "Focusing Illusion Would You Be Happier If You Were Richer?" *Science*, 312, 1908–1910.

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

Kuriyama, K. (2008): "Economic analysis of biases in the stated preference data (in Japanese)," *Environmental Economics and Policy Analyses*, 1, 51–63.

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

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

McBride, M. (2001): "Relative-income effects on subjective well-being in the cross-section," *Journal of Economic Behavior & Organization*, 45, 251–278.

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

Olsen, S. (2009): "Choosing Between Internet and Mail Survey Modes for Choice Experiment Surveys Considering Non-Market Goods," *Environmental & Resource Economics*, 44, 591–610.

Revelt, D. and K. Train (1998): "Mixed Logit With Repeated Choices: Households' Choices Of Appliance Efficiency Level," *The Review of Economics and Statistics*, 80, 647–657.

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

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

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

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

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

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

Stiglitz, J. E., A. Sen, and J.-P. Fitoussi (2009): "Report by The Commission on the Measurement of Economic Performance and Social Progress," Tech. rep., The Commission on the Measurement of Economic Performance and Social Progress.

Train, K. (2003): *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge.

Train, K. E., D. L. McFadden, and M. Ben-Akiva (1987): "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices," *RAND Journal of Economics*, 18, 109–123.

Tricomi, E., A. Rangel, C. F. Camerer, and J. P. O'Doherty (2010): "Neural evidence for inequality-averse social preferences," *Nature*, 463, 1089–1091.

van de Stadt, H., A. Kapteyn, and S. van de Geer (1985): "The Relativity of Utility: Evidence from Panel Data," *The Review of Economics and Statistics*, 67, 179–87.

van Praag, B. M. and P. Frijters (1999): "The measurement of welfare and well-being; the Leyden approach," Paul frijters discussion papers, School of Economics and Finance, Queensland University of Technology.

van Praag, B. M. S. (1991): "Ordinal and cardinal utility," *Journal of Econometrics*, 50, 69–89.

Viscusi, W. K., J. Huber, and J. Bell (2008): "Estimating discount rates for environmental quality from utility-based choice experiments," *Journal of Risk and Uncertainty*, 37, 199–220.

Figure 1: Distribution of Elapsed Time in Completing the Survey

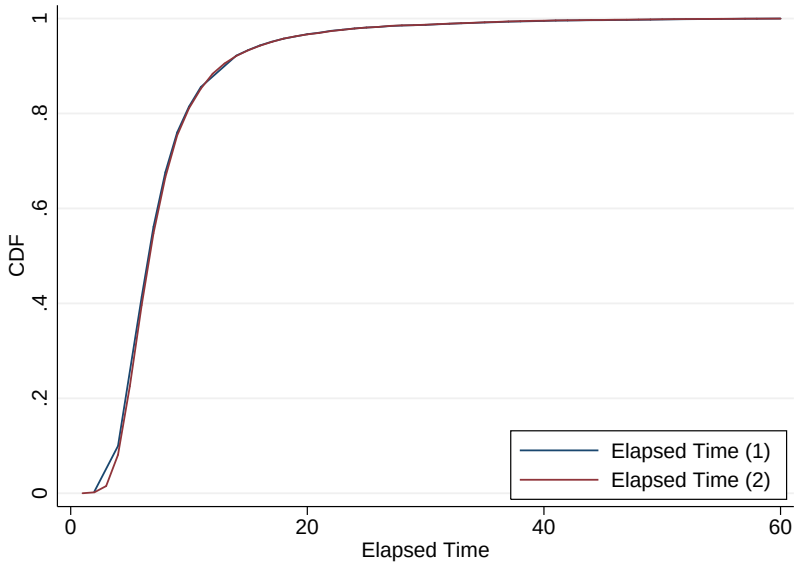

Figure 2: Distributions of Relative Utility Parameters (social average, friend, and colleague)

