Are Americans Really Less Happy With Their Incomes?

Arie Kapteyn, James P. Smith, RAND
Arthur van Soest, Tilburg University, Netspar and RAND

May 2008

Abstract

Recent economic research on international comparisons of subjective well-being suffers from several important biases due to the potential incomparability of response scales within and across countries. In this paper we concentrate on self-reported satisfaction with income in two countries: The Netherlands and the U.S. The comparability problem is addressed by using anchoring vignettes. We find that in the raw data, Americans appear decidedly less satisfied with their income than the Dutch. It turns out however that after response scale adjustment based on vignettes the distribution of satisfaction in the two countries is essentially identical. In addition, we find that the within-country cross-sectional effect of income on satisfaction- a key parameter in the recent debate in the economic literature- is significantly under-estimated especially in the US when differences in response scales are not taken into account.

JEL codes: I30, J30
Keywords: happiness, life satisfaction, vignettes, reporting bias

This research was supported by grants from the National Institute on Aging to RAND.

1. Introduction

Economists have discovered happiness or at least research on subjective well-being and its economic correlates. The rapidly growing subsequent research has touched on several important themes. These have included the so-called Easterlin paradox whereby average happiness remains relatively constant over time in spite of large increases in income per capita (Easterlin, 1974, 1995). In contrast, within country cross-sectional and panel data almost always shows that rising incomes 'buy' additional satisfaction, although the magnitude of the within country cross-sectional effect of income on satisfaction is under dispute (Blanchflower and Oswald, 2004, Di Tella et al, 2007 and Stevenson and Wolfers, 2008). Resolving this paradox, which is often interpreted as a fundamental challenge to the conventional economic theory of utility maximization, has generated a substantial amount of subsequent research attempting to reconcile this finding with the normally positive correlation between income and subjective-well being based on within country estimates.

This reconciliation has included adding relative incomes (of others or of oneself in the past) in the utility function (Van de Stadt et al. 1985, Clark et al, 2008) or a sometimes rapid process of adaptation to new circumstances (Di Tella et al, 2003) often labeled the 'hedonic treadmill'(Di Tella et al, 2007). A recent contrary view is provided by Deaton (2008) who documents that if one considers a much wider range of countries arrayed by their level of economic development, the normally positive association of income with subjective life satisfaction reappears. His work also claims that the effect of income estimated across countries on life satisfaction is if anything higher in the high income countries. Stevenson and Wolfers (2008) revisit much of the earlier evidence and look at new data to reach similar conclusions.

A considerable amount of research has focused on cross-country differences in subjective well-being, in particular comparing Europe and the U.S. where the US appears to rank lower in satisfaction than many European countries with lower per capita incomes (Alesina et al, 2004, Di Tella et al, 2003, and Blanchflower and Oswald, 2004). For instance, Europeans apparently exhibit a stronger distaste for inequality than do Americans that may be partly explained by a perception of greater mobility in the US (Alesina et al, 2004). Blanchflower and Oswald (2004) study trends in well-being over time in the UK and the US and find that reported levels of well-being have been dropping over time in the US and have been flat in the UK, despite the fact that in both countries average incomes have grown substantially over the last couple of decades.

A fundamental problem in international comparisons, cross-sectional and time series analyses of subjective well-being is that one has to assume that somehow response scales are the same across countries, across time and across groups of respondents within a country. This critical and largely untested assumption becomes even more tenuous if question phrasings change or differ across surveys, as is often the case (see Stevenson and Wolfers, 2008). Here we address these problems head on. In view of the specific interest of economists in the relation between life satisfaction and income, we focus specifically on satisfaction with income.

The population distribution of income satisfaction in a country will depend in the first instance on levels and distribution of incomes. Residents of alternative countries can however differ in the way they translate any given level of income into a subjective level of satisfaction with that income. Moreover, residents of countries may differ in the subjective thresholds that they use in demarcating income satisfaction into discrete categories such as very satisfied or not satisfied. Income distributions, the translation from income to income satisfaction and the demarcation thresholds can all affect differences observed within and between countries in their

distribution of stated level of income satisfaction. These distinct factors are often confused in the existing literature on life satisfaction and happiness. In our research, we have created unique data sources in two countries—the United States and the Netherlands—and developed a statistical methodology that allows us to separate out these distinct factors.

Our research indicates that the biases that flow from not taking into account differences in response scales are very large. In the raw data, Americans are decidedly less satisfied with their income than the Dutch. However, after response scale adjustment based on vignettes the distribution of satisfaction in the two countries is essentially identical. In addition, we find that the within-country cross-sectional effect of income on satisfaction- a key parameter in the recent debate in the economic literature- is significantly under-estimated in the United States when differences in response scales are not taken into account.

This paper is divided into four parts. Section 2 describes the data sources that we developed and will use in this analysis. In the next section, we first summarize the vignette methodology that serves as the basis of our analysis and then sketch our statistical model that corrects for response scale differences across countries. Section 4 presents the empirical results and their implications for interpreting observed differences in income satisfaction in the two countries. In section 5, simulations based on our estimated model are used to ascertain what the Dutch distributions of income satisfaction would be if the Dutch had American thresholds rather than their own. The final section highlights our main conclusions.

