GCOE Discussion Paper Series

Global COE Program

Human Behavior and Socioeconomic Dynamics

Discussion Paper No.162

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

Katsunori Yamada and Masayuki Sato

October 2010; revised December 2010; revised November 2011

GCOE Secretariat
Graduate School of Economics
OSAKA UNIVERSITY
1-7 Machikaneyama, Toyonaka, Osaka, 560-0043, Japan

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

Katsunori Yamada

Corresponding author: Institute of Social and Economic Research, Osaka University

Email: kyamada@iser.osaka-u.ac.jp

Tel/Fax +81-6-6879-8566, Mihogaoka 6-1, Ibaraki 567-0047, Japan

Masayuki Sato

Field Science Education and Research Center, Kyoto University

Kitashirakawa Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan.

Abstract

We propose a new avenue in the literature of the relative utility hypothesis, along which researchers can clearly investigate the changing effects of income comparison in accordance with different features of reference persons. We test our methodology of hypothetical discrete choice experiment with an original large-scale survey data. We show that the distribution as well as the average intensity of relative utility are different across specific comparison benchmarks, and across types of reference groups people are facing in the experiments. Our results suggest that features and the distribution of true reference groups seriously affect the outcome of empirical investigations of the Easterlin paradox.

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

JEL classifications: C2; D1; D3

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 a new approach to investigate the social comparison effects via hypothetical discrete choice experiment. Our approach estimates parameters of social comparison in *decision utility*, as named by Kahneman et al. (1997), while majority of previous studies of the relative utility hypothesis exploited information of *experienced utility*. With this new approach, we can focus on the changing effects of social comparison derived from differences in comparison benchmarks, which subjects perceive in mind, and derived from differences in comparison targets, which subjects face in experiments. We conducted an internet-based original large-scale survey, bringing a new data set for testing the relative utility hypothesis with an extended view that "when you compare different persons, you feel differently."

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.[1]

Despite all these new innovations along the line of happiness study, some alternative approach for testing the relative utility hypothesis will be necessary. One reason 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.[2] Another reason will be that the use of sub-

[1]Similar to Clark and Senik (2010), Mayraz et al. (2009) studied the effect of income comparisons on SWB with subjective information of income relative to comparisons groups of interest. They find that workplace comparisons are particularly important, while they find little or no evidence for comparisons with friends or neighbors.

[2]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. (2010). 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.

jective well-being (SWB) information is sometimes subject to criticisms by economists of other fields, even though there has been an well-established concensus that SWB information is valid (Hollander 2001, Kahneman and Krueger 2006, Oswald and Wu 2010). It is actually true that SWB measures of respondents are influenced by many aspects of their lives, including individual characteristics and mood on the day of the survey (Kahneman et al. 2006). Hence, it is beneficial if we re-examine and reinforce previous findings from happiness studies on the relative utility hypothesis with our new avenue. Nice merits of this alternative approach includes; (i) direct and cardinal measures of reference income are available, (ii) subjects can see the features of reference groups in situation choice tasks, and (iii) they choose discrete choices of less-cognitive burden rather than evaluate their lives with happiness scores, which is sometimes difficult.

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. Results from our original web-based large scale survey suggest that the distribution as well as the intensity of relative utility are different across specific comparison benchmarks, and across types of reference groups people are facing in the experiments. They also suggest that the relative utility effects among Japanese respondents may not be as strong as can validate the Easterlin paradox. We also provide empirical evidence, which nuances that reference groups are chosen endogenously using a nested logit model framework.

Before closing the introduction, we mention biases which are normally associated with *stated preference methods*. Dolan and Kahneman (2008) critically summarized biases associated with stated preference methods including hypothetical discrete choice experiments. They then encouraged happiness (or experience utility) research, when researchers would like to infer the market values of non-market goods. Note, however, that footnote 4 in Dolan and Kahneman (2008) holds that "[their] critique is focused on the use of measures of decision utility to elicit values of this kind, rather than their usefulness in other contexts, such as predicting behaviour." The purpose of this paper is to elicite the sign and the intensity of comparison effects which affect human behavior. Potential biases in our survey are listed in Section Appendix C, and we explain how we tried to cope with them. Given the potential drawbacks in the survey, obviously, we do not insist that our results are free of those biases and we ask readers for cautions in interpreting estimated parameters. We, instead, insist that the main contribution of the paper is provision of new

avenue of testing social comparison effects, with which we can investigate the changing effects of income comparison in accordance with different features of reference persons, and with comparison benchmark.