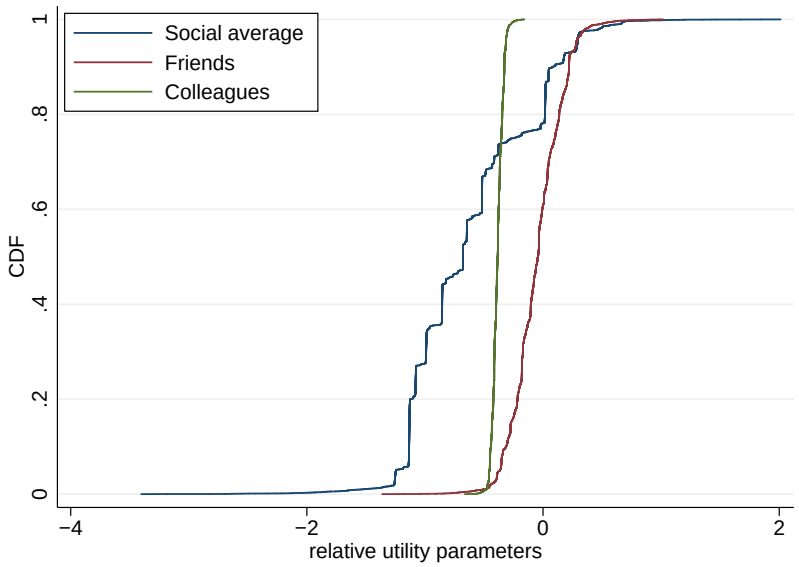

Table 1: Descriptive Statistics

	(1) Whole sample		(2) Study sample		(3) NIKKEI	(4) National Data (b)	
	Male	Female	Male	Female	Pooled	Male	Female
Age category							
20s	18.20	22.18	16.98	21.94	13.73	19.48	18.75
30s	24.53	22.90	25.01	23.14	36.42	24.48	24.06
40s	20.53	24.47	21.37	25.26	30.07	21.78	21.73
50s	23.07	18.61	25.22	19.16	13.88	22.11	22.64
60s	13.68	11.83	11.43	10.50	5.89	12.15	12.82
Education (a)					N.A.		
Middle school	1.03	0.97	0.94	0.90		18.18	20.80
High school	18.58	25.51	18.38	24.24		41.60	43.39
Some college	10.72	32.75	10.94	33.31		11.36	24.54
College	69.67	40.80	69.77	41.54		28.33	10.67
Marital Status							
Single	32.69	26.5	31.65	27.47	29.99	32.00	23.40
Married	63.95	67.24	64.95	66.18	60.66	61.80	57.60
Divorced/separated	3.36	6.26	3.40	6.35	9.35	6.20	19.00
Region							
Hokkaido	4.31		4.53		3.97	4.3	
Tohoku	4.21		4.37		4.06	7.4	
Kanto	45.32		44.79		46.94	32.9	
Koshinetsu	3.95		4.14		3.67	6.7	
Chubu	10.09		9.98		9.45	11.9	
Kansai	20.23		20.33		19.73	16.3	
Chugoku	3.92		3.85		3.82	6.0	
Shikoku	1.84		1.88		1.90	3.1	
Kyushu	6.14		6.14		6.45	11.4	
Female [0.1]	52.57		53.03		56.64	51.27	
Student [0,1]	3.60		3.16		N.A.	7.60	
Unemployment	4.05		2.91		N.A.	4.90	

All figures are in percent 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 is from the Employment Status Survey (Table 3; 2007) by the Statistics Bureau, Marital Status is from the Population Statistics of Japan (Table 6.21; 2008) by National Institute of Population and Social Security Research, Region is from the Population Statistics of Japan (Table 9.5; 2008), and the unemployment rate is from the Labour Force Survey (Feb. 2010) by the Statistics Bureau.

Table 2: "Whose income would you be most likely to compare your own with?"

	Family	Neighbors	Friend	Colleague	Don't compare	Others
Observations	483	578	4279	2024	2592	247
%	4.73	5.67	41.94	19.84	25.40	2.42

Table 3: The Intensity and Direction of Income Comparisons (%)

Comparisons	Family	Neighbors	Friend	Colleague	Don't compare	Others	Total
Not at all important							
1	2.99	0.47	13.23	8.03	74.33	0.94	100.00
	3.93	0.52	1.96	2.52	18.21	2.43	6.22
2	4.14	2.76	24.87	15.17	51.72	1.34	100.00
	21.12	11.76	14.33	18.84	49.19	13.36	24.16
3	5.46	5.23	40.16	20.01	26.85	2.28	100.00
	33.75	26.99	28.00	29.50	30.90	27.53	29.24
4	4.67	8.39	58.29	24.20	1.10	3.35	100.00
	35.20	52.77	49.54	43.48	1.54	49.39	35.65
5	6.00	9.52	54.66	25.26	0.83	3.73	100.00
	6.00	7.96	6.17	6.03	0.15	7.29	4.73
Important	4.73	5.67	41.94	19.84	25.40	2.42	100.00
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00

The first figure in each cell refers to the row percentage and the second to the column percentage.