2. Data Sources and Vignettes

Our analysis in this paper is based on information obtained from two Internet surveys, which we designed and implemented in both the Netherlands and the United States. For the Netherlands, we use the CentERpanel, which includes about 2,250 households who have agreed

to respond to questions every weekend over the Internet. This Dutch sample is not restricted to households with their own Internet access. Instead, respondents are initially recruited by telephone. If they agree to participate and do not have Internet access, they are provided with Internet access (and if necessary, a set-top box). Thus, CentERpanel is representative of the adult Dutch population except the institutionalized. The Internet infrastructure makes CentERpanel an extremely valuable tool to conduct experiments, with possibilities for randomization of content. Production lags are very short, with about one month between module design and data delivery.

From multiple waves collected in the past, CentERpanel has a rich set of variables on demographic, health, and economic characteristics of respondents. In 2006, we collected vignette evaluations concerning several domains of life satisfaction including their subjective satisfaction with their own income (described below).

Our Internet survey for the United States is the RAND American Life Panel (ALP). This panel was initially recruited from respondents age 40 plus in the Monthly Survey (MS) of Michigan's Survey Research Center but has been subsequently supplemented with younger respondents.[1] Similar background information was collected for these respondents as was available for Dutch respondents. The American sample that we use for estimation consists of 1,113 respondents interviewed during 2006-2007.

In both the Dutch and American samples, respondents were given vignettes that cover the four life domains that have figured prominently in the happiness and life satisfaction literature—income, family relations, work, and health. In each domain, they were asked to rate themselves on the same five point scale as they rate the vignette person. The scale that is used is the same for

[1]The MS, the leading consumer sentiments survey, produces the widely used Index of Consumer Attitudes. MS respondents are asked if they have Internet access and, if yes, if they are willing to participate in Internet surveys. Those who agree are added to our household panel to be interviewed regularly over the Internet. As with the CentERpanel, respondents who do not have Internet access are provided with a set top box (an MSN Web TV) that allows them to browse the Internet and send and receive email.

all domains: (*very satisfied, satisfied, not satisfied or dissatisfied, not satisfied, and very satisfied*). In this paper we will investigate satisfaction with income.

Using the subjective scale, individual respondents are first asked *"How satisfied are you with the total income in your household?"* The vignette questions that were asked after that have the following form:

" (Name) is married and has two children; the total after tax household income of his/her family is ($Income_i$)

How satisfied do you think (Name) is with the total income of (his/her) household?"

Once again, the response categories are *very satisfied, satisfied, not satisfied or dissatisfied, not satisfied, and very satisfied. Name* can be either a male of female name assigned randomly across vignettes. $Income_{i,}$, $i=1,\ldots4,$ can take four different values corresponding to respectively half the median, the median, twice the median or four times the median income in the country where the respondent is located. These incomes are also assigned randomly across vignettes. Specifically:

$Income_1$	€15,000	$23,000;
$Income_2$:	€30,000	$46,000;
$Income_3$:	€60,000	$92,000;
$Income_4$:	€120,000	$184,000

The second income amount in each country is chosen equal to the median household income in that country. Thus the amounts are half the median, the median, twice the median and four times the median in each country.

3.1. The Theory of Vignettes

In this section, we first provide an intuitive description of the use of vignettes for identifying response scale differences and then sketch in detail our statistical approach. The basic

idea behind the use of vignettes is illustrated in Figure 1, which presents the distribution of satisfaction (or rather dissatisfaction) with income in two hypothetical countries.

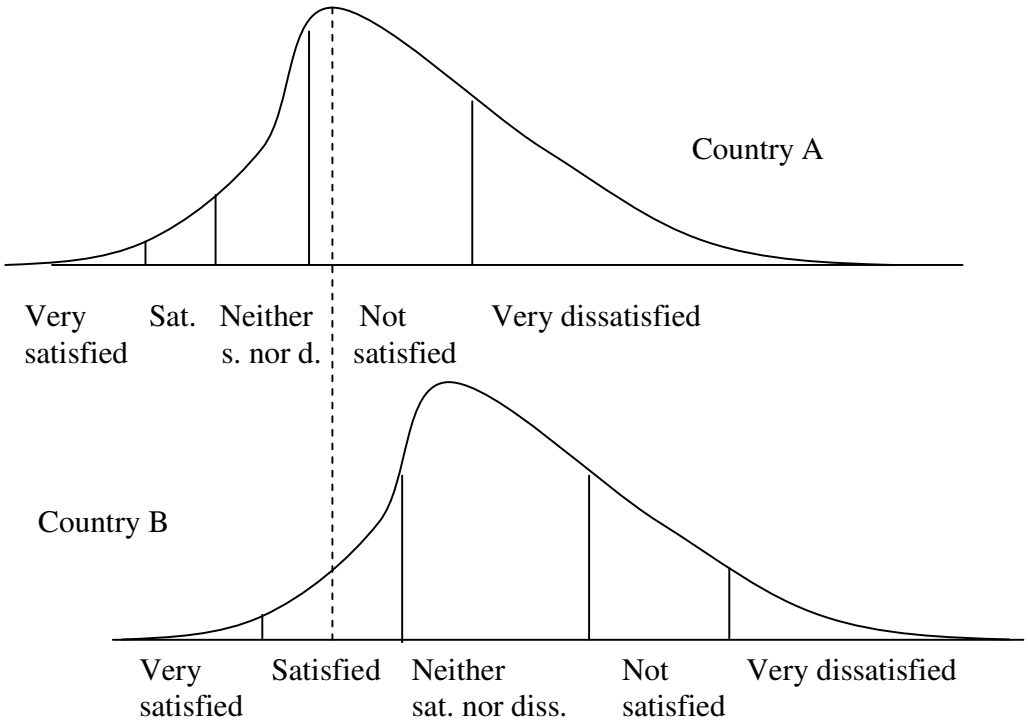
The starting point for the population distribution of satisfaction with income begins with the underlying income distribution. Countries can and do differ in the level and spread of this income distribution. Income distributions are translated by country residents into distributions of income satisfaction by some country and perhaps individual specific translation parameters.