The paper is organized as follows. Section 2 presents a brief review on alternative methodologies in testing the relative utility hypothesis for economics of happiness. Section 3 introduces our data and explains 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 and data construction

Our data set is taken from an original Web-based survey conducted in March 2010.[3] 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 the survey, 60,482 Japanese subjects between the ages of 20 and 65 were chosen with stratified random sampling from the Nikkei Database (total number of the registered subjects is over 160,000), such that the cohort profile of our sample mirrored the Japanese census statistics. 14,370 subjects completed the survey. The observations dropped either because of too short elapsed time in completing the survey, of missed information for some of the variables used in the empirical analysis below, or of inconsistency in the answers. At the final stage, we are left with 10,203 respondents. The 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. Detailed descriptions on the data collection and study sample construction issues are in the appendix.

This response rate of 23.8 percent (14,370/60,482) is smaller than might have been

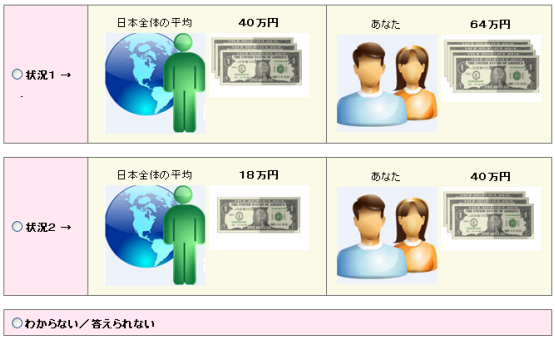
[3]We conducted three pre-tests before the main test. We then designed the questions used in conjoint analyses, which take into account differences in reference groups. For detail, see the conjoint questions explained in section Appendix B. The use of the Internet for social surveys is new and is becoming 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).

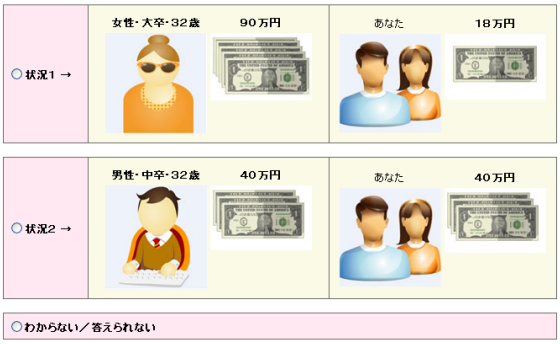
desired. However, it should be noted that the un-weighted average of the response rates for similar academic conjoint surveys conducted by the same company in 2008 and 2009 (7 projects) is 28.8 percent. Actually, the response rate is not a problem since we have an apparent merit that we can collect a large number of samples that cover varieties of individual characteristics from throughout Japan. As is seen in Table 1, presenting a comparison of the sample characteristics to overall statistics for the Japanese population, we find that our data set captures significant features of the Japanese people. This is difficult to pursue with small sample, face-to-face surveys. Also, as documented in Ida and Kuroda (2006), broadband Internet services are pervasive in Japan, thus the sample selection problem associated with a web-based survey will not be serious.

2.2 Survey structure

The survey consisted of conjoint questions and questions about individual characteristics.

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. Conjoint questions are made of three types of situation choice tasks which are characterized by different reference groups. The following figure images are the ones used for the survey (hence, characters are in Japanese). As it can be seen, subjects are visually informed about the difference in reference groups across tasks (left: social average; right: Leyden fashion reference groups), aiming at picking up changing effects of social comparison in accordance with reference groups chosen.





Like Clark and Senik (2010), we also see if the intensity and the sign of comparison effects are affected by specific comparison benchmarks, information of which is collected with a "who compares to whom" question in our survey explained below. We also would like to investigate if reference groups are determined endogenously, probably in accordance with the frequency of daily contacts. This analysis 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. Detailed methodological explanations on the actual procedure we designed the choice tasks are given in section Appendix B.2.

2.2.1 Social average task

The first choice task is a benchmark task and we call it *social average task*. The specification here is that the reference group is given by "the Japanese social average." The Easterlin paradox was found when proxies of reference income of the national averages were imposed on subjects by econometricians in the literature of *experienced utility*. Hence, despite its simplicity, this task will provide a useful alternative method in investigating the paradox, with which we estimate parameters of *decision utility*. Clear merits in our method are that in making choices subjects recognize that they are competing with the Japanese average, and that economic situations are explicitly shown to them.

This task will be also useful for theoretical macro-economists. Previous theoretical studies such as Abel (1990), Gali (1994), Liu and Turnovsky (2005), and
Garcia-Penalosa and Turnovsky (2008) were conducted without estimating important parameters of the relative utility. And they 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.

2.2.2 Leyden task

The second choice 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 (van Praag and Frijters 1999). 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 demographic characterizations of rivals.

2.2.3 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 in a question explained in section Appendix B that these two are dominant among Japanese respondents (and also among European people, as was shown in

Clark and Senik 2010).[4] In this task we addressed 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.

2.2.4 Individual characteristics questions

Individual characteristics questions include a happiness question, questions about the intensity and directions of social comparisons, and questions about demographic variables of age, sex, educational background, job, marital status, types of residence, residence area, and annual pre-tax personal income in 2009. Among them, the intensity and directions of social comparisons are important for this study.

