Reversing the Gender Gap in Happiness: Validating the Use of Life Satisfaction Self-Reports Worldwide*

Job market paper

Mallory Montgomery†

November 2016

Abstract

Life satisfaction surveys are increasingly being used as a measure of welfare (Stiglitz et al., 2009), and even proposed as a primary measure (Layard, 2005). On average worldwide, surveys consistently find that women report higher life satisfaction than men. Yet, women are worse off in many ways: less educated, lower incomes, worse self-reported health, fewer opportunities, etc. Why do they report higher life satisfaction? Using recent data from the Gallup World Poll, I show that women are not happier than men; using anchoring vignettes, I show that the gap is due to women and men systematically using different response scales, and that once these scales have been normalized, women are *less* happy than men on average. On the other hand, the effects of other characteristics commonly studied (income, education, marital status, etc.) are robust to vignette adjustment, reinforcing previous findings. Before we can use life satisfaction to evaluate welfare and policy, we need to be sure our measures are well suited to the task.

1 Introduction

Women report higher life satisfaction[1] than men; much higher. In my data, all else equal, just being a woman increases life satisfaction reports more than moving more than a decile higher in the income distribution.[2] Where other studies have failed to explain the gender happiness gap well, this paper explains it entirely, and even reverses it by about 50%. Using anchoring vignettes as well as self-reports from the Gallup World Poll, I show that the gap is due to men and women using different response scales. Specifically, women use lower thresholds for each life satisfaction rating, so that for the same "objective" life satisfaction level, they give a higher rating. This has large implications for policy, and for life satisfaction research as a whole.

Interest in measures of subjective well-being is increasing. In 2009, Joseph Stiglitz, Amartya Sen, and Jean-Paul Fitoussi published the Report by the Commission on the Measurement of Economic Performance and Social Progress (Stiglitz et al., 2009). It found that GDP alone was a poor measure of human progress and welfare, and suggested incorporating other relevant measures – sustainability and measures of well-being, including life satisfaction. Layard (2005) proposes the

*This research was supported by grant 1 R01 AG036784-01 from the National Institute on Aging.

†University of Southern California, m.montgomery@gmail.com

[1]Though there are some differences, I will follow much of the literature and use "life satisfaction," "subjective well-being," and "happiness" interchangeably; in all cases, I am referring to satisfaction with one's life as a whole.

[2]In the U.S., that corresponds to an increase in annual per-capita income of between $5,000 and $10,000.

utilitarian ideal of getting the greatest happiness for the greatest number of people as the "ultimate" policy goal, rejecting GDP as an inadequate measure of welfare, and Helliwell et al. (2015) promote evaluating policies in terms of their impact on happiness, such as having a "critical level of extra happiness" per dollar to justify expenditures.

In order to use life satisfaction as a central indicator of social progress, we must be confident that it is measuring what it purports to; thus methodological studies are needed, to validate the use of life satisfaction surveys and to find where they yield systematically biased results.

At the same time, women's rights and well-being are central policy concerns, for international organizations (e.g. the U.N. Entity for Gender Equality and the Empowerment of Women) as well as for governments and other agencies around the globe (e.g. the White House Council on Women and Girls). These agencies attempt to improve life for women, who are often granted fewer rights, get worse representation, experience discrimination, and are more frequent victims of violence (UN Women, 2015). In life satisfaction studies, research consistently finds that education, income, and health are strong predictors of happiness; women tend to be less educated, have lower incomes, and have worse self-reported health. Yet, on average women report higher life satisfaction than men. Although this pattern does not exist everywhere, it is true in much of the world, and of the world on average, as we will see later. Why do women report greater happiness when they are objectively worse off in so many ways?

In this paper, I reveal gender differences in life satisfaction as one area where we can now see that our international survey responses have been biased. When women and men use the same scale, women are less happy on average (although, as we will see, the marginal effect of being female is much smaller but still positive). And the method described here to show this bias, the use of anchoring vignettes, can be used similarly to root out other biases, to adjust self-reports to give them more validity, and thus make them a better basis for policy decisions.

Anchoring vignettes have been used to study different response scales in many applications, particularly health, but also political efficacy, affect, corruption, job satisfaction, life satisfaction, and more. By asking respondents to rate hypothetical vignette characters on the same scale as their self-reports, we "anchor" their scales, making their responses comparable across heterogeneous groups.

Although I focus here on gender, which has a very different effect in an unadjusted than in a vignette-adjusted model, another contribution of this paper is showing how consistent the effects of other variables are in both cases. Education, marital status, employment status, urban/rural, self-reported health, and income, remain significant determinants of life satisfaction, with coefficients and marginal effects of similar magnitudes, whether I use a vignette-adjusted model or not.

1.1 The gender happiness gap and anchoring vignettes in the literature

Women are frequently found to report higher happiness levels than men (Helliwell et al., 2015). Often, the finding is noted in a single line in the text or coefficient on a table meant to examine some other variable's impact. The gap has been well documented, particularly in developed countries, using several data sources such as past rounds of the Gallup World Poll, the World Values Survey, and others (Graham and Chattopadhyay, 2012; Zweig, 2014; Stevenson and Wolfers, 2009).

Few studies have attempted to explain the gap; among those that have, evidence is mixed. Similar to this paper, Arrossa and Gandelman (2016) use several major data sources and find that marginal effect of being female is much larger than it appears, given observed characteristics. Depending on the study, the gap is increasing with GDP (Graham and Chattopadhyay, 2012) or decreasing with GDP (Lima, 2011); decreasing over time (Stevenson and Wolfers, 2009) or not changing (Herbst, 2011). Some find trends in the gap according to religion, whether the country has a Communist past, and occupational patterns (Meisenberg and Woodley, 2015), or with age (Plagnol and Easterlin, 2008; Easterlin, 2003); others find no patterns at all (Zweig, 2014).

Some studies have found evidence that life satisfaction declines among women following improvements in gender equality (Graham and Chattopadhyay, 2012; Stevenson and Wolfers, 2009). Stevenson and Wolfers (2009) propose that socioeconomic changes, changes in what the measures capture due to large social shifts (e.g. women previously considered happiness with their home lives, and now consider happiness with their home and professional lives), or changes in reference group (e.g. women may be unsatisfied with a wonderful home if they feel they have failed professionally) could explain the decline. In a similar vein, other authors have proposed that life satisfaction self-reports capture more than just true life satisfaction; they incorporate things like aspirations (Zweig, 2014; Plagnol and Easterlin, 2008) and optimism (Arrossa and Gandelman, 2016).

Considering different domains of life, having different reference groups, different aspirations, or different levels of optimism could all lead to men and women using different response scales when answering life satisfaction self-reports. By using an anchoring vignette-adjusted model, I can identify scale differences, and estimate to what extent they are driven by gender.

Anchoring vignettes allow researchers to consider not only an individual's self-report, but also the scale she uses to report it, by asking her to rate hypothetical vignette subjects on the same scale as her self-report. By using the same vignettes for many people, we can see how rating scales systematically differ with individual characteristics such as gender, education level, or country. Anchoring vignettes are commonly used when self-reports are subjective, and an objective measure is unavailable or would be too difficult to obtain (King et al., 2004).

Kapteyn et al. (2007) used anchoring vignettes about work disability, an objective measure of which is difficult or impossible to obtain, and found that people in the Netherlands were more likely to say someone was too disabled to work than people in the United States, given the same objective description of disability. Much research has looked at cross-country differences in health, e.g. Bago d'Uva et al. (2008), King et al. (2004), Grol-prokopczyk et al. (2011). For example, Molina (2016) considers self-reports of health such as shortness of breath and memory loss, which are possible but costly to measure, to examine cross-country differences. (She also finds that gender gaps in health self-reports are reduced once responses are adjusted for vignettes.) Other domains include political efficacy (King et al., 2004), job satisfaction (Kristensen and Johansson, 2008), drinking behavior (Van Soest et al., 2011), and corruption (Grzymala-Busse, 2007).

Little work on life satisfaction using anchoring vignettes has been published thus far. In one paper, the authors study the relationship between income and life satisfaction in the United States and the Netherlands using surveys from each country (Kapteyn et al., 2010); in another, the authors use the Survey of Health, Ageing and Retirement in Europe (SHARE) to see how response styles change with age (Angelini et al., 2012).

This paper proceeds as follows. Section 2 describes my data. Section 3 establishes the gender life satisfaction reporting gap and its magnitude. Section 4 considers three pieces of evidence that women may not actually be happier than men, prompting Section 5, which describes intuitively and then formally the vignette-adjustment model. Section 6 estimates the vignette-adjusted model and compares it with the unadjusted model, for the global sample and separately by country, and looks for trends in the gender life satisfaction gap by region, GDP per capita, gender equality, and religion. Finally, Section 7 discusses the results concludes.