To illustrate and simplify our point, assume for the moment that the two countries (A and B) represented in Figure 1 share an identical income distribution. The density of the continuous income satisfaction variable in country B is to the right of that in country A, implying that on average, people in country B are less satisfied with their income than in country A. Thus, residents of country A translate the same income into a higher level of income satisfaction on a continuous scale. However, residents of the two countries also differ in another important sense. They use very different response scales if asked to report their satisfaction on the standard five-point scale that we introduced above.

In the example in Figure 1, people in country B use much more positive labels to express their income satisfaction than do people in country A. Someone in country B with the satisfaction indicated by the dashed line would report to be satisfied with his or her income, while a person in country A with the same actual income would report “not satisfied.” The frequency distribution of self-reports in the two countries would suggest that people in country B are more satisfied (or less dissatisfied) with their income than those in country A—the opposite of the true income satisfaction distribution. In the literature, the phenomenon that different countries or socio-economic groups use different response scales is called Differential Item

Functioning (DIF; see, e.g., King et al., 2004). Correcting for these differences in the response scales is essential to compare the actual income satisfaction distributions in the two countries.

Figure 1. Comparing self-reported income satisfaction



Vignettes can be used to do the correction and to obtain an unbiased estimate of the translation from income to satisfaction. A vignette question describes the satisfaction of a hypothetical person and then asks the respondent to evaluate the satisfaction of that person on the same five-point scale that was used for the self-report of their satisfaction. Since the vignette descriptions are the same in the two countries, the vignette persons in the two countries have the same actual income satisfaction. For example, respondents can be asked to evaluate the income satisfaction of a person whose satisfaction is given by the dashed line. In country A, this will be

evaluated as “not satisfied.” In country B, the evaluation would be “satisfied.” Since the actual level of satisfaction is the same in the two countries, the difference in the country evaluations must be due to a different threshold (DIF).

Vignette evaluations thus help to identify differences between the response scales. Using the scales in one of the two countries as the benchmark, the distribution of evaluations in the other country can be adjusted by evaluating them on the benchmark scale. The corrected distribution of the evaluations can then be compared to that in the benchmark country—they are now on the same scale. In the example in the figure, this will lead to the correct conclusion that, on average, people in country A are more satisfied with their incomes than people in country B.

The assumptions underlying the vignette corrections are twofold. The first is *response consistency:* a given respondent uses the same scale for self-reports and vignette evaluations. King et al. (2004) and Van Soest et al (2007) provide evidence supporting this assumption for vignettes on vision and drinking behavior, respectively, by comparing vignette corrected self-reports with an objective measure. The second assumption is *vignette equivalence*: no systematic differences in the interpretation of a given vignette between the different groups of respondents (so that systematic differences in evaluations are due to DIF only). Since we have given the vignette households an income that relates to the country specific median, this assumption is valid if respondents evaluate on the basis of relative income compared to the country median (in line with relativity of satisfaction; e.g. Van de Stadt et al, 1985). At the time of the survey the median income in the US was, in terms of purchasing power parity, higher than the median income in the Netherlands. As a consequence, it may be the case that US vignettes are interpreted more positively than Dutch vignettes if absolute income level also plays a role. In the

concluding section, we will discuss what the direction and magnitude of this bias would imply for our results.

3.2. Econometric Model

We will apply the vignette approach to income satisfaction, using vignettes not only to obtain international comparisons corrected for DIF, but also for comparisons of different groups within a given country. Our model explains respondents' self-reports on satisfaction with their own household incomes as well as their reports on income satisfaction of hypothetical vignette persons. Self-reports are modeled as a function of respondent characteristics X_i (including household income, a country dummy and interactions of all characteristics with that dummy) and an error term ε_i by the following ordered response equation:

$$Y_i^* = X_i\beta + \varepsilon_i;\ \ \varepsilon_i \sim N(0,\sigma^2),\ \varepsilon_i \text{ independent of } X_i \tag{1.1}$$

$$Y_i = j \text{ if } \tau_i^{j-1} < Y_i^* \le \tau_i^j,\ \ j = 1,...5 \tag{1.2}$$

The thresholds τ_j^i between the categories are given by

$$\begin{aligned} &\tau_i^0 = -\infty,\ \tau_i^5 = \infty,\ \tau_i^1 = \gamma^1 X_i + u_i,\ \tau_i^j = \tau_i^{j-1} + \exp(\gamma^j X_i),\ j = 2,3,4 \\ &u_i \sim N(0,\sigma_u^2),\ u_i \text{ independent of } X_i \text{ and the other error terms in the model} \end{aligned} \tag{1.3}$$

As noted before, the fact that different respondents use different response scales τ_i^j is called "differential item functioning" (DIF). The term u_i introduces an unobserved individual effect in the response scale. It implies that evaluations of different vignettes are correlated with each other and with self-reports (conditional on X_i), since some respondents will tend to use high thresholds and others will use low thresholds in all their evaluations.