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 colleague*. 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 2.2.3. 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 2.2.3, where we compare the reference groups of friend and work colleague.

[4]In the Japanese social backgrounds, the two reference groups (friends/classmates and work-related) may not be mutually exclusive. The nested-logit regerssions documented in Section 3.4.1, however, show that respondents distinguish these two reference groups clearly. We thank Charles Yuji Horioka for pointing out this potential flaw of the structure of the choice experiment.

3 Empirical Results

3.1 Random utility model

Here we describe our theoretical foundation of the concept of conjoint analysis, but only for the case of social average 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. 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 2009, 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 2009)

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.[5]

3.2 Social average task

3.2.1 Results of an alternative test on the Easterlin paradox

The first task is the benchmark case where the reference group is defined as the social average, and can be thought of as an alternative avenue for testing the Eastelrin paradox which was originally found by exlpoiting information of experienced utility. We instead estimate a parameter of utility function (decision utility) which captures the intensity and direction of social comparison, γ, using the choice experiment strategy and random utility model. Table 4 summarizes the results for this benchmark analysis.

First, 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 al-

[5]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.

ternative 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.[6]

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 as strong as that which validates the Easterlin paradox which requires $\gamma = -1$. To paraphrase Easterlin (1995), increasing the income of all increases the happiness of all to some extent, or to paraphrase Blanchflower and Oswald (2004), money buys happiness.

According to Figure 3 in Easterlin (1995), the Easterlin paradox evidently holds for Japan from the period 1958–1987. Although his argument was refuted by
Stevenson and Wolfers (2008) in the Japanese case, the paradox is still influential in the literature of behavior economics so that we mention some potential reason why the comparison intensity we have estimated has fallen short of the level validating the paradox. A clear reason will be that social comparison effect is just one of many explanations to the Easterlin paradox. Adaptation and habit formation, for example, explain the paradox as well. And so, it might be that the parameters obtained from social comparison alone may not be able to explain the relative utility effect fully.

While our result from decision utility framework stands in the middle of two extremes in experienced utility frameworks: fully relative utility function of Easterlin (1995) and only absolute utility function of Stevenson and Wolfers (2008), it is noteworthy that a recent finding in the literature of happiness study by de la Garza et al. (2010) reached at the same conclusion that money buys happiness to some extent in Japan, using direct and cardinal measures of reference income. Also, it should be noted that our result is in

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

fact perfectly in line with the studies discussed based on hypothetical choices between different societies such as Solnick and Hemenway (1998); Johansson-Stenman et al. (2002); Alpizar et al. (2005) and Carlsson et al. (2007).

3.2.2 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. It is interesting to see how people change the intensity of comparison in combination with their relevant reference groups.

In this task, the reference group is fixed as the social average. The differences of relative utility effects derived in this task will reflect a 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 γ across the sub-groups. In doing so, we need information of variances of estimated γ by sub-groups of comparison benchmark to examine statistical significance of the differences in estimated parameters. Using the delta method, we confirmed that the differences in relative utility parameters are significant at least at the 5 percent level. Columns (2)–(7) in table 4 provide the results.

Column (2) of Table 4 shows the estimate of γ for those whose comparison bench mark is 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 comparison bench mark 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 comparison bench mark, friends, provides the second strongest jealousy among reference groups, as is documented in column (4). On the other hand, the second most cited comparison bench mark 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 the quality of our data set.

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 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 weaker intensity of relative utility for those whose comparison bench mark is another close contacts of work colleague, in column (6). Regarding this finding, it may make sense to appeal to the tunnel effect conjecture by Hirschman and Rothschild (1973). He 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 tunnel effect here is not so strong as providing *positive* relative utility, as was found by Senik (2004) for a Russian data set.

3.2.3 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 (8) 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 γ.

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.[7] Also, it is natural to expect that sign and intensity of relative utility depend on individual characteristics, just as we have confirmed about comparison benchmarks.

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 and demographic variables in conditional logit models.[8] 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 "do not compare" dummy, very happy dummy, and very comparison conscious dummy. Table 5 reports the results with various dummy interactions. Column (2) of table 5 is the benchmark case, and columns (3)–(6) are presented for robustness checks.[9] 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

[7]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. Viscusi et al. (2008) showed that eco-conscious individuals have a lower rate of time discounting than those who are not eco-friendly. Ida and Goto (2009) showed that smokers are endowed with a higher value of time discounting and a lower value of risk aversion than nonsmokers.

[8]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, as is clearly explained at page 22 in Train (2009). See Ai and Norton (2003) for interpretations of *marginal effects* of dummy interaction terms in logit models. As long as one can interpret the coefficients as marginal utilities, as we do in a random utility model framework, Ai and Norton's point is not relevant.

[9]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.

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 report that they do not compare 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). Thus, we can confirm that heterogeneity plays a role in determining the intensity of social comparison, just as previous behavioral economics studies have found in other fields.