2 Data[3]

Since 2005, the Gallup World Poll continually surveys residents in over 150 countries, interviewing about 1,000 randomly sampled individuals in each country. World Poll questions measure opinions about national institutions, corruption, youth development, community basics, diversity, optimism, violence, religiosity, and other topics. The World Poll questionnaire is translated into major languages of each country. The translation process starts with an English, French, or Spanish version, depending on the region. A translator proficient in both original and target languages translates the survey into the target language. A second translator reviews the language version against the original version and recommends refinements.

With some exceptions, all samples are probability-based and nationally representative of the resident population aged 15 and older. The coverage area is the entire country including rural areas, and the sampling frame represents the entire civilian, non-institutionalized, aged 15 and older population of each country. Exceptions include areas where safety of interviewing staff is threatened, scarcely populated islands, and areas interviewers can reach only by foot, animal, or small boat. Specifically, sampling in the Central African Republic, Democratic Republic of the Congo, Lebanon, Pakistan, India, Syria, Azerbaijan, Georgia, Morocco, Myanmar (Burma), Chad, Madagascar, Moldova, and Sudan was affected by security; some of these as well as Canada, China, Laos, and small parts of Japan had non-representative sampling of some geographic regions. In Arab countries (Bahrain, Kuwait, Saudi Arabia), sampling was of citizens (including Arab expatriates) and those who could complete the survey in Arabic or English; in the United Arab Emirates, all non-Arabs were excluded, i.e. more than half of the population. In the Philippines, urban areas were over-sampled. Israel excludes East Jerusalem (Gallup reports Palestinian Territories separately).[4] Running the analysis without these sampling-compromised countries would seriously restrict my available data. I therefore present the results including these countries; still the main results remain broadly the same if I remove them (see appendix).

Telephone surveys are used in countries where telephone coverage represents at least 80% of the population or is the customary survey methodology. In Central and Eastern Europe and most of the developing world, an area frame design is used for face-to-face interviewing. In some countries, over-samples are collected in major cities or areas of special interest. In some large countries, such as China and Russia, samples of at least 2,000 are collected.

Gallup has created a worldwide data set with standardized income and education data. To make education comparable across countries, education descriptions are recoded into one of three rele-

[3] Much of this description is also in another (coauthored) paper of mine using the same dataset, not yet published, "Life Satisfaction Within and Across Countries: Societal Capital and Relative Income."

[4] http://www.gallup.com/services/177797/country-data-set-details.aspx

vant categories: "Elementary": completed elementary education or less (up to eight years of basic education); "Secondary": completed some education beyond elementary education (9 to 15 years of education); "Tertiary": completed four years of education beyond "high school" and/or received a four-year college degree. Similarly, annual household income in international dollars is calculated using the Individual Consumption Expenditure corrected for the Household PPP ratio from the World Bank. These PPP-corrected values correlate strongly (r=0.94) with the World Bank estimate of per-capita GDP (PPP-corrected). The result is a household income measure that is comparable across all respondents, countries, and local and global regions.

Response rates are calculated according to AAPOR Standard Definitions (Callegaro and Disogra, 2008), and reported figures include completed and partial interviews, refusals, non-contacts, and unknown households. Gallup World Poll response rates vary by mode of survey and region. Response rates in Sub-Saharan Africa are higher than other world regions, ranging from a high of 96% in Sierra Leone to a low of 54% in Nigeria, with an average response rate of 80%. Average response rates for the Middle East, Asia, South America and former Soviet Union countries are 63%, 56%, 43%, and 50%, respectively.

As part of a National Institute on Aging supported project,[5] Gallup added a module on the international comparison of well-being to surveys conducted in 109 countries conducted during 2011-2014. These countries provide the sample for the current paper. Eighteen countries were interviewed in 2011, 39 in 2012, 26 in 2013, and 26 in 2014.

Most countries have approximately 1,000 observations, with the exceptions of Russia (1,500), India (5,000), China (4,500), Germany (3,000), United Kingdom (3,000) and Haiti (500), for a total of about 120,000 observations.

My primary interest is in the answers to the following question (the so-called Cantril ladder):

> *"Please imagine a ladder with steps numbered from zero at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?"*

In addition to a self-report, subjects were asked to rate the life satisfaction of each person in a set of six vignettes. The interviewer randomly asked about one of two possible vignette sets, set A or set B. Although respondents used a 0-10 scale, because there are relatively few responses at the top and particularly the bottom of the scale, I combine ratings into a five-point scale: ratings 0-2 are recoded as 1, ratings 3-4 are recoded as 2, ratings 5-6 are recoded as 3, ratings 7-8 are recoded as 4, and ratings 9-10 are recoded as 5. I do this for both self-reports and vignette ratings. As discussed in Section 5.3.2, the B set of vignettes fails tests for one of the crucial assumptions of vignette adjustment models (vignette equivalence), so I restrict my analysis to respondents who answered the A set. Of those, approximately 200 observations are dropped because respondents didn't answer at least two vignette questions.[6] After dropping observations that are missing an essential variable (life satisfaction, household income, etc.), I end up with 45,731 observations from 103 nations. Summary statistics for those respondents are given in Table 1. About 15% of the

[5] Awarded to Arie Kapteyn, James P. Smith, and Arthur van Soest.

[6] Results are similar if I only use respondents that answered all six vignette questions, although it reduces the sample by about 1,000 observations.

sample (18,956 observations) was missing household income, the most commonly missing variable. Georgia, Singapore, and Ecuador were missing data on unemployment. Egypt and Lebanon were missing data on health.

Thus my sample contains 45,731 respondents, all of whom answered at least two A-set vignette questions, as well as providing all other key variables.

3 Do women report higher life satisfaction?

In short, based on the self-reports, yes. As noted, the existing literature certainly says so. Here I'll show it using the new 2011-2014 Gallup World Poll data. As shown in Table 1, the overall female minus male reported life satisfaction gap is 0.030, significant at the 1% level. Next I show the effect of being female on life satisfaction reporting at the global and then regional level.

Table 2 shows the results of ordered probit regressions of life satisfaction on personal characteristics: column 1 includes just an indicator for female, age, and age^2, while column 2 adds an indicator for urban, marital status indicators, employment status indicators, education level indicators, whether they have health problems,[7] and log of equivalized income. The reference categories are single, employed full time for an employer, and low education (elementary or less). To get log equivalized household income, I took reported household income and weighted each member according to OECD weights (1 for the first adult, 0.5 for each additional adult, and 0.3 for each child). This may overstate purchasing power of women, because income is reported at the household level. Both columns include country fixed effects.

The coefficients in both columns fit well with the existing literature. Life satisfaction follows a U-shaped function with age, with minimums at 52 and 53 years of age respectively. Being married increases life satisfaction (compared with being single), while being separated, divorced, or widowed reduces it. Working less than one would like (unemployed or underemployed) leads to lower life satisfaction. Education and income increase life satisfaction, and having health problems decreases it.

The effect of being female is positive in both models. In the first column, the coefficient on the female indicator is 0.028, significant at the 5% level, showing that unconditionally, women report higher happiness than men. In the second column, adding the other characteristics increases the coefficient on the female indicator – women are worse off in most of those characteristics, so in order to maintain the positive gap in life satisfaction, there must be a large positive effect of being female.

Table 3 shows the marginal effects of changing individual characteristics. The first column shows estimated life satisfaction self-reports with that characteristic, holding all other characteristics constant, and the second shows the estimated change in life satisfaction reports versus the listed comparison. The marginal effect of changing only gender, from male to female, appears in the last row. Because it holds all other variables constant, the marginal effect is much larger than the raw gap, at 0.076. The third column compares that row's change with the gender effect. For example, moving from the 30th percentile of one's country's equivalized income distribution to the 40th percentile

[7]The health measure is the yes/no answer to the question "Do you have any health problems that prevent you from doing any of the things people your age normally can do?"

increases life satisfaction reports by 0.047, which is 0.66 times as large an effect as the gender effect.

Moving up a decile in the equivalized income distribution increases life satisfaction reports by about 0.5 to 1 times as much as the gender effect. In the U.S., moving up a decile is equivalent to increasing equivalized income by $5000 (10th to 20th percentile) to $14,000 (80th to 90th percentile). Other characteristic changes have even larger effects than gender – compared to not having health problems, having them on average reduces life satisfaction by 0.215, nearly three times as much as the gender effect, and unemployment (versus working full time for an employer) is similar. Still, even the largest magnitude change is less than three times the gender effect.

Table 4 reports the coefficients on female dummies in ordered probit regressions of life satisfaction on the same variables included in Table 2 at the regional level, as well as the marginal gender effect. All female coefficients are significantly positive, significant at the 10% level except for Latin America & the Caribbean. Marginal effects range from 0.045 in the transition economies to 0.160 in Australia/New Zealand/United States/Canada. Even the smallest marginal effect reported is quite large, given the comparisons in Table 3.