Define a benchmark respondent with characteristics $X_i = X(B)$. The DIF adjustment involves comparing Y_i^* to thresholds τ_B^j rather than τ_i^j, where τ_B^j is obtained in the same way as

τ_i^j but using *X(B)* instead of X_i. A respondent's reported satisfaction is computed using a benchmark scale instead of a respondent's own scale. This does not give an adjusted score for each individual (since Y_i^* is not observed) but it can be used to simulate adjusted *distributions* of Y_i for the whole population or conditional upon some of the characteristics in X_i.

Using self-reports on own income satisfaction only, parameters β and γ^1 are not separately identified, only the difference between β and γ^1. For example, consider country dummies: people in two different countries can have systematically different income satisfactions, but if the scales on which they report their income satisfaction can also differ across countries, then self-reports are not enough to identify the income satisfaction difference between the countries. The vignettes will be used to identify β and γ^1 separately.

The evaluations Y_{li} of vignettes $l=1,\ldots,L$ are modeled using similar ordered response equations:

(1.4) $$Y_{li}^* = \theta_l + \varepsilon_{li}$$

(1.5) $$Y_{li} = j \text{ if } \tau_i^{j-1} < Y_{li}^* \le \tau_i^j,\ j = 1,\ldots 5$$

(1.6) $$\varepsilon_{li} \sim N(0, \sigma^2),\ \text{independent of each other, of } \varepsilon_{ri} \text{ and of } X_i$$

Thus the vignette evaluations are simply modeled as dummies – they do not depend on respondent characteristics X_i (the assumption of *vignette equivalence)*. Since the only variation across vignette descriptions is the level of income (see above), one can interpret the dummies as indicating how the income of the vignette person is translated into satisfaction by the respondent. The translation of the satisfaction into verbal labels follows the same scheme as for self-reports.

The maintained assumption here is that of *"response consistency"*, meaning that the thresholds τ_i^j are the same for self-reports and the vignettes.

With these assumptions, it is clear how vignette evaluations can separately identify β and γ $(=\gamma^1,...,\gamma^5)$: From the vignette evaluations alone, γ, θ, $\theta_1,...\theta_5$ can be identified (up to the usual normalization of scale and location). From self-reports, β can then be identified in addition. Thus the vignettes can be used to solve the identification problem due to DIF

4. Empirical Results

This section highlights our main empirical findings. We first describe what our data imply for satisfaction with own income in the US and the Netherlands and next summarize the distribution of answers given by Dutch and American respondents to the income vignette questions. The third subsection discusses our main parameter estimates determining the level of own satisfaction with income and well as the threshold parameters in both countries.

4.1. Levels of self- reported income satisfaction

Respondents are first given a set of questions asking them to rate their own level of satisfaction with their incomes on the five point very satisfied to very dissatisfied scale. Table 1 summarizes responses obtained from the Dutch and American samples for income satisfaction. In spite of the fact that on average incomes are higher in the United States than in the Netherlands (compared in terms of purchasing power parity), Americans appear to be much less satisfied with their incomes than the Dutch are. Sixty-four percent of Dutch respondents say that they are either satisfied or very satisfied with their total household income. The comparable fraction for Americans is 46%—eighteen percentage points lower than the Dutch. Similarly, a much larger fraction of Americans respond that they are either not satisfied or very dissatisfied—a third of Americans compared to 13% amongst the Dutch. This avoidance of the extremes and

rush to the middle is a common feature of Dutch responses to subjective scale questions and is similar to what we have documented in prior work on other outcomes (Kapteyn et al., 2007).

4.2. Responses to Vignette questions on satisfaction

We gave the Dutch and US Internet respondents vignettes in the domain of income satisfaction. All vignettes were presented with either a female or male name, which was randomized across respondents. Vignettes were presented in random order to eliminate any order effects. Comparing the rank ordering of vignette evaluations across respondents shows that different respondents tend to order vignettes in the same way – in less than 0.5% (13 cases) of all pairs of vignette evaluations evaluated by the same respondent, the evaluation of the higher income vignette is worse than that of the lower income vignette.

Table 2 summarizes responses obtained for both countries per vignette. There are four vignettes which we index one to four with the lowest number representing the lowest income used in the vignette. As with the self-ratings of their own incomes, Americans are less satisfied than the Dutch with the incomes of hypothetical vignette persons at all four income levels. Note that in terms of PPP comparison, the American vignette households have higher incomes than the Dutch, making the fact that the Americans evaluate the vignette incomes as less satisfactory even more striking.

These between country differences narrow substantially as we increase incomes of the hypothetical vignette person. For example, at both twice and four times the median country incomes, there are relatively small differences between the Americans and the Dutch. Much larger differences appear at lower country incomes. Forty-five percent of Americans say that the vignette person should not be satisfied with the median income compared to 22% of Dutch

respondents. If the vignette person was at the bottom point of half the median income, 47% of Americans say that they should be very dissatisfied compared to 27% of the Dutch.

4.3. Parameter Estimates

The model presented in Section 3.2 was estimated using the self-evaluations and vignettes in the Dutch CentERpanel and the RAND American Life Panel. The equations for income satisfaction and for the thresholds corresponding to that domain include a complete set of interactions with a country dummy for the United States. We also estimated the simpler model that does not allow for DIF, which for the self-assessments is similar to a standard ordered probit for self-assessed income satisfaction that would be obtained in the current literature.

Table 3 lists estimated parameters and associated standard errors for satisfaction with own income. We estimated an ordered probit model for satisfaction with own income both with (DIF) and without (no DIF) the correction for different thresholds used by respondents within and across both countries. Differences between these two model estimates inform us about the impact of the threshold correction on the parameter estimates in the income satisfaction equation. Table 4 summarizes the impact of co-variates in both countries on the placement of each of the five thresholds in the income satisfaction domain.