3.2.4 Implications on the loss aversion theory

In the previous sections we analyzed changing effects of social comparison derived from differences in comparison benchmarks, and derived from differences in individual characteristics. In the literature of behavior economics, as pointed out by Tversky and Kahneman (1991), many researchers believe that the sign of reference income in one's utility depend on whether they earn more or less than the rivals. Hence, introducing inflection points at $y = \bar{y}$ in choice task settings, we analyze changing effects of social comparison derived from differences in comparison situations between downward comparison (where $y \geq \bar{y}$ stands) and upward comparison (where $y \leq \bar{y}$ stands).

In doing so, we create two dummy variables of μ and ν. μ (ν) takes the value of one when the choice situation is downward comparison (upward comparison), while turning zero otherwise. We then add those dummy variables into the random utility model of equation (3) either in an additionally-separable fashion, or with interactions with $\bar{y}$. Table 6 shows the results.

Column (1) of Table 6 recaps the benchmark regression in Table 4 for readers' reference. Column (2) suggests that downward comparison (upward comparison) provides positive (negative) marginal utility independently of marginal utility from self and reference income levels. This finding is, however, reversed when we further add interaction terms of downward and upward comparison dummies with reference income levels. The main effect of upward comparison in column (3) is found to be positive and significant, while that of downward comparison turns insignificant. More interestingly, the sings of interaction terms suggest the existence of egalitarianism, similarly to previous experi-

mental studies of Fehr et al. (2008), Bartling et al. (2009), and Tricomi et al. (2010): the marginal utility from the increase in reference income tends to become altruistic when their rivals earn less than you (namely, downward comparison), while it gets even more jealous when you are left behind of the rivals (namely, upward comparison). The above exercises will confirm the loss aversion effects in Japan.

3.3 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. Following the spirit of the Leyden school, we consider reference groups with individual characteristics of age, gender, and educational background.

Table 7 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"). 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 task, thereby examining how this additional information on reference groups affects social comparision effects.

In column (1) of Table 7, 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 previous task. The partial effect of reference income is picked up as negative and significant, which is also the same for the case of social average task.

The significance in this analysis is that we can see the changes in sign and magnitude of relative utility effects toward specific types of reference group. As column (1) 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.[10] Finally, in 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 7, 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 social average task regressions in columns (2) to (6) of Table 5.[11] 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 subgroups of males and females. Column (7) of Table 7 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 in the Japanese society.

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,

[10] 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.

[11] See for detail footnote 9.

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.

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

3.4.1 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 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. Detailed explanations on the structure and construction of the choice task are in the appendix.

The results from the nested logit model are as follows (not shown in a table). 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.[12] In our choice setting, the outcome implies that subjects regard both

[12]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.

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.

3.4.2 Different relative utility effects toward friend and colleague

In order to investigate the difference in relative utility effects toward friend and colleague, we first employ a conditional logit model framework as is shown in Table 8. In this task the true parameters of relative utility, γ_f and γ_w, 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 8 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 friend disappears in some cases. Especially, from column (4), for those who think that their reference group is friend, the relative utility effect to friend is not significantly different from zero; 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 that they do not compare is the weakest among sub-groups (column 6). Together with the same finding in the social average 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 8. Firstly, we obtain very similar magnitudes (with respect to the means of relative utility effect distributions) 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). 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 reference income level of friend. With this finding, we guess that the mean estimate of reference income of friend that is close to zero reflects various attitudes among subjects toward their friends. In order to obtain intuition on this point, we provide figure 1 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.[13] The figure clearly shows that the distributions of these true magnitudes of relative utility effects exhibit different patterns with each other. The CDF of γ_w shows that all the subjects in our study sample have negative relative utility against work colleague, while the CDF of γ_f 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.

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 with the mixed logit model framework. Results are shown in Table 9. 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 friend are found to be significantly negative, except for sub-groups with comparison benchmark of neighbors who are found to feel *altruism* toward their friend on average. For the sub-groups with comparison benchmark of family, the mean relative utility effect to friend's income 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 to friend are small. Another 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 the one for comparison benchmark of work colleague.

The mean relative utility effects toward work colleague's income level are found to be

[13] γ_a is obtained in the social average task. We obtained individual parameters of relative utility using Bayesian Reverse Formula after mixed logit model estimations (Train 2009). See also footnote 20.

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 comparison benchmark of friend (column 3) and those with work 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.[14]

4 Concluding remarks

This paper provides evidence from an internet-based, large-scale orignal survey data set of hypothetical choice experiment on the relative utility hypothesis. We provided a new approach to investigate the social comparison effects via hypothetical discrete choice experiment, with a view that "when you compare different persons, you feel differently." Our approach estimates parameters of social comparison in *decision utility* rather than exploiting information of *experienced utility*. We focused on the changing effects of social comparison derived either from differences in comparison benchmarks, from differences in comparison targets, or from choice situations.