4 Why should women be happier? Perhaps they are not.

Despite the strong findings of the previous section that women report higher life satisfaction, there are many reasons for women to be less satisfied with their lives than men. In this section, I will provide evidence using vignettes that women tend to rate the same objective life circumstances more highly than men do. I will also review gender differences in responses to affect questions, which ask respondents about their feelings during the previous day. These responses will indicate that women are more likely to report all negative affects, and less likely to report some positive affects. Finally, I will attempt to decompose the gap in life satisfaction responses into the portion explained by differential characteristics between women and men (e.g. education), by how they differentially value those characteristics (e.g. how education influences life satisfaction for women vs. for men), and the gap explained by neither – that is, the portion of that gap that is just due to one's gender. As it turns out, gender alone must explain a large gap.

4.1 Vignette ratings

In addition to asking respondents on which step of the ladder they current stand, Gallup also asked them to state on which step of the ladder six hypothetical people stand. The exact vignettes were as follows:

> A1: *Think of a female who is 40 years old and happily married with a good family life. Her monthly family income is about [median income]. She has severe back pain, which keeps her awake at night. On which step of the ladder do you think this person stands?*
>
> A2: *Think of a male who is 50 years old and divorced. He has a daughter with whom he has a good relationship. He has a secure job that pays about [twice median income] per month. He has no serious health problems. On which step of the ladder do you think this person stands?*
>
> A3: *Think of a male who is 25 years old and single without many friends. He makes about [half median income] per month. He feels he has little control over his job and*

worries about losing it. He has no health problems but feels stressed sometimes. On which step of the ladder do you think this person stands?

A4: *Think of a female who is 35 years old and married, with no children. Her monthly family income is about [median income]. Her work is a bit dull sometimes, but it is a very secure job. On which step of the ladder do you think this person stands?*

A5: *Think of a female who is a 70-year-old widow. She receives about [half median income] in income each month. She has many friends. Lately, she suffers from back pain, which makes housework painful. On which step of the ladder do you think this person stands?*

A6: *Think of a male who is 60 years old. His is single but has many friends his age. He no longer works but is comfortable with his decision to stop working. He receives about [twice median income] in income each month. He is very physically active. On which step of the ladder do you think this person stands?*

The income levels (median, half the median, or twice the median) were filled in with the the appropriate value for the respondent's country. Every respondent saw the same six vignettes; they were not randomized by e.g. gender or income.

Table 5 considers gender differences in ratings for each vignette. Women rate every single vignette higher, significantly so. This implies that women tend to give higher ratings for the same life circumstances; this is true whether they are rating vignettes of men or women. And, the differences between men's and women's vignette ratings – 0.03, 0.05, 0.04, 0.03, 0.02, 0.06 – are similar in magnitude to the difference between women's and men's self-reports, 0.03.

4.2 Feelings day-to-day

Beyond the Cantril's ladder questions, respondents were also asked about their feelings during the previous day. Specifically they were asked, "Did you experience the following feelings during a lot of the day yesterday?" followed by several questions (e.g., "How about enjoyment?" and "How about worry?"). In all, about 99% of the sample was asked if they experienced physical pain, worry, sadness, stress, anger, enjoyment, whether they felt well rested, whether they were treated with respect, whether they learned something, and whether they smiled or laughed. All countries were asked these questions, though a small number in many countries was not.

Table 6 compares men's and women's responses to these questions. Women are more likely to experience negative feelings. For positive feelings, women were more likely to feel respected[8] and to smile/laugh, while men were more likely to feel well rested and to have learned something; neither men nor women were more likely to experience enjoyment.

Overall, women were more likely to say they experienced all negative feelings, while men were more likely to say they experienced some positive ones. It seems that although women report having higher life satisfaction than men on average, their day-to-day lives are not more enjoyable. In itself, this does not contradict the finding that women report higher overall life satisfaction – any parent (or graduate student) understands taking on extra stress or worry because a child (or dissertation!) brings a deeper joy. But it is one piece of evidence.

[8]The question did not specify by whom they felt respected; there may be gender differences in whose respect came to mind when answering this question (e.g. spouse, children, coworkers, neighbors, etc.).

4.3 Decomposition analysis

Perhaps women are happier despite being objectively disadvantaged, because they subjectively value their own characteristics more than men do. For example, perhaps women value being married more than men do, while men value being educated more than women. In that case, we may see that women's life satisfaction ratings are high if many are married, even if they are not highly educated. In this section I show that this is not the case, and that in fact, women's self-reports are *much* higher than they should be, given their characteristics and how they value them.

To address this, I attempt to decompose the observed differences, using a two-fold Blinder-Oaxaca decomposition (Blinder 1973, Oaxaca 1973), in the gap between women's and men's life satisfaction ratings. This method is commonly used to decompose gender gaps in wages, to determine how much is due to men having more desirable attributes such as additional work experience, and how much is due to rewarding those attributes less when women have them. Rather than looking at the difference between men's and women's wages, I decompose the difference between women's and men's life satisfaction. To do so, I report the portion of the gap attributable to differences in observed characteristics (e.g. education level, health) and the portion attributable to differences in how those characteristics translate into life satisfaction (i.e. the differences in coefficients on those observed characteristics). The portion explained by differences in characteristics is often called the explained portion, while the portion explained by differences in coefficients is called the unexplained portion.[9].

For a linear model, the two-fold Blinder-Oaxaca decomposition splits the observed difference in the mean as follows:

$$\overline{LS_f} - \overline{LS_m} = [E(X_f) - E(X_m)]'\beta^* + [E(X_f)'(\beta_f - \beta^*) + E(X_m)'(\beta^* - \beta_m)] \quad (1)$$

where X_f and X_m are female and male characteristics respectively, β_f and β_m are female and male coefficients on those characteristics in separate regressions, and β^* is coefficients from a pooled regression.

The first term on the right hand side is the portion of the mean attributable to differences in characteristics, while the second term is the portion attributable to differences in coefficients. Thus the first term will capture whether women are happier because they have some advantage in their life circumstances (e.g. if they have higher incomes) and the second will capture whether women are happier because they differentially value the characteristics they have (e.g. if education improves life satisfaction more for women than for men).

Several of the included observed characteristics are categorical (e.g. high, medium, or low education). The original decomposition proposed by Oaxaca (1973) and Blinder (1973) gave different estimates for the impacts of the individual characteristics and coefficients depending on which category was the reference category. Yun (2005) improved the method by incorporating normalized regression into the model, essentially requiring that the coefficients on the categorical dummy variables sum to zero. Normalizing the regression coefficients on categorical variables does not change the overall explained and unexplained portions, nor the coefficients on other variables (Yun, 2005). I normalize the coefficients on all of the included categorical variables (marital status, employment

[9]Because there is no reason to presume that the "true" coefficients are from a regression using only males or only females, the reference coefficients are from a pooled regression including both groups, as proposed by Oaxaca and Ransom (1994)

status, education level). Standard errors are clustered at the country level.

Results of the linear decomposition are shown in Table 7. (Linear models are generally good approximations for ordered probit models, at least in life satisfaction surveys (Ferrer-i Carbonell and Frijters, 2004))[10]. The total gap in self-reports is 0.030. The explained portion of the gap is -0.076, and the total unexplained portion is 0.105; that can further be broken down into the unexplained portion due to coefficients on included variables, 0.025, and the constant term, 0.080, which measures the remaining unexplained gap (the portion due to being female). In other words, women's characteristics tend to be worse, so the contribution of these differences to the overall gap is actually negative; the gap increases once I include them. Women are more likely to be separated, divorced, or widowed, less likely to work full time for an employer or be self-employed, more likely to have low levels of education, more likely to report health problems, and have lower incomes. There is no characteristic here that explains the gap; all significant characteristics have negative contributions, increasing it further.

On the unexplained side, women have different curvature of their age functions which leads to higher life satisfaction for them, that is, their life satisfaction reports evolve somewhat differently over the life cycle than men's, which explains some of the gap. Women also mind less being separated, unemployed, or out of the workforce. On the other hand, women benefit less from working full time for an employer. Overall, differences in how women and men translate their characteristics into life satisfaction reduce the gap somewhat. The final gap, the effect of being female, all else equal, is 0.080, about 2.5 times the raw gap. Women have worse observable characteristics, which widens the gap. And although women do value some characteristics differently than men do, it is not nearly enough to close the gap.

Interpreting the constant term in the unexplained portion as the value of "group membership" for women requires that I assume that everything else relevant to life satisfaction determination is included in the model. Of course, that is not true; there are many important variables that determine one's happiness, and it is impossible to include them all. The large group membership bonus does imply that when considering some of the most consistently found determinants of happiness – employment status, education, health, income – they point to reasons that women should be *less* happy than men, not more. This suggests that something else is at play.