Table 3 lists parameter estimates for ordered probits for self satisfaction with income where the scale is from good to bad (1: very satisfied, …, 5: very dissatisfied). Demographic variables include dummy variables for whether the respondent is female, married, and dummies for age categories 40-50, 51-64, 65+ (the omitted group is under 40 years old). Education is separated into three groups—low, medium or high with the low education group as the omitted category. In the United States, High school Graduate or less is coded as low education, and Post College degrees as high education, with the medium group including all others between these

two. In the Netherlands, the medium group has intermediate vocational or general training, and the high education group has higher vocational training or any university degree. The low education category has everyone with primary school only or lower vocational training.

Income is measured as the logarithm of equivalized family income, defined as the logarithm of family income minus the logarithm of family size. Dutch incomes are transformed to US$ incomes using the same transformation as in the income vignettes – based upon equalizing median incomes in the two countries[2]. Log family size is a separate regressor in part to test for the impact of this functional form of an equivalence scale. Finally, a dummy variable is included indicating whether the respondent is working.

Our preferred model—the one with DIF (adjusting for threshold differences) — is contained in the last two columns. Our estimates indicate relatively small, but statistically significant, differences in income satisfaction among the Dutch in terms of most observable attributes. To illustrate, in the Dutch sample there are no differences in satisfaction with own income by gender, age or work status. Higher income does make the Dutch more satisfied with their income situation, a relationship that is highly statistically significant. Conditional on own income, higher education also makes the Dutch more satisfied with their income; one interpretation of this is that this is a permanent income effect. Individuals with higher education enjoy higher permanent income. An alternative explanation is that the self-reported household income measure we use is imperfect, and education proxies the deviation between this measure and actual family income. Finally, conditional on the equivalized income and family size, married Dutch respondents are more satisfied with their income. One interpretation of this

[2]Since incomes are entered in log-form this only affects intercepts in estimated equations

finding is that the economic benefits of living as two instead of one are greater than the extra income married people receive in terms of holding equivalized income constant.

Turning to our estimates of differences in parameters between the two countries and thus implicitly the US parameters, there are many more differences in income satisfaction in terms of the attributes included in the model among Americans compared to the Dutch and many across country differences as well. Similar to the Dutch, there are no statistically significant gender differences in income satisfaction among Americans and the estimated age and education patterns in income satisfaction are not all that different. While working had no effect on income satisfaction for the Dutch, American workers are somewhat less satisfied with their incomes than American non-workers. There also appear to be fewer economies of scale with marriage among the Americans but a lot more economies of living together with children given the sharply negative estimate for log family size among the Americans.

The most important variable for comparing the two countries is income. The impact of log income on income satisfaction is much more pronounced in the US compared to the Netherlands (twice as large in the US). Americans become much more satisfied with their incomes at high incomes levels and much less satisfied than the Dutch at low incomes levels. Conditional on income, higher education makes Americans even more satisfied than it does for the Dutch, but the differences in the education effects are not significant.

A central question is how important the corrections for threshold differences within and across countries are in our interpretation of these relationships with income satisfaction. It turns out to be quite important. This question is first addressed by comparing the parameter estimates in the model without DIF to the model with DIF. While the Dutch main effects mostly have the same sign, the estimated magnitudes are substantially different in several instances. For example,

for the Dutch the DIF estimate of high education is -0.42 while it is -0.28 with DIF. Similarly for the Dutch the estimated impact of marital status is 50% larger when taking into account threshold differences than when not. Most important, the estimated effect of income on income satisfaction in the United States is fifty percent higher when differences in response scales are taken into account. (-1.21 compared to -0.81).

The (interaction of the) US dummy (with the constant term) is difficult to interpret, since other regressors do not have mean zero. Instead, it is better to look at predicted systematic parts for 'average' Dutch and US respondents. According to results allowing for DIF, these predictions are very similar using Dutch and US parameters (1.146 and 1.155 for the Dutch average and 1.291 and 1.318 for the US average; both differences are insignificantly different from 0). This indicates that, once a correction for DIF is made, the own income evaluation of the average respondent in both countries hardly depends on the scale he or she uses.

Not correcting for DIF, however, would have lead to a significantly different conclusion. With the Dutch and US parameters of the self-evaluation, the average Dutch respondent's predicted systematic part would be 1.230 and 1.625, respectively, and for the average US respondent the numbers would be 1.331 and 1.767, suggesting that in the US, much higher incomes are needed to be equally well off. In fact, this is the interpretation in the existing literature. The result allowing for DIF shows that this is completely due to the fact that the average US respondent uses a less positive response scale than the average Dutch respondent.[3]

Differences between standard ordered probits and the model with a DIF correction shows how misleading simple cross country comparisons can be, even with seemingly identically

[3]An equivalent way of showing this would be to define the regressors in deviations to their (overall, Dutch or US) means. The US dummy would then become small and insignificant in the model with DIF, and positive and significant (and equal to about 0.43) according to the model without DIF.

worded questions. We find that the DIF correction changes our conclusion with respect to levels of income satisfaction in both countries. Moreover we find that variation in income satisfaction is much steeper in the US as income rises than conventional models would indicate.