Table 4: Conditional Logit and Mixed Logit Estimates (Representative Agent Task)

Comparison benchmarks	(1)	(2) family	(3) neighbor	(4) friend	(5) colleague	(6) N.A.	(7) others	(8)	
	Conditional logit							Mixed logit	
								Mean	SD
Own income	0.048***	0.045***	0.056***	0.052***	0.050***	0.041***	0.054***	0.097***	0.077***
	(0.001)	(0.003)	(0.003)	(0.001)	(0.002)	(0.001)	(0.005)	(0.002)	(0.002)
Reference income	-0.022***	-0.021***	-0.031***	-0.025***	-0.022***	-0.017***	-0.024***	-0.044***	0.081***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.004)	(0.001)	(0.002)
Estimated γ	-0.458	-0.467	-0.554	-0.481	-0.440	-0.415	-0.444		
Observations	48172	2255	2739	20442	9581	11982	1173	48172	
Pseudo R-squared	0.249	0.228	0.320	0.279	0.253	0.190	0.284		

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$. Standard errors clustered by subject in parentheses.
Sub-groups of comparison benchmark: (1) whole sample, (2) family, (3) neighbors, (4) friends, (5) colleagues, (6) do not care, (7) others, and (8) whole sample

Table 5: Heterogeneity of Preference Paremeters (Representative Agent Task)

	(1)	(2)	(3)	(4)	(5)	(6)
			Conditional logit model			
Own income	0.048***	0.037***	0.035***	0.036***	0.038***	0.036***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Reference income	-0.022***	-0.009***	-0.007***	-0.009***	-0.009***	-0.006***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Interaction terms						
Own income ∗						
High income dummy		0.011***	0.012***	0.010***	0.012***	0.012***
		(0.003)	(0.003)	(0.002)	(0.003)	(0.002)
Female dummy		-0.001	-0.001	0.000	-0.001	0.001
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
High age dummy		-0.012***	-0.011***	-0.012***	-0.009***	-0.009***
		(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
High education dummy		0.014***	0.014***	0.013***	0.013***	0.013***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Urban dummy		0.005***	0.005***	0.005***	0.005***	0.005***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Married dummy		0.006***	0.006***	0.005***	0.006***	0.006***
		(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
No ref. group dummy		-0.006***	-0.003	-0.006***	-0.007***	-0.003
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Very happy dummy		-0.003	-0.005***	-0.004	-0.004	-0.006***
		(0.003)	(0.001)	(0.003)	(0.003)	(0.001)
Very conscious dummy		0.014***	0.008***	0.015***	0.014***	0.008***
		(0.004)	(0.002)	(0.004)	(0.004)	(0.002)
Reference income ∗						
High income dummy		-0.005**	-0.006**	-0.005***	-0.005**	-0.005***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Female dummy		-0.005***	-0.004***	-0.005***	-0.005***	-0.005***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age		0.001	0.001	0.001	0.001	0.000
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
High education dummy		-0.005***	-0.005***	-0.004***	-0.005***	-0.004***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Urban dummy		-0.000	-0.000	-0.000	-0.000	-0.000
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Married dummy		-0.004***	-0.004***	-0.004***	-0.004***	-0.004***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
No ref. group dummy		0.007***	0.005***	0.007***	0.007***	0.004***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Very happy dummy		-0.001	-0.000	-0.000	-0.001	0.000
		(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
Very conscious dummy		-0.012***	-0.007***	-0.012***	-0.012***	-0.007***
		(0.003)	(0.001)	(0.003)	(0.003)	(0.001)
Observations	48172	48172	48172	48172	48172	48172
Pseudo R-squared	0.249	0.279	0.279	0.279	0.278	0.279

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$. Standard errors clustered by subject in parentheses.
Other controlled dummy interactions; survey pattern dummies, student dummies, and unemployment dummies.