5 Anchoring vignette-adjusted model

5.1 Intuition

Previous studies of life satisfaction ask each respondent for their self-report, and find correlates with these ratings. But in comparing life satisfaction ratings across individuals (in this case, between men and women), we are assuming that they are using comparable scales when rating their own lives. This is not obvious. As noted, women rated every vignette subject more highly than men

[10] In a nonlinear (ordered probit) model, (1) is replaced by the similar equation:

$$\overline{LS_f} - \overline{LS_m} = [E_{\beta^*}(Y_{if}|X_{if}) - E_{\beta^*}(Y_{im}|X_{im})] + [(E_{\beta_f}(Y_{if}|X_{if}) - E_{\beta^*}(Y_{if}|X_{if})) + (E_{\beta^*}(Y_{im}|X_{im}) - E_{\beta_m}(Y_{im}|X_{im}))] \quad (2)$$

The results are similar: the gap to be explained, 0.023 (100%), the explained portion, -0.056 (-244%), and the unexplained portion, 0.079 (+344%), tell the same story as the linear case.

did. This begs the question, is the scale men are using systematically different from women's?

Consider Figure 1. Imagine that men actually have higher average life satisfaction than women, as indicated by men's distribution being shifted right of women's, but that they are less generous in their ratings, as indicated by the right-shifted thresholds. Now imagine a hypothetical person with true life satisfaction at the dashed line. On the men's scale, that's a 2 out of 5. On the women's scale, because the thresholds are shifted left, it's a 4 out of 5. If men and women are using these different response scales in their self-evaluations, it will appear that women are happier, when in fact men are.

To account for this, I want to put men's and women's responses on the same scale, by evaluating everyone's self-reports using the same set of thresholds, in this case, the men's. To do so, Gallup asked every individual to rate the same sets of vignettes. Because the vignettes shown to men and women are the same, they have the same "true" underlying life satisfaction level. If men and women give systematically different ratings, it must be because of differential item functioning, or DIF, otherwise known as differing response scales (King et al., 2004).

Once men and women are using the same adjusted response scale, I can reevaluate the life satisfaction gender gap by simulating women's life satisfaction ratings when using men's scales. This analysis depends on two key assumptions. First, *response consistency* means that individuals are using the same scale to rate their own life satisfaction as the vignette subjects'. King et al. (2004) and Van Soest et al. (2011) have provided good evidence that response consistency holds in other domains. Second, *vignette equivalence* means that men and women interpret the vignettes in the same way. Because my vignettes include incomes based on the median income, this assumes that people rate their life satisfaction in part based on their income relative to their country's median; Kapteyn et al. (forthcoming) provide evidence that this relative measure is appropriate.

5.2 Formal model

To do the vignette adjustments formally, I use a HOPIT (hierarchical ordered probit) model. HOPIT operates similarly to ordered probit, except that in addition to affecting the underlying latent variable for life satisfaction, individual characteristics can *also* affect the thresholds between response levels. Each respondent answers the life satisfaction question for themselves, as well as for six hypothetical individuals.

Y_{ri} is a respondent i's self-report, and it is a function of their characteristics as well as an error term ϵ_{ri} which is normally distributed with zero mean and is independent of X_i. As in an ordered probit model, Y^*_{ri} is a latent variable such that $Y_{ri} = \text{j}$ if Y^*_{ri} is above the threshold τ_i^{j-1} and below the threshold τ_i^j:

$$\begin{aligned} Y^*_{ri} &= X_i\beta + \epsilon_{ri}, \\ \epsilon_{ri} &\sim N(0, \sigma_r^2) \end{aligned} \tag{3}$$

$$Y_{ri} = j \text{ if } \tau_i^{j-1} < Y^*_{ri} \leq \tau_i^j \tag{4}$$

What makes this model different from an ordered probit model is how the thresholds are determined. They are also functions of the subject's characteristics, as well as an idiosyncratic error u_i, which is independent of ϵ_{ri} and X_i:

$$\begin{aligned} \tau_i^0 = -\infty, \tau_i^5 = \infty, \tau_i^1 = \gamma^1 X_i + u_i \\ \tau_i^j = \tau_i^{j-1} + \exp(\gamma^j X_i),\ j = 2, 3, 4 \end{aligned} \tag{5}$$

The individual thresholds τ_i^j represent DIF.

To make self-evaluations comparable, take one respondent's scale as the benchmark scale. That respondent has characteristics $X_i = X(B)$, and thus has thresholds τ_B^j. I can now compare all other individuals using thresholds τ_B^j. Because the latent variable Y_{ri}^* is not affected by the thresholds, this does not imply a new level of the latent variable. But it does imply a new rating, Y_{ri}, for which it is possible to simulate the adjusted distribution.

Not all parameters are identified; namely, only the difference between β and γ^1 can be determined. Only looking at self-reports, I encounter the problem in my simple example: men and women are using different scales. To identify them separately, I use vignette ratings. Each respondent rated a set of six vignettes. Y_{li} is the rating given by i to vignette l:

$$Y_{li}^* = \theta_l + \epsilon_{li} \tag{6}$$

$$\begin{aligned} Y_{li} = j \text{ if } \tau_i^{j-1} < Y_{li}^* \leq \tau_i^j,\ j = 1, ..., 5, \\ \epsilon_{li} \sim N(0, \sigma^2), \text{ independent of } \epsilon_{li}, \epsilon_{ri}, \text{ and } X_i. \end{aligned} \tag{7}$$

where θ_l is a dummy variable indicating vignette l. Notice that the equation for Y_{li}^* does not include any personal characteristics X_i, in line with the assumption of vignette equivalence. The assumption of response consistency means that the τ_i^j here are the same as those used with Y_{ri}. With this, I can identify β, $\gamma_1, ..., \gamma_5$, $\theta_1, ..., \theta_6$ up to a normalization of scale and location.

5.3 Testing vignette equivalence

To check vignette equivalence, I perform two tests. First, I investigate whether each individual's vignette ordering matches the global ordering; if each individual is interpreting the vignettes in the same way, they should order them the same way (with some error). For this I follow Murray et al. (2003) by calculating adjusted Spearman rank order correlations between individual vignette orderings and the global ordering, as explained below. Second, following D'Uva et al. (2011), I test whether individual characteristics significantly determine vignette evaluations. If the vignettes are interpreted the same way by everyone, then the latent vignette rating Y_{li}^* should be determined entirely by the vignette fixed effect, and any difference in the observed Y_{li} should come from differing thresholds τ_i^j, which vary by individual characteristics (e.g. female vs. male).

In the first test, the B vignettes underperform relative to the A vignettes. In the second test, the B vignettes fail entirely, and thus those respondents that received the B set of vignettes, roughly half, are removed from my sample.

5.3.1 Vignette ordering

First, I quantify to what extent respondents deviated from the global vignette ordering in their individual orderings. The "global" vignette ordering is defined as the ordering when all respondents' ratings are averaged. (The global ordering is the same if I define it as the mode ordering.) Following Murray et al. (2003), I call the benefit of the doubt (Spearman) rank order correlation

coefficient (BDROCC) the Spearman correlation with ties resolved to favor the overall (global) ordering. Each respondent rated the vignettes on the same 0-10 scale, and thus, they could give the same rating to multiple vignettes. In finding which orderings are consistent with the global ordering, I resolve ties as if they matched the global order. In the extreme case, if an individual rated every vignette the same way, it would be perfectly consistent with the global ordering. (In practice, well under 1% of individuals did this in the A set, and about 1% did this in the B set.) A high BDROCC (near 1) means it is close to the global ordering.

Table 8 summarizes the BDROCC (see the appendix for a full list of every country) for both sets of vignettes. Four individual countries were especially problematic: Chad (median BDROCC = -0.337 for A set), Palestinian Territories (median BDROCC = -0.143 for B set), Japan (median BDROCC = 0.086 for B set), and United Arab Emirates (median BDROCC = 0.029 for B set). Chad is removed from all A-set analysis that follows.

Correlations are higher in the A set. The median BDROCC for the B set is just 0.543, compared with 0.829 with the A set. While the majority of respondents (59%) were very close to the global ordering in the A set (perfect match, one single-rank inversion, two single-rank inversions, or one double-rank inversion), only about a third were very close in the B set. Looking back at Table 5, there is less variation in the average rankings for the B set than the A set; the range of averages is 3.95-6.60 for the A set and just 3.95-5.14 for the B set. This could reflect any number of factors, including how the vignettes were written (they may not be different enough), how the individuals interpreted the questions (individuals may emphasize different characteristics while developing their rankings, so that a vignette with a high income but bad health may be highly ranked by one person and poorly ranked by another), or how the survey was administered (the surveyors may have made mistakes in asking the questions, Gallup may have made a mistake in the materials they used, or the data may have been recorded incorrectly). For whatever reason, the people rating the B-set were not as consistent, and thus vignette equivalence may not hold for them. There is no theoretical cut-off for how high the BDROCC can be while still assuming vignette equivalence. Clearly, however, the B set is much less consistent than the A set.