The reasons for the impact of threshold differences on how we interpret satisfaction with income can be seen if we examine parameter estimates for thresholds contained in Table 4. In these models a negative coefficient means that a respondent sets a tougher standard on income satisfaction—that is, it takes a higher income to be satisfied with one's income. Because of this, if an estimated coefficient is negative fewer people with this trait will be satisfied with their income. The opposite is true for an estimated positive coefficient—people with this trait will tend to move their threshold to the right (so that it is less demanding) and more of them will be satisfied with their incomes.

To illustrate the implications of these estimated coefficients on the thresholds, examine first the set of estimated coefficients that apply to the first threshold—that is the threshold that separates the "Very Satisfied" from the "Satisfied". Judging by the estimated main (Dutch) coefficients, there does not appear to be a great deal of heterogeneity amongst the Dutch in how they set this threshold. In particular, Dutch income levels do not seem to alter the placement of this threshold very much.

There are many significant American interactions on attributes, implying that Americans are much more heterogeneous in the threshold they use to distinguish between an income that is very satisfactory from one that is just satisfactory. In particular, the negative coefficient on the income interaction with being an American indicates that higher income makes Americans more demanding on this particular threshold. Higher income Americans are less likely to say they are very satisfied with the same level of high income than are low income Americans. Two other

attributes that appear to matter are family size and education where an increase in family size or in education (controlling for equivalized income) makes an American respondent more demanding on this threshold.

The other thresholds are more difficult to interpret since they are all adjustments relative to the first threshold just discussed. But in general Dutch attributes appear to matter much less than American attributes in setting these thresholds and increases in income make Americans more demanding. An interesting contrast takes place on the fourth threshold—the one between dissatisfied and very dissatisfied. In this case, higher income makes the Dutch less demanding but the Americans more demanding.

The threshold predictions for the average Dutch or US respondents confirm what we already saw in the data: the thresholds using the US parameters are always significantly lower (and very similar using the average US or the average Dutch sample characteristics), indicating that response scales of US respondents are more demanding than response scales of Dutch respondents, keeping characteristics fixed at some average level.

5. Model Simulations

A transparent way of understanding the implications of our approach is to simulate the distribution of satisfaction with income in the two countries for different parameter values. Essentially we first simulate the Dutch distribution of self-reported income satisfaction and then replace various sets of parameters by the corresponding American values. Table 5 presents the results of these simulations by four age groups—those less than 40, 40-50 years old, 50-64 years old, and at least 65 years old. The first row for each age group summarizes the distribution of satisfaction with income for the Dutch using their own parameters. The second row simulates the Dutch distribution if we replace the parameters in the Dutch satisfaction equation (i.e. Table 3

with DIF) by the American parameters. The third row replaces Dutch thresholds by American thresholds (cf. Table 4). The fourth row replaces all Dutch parameters by American parameters. The fifth row simulates distributions for the American sample using American parameters.

To illustrate, in the age group of those less than 40 years old, comparing the first and fifth row shows that 10% of the Dutch claim that they are very satisfied with their incomes, compared to 6% of Americans. The differences between the countries are even larger at other satisfaction levels, where for example 15% of Dutch respondents say that they are dissatisfied with their income compared to 35% of Americans. In general when their own country specific thresholds are used, more Dutch than Americans are at least satisfied with their incomes and fewer of them are dissatisfied with their incomes.

The second row in each panel shows that replacing Dutch parameters in the own income satisfaction equation by American parameters does not really change that conclusion. The distributions in the second row of each panel are very similar to those in the first row. Apparently, the source of the difference between the Dutch and the Americans does not fundamentally lie in differences in their respective income satisfaction equations.

However, if the Dutch had American thresholds instead of their own (the third row of each panel), the situation would be quite different in that the Dutch distribution of happiness with their income in all age groups looks almost identical to the American distribution of income satisfaction. That conclusion does not change appreciably if we also assign the American satisfaction parameters to the Dutch, as one would expect in view of the comparison between the first and second row. Thus, the results strongly suggest that most of the observed differences in the raw data between the Dutch and the Americans lie in the scales they use (the thresholds separating the various verbal labels).

We will illustrate why this happens by first considering the highest 'very satisfied' threshold. Our estimates indicate both that own income increases overall income satisfaction (Table 3) and that high income Americans have more demanding standards than the Dutch on what income is necessary to be very satisfied with income (Table 4).

Since income satisfaction is increasing in income, attributes of respondents around the threshold between very satisfied and satisfied are those of higher income respondents. Thus, we should be using the comparative thresholds of higher income Americans and higher income Dutch in making the Dutch adopt the American thresholds. Our estimates show that higher income Americans are more demanding than higher income Dutch so having the Dutch look like Americans at the very satisfied threshold basically makes the Dutch set a higher standard (higher income) for claiming to be very satisfied with their incomes. Consequently, fewer Dutch will claim that they are very satisfied with their incomes. Moreover this effect is strong enough that the hypothetical Dutch distribution of very satisfied is almost identical to the American one.

Next examine the other end of the income satisfaction distribution—the threshold between dissatisfied and very dissatisfied. The positive association of own income with income satisfaction now implies that on average attributes of respondents around this threshold are those of lower income Dutch and American respondents. The estimated steeper effect of income on this threshold now implies that Americans would be less demanding than the Dutch. That is, they would be less likely to translate a satisfaction level into the verbal category "very dissatisfied". On the other hand however the coefficient on income in the satisfaction equation is larger for Americans, which implies a lower level of satisfaction with income at low levels. Indeed we see (comparing the first two rows in each panel) that giving the Dutch the US satisfaction parameters leads to an increase in the number of Dutch who are classified as very dissatisfied.