In addition to a basic finding that the intensity of negative relative utility among Japanese people may not be as strong as that which validates the Easterlin paradox, we show that the distribution as well as the intensity of relative utility effects are in fact different across specific comparison benchmarks, and across types of reference groups people are facing in the experiments.

[14] As it is pointed out in section 3.2.2, regarding work colleague, the tunnel effects of signaling future prospects may change the sign of relative utility toward work colleagues. 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..

From the benchmark task, we find that those who compare neighbors have the strongest intensity of relative utility. Because a feature of reference group of neighbors will be characterized by close contacts, this result seems natural if we accept that people endogenously choose their reference groups from groups of close contacts. It is also interesting to see a weaker intensity of relative utility for those whose comparison benchmark is another close contacts of work colleague. We will be able to appeal to the tunnel effect conjecture by Hirschman and Rothschild (1973) that the increase of work colleagues' income could be interpreted as a positive signal regarding likely future outcomes. From the same task we also find that the marginal utility from the increase in reference income tends to become altruistic when their rivals earn less than you (namely, downward comparison), while it gets even more jealous when you are left behind of the rivals (namely, upward comparison). This finding will confirm the loss aversion effects in Japan.

From the second task where reference groups are not just the social average but characterized by individual characteristics of age, educational attainment, and gender, it is found that rivals' characteristics do affect intensity and sign of comparison attitudes. Specifically, it is suggested that males are the target of stronger jealousy both from men and from women in the Japanese society. Those with a higher educational background are a target of higher jealousy, and elderly people tend to draw altruism (or admiration).

In the final task, we examine a "who compares to whom" setting via the hypothetical choice experiment. We find that reference groups of friends and work colleagues are different, not only in the sense of the average intensities of relative utility effects they draw, but also in the sense of the distributions of relative utility across sub-sample groups of specific comparison benchmarks. On one hand, those who think that their rivals are friends have variety of attitudes toward the relevant rivals, even including altruism. On the other hand, those whose comparison benchmark is work colleague uniformly feel jealous toward the relevant rivals.

All in all, these findings 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 and the distribution of true reference groups seriously affect the outcome of empirical investigations of the paradox.

We suggest a future research agenda. 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 (decision) utility function with relative concerns. Because usually subjective well-being information is strongly influenced by country

fixed effects or by social norms, hypothetical choice experiment frameworks will be useful in conducting international comparisons of the relative utility effects. Our future tasks will be re-doing of the proposed new avenue via hypothetical choice experiment frameworks for various countries.

Finally, as we mentioned, our survey is not free from biases which are normally associated with stated preference methods, and with internet-based surveys. Given the drawbacks, we ask readers for cautions in interpreting estimated parameters. It is important to disentangle directions of various biases in the social comparison experiment via the hypothetical choice experiment framework, and we call for further contributions exploiting the new avenue of the anatomy of income comparisons using different survey settings.

Appendix A Data issues

Appendix A.1 Collection of data

In the survey, 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. In order to provide highly reliable research data, Nikkei Database's registered subjects are subject to monthly screenings, which keep information up to date and exclude double registrations. 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 those subjects who can access the Internet only 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.

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 as it is shown in Table 1. 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.

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.

Hence, since our sample is limited to registered subjects of Nikkei Database, and since the sample distribution is not identical to the national statistics, this study does not intend to provide an accurate depiction of the representative Japanese people. 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.

Appendix A.2 Construction of study sample

In our experiment, 14,370 subjects out of 60,482 completed the survey. 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.

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.[15]

For subjects, the easiest way to finish the survey is to provide the same answers for conjoint questions. 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) after we confirmed 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).[16] To ensure that this 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.

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

[16]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. With the exchange rate in March 2009, 12 million Japanese Yen is around 130,000 US dollars.

Appendix B Survey design

Appendix B.1 Warm up questions

In the beginning, 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 are you concerned, anxious or envious about other persons' amounts of income?". The respondents were asked to choose from five response options, where 1 corresponded to "not at all," and 5 denoted "very much". The third 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 colleague*. 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 2.2.3. 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 2.2.3.

Appendix B.2 Conjoint questions

Our research strategy was to elicit respondents' changing preference toward social comparison across comparison benchmark and across reference groups through hypothetical choice experiments (stated choice method). This methodology was developed in order to analyse the preference on choice of multi-attribute goods or situations. 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 attributes of alternatives in stated choice situation.

In this study, to ease the understandings of respondents, we provided them with figure images of the choice tasks. 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 random utility function; 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.

Appendix B.2.1 Social average task

The reference group in this task is described by the Japanese social average. 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."

It was also repeatedly stressed that price levels were the same in these two situations. Respondents repeat five choice situations.