5.3.2 The effect of gender on vignette ratings

As a second check of vignette equivalence, I consider a slightly modified version of the model, replacing equation (6) with the following:

$$\begin{aligned} Y_{l1}^{*} &= \theta_1 + \epsilon_{l1} \\ Y_{li}^{*} &= \theta_l + \lambda_l X_i + \epsilon_{li}, l \neq 1 \end{aligned} \tag{8}$$

where l counts $1, ..., 12$ when both sets of vignettes are included and $1, ..., 6$ when only using one set. For identification, I must omit $\lambda_l X_i$ from one vignette equation. If vignettes are equivalent, then coefficients λ_l should be all equal to zero.

Table 9 shows coefficients on female in (8). Using the pooled sample (columns 1 and 2), the A vignettes do not indicate a vignette equivalence violation, but the B vignettes do. In the separate samples, for both the full model and the female and only model, the pattern holds: female does seem to influence interpretation of the vignettes in the B set, but not the A set.

Because of the evidence here that the B vignettes violate the vignette equivalence assumption, I focus my analysis on the sample that answered the A vignettes.

6 The gender gap with vignette adjustments

To see how much of the gender gap is due to DIF, I compare results for ordered probit regressions (no vignette adjustments) to results for HOPIT regressions (with vignette adjustments). Likelihood ratio tests show that the HOPIT model is preferred in both cases.

Table 10 shows coefficients in HOPIT regressions of the same specifications as Table 2: the first column regresses life satisfaction on a female indicator, age, and age^2; the second column adds marital status, employment status, education level, health problems, and log equivalized household income. Both columns include country fixed effects. Each specification in Table 10 has two sets of coefficients: the left two columns are coefficients affecting the life satisfaction rating, and the right two columns re coefficients affecting τ^1, the threshold between the lowest and second-lowest response categories. Here I only report the first threshold equation; see the appendix for coefficient estimates on all four response thresholds.

A positive coefficient in the life satisfaction equation means increasing that variable increases life satisfaction. A positive coefficient on τ^1 means that increasing that variable is moving the threshold toward higher values of Y_i^*, i.e., making a respondent with Y_i^* just above the threshold change their response to the lowest response category. A negative coefficient on τ^1 means the opposite.

The first specification shows a striking difference in the effect of "female" in the ordered probit vs. HOPIT regressions. Because this specification includes only female and age as regressors, it shows the overall impact of being female, including all gender differences in characteristics. The second model includes additional individual characteristics, so that it isolates the impact of being female holding those factors constant. In specification (1), Table 2 shows that women have significantly higher life satisfaction, while the HOPIT model in Table 10 shows that there is essentially no effect. The coefficient on τ^1 explains the difference: women have lower τ^1 thresholds than men. As a result, it takes less for them to move up to a higher response category (or equivalently, it takes a worse situation for them to report a lower response category).

The next specification tells a similar story. The coefficient on female in Table 2 is positive and strongly significant (0.087). In Table 10, the coefficient on female is still significant, but it is much smaller (0.043). Once again, we see that the coefficient on female in τ^1 equation is negative, implying that women use more generous thresholds.

The rest of the coefficients in the HOPIT life satisfaction equation match the results of the ordered probit regressions in Table 2 in their signs, significance levels, and for the most part, even in magnitude. This tells us that these effects are not due simply to DIF, and that for example, education really increases life satisfaction – it isn't that more educated people are just more generous in their self-reports.

Looking at the coefficients in the τ^1 equation of Table 10, being female reduces the first threshold, as expected. Being more educated reduces the first threshold, while having health problems increases it, as does being divorced.

Table 11 shows the marginal effects of changing individual characteristics in the HOPIT equation for life satisfaction alongside the ordered probit marginal effects from Table 3. The marginal effects are similar for many variables, though there are some differences. Moving up a decile in the

equivalized income distribution has a somewhat larger effect in the vignette-adjusted model, as does moving from low education to high education. The benefit of marriage is doubled, but the effects of being divorced or widowed are somewhat smaller. The harm from being unemployed or underemployed are even larger. In the last row, the marginal effect of gender is more than twice as large without vignette adjustment as with it. The ordered probit model overestimates the effect of being female because it fails to consider that women use more generous response scales as well. The HOPIT model in Table 11 isolates the improvement in life satisfaction for women, holding their response thresholds constant. Although being female still improves life satisfaction, all else equal, the improvement is smaller in level as well as compared with other possible characteristic changes.

Although the marginal effect of being female is still positive, because women are worse off in their other characteristics, their overall life satisfaction is lower than men's if they use the same scale as men's. Table 12 shows simulations of life satisfaction ratings for each specification shown in Table 10. The top panel shows estimates using each respondent's own gender's thresholds: women's estimated life satisfaction, men's estimated life satisfaction, the gap, and the simulated distribution of women's and men's responses. The bottom panel shows the same calculations, but adjusts women's thresholds to match men's (that is, to set the coefficient on "female" to zero in the equations for τ^1 through τ^4).[11]. The last line shows the percentage of the top panel gap that is eliminated in the bottom panel.

In both models, evaluating women's self-reports using men's thresholds completely reverses the gap, by about 150%. If women used the same response scale as men, they would have *lower* life satisfaction than men; the gap would be about half as large as it is now, in the other direction. Using men's thresholds, women are more likely to put themselves in the highest response category and less likely to put themselves in the lowest response category.

At the global level, women on average report higher life satisfaction than men. At the individual country level, there is heterogeneity. Table 13 shows simulations of the gender gap using respondents' own thresholds (Gap_{own}) and using men's thresholds (Gap_{m}), as in Table 12, but for models run at the individual country level. It also shows the number of men and women in the sample (N_{m}) and $\text{N}(_{\text{f}})$, the size of the scale adjustment (Avg. adj.), and the marginal effect of changing from male to female (Marg. eff.). To measure the size of the scale adjustment, I use the average adjustment in women's life satisfaction when women in that country move from their own thresholds to men's thresholds: for each female observation, I subtract their expected life satisfaction rating using men's thresholds minus their expected life satisfaction rating using their own thresholds. A negative number means their self-reports are lower with men's thresholds, a positive number means the opposite. Specifically, the average life satisfaction adjustment for women when I set female equal to zero in the equations for τ^1 through τ^4. (Subtracting Gap_{m} - Gap_{own} yields the same value.)

Overall, the individual country results are similar to the global results, with some key differences. First, when estimated separately by country (and weighting each country equally), the average gender life satisfaction gap is smaller when respondents use their own thresholds, 0.020, and the gap reversal is much stronger, all the way to -0.026 (230% decrease). The overall marginal effect, 0.038, is somewhat larger than in the global results.

[11]This method does not account for the fact that women also have other characteristics that are different from men's (e.g. men's higher education levels should shift their τ^1 threshold left, as shown in Table 10) Overall, the results in the bottom panel are similar if I set all coefficients in the threshold equations to men's values.

The next section investigates patterns in Gap_{own} Gap_{m}, using the estimates from these individual country regressions.

6.1 Patterns in the gender gap

To compare with the existing literature, here I consider trends in the gender gap when women use their own thresholds and when they use men's thresholds, to see if they are associated with geographical region, GDP per capita, gender equality, or religion. All values reported here are based on the simulations reported in Table 13.

Not all countries and individuals are included in the following tables. First, people who were not asked (or did not answer) their religion are not included in the religion table. Many countries are missing at least one component in the women's rights and status table, and some are missing GDP per capita.[12] And, because these tables are all based on the adjustment at the national level, two countries that had a problem with fitting the HOPIT model are also excluded from all tables.[13]

First, regionally. Table 14 shows the estimated gender gap when women use their own and men's thresholds respectively, in Gap_{own} and Gap_{m}, and reports the average adjustment for women when switching from their own to men's thresholds. Significance stars reflect t-tests comparing men's and women's average simulated life satisfaction. Every region saw a decrease in the gender gap – the largest is in the Middle East and North Africa, where the gap reduced from 0.02 to -0.12. The smallest adjustments were in East Asia and Sub-Saharan Africa. Gap_{own} is most negative in Latin America and the Caribbean, East Asia, and Sub-Saharan Africa, where men have higher relative life satisfaction on average; on the other side, South Asia and Australia/New Zealand/Canada/United States have the largest positive Gap_{own}. Gap_{m} remains most positive in Australia/New Zealand/Canada/United States, though at a much smaller magnitude, and the Middle East and North Africa has the most negative gap, a very large -0.120.