6. Conclusions and interpretation

In this paper, we have used vignettes to disentangle determinants of income satisfaction within and across countries from the verbal scales people use to express this satisfaction. We find that the verbal scales are different across the Netherlands and the U.S. and also among country residents, particular in the United States. Correcting for the differences leads to very similar distributions of income satisfaction in sharp contrast to the large differences between the two countries in the raw data. In this case, not adjusting for response scale differences between countries can lead to very misleading inferences about cross-country differences in life satisfaction. At the same time, applying the DIF correction has an appreciable impact on some of the parameters in the income satisfaction equation. For instance, the effect of own income on income satisfaction of Americans increases substantially when we estimate the model with DIF.

Anchoring vignettes provide additional information but making use of this requires additional assumptions, as discussed in section 3. One assumption is vignette equivalence: it implies that a hypothetical family with US median income is evaluated in the US in the same way as a family with the Dutch median income in the Netherlands, keeping response scales constant. As discussed, above, a deviation from this assumption in the sense that absolute income matters, would reinforce our findings, since the ppp adjusted median income in the US is higher than in the Netherlands. Therefore, none of our qualitative conclusions would change.

In the debate about the cross-national relation between income and income satisfaction the incomparability of response scales has been long recognized. Vignettes are an obvious instrument to get at the incomparability issue. Our results suggest that after the vignette corrections the distributions of income satisfaction are not all that different across the two countries. It will be of interest to expand the analysis to more countries to see what light vignette corrections shed at the Easterlin paradox.

References

Alesina, Alberto. Rafael Di Tella and Robert MacCulloch. 2004. "Inequality and Happiness; Are Europeans and Americans Different?" *Journal of Public Economics,* 88(9-10), 2009-2042.

Blanchflower, David G. and Oswald, Andrew J. 2004. 'Well-Being Over Time in Britain and the USA." *Journal of Public Economics,* 88(7-8), 1359-1386.

Clark, Andrew E., Paul Frijters, and Michael Shields. 2008. "Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles." *Journal of Economic Literature*. 46(1), 95-144.

Deaton, Angus. 2008. "Income, Aging, Health and Well-Being Around the World: Evidence from the Gallup World Poll." *Journal of Economic Perspectives*. 22(2), 53-72.

Di Tella, Rafael, MacCulloch, Robert J and Blanchflower, David G. 2003. "The Macroeconomics of Happiness." *Review of Economics and Statistics*. 85(4), 809-827.

Di Tella, Rafael, John Haisken–DeNew, and Robert MacCulloch, 2007, "Happiness adaptation to income and to status in an individual panel," processed, October

Easterlin Richard A. (1974), "Does Economic Growth Improve the Human Lot? Some Empirical Evidence", in: R. David and M. Reder (eds.) Nations and Households in Economic Growth: Essays in honor of Moses Abramowitz, New York, Academic Press, 89-125.

Easterlin, Richard A. 1995. "Will Raising the Incomes of All Increase the Happiness of All?" *Journal of Economic Behavior and Organization*, 27(1), 35-48.

Kapteyn, Arie., Smith, James. P. & van Soest, Arthur. 2007. "Vignettes and self-reports of work disability in the U.S. and the Netherlands." *American Economic Review,* 97(1), 461-473.

King, Gary; Murray, Christopher; Salomon, Joshua and Tandon, Ajay. 2004. "Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research." *American Political Science Review,* 98(1), 567-583.

Layard, Richard. 2005. *Happiness: Lessons from a New Science,*. Penguin Books: London.

Van de Stadt, Huib, Arie Kapteyn & Sara van de Geer. 1985. "The Relativity of Utility: Evidence from Panel Data." *The Review of Economics and Statistics,* 67, 179-187.

Van Soest, Arthur., Liam Delaney, Liam, Harmon, Colm, Arie Kapteyn Arie., Smith, James P. 2007. "Validating the Use of Vignettes for Subjective Threshold Scales." RAND Labor and Population working paper WP-501.

Stevenson, Betsey and Wolfers, Justin. 2008, "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox," Working Paper, Wharton School, University of Pennsylvania, prepared for *Brookings Papers on Economic Activity*, Spring 2008.

Table 1
Self Satisfaction with Income

Self report: How satisfied are you with the total income of your household?	Country	
	NL	US
Very satisfied	9.9	6.5
Satisfied	53.6	39.4
Not satisfied or dissatisfied	23.6	21.5
Not satisfied	10.3	27.4
Very dissatisfied	2.7	5.2

Table 2

Vignette Evaluations for Income Satisifaction in United States and Netherlands

Income Satisfaction Vignettes Income in Vignette	1 Half Median		2 Median		3 Twice Median		4 Four Times Median	
	NL	US	NL	US	NL	US	NL	US
Very satisfied	0.8	1.1	5.1	2.7	39.6	38.0	74.9	69.5
Satisfied	6.2	0.7	32.2	23.3	50.0	46.0	20.1	22.7
Not satisfied or dissatisfied	15.4	9.1	40.3	28.9	7.7	10.9	3.1	4.9
Not satisfied	50.7	42.4	20.0	37.9	2.2	4.7	1.0	2.2
Very dissatisfied	27.0	46.8	2.4	7.1	0.6	0.4	1.0	0.7