Each situation is described by two *attributes* of one's own monthly pre-tax income and the monthly pre-tax income of the reference group. The levels of attributes are chosen

from the following situation variations: JPY 180 thousands, JPY 240 thousands, JPY 400 thousands, JPY 640 thousands, and JPY 900 thousands. We determined the levels of these attributes after three pretests with reference to the Japanese current income distribution..

We have two attributes and five levels for each attribute, providing 25 potential variations of choice situations. Louviere et al. (2000) introduced some ways of pairing attributes in making a situation, which is called an *alternative*. In this study, we followed the orthogonal planning in making alternatives. This method effectively pairs multi-dimensional and multiple-level variables, and offers experiment plans with the greatest amount of information with the least observations. Employing the orthogonal planning, each attribute vector becomes orthogonal, and we can avoid the multi co-linearity problem in regressions of the random utility model framework.

We used SPSS Conjoint (ver. 15.0) for orthogonal design in this study. We generated an orthogonal matrix (a set of alternatives) consisting of 25 pairs of own income levels and reference income levels from 25 random draws in the orthogonal design process. We replicated this procedure again to obtain two sets of alternative vectors.

Two alternatives chosen from the alternative vectors are paired to form a *choice set* including a no-choice option, and the pairing strategy is the researcher's choice to determine. As is documented in Huber and Zwerina (1996) and in Viscusi et al. (2008), the choice design should be paired so as to balance the utility of each alternative. Because the orthogonality in an alternative matrix is maintained for row permutation, we can arbitrarily pair alternatives to meet the requirement.[17] A difficulty in the choice experiment of relative utility, however, is that it does not necessarily mean that an increase (decrease) in a reference group's income leads to decline (increase) 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:

[17]In the survey, the respondents actually 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 out of the five questions 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.

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.

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.

Appendix B.2.2 Leyden task

At the beginning of this task, respondents are 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. 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.

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.

The general method of making the choice situations is the same as in the previous task and we addressed the orthogonal planning. However, in this task, reference group is not solely the social average but is characterized by gender, age category, and educational backgrounds.

Hence, situations in this task were defined along five dimensions in total. 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.

Potential variations in sets of these attributes are 1,000. The computer algorism of orthogonal planning in SPSS Conjoint provided 25 sets of alternatives randomly. We

replicated this procedure to obtain two sets of alternative vectors. Our pairing strategy of the alternatives for this task is the same as in the *social average task*, except that information on three attributes of the socio-economic characteristics was not taken into account. We also added the no-choice options as in the previous task.

Appendix B.2.3 Who compares to whom task

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

We 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.

As before, the levels of three attributes have five variations of JPY 180 thousands, JPY 240 thousands, JPY 400 thousands, JPY 640 thousands, and JPY 900 thousands. Since we have three attributes in this task, there are 125 potential variations in alternative. Again, orthogonal design was used to pick up 25 out of the 125 to make a vector of alternatives. 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.

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 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. 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.

Appendix B.3 Demographic questions

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. Descriptive statistics for major variables is presented in Table 1.

Appendix C Discussion about biases

This appendix draws attention to possible biases.

Appendix C.1 Biases accompanying with stated choice methods

As Hausman (1993) and Carson et al. (2001) pointed out, there are some potential biases in stated choice methods. Following Bateman et al. (2002), those biases could be summarized into three broad categories of (i) incentives to misrepresent responses, (ii) implied value cues, and (iii) scenario misspecification. The first category is related to false answers to the survey questions. This bias arises when the questions and scenario settings are not well-designed. (ii) and (iii) are the biases arising from respondents' misunderstandings of the survey questions. These are also considered to be cognitive psychological effects, and descriptions of the survey are often the sources of these biases.

In order to avoid these potential biases, researcher should deliberately design choice tasks through pilot survey, repeated pre-tests and close investigations of the pre-results. In the current study, we conducted pilot surveys and three times of pre-tests for fine-tuning of our questionnaire. The most substantial change in our main test from a pre-test was in introducing figure images for the choice situation tasks. By introducing figure images, cognitive burdens of subjects could be decreased, with the average elapsed time in finishing the survey halved from the case without those figure images. Variances in logit

model estimations were also reduced. From pre-tests, we also find that the ordering of the questions and selection of questions other than the conjoint questions in the current study do not seem to affect seriously the main results of our study. Nonetheless, we mention some potential framing effects below.

The other drawback in stated choice methods often mentioned is their artificial nature of the questions, and incentive in-compatibility for subjects in making choices. Regarding the issue, Lusk and Schroeder (2004) showed that stated choice methods provided similar results of marginal effects with the ones from non-hypothetical settings. They held that deliberate design of the survey is the key issue.

Appendix C.2 Framing effects

In our survey, the questionnaire started with questions about income satisfaction and about comparison attitudes. Because these "warm-up" questions are followed by the hypothetical choice tasks, it is likely that the question order makes individuals more pecuniary-comparison-conscious. We also stated in the instruction page of the web-based survey to the subjects that "This survey is for an academic purpose, and the main topic is about comparison in a social life context."