Table 6: Conditional Logit Estimates with Characterized Reference Groups (Leyden Task)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Conditional logit model				
Own income	0.034***	0.028***	0.027***	0.027***	0.027***	0.026***	0.029***	0.020***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Reference income	-0.005***	-0.005***	-0.005***	-0.005***	-0.005***	-0.005***	-0.007***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Type of reference group interactions								
Reference group of								
Different sex	-0.002***	0.001	0.001	0.001	0.001	0.001	0.003***	-0.007***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Higher age	0.008***	0.009***	0.008***	0.009***	0.009***	0.008***	0.011***	0.010***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Same age	0.007***	0.007*	0.006	0.007**	0.006	0.005	0.008	0.007**
	(0.001)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.005)	(0.004)
Higher education	-0.009***	-0.004***	-0.003***	-0.003***	-0.003***	-0.002*	-0.004***	-0.001
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Same education	-0.003***	0.000	0.000	0.000	0.001	0.001	0.000	0.000
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Interactions with own individual characterstics	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	45554	45554	45554	45554	45554	45554	20328	25226
Pseudo R-squared	0.194	0.213	0.213	0.213	0.213	0.213	0.253	0.186

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$. Standard errors clustered by subject in parentheses.
(a) Individual characteristics are controlled for the own income term, and for all the terms of reference income. The same set of individual characteristic variables cotrolled in the Representative-Agent-Task-regressions is taken into account.

Table 7: Conditional Logit and Mixed Logit Estimates (Who Compares to Whom Task)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
	Conditional logit							Mixed logit	
								Mean	SD
Own income	0.028***	0.028***	0.031***	0.028***	0.030***	0.025***	0.032***	0.035***	0.029***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.002)	(0.000)	(0.001)
Reference income (Friend)	-0.001***	-0.001	0.000	0.000	-0.003***	-0.002***	-0.003**	-0.002***	-0.011***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.002)	(0.000)	(0.000)
Reference income (Colleague)	-0.011***	-0.012***	-0.012***	-0.012***	-0.011***	-0.007***	-0.013***	-0.011***	0.006***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.002)	(0.000)	(0.001)
Estimated γ_f	-0.036	0	0	0	-0.100	-0.080	-0.094		
Estimated γ_w	-0.393	-0.429	-0.387	-0.429	-0.367	-0.280	-0.406		
Observations	47180	2226	2664	20131	9387	11639	1133	47180	
Pseudo R-squared	0.140	0.143	0.168	0.147	0.158	0.111	0.178		

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$. Standard errors clustered by subject in parentheses.
Sub-groups of comparison benchmark: (1) whole sample, (2) family, (3) neighbors, (4) friends, (5) colleagues, (6) do not care, (7) others, and (8) whole sample

Table 8: Mixed Logit Estimates for Sub-groups of Comparison Benchmarks (Who Compares to Whom Task)

	(1)		(2)		(3)		(4)		(5)		(6)	
	Mixed logit											
	Mean	SD	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***	0.043***	0.037***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.004)
Reference income (Friend)	-0.002*	0.013***	-0.000	0.006*	-0.001**	0.012***	-0.004***	0.011***	-0.003***	0.008***	-0.005***	0.009**
	(0.001)	(0.002)	(0.001)	(0.003)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.002)	(0.004)
Reference income (Colleague)	-0.013***	0.010***	-0.012***	0.006**	-0.013***	0.007***	-0.011***	0.003	-0.007***	0.003	-0.014***	0.013***
	(0.001)	(0.003)	(0.001)	(0.003)	(0.000)	(0.001)	(0.001)	(0.005)	(0.000)	(0.002)	(0.002)	(0.004)
Observations	2226		2664		20131		9387		11639		1133	

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$. Robust standard errors in parentheses.
Sub-groups of comparison benchmars: (1) family, (2) neighbors, (3) friends, (4) colleagues, (5) do not care, and (6) others

Figure 3: Supplementary: Images for Hypothetical Choice Experiments