Evidence in the literature has been mixed about the relationship between the gender gap and GDP per capita. Table 15 shows results of regressions of Gap_{own} (column 1) and Gap_{m} (column 2) on log GDP per capita and log GDP per capita squared, while Figure 2 shows the same regressions graphically with 95% confidence intervals. The gap when women use their own thresholds has no significant relationship with GDP per capita. When women use men's thresholds, there is somewhat of an inverse-U shaped relationship, with the remaining gap being at its highest among middle-income countries, implying that men's average happiness advantage is greatest among low- and high-income countries. However, the coefficients are only marginally significant, and graphically, the inverse-U shape does not appear to be a strong relationship.

Table 16 considers the effects of some measures of women's rights and status. Data on the percent of parliament made up of women comes from the World Bank Development Indicators, while the other measures come from the OECD Development Centre's Social Institutions & Gender Index (SIGI). SIGI rates countries' gender equality on five components: discriminatory family code (e.g. inheritance rights of widows/daughters), restricted physical integrity (e.g. laws on rape, prevalence of female genital mutilation), son bias, restricted resources and assets (e.g. women's access to financial services), and restricted civil liberties (e.g. political representation, freedom of movement

[12] Argentina, Myanmar, Syria, Taiwan, Yemen.

[13] Afghanistan and Senegal; these will be updated in a future draft.

for women); some measures are missing for some countries. SIGI also creates an overall composite score including all five components. In all SIGI measures, which range from zero to one, a low value indicates better gender equality. Table 16 regresses Gap_{own} and Gap_{m} on each of these measures separately at the country level. Most measures indicate that having better gender equality reduces Gap_{own}, toward higher men's life satisfaction relative to women's; son bias has the largest effect in both Gap_{own} and Gap_{m}. A discriminatory family code and women's restricted access to resources and assets are also somewhat predictive of Gap_{m}.

Finally, I consider religion, both as reported at the individual level (Table 17 and the "main religion" of each country (Table 18). Individual religion was reported by most respondents; for completeness I show results from those that did not answer. Main religion data comes from the Pew Research Center's Global Religion Landscape survey. A country has a main religion if at least 50% of the country identifies with that religion. "Unaffiliated" means not affiliated with any religion, while "no main religion" means that no religion has at least 50% in that country.

In Table 17, Gap_{own} is the most negative among Jewish respondents, and Gap_{m} is even more negative. Christians, the largest reporting group, have a significantly positive Gap_{own} and an insignificantly negative Gap_{m}. Besides the relatively small number of Jewish respondents (who are mostly located in Israel), the largest adjustments are among Hindus and Muslims, who move from having a significantly positive Gap_{own} to a significantly negative Gap_{m}. The only group that maintains a positive Gap_{own} is the secular/non-religious. They, along with Christians and those who did not answer the question, had the smallest magnitude adjustments.

Considering each country's main religion, in Table 18, the findings are similar. Israel, the only Jewish-majority country, has similar Gap_{own} and Gap_{m} as in the individual religion table. Similarly, most Hindus in Table 17 are in India and Nepal, the two Hindu-majority countries. The main differences between Tables 17 and 18 are in religions that are less concentrated nationally: Christians, Buddhists, and "others." Individual Christians have a significantly positive Gap_{own} and no significant Gap_{m}, while Christian-majority countries have no significant Gap_{own} and a significantly negative Gap_{m}; Buddhists show a similar pattern. In both cases, the average adjustment is similar in both tables.

6.2 Discussion

I agree with Stiglitz et al. (2009) – GDP *isn't* an adequate measure of human welfare, and improvements in GDP per capita are not the best way to evaluate development. Life satisfaction surveys capture dimensions of well-being that income alone cannot. Indeed, the kind of well-being this paper focuses on, evaluative well-being, is just one of three main kinds recently studied (the other two are experiential – how one feels at any moment – and eudemonic – how fulfilling and meaningful one's life is). All are vital to understanding human welfare. Yet, simply asking people how satisfied with their lives they are may not show us their underlying "true" life satisfaction. There are good reasons that GDP has been used for so long, despite most people (including the general public, not just social scientists) agreeing that it is a flawed metric: it is relatively consistently measured, and relatively easy to interpret. To begin considering a subjective well-being measure when examining potential policy decisions, we need to thoroughly validate that measure. Asking the incredibly diverse world population the same life satisfaction question and comparing their responses is not enough.

The World Happiness Report summarizes findings related to happiness around the world, and describes how happiness surveys can be used to evaluate progress. In a chapter of the 2015 report, titled "How to Make Policy When Happiness is the Goal," Richard Layard and Gus O'Donnell propose assessing policy options by determining a critical happiness per dollar of cost, and enacting policies that have the highest happiness per dollar above that threshold (Layard and O'Donnell, 2015). To apply this methodology, we *need* valid happiness measures. Consider Graham and Chattopadhyay (2012)'s finding that women's happiness actually decreased following improvements in gender equality. It may be that their response scales changed based on their new life experiences, or they may have become less happy for any number of reasons. In any case, improvements in gender equality should not be avoided because of this. Using only self-reports as a basis can lead to promoting policies that encourage higher happiness *reporting*, rather than genuine improvements in well-being.

Anchoring vignettes are a useful tool to expose response scale differences, though there are downsides. Even if survey data passes tests of vignette equivalence, there may still remain some differences. For example, if there are systematic differences in how respondents empathize with the vignette subjects, as proposed by Deaton (2011). This is a real concern, and thus far, it has not been fully answered. Still, given the findings of this and other studies, it is hard to argue that unadjusted models are a better alternative than vignette-adjusted models. The motivating evidence of DIF herein cannot be ignored; nor can we ignore real legal and cultural factors that frequently make women objectively less capable than men. Given all of this, the evidence is strong that what we are observing is not women with higher life satisfaction than men, but women who are *reporting* higher life satisfaction than men.

7 Conclusion

Although there is heterogeneity, on a global scale, women report being more satisfied with their lives than men – considering self-reports only, the marginal effect of being female rather than male is equivalent to moving up one's country's income distribution by more than a decile. Yet, my analysis suggests that much of the gap is due to men and women using the life satisfaction response scale differently, that the true marginal gender effect is much smaller, and that on average, women have lower life satisfaction than men do.

In this paper, I present evidence that women should not be more satisfied, and not only because of their reduced rights, opportunities, and representation. I show three indications that women may not be more satisfied with their lives than men: women rate the lives of hypothetical vignette subjects systematically higher than men do, and that the gender difference in those ratings is similar in magnitude to the gender gap in self-reports; women are more likely to have negative feelings during much of the day, and less likely to feel well rested or like they learned something; using decomposition analysis, I show that women's unfavorable individual characteristics imply that their life satisfaction should be much lower than it is, and that the way they translate their characteristics into life satisfaction does not close the gap, either – women's higher self-reports are attributed to just a difference in gender.

Given the evidence that women may not have higher life satisfaction, I use an anchoring vignette-adjusted model. In that model, individual characteristics can affect not only an individual's life satisfaction, but also the thresholds between ratings on the response scales, moving them up (so

that the same underlying life satisfaction level results in a lower rating) or down (so that the same underlying life satisfaction level results in a higher rating).

I find that the marginal effect of gender is still positive, but much smaller, less than half as large as the unadjusted model estimates. And by simulating each woman's life satisfaction self-report using her own thresholds as well as men's thresholds, I find that putting men and women on the same scale completely reverses the gap, that in the end, women are *less* happy, and that the reversed gender gap is about half as large as the original. The reason the unadjusted model exaggerates the gender effect is that it cannot distinguish between an increase in life satisfaction and a lowering of response thresholds. There is a positive effect of being female on life satisfaction, but more importantly, women give a higher rating for the same underlying life satisfaction, because their response thresholds are lower.

My results are consistent with several previous papers' findings on the gender gap in life satisfaction. Arrossa and Gandelman (2016) did a similar analysis as in my Oaxaca-Blinder decomposition, and with the same conclusion: the life satisfaction gap should be larger given women's observed characteristics and their preferences over those characteristics. Indeed, the authors concluded that women must be more optimistic than men in their self-reports, which is precisely what my results show.

Stevenson and Wolfers (2009) found that women's happiness was declining, in absolute value and relative to men's, in the United States and Western Europe over the last several decades. It may be that this is true; it may also be explained, at least in part, by women's response scales converging with men's. As other gender gaps have closed or narrowed, gender differences in happiness self-reports have likely narrowed as well.

It is also worth noting, to the credit of existing life satisfaction research, that other variables (marital status, employment status, education, health, income) influence life satisfaction in the same direction and in most cases in similar magnitudes in vignette-adjusted and unadjusted models. These results appear to be quite robust across many contexts.