The fact that the subjects were reminded about social comparisons does not necessarily lead to a bias which over-estimates or under-estimates the true effect. On one hand, after the instruction subjects may be motivated to "beat" the reference persons in the hypothetical choices, which will over-estimates the true effects of social comparison. On the other hand, one can equally well think about mechanisms in the other direction, which is that people dislike to think about themselves as status-seeking persons, and that they therefore *underestimate* the degree that they themselves care about social comparisons. This story resembles purchase of moral satisfaction in Kahneman and Knetsch (1992).

The purpose of the reminder was to keep the subjects clear about on what they are working and to have them focused on the survey. This strategy is different from the *behind the veil of ignorance* strategy that researcher can typically apply for face-to-face surveys, where subjects are obliged to be focused in front of interviewers.

Appendix C.3 Bias associated with sample distribution

In our data set, biases may arise with respect to sample distribution since people with higher educational backgrounds and those living in urban cities are over-sampled. The

direction of bias due to the sample distribution issue 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.

From Social average task, we found that those with higher educational backgrounds tend to become comparison conscious. This result is the same as evidence from Europe in Clark and Senik (2010). On the other hand, we found that city dwellers are just as comparison conscious as people living outside of urban cities are. This result is different from European evidence in Clark and Senik (2010).

Appendix C.4 Disentangling actual consumption from the hypothetical one

In Johansson-Stenman et al. (2002), Alpizar et al. (2005), Carlsson et al. (2007), and Andersson (2008), 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, disentangling their actual consumption from the hypothetical consumption choices in the survey.

In this study, we instead asked about respondents' own interests. This is because we would like to know the *current* situations within Japanese society. Parameters thus estimated will be more useful for drawing current policy implications.

Bias associated with ignoring the previous strategy will not be so severe, because we can control individual fixed effects, as we asked respondents to make repeated, five situation choices for each task not likely to those studies. Our strategy here was also motivated by Dolan and Kahneman (2008), who showed a critical view on having subjects make hypothetical choices based on future expectations and on past memory.

Appendix D Literature review

Possible complementary research agenda for happiness study on the relative utility hypothesis is actually threefold. The first agenda is incentivized choice experiments such

as Fehr and Schmidt 1999 and 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. Also, the 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.

The third alternative methodology of *hypothetical choice experiments* may be able to overcome these issues, and will be 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. In this study, 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 in the choice tasks. 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. 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 in the appendix.

Similar to the present analysis, Solnick and Hemenway (1998); Johansson-Stenman et al. (2002); Alpizar et al. (2005), Carlsson et al. (2007), and Andersson (2008) investigated the intensity of social comparisons by addressing methodologies of hypothetical choice experiments. These studies, however, considered that reference incomes were merely the national average income levels. Another issue among Johansson-Stenman et al. (2002), Alpizar et al. (2005), and Andersson (2008) 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. Solnick and Hemenway (1998) and Carlsson et al. (2007) did not use iterative choices. Instead, each respondent only made one single choice between two alternatives related to relative income. With these strategies, we cannot apply mixed logit framwework with which we can estimate the distribution of a parameter of relative utility. Also, *degree of positionality* inferred by these previous studies, except for Solnick and Hemenway (1998), encompassed measurement errors, since assigned value of degree of positionality was given arbitrarily. Finally, respondents in Solnick and Hemenway (1998); Johansson-Stenman et al. (2002), and Alpizar et al. (2005) were just students, those in Andersson (2008) are people in academia while Carlsson et al. (2007) and this study have given the social representative survey.

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 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.[18] This suggestion motivates studies in which we consider reference groups

[18]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

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).

Appendix E Welfare implications

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. We employ the results in the social average task.

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 $\tilde{y}$ 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

individuals' abilities, and that people thus tend to compare themselves to similar others for the purpose of self-improvement.

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 2009).[19] 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.[20] 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 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

[19] 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.

[20] 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.

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.

Acknowledgements

We appreciate helpful comments from Andrew Clark, David Flath, Ronald Harstad, Arne Risa Hole, Takahiro Ito, Louis Levy-Garboua, Fumio Ohtake, Claudia Senik, seminar participants at Kyoto University, Osaka University, Paris School of Economics, and 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. An earlier version of the paper was circulated as "The Easterlin Paradox and Another Anatomy of Income Comparisons: Evidence from Hypothetical Choice Experiments."

References

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

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

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

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

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

Bartling, B., E. Fehr, M. A. Marechal and D. Schunk, "Egalitarianism and Competitiveness," *American Economic Review* 99 (May 2009), 93–98.

Bateman, I., R. Carson, B. Day, M. Hanemann, N. Hanley, T. Hett, M. Jones-Lee, G. Loomes, S. Mourato, E. Ozdemiroglu, D. Pearce, R. Sugden and J. Swanson, *Economic Valuation with Stated Preference Techniques* (Edward Elgar, 2002).