Table 3
Satisfaction with Own Household Income

	Model without DIF		Model with DIF	
	β	s.e.	β	s.e.
Constant	4.73*	0.27	5.38*	0.40
Female	-0.03	0.05	-0.04	0.06
Married	-0.50*	0.09	-0.74*	0.10
ln family size	-0.08	0.08	0.09	0.09
Age 40-50	-0.06	0.07	-0.06	0.08
Age 51-64	0.07	0.07	0.09	0.08
Age 65+	-0.19*	0.09	-0.11	0.11
Ed med	-0.01	0.06	-0.06	0.07
Ed high	-0.28*	0.06	-0.42*	0.08
Working	-0.02	0.06	-0.04	0.07
ln eq income	-0.30*	0.03	-0.36*	0.03
Interactions with dummy US				
Constant	5.49.*	0.50	7.73*	0.66
Female	0.05	0.08	-0.10	0.10
Married	0.27*	0.12	0.48*	0.15
ln family size	-0.48*	0.12	-0.94*	0.15
Age 40-50	0.12	0.11	0.26+	0.14
Age 51-64	-0.11	0.12	0.00	0.14
Age 65+	-0.26+	0.15	-0.11	0.19
Ed med	0.05	0.11	-0.12	0.14
Ed high	0.07	0.11	-0.22	0.15
Working	0.22*	0.09	0.23+	0.12
ln eq income	-0.51*	0.05	-0.74*	0.07

* indicates significance at the 5% level and + indicates significance at the 10% level.

Table 4

Thresholds of Estimated Equation for the Self Assessment: Income

	Threshold 1		ln (Threshold 2 – Threshold 1)		ln (Threshold 3 – Threshold 2)		ln (Threshold 4 – Threshold 3)	
	β	s.e.	β	s.e.	β	s.e.	β	s.e.
Constant	0.00	0.00	0.43+	0.23	0.67*	0.37	-1.02*	0.43
Female	0.04	0.06	-0.04	0.03	-0.01	0.05	0.64	0.05
Married	-0.14	0.09	-0.03	0.06	0.01	0.09	0.07	0.09
ln family size	0.96	0.08	0.06	0.05	-0.01	0.08	0.13	0.08
Age 40-50	0.09	0.08	-0.08+	0.05	0.11	0.07	-0.00	0.07
Age 51-64	0.08	0.08	-0.10*	0.05	0.23*	0.07	-0.04	0.07
Age 65+	0.16+	0.10	-0.10	0.06	0.30*	0.09	0.03	0.10
Ed med	0.05	0.07	-0.08+	0.04	-0.04	0.06	0.16*	0.06
Ed high	-0.08	0.07	0.00	0.04	-0.09	0.06	0.20*	0.07
Working	-0.01	0.07	-0.01	0.04	-0.01	0.06	-0.02	0.06
ln eq income	-0.04	0.03	0.02	0.02	-0.08*	0.03	0.11*	0.04
Interactions with dummy US								
Dummy	2.14*	0.59	-0.84+	0.51	-0.94	0.58	2.13*	0.62
Female	-0.16+	0.09	0.07	0.06	-0.29*	0.09	-0.03	0.08
Married	0.17	0.13	-0.04	0.09	0.11	0.13	-0.11	0.11
ln family size	-0.34*	0.13	-0.01	0.09	-0.09	0.13	-0.18	0.11
Age 40-50	0.15	0.13	0.01	0.09	-0.09	0.12	-0.09	0.10
Age 51-64	0.08	0.13	0.09	0.09	-0.26*	0.13	0.00	0.10
Age 65+	0.18	0.18	0.04	0.12	-0.27	0.17	-0.03	0.15
Ed med	-0.41*	0.13	0.16+	0.09	0.15	0.12	-0.17	0.10
Ed high	-0.43*	0.13	0.04	0.09	0.21+	0.12	-0.23*	0.10
Working	-0.03	0.11	0.03	0.07	0.04	0.11	0.11	0.09
ln eq income	-0.20*	0.06	0.06	0.05	-0.11+	0.06	-0.17*	0.06

* indicates significance at the 5% level.

+ indicates significance at the 10% level.

Table 5

Simulations from Model: Percent Distribution of Income Satisfaction

	Very Satisfied	Satisfied	Not Satisfied/ Dissatisfied	Dissatisfied	Very Dissatisfied
Age group younger than 40					
Dutch using own thresholds	10	49	24	15	2
Dutch using US self- assess equation	12	48	21	13	6
Dutch using US thresholds	5	32	26	33	4
Dutch using all US parameters	5	32	25	34	4
US using US thresholds	6	31	21	35	8
Age group 40-50					
Dutch using own thresholds	12	49	25	12	2
Dutch using US self- assess equation	12	45	25	14	5
Dutch using US thresholds	9	36	26	25	4
Dutch using all US parameters	9	36	25	27	3
US using US thresholds	9	35	22	29	5
Age group 50-64					
Dutch using own thresholds	11	46	29	12	2
Dutch using US self- assess equation	13	45	26	12	4
Dutch using US thresholds	8	37	24	27	4
Dutch using all US parameters	8	37	23	28	3
US using US thresholds	10	37	20	27	6
Age group 65 and older					
Dutch using own thresholds	16	51	25	8	1
Dutch using US self- assess equation	12	48	21	13	6
Dutch using US thresholds	5	32	26	33	4
Dutch using US thresholds	16	43	21	19	1
US using US thresholds	17	38	21	21	3
All age groups					
Dutch using own thresholds	12	48	26	12	2
Dutch using US self- assess equation	15	46	23	11	4
Dutch using US thresholds	9	36	25	26	3
Dutch using all US parameters	11	36	22	26	5
US using US thresholds	9	34	22	29	6