Unfortunately, the scale adjustments identified in this study are not enough to solve the problem of scale differences going forward; we cannot simply ask for life satisfaction ratings and then scale women's responses down a little. Scale differences clearly vary geographically, as demonstrated here, and they likely vary over time as well.

It is not trivial to add anchoring vignettes to a survey – respondents incur an effort cost for each additional question, and survey designers have a difficult task of creating vignettes that will result in both response consistency (respondents rate the vignettes on the same scale as they rate themselves) and vignette equivalence (respondents all interpret the vignettes in the same way). As shown in Section 5.3.2, vignette equivalence is not always easy to achieve.

Here I've shown that looking only at self-reports leads to a misunderstanding of gender differences. Any comparison between groups residents of different countries, people with different ethnic backgrounds, etc. may suffer from a similar bias. Unfortunately, we cannot know what scale differences exist without including anchoring vignettes in surveys. Although it can be costly to add anchoring vignettes, it is worth the expense to improve the validity of the results.

References

Angelini, V., Cavapozzi, D., Corazzini, L., and Paccagnella, O. (2012). Age, Health and Life Satisfaction Among Older Europeans. *Social Indicators Research*, 105(2):293–308.

Arrossa, M. L. and Gandelman, N. (2016). Happiness Decomposition: Female Optimism. *Journal of Happiness Studies*, 17(2):731–756.

Bago d'Uva, T., van Doorslaer, E., Lindeboom, M., and O'Donnell, O. (2008). Does reporting heterogeneity bias the measurement of health disparities? *Health economics*, 17(3):351–375.

Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, 8(4):436–455.

Callegaro, M. and Disogra, C. (2008). Computing response metrics for online panels. *Public Opinion Quarterly*, 72(5):1008–1032.

Deaton, A. (2011). Comment on "Work Disability, Work, and Justification Bias in Europe and the United States". In Wise, D. A., editor, *Explorations in the Economics of Aging*, pages 312–314. University of Chicago Press, Chicago.

D'Uva, T. B., Lindeboom, M., O'Donnell, O., and van Doorslaer, E. (2011). Slipping Anchor? Testing the Vignettes Approach to Identification and Correction of Reporting Heterogeneity. *The Journal of human resources*, 46(4):875–906.

Easterlin, R. A. (2003). Happiness of Women and Men in Later Life: Nature Determinants and Prospects. In *Advances in Quality-of-Life Theory and Research*, volume 20, pages 13–26. Springer Netherlands.

Ferrer-i Carbonell, A. and Frijters, P. (2004). How Important is Methodology for the Estimate of the Determinants of Happiness? *The Economic Journal*, 114(497):641–659.

Graham, C. and Chattopadhyay, S. (2012). Human Capital and Economic Opportunity: Gender and Well-Being around the World: Some Insights from the Economics of Happiness. *Brookings Institution*.

Grol-prokopczyk, H., Freese, J., and Hauser, R. M. (2011). Using Anchoring Vignettes to Assess Group Differences in General Self-Rated Health. *Journal of Health and Social Behavior*, 52(2):246–261.

Grzymala-Busse, A. (2007). *Party competition and state exploitation in post-communist democracies*. Cambridge University Press.

Helliwell, J., Layard, R., and Sachs, J., editors (2015). *2015 World Happiness Report*. United Nations, New York.

Herbst, C. M. (2011). 'Paradoxical' decline? Another look at the relative reduction in female happiness. *Journal of Economic Psychology*, 32(5):773–788.

Kapteyn, A., Smith, J. P., and Van Soest, A. (2007). Vignettes and Self-Report of Work Disability in the United States and the Netherlands. *American Economic Review*, 97(1):461–473.

Kapteyn, A., Smith, J. P., and Van Soest, A. (2010). Life Satisfaction. In Diener, E., Kahneman, D., and Helliwell, J., editors, *International Differences in Well-Being*, chapter 4, pages 70–104. Oxford University Press, Oxford, UK.

King, G., Murray, C. J. L., Salomon, J. A., and Tandon, A. (2004). Enchancing the Validty and Cross-Cultural Comparability of Measurement in Survey Research. *American Political Science Review*, 98(1):191–207.

Kristensen, N. and Johansson, E. (2008). New evidence on cross-country differences in job satisfaction using anchoring vignettes. *Labour Economics*, 15(1):96–117.

Layard, R. (2005). Rethinking Public Economics: The Implications of Rivalry and Habit. *Economics and Happiness: Framing the Analysis*, March.

Layard, R. and O'Donnell, G. (2015). How to Make Policy when Happiness is the Goal. In *2015 World Happiness Report*, chapter 4, pages 76–85. United Nations.

Lima, S. V. (2011). A Cross-Country Investigation of the Determinants of the Happiness Gender Gap. *Mimeo*.

Meisenberg, G. and Woodley, M. A. (2015). Gender Differences in Subjective Well-Being and Their Relationships with Gender Equality. *Journal of Happiness Studies*, 16(6):1539–1555.

Molina, T. (2016). Reporting Heterogeneity and Health Disparities Across Gender and Education Levels: Evidence From Four Countries. *Demography*, 53(2):295–323.

Murray, C. J., Ozaltin, E., Tandon, A., Saloman, J. A., Sadana, R., and Chatterji, S. (2003). Empirical Evalation of the Anchoring Vignette Approach in Health Surveys. In *Health Systems Performance Assessment: Debates, Methods and Empiricism*, pages 369–399.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3):693–709.

Oaxaca, R. L. and Ransom, M. R. (1994). On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, 61(1):5–21.

Plagnol, A. C. and Easterlin, R. A. (2008). Aspirations, attainments, and satisfaction: Life cycle differences between American women and men. *Journal of Happiness Studies*, 9(4):601–619.

Stevenson, B. and Wolfers, J. (2009). The Paradox of Declining Female Happiness. *Economic Policy*, 1(2):190–225.

Stiglitz, J., Sen, A., and Fitoussi, J. P. (2009). Report by Stiglitz Commission on the Measurement of Economic Performance and Social Progress.

UN Women (2015). Progress of the world's women 2015-2016: Transforming economies, realizing rights. Technical report, New York.

Van Soest, A., Delaney, L., Harmon, C., Kapteyn, A., and Smith, J. P. (2011). Validating the use of anchoring vignettes for the correction of response scale differences in subjective questions. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 174(3):575–595.

Yun, M. S. (2005). A simple solution to the identification problem in detailed wage decompositions. *Economic Inquiry*, 43(4):766–772.

Zweig, J. S. (2014). Are Women Happier than Men? Evidence from the Gallup World Poll. *Journal of Happiness Studies*, 16(2):515–541.

Tables and figures

Table 1: Summary statistics

	Overall	**Female**	**Male**	**Difference**	**Std. Err.**
Life satisfaction	2.966	2.980	2.950	0.030***	0.010
Equivalized household income	7,058.70	6,397.51	7,838.88	-1,441.37***	115.93
Primary education or less	34.8%	37.6%	31.4%	6.2%***	0.4%
Secondary education	50.7%	49.1%	52.7%	-3.6%***	0.5%
Tertiary education & up	14.5%	13.3%	15.9%	-2.5%***	0.3%
Single/never married	27.2%	23.6%	31.4%	-7.8%***	0.4%
Married/domestic partner	60.3%	59.6%	61.1%	-1.6%***	0.5%
Separated/divorced/widowed	12.5%	16.9%	7.5%	9.4%***	0.3%
Health problems	23.9%	25.9%	21.4%	4.5%***	0.4%
Unemployed	6.7%	6.8%	6.6%	0.2%	0.2%
N	45,731	24,753	20,978	3,775	-

Table 2: Ordered probit regression

	(1)	(2)
Female	0.028**	0.087***
	(0.014)	(0.015)
Age	-0.012***	-0.015***
	(0.002)	(0.002)
Age^2	0.000***	0.000***
	(0.000)	(0.000)
Urban		0.054***
		(0.021)
Married		0.050***
		(0.019)
Separated		-0.120***
		(0.036)
Divorced		-0.191***
		(0.034)
Widowed		-0.127***
		(0.032)
Domestic partner		-0.042
		(0.032)
Employed full time for self		0.008
		(0.018)
Employed part time do not want full time		0.029
		(0.032)
Unemployed		-0.210***
		(0.030)
Employed part time want full time		-0.080***
		(0.029)
Out of workforce		-0.006
		(0.024)
Secondary education		0.147***
		(0.021)
Tertiary education		0.302***
		(0.029)
Health problems		-0.246***
		(0.032)
Log equivalized household income		0.236***
		(0.015)
Observations	45,731	45,731
Pseudo R-squared	0.082	0.108

Ordered probit regression of life satisfaction on listed variables as well as country fixed effects (103 countries included).