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

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

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

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

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

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

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

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

Carson, R., N. Flores and N. Meade, "Contingent Valuation: Controversies and Evidence," *Environmental & Resource Economics* 19 (June 2001), 173–210.

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

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

Clark, A. E., N. Kristensen and N. Westergard-Nielsen, "Economic Satisfaction and Income Rank in Small Neighbourhoods," *Journal of the European Economic Association* 7 (04-05 2009a), 519–527.

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

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

de la Garza, A., G. Mastrobuoni, A. Sannabe and K. Yamada, "The Relative Utility Hypothesis With and Without Self-reported Reference Wages," Iser discussion paper, Institute of Social and Economic Research, Osaka University, 2010.

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

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

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

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

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

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

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

Fehr, E., H. Bernhard and B. Rockenbach, "Egalitarianism in young children," *Nature* 454 (August 2008), 1079–1083.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Tversky, A. and D. Kahneman, "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *The Quarterly Journal of Economics* 106 (November 1991), 1039–61.

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

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

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

Figure 1: Distributions of Relative Utility Parameters (social average, friend, and colleague)

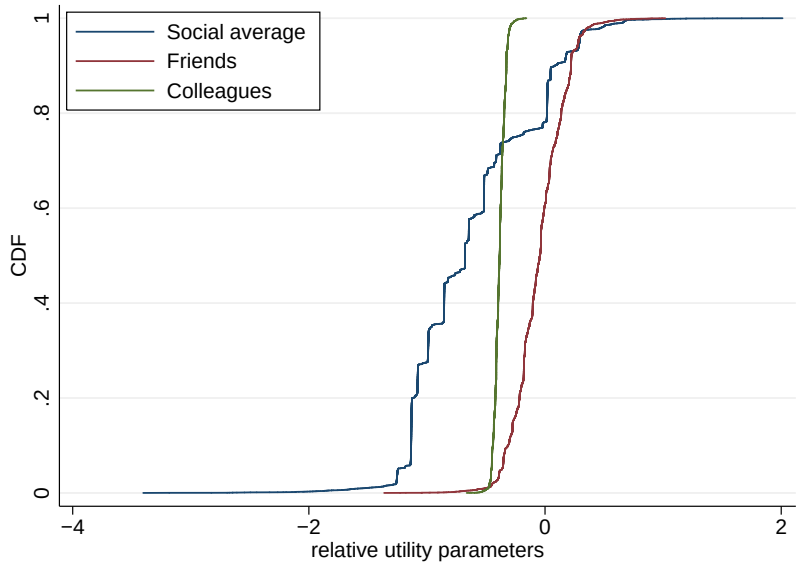


Table 1: Descriptive Statistics

	(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 (Social Average Task)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Comparison against: Social Average									
Comparison benchmarks		family	neighbor	friend	colleague	N.A.	others		
	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		
		(0.049)	(0.018)	(0.064)	(0.073)	(0.026)	(0.026)		
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. For estimates of γ, we report standard errors constructed via the delta method.

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 (Social Average 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: Test on Loss aversion Theory (Social Average Task)

	(1)	(2)	(3)
	Conditional logit		
Own income	0.048***	0.034***	0.024***
	(0.001)	(0.001)	(0.001)
Reference income	-0.022***	-0.012***	0.012***
	(0.001)	(0.001)	(0.001)
Downward comparison ($\mu = 1$)		0.107***	-0.091
		(0.029)	(0.057)
Upward comparison ($\nu = 1$)		-1.625***	0.434***
		(0.045)	(0.064)
Reference income * Downward comparison			0.025***
			(0.002)
Reference income * Upward comparison			-0.044***
			(0.001)
Observations	48172	48172	48172
Pseudo R-squared	0.249	0.290	0.341

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

Table 7: 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)
Interactions of types of ref. group and ref. income levels:								
Reference group of								
Different sex $*\bar{y}$	-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 $*\bar{y}$	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 $*\bar{y}$	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 $*\bar{y}$	-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 $*\bar{y}$	-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 of own individual characterstics								
and income levels (a)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	45554	45554	45554	45554	45554	45554	20328 (male)	25226 (female)
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) If Yes, individual characteristics are controlled with interaction terms for the own income term, and for reference income. The same set of individual characteristic variables cotrolled in the Social-Average-Task-regressions is taken into account.

Table 8: Conditional Logit and Mixed Logit Estimates (Who Compares to Whom 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	0.028***	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		
		(0.039)	(0.030)	(0.013)	(0.017)	(0.017)	(0.048)		
Estimated γ_w	-0.393	-0.429	-0.387	-0.429	-0.367	-0.280	-0.406		
		(0.047)	(0.036)	(0.015)	(0.019)	(0.019)	(0.059)		
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. For estimates of γ, we report standard errors constructed via the delta method.

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 9: 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