Table 3: Ordered probit marginal effects

Characteristic	**Comparison**	**Life sat.**	**Change**	**x(gender eff.)**
Income at 10th percentile	-	2.746	-	-
Income at 20th percentile	10th percentile	2.826	0.080	1.05
Income at 30th percentile	20th percentile	2.881	0.055	0.72
Income at 40th percentile	30th percentile	2.927	0.047	0.61
Income at 50th percentile	40th percentile	2.970	0.043	0.57
Income at 60th percentile	50th percentile	3.011	0.041	0.54
Income at 70th percentile	60th percentile	3.056	0.045	0.59
Income at 80th percentile	70th percentile	3.110	0.053	0.70
Income at 90th percentile	80th percentile	3.183	0.074	0.97
Low education	-	2.862	-	-
Medium education	Low	2.992	0.129	1.70
High education	Low	3.127	0.264	3.47
No health problems	-	3.017	-	-
Health problems	No health problems	2.802	-0.215	-2.83
Single	-	2.960	-	-
Married	Single	3.003	0.043	0.57
Separated	Single	2.855	-0.104	-1.37
Divorced	Single	2.793	-0.167	-2.19
Widowed	Single	2.849	-0.111	-1.45
Domestic partner	Single	2.923	-0.036	-0.48
Employed full time for employer	-	2.982	-	-
Employed full time for self	Employed full time for emp.	2.989	0.007	0.09
Employed part time, don't want full time	Employed full time for emp.	3.008	0.026	0.34
Unemployed	Employed full time for emp.	2.799	-0.183	-2.40
Employed part time, want full time	Employed full time for emp.	2.912	-0.070	-0.92
Out of the workforce	Employed full time for emp.	2.977	-0.005	-0.07
Rural	-	2.947	-	-
Urban	Rural	2.994	0.047	0.62
Male	-	2.925	-	-
Female	Male	3.001	0.076	1.00

Full sample (N = 45,731). Marginal effects of changing various characteristics, based on ordered probit regression coefficients in Table 2, column 2. Income percentiles are from respondent's own country. "Life sat." shows estimated life satisfaction with that characteristic, all other characteristics held constant. "x(gender eff.)" compares the magnitude of that marginal effect with the gender effect in the last row. Low education is elementary or less, medium education is secondary up to 3 years tertiary, high education is 4 years tertiary or more.

Table 4: Ordered probit regressions, by region

Region	N	Coef.	Std. err.	p-value	Marg. eff.
Transition economies	7,683	0.052**	0.026	0.041	0.045
Latin America & the Caribbean	4,791	0.052	0.034	0.119	0.049
East Asia	3,551	0.066*	0.038	0.082	0.054
Sub-Saharan Africa	8,677	0.075***	0.024	0.002	0.066
Central America	2,503	0.077*	0.045	0.089	0.080
Southeast Asia	3,459	0.104***	0.039	0.007	0.089
Western Europe	3,215	0.115***	0.039	0.004	0.091
South Asia	4,579	0.153***	0.040	0.000	0.132
Aus/NZ/US/Can	1,655	0.213***	0.056	0.000	0.160
Middle East & North Africa	5,618	0.163***	0.032	0.000	0.161

Ordered probit regressions of life satisfaction on female, age, age^2, urban indicator, marital status indicators, employment status indicators, education level indicators, health problems indicator, and log equivalized household income. Coefficient on female indicator.

Table 5: Vignette ratings, by gender

Vignette	N	F avg	M avg	Difference	p-value	Result
Vignette A1	45,411	2.524	2.493	0.031***	0.000	Women higher
Vignette A2	45,518	3.474	3.422	0.052***	0.000	Women higher
Vignette A3	45,435	2.325	2.286	0.039***	0.000	Women higher
Vignette A4	45,454	2.915	2.886	0.029***	0.002	Women higher
Vignette A5	45,426	2.273	2.251	0.022***	0.002	Women higher
Vignette A6	45,403	3.572	3.513	0.059***	0.000	Women higher

t-tests comparing men's and women's average ratings. Result column is based on significance at the 10% level.

Table 6: Feelings yesterday, by gender

	N	F avg	M avg	Diff	p-value	Result
Pain	45,542	0.315	0.274	0.040***	0.000	Women more
Worry	45,401	0.374	0.341	0.033***	0.000	Women more
Sadness	45,369	0.238	0.193	0.046***	0.000	Women more
Stress	45,300	0.323	0.312	0.012***	0.004	Women more
Anger	45,362	0.199	0.190	0.008**	0.013	Women more
Enjoyment	44,770	0.716	0.719	-0.003	0.771	Neither
Well rested	45,415	0.680	0.693	-0.013***	0.001	Men more
Treated with respect	44,808	0.870	0.863	0.007**	0.014	Women more
Learned something	44,808	0.502	0.528	-0.026***	0.000	Men more
Smiled or laughed	44,935	0.739	0.725	0.015***	0.000	Women more

Yes/no responses to "Did you experience the following feelings during a lot of yesterday? How about ...?"

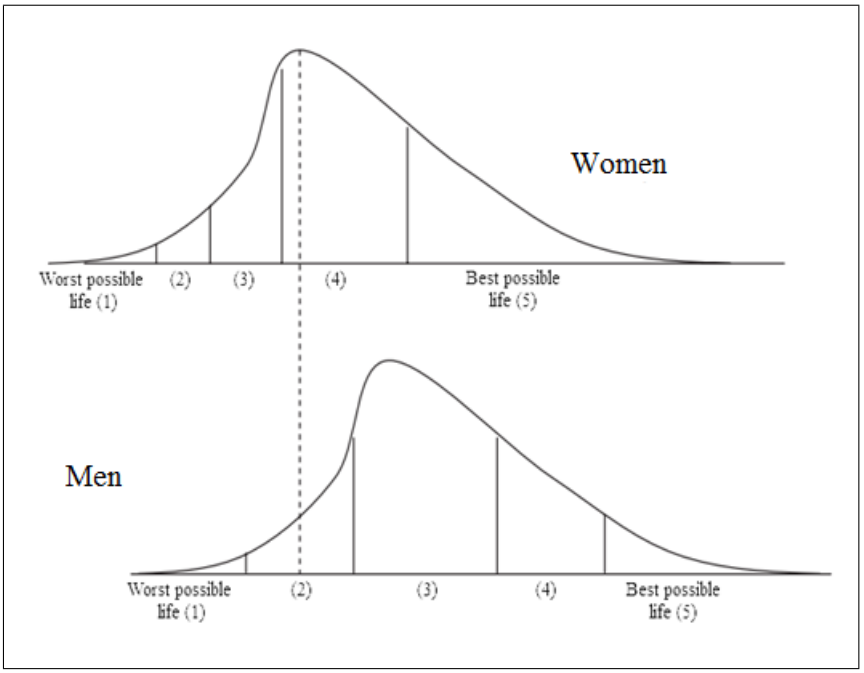
Table 7: Linear Blinder-Oaxaca decomposition of gender life satisfaction gap

Women average	2.980	Overall gap	0.030	100%
Men average	2.950	Explained	-0.076	-253%
Observations	45,731	Total unexplained	0.105	350%
		Observed coefficients	0.025	83%
		Group membership	0.080	267%

	Explained		Unexplained	
Urban	0.001	3%	0.016	53%
	(0.001)		(0.010)	
Age	0.005	17%	0.140**	467%
	(0.004)		(0.065)	
Age^2	-0.003	-10%	-0.103**	-343%
	(0.004)		(0.041)	
Single	-0.004**	-13%	-0.001	-3%
	(0.002)		(0.010)	
Married	-0.001	-3%	-0.010	-33%
	(0.001)		(0.015)	
Separated	0.000	0%	0.003**	10%
	(0.000)		(0.001)	
Divorced	-0.003***	-10%	-0.001	-3%
	(0.001)		(0.002)	
Widowed	-0.010***	-33%	-0.002	-7%
	(0.002)		(0.002)	
Domestic partner	0.002*	7%	-0.001	-3%
	(0.001)		(0.002)	
Employed full time for employer	-0.009***	-30%	-0.023***	-77%
	(0.003)		(0.007)	
Unemployed	-0.000	0%	0.005**	17%
	(0.001)		(0.002)	
Self employed	-0.004**	-13%	-0.000	0%
	(0.002)		(0.005)	
Employed part time (don't want full time)	0.000	0%	-0.002	-7%
	(0.000)		(0.003)	
Employed part time (want full time)	0.000	0%	-0.002	-7%
	(0.000)		(0.002)	
Out of the workforce	0.005	17%	0.020**	67%
	(0.004)		(0.008)	
Elementary education or less	-0.009***	-30%	-0.008	-27%
	(0.002)		(0.007)	
Secondary education	-0.001*	-3%	0.008	27%
	(0.000)		(0.007)	
Tertiary education	-0.003***	-10%	0.001	3%
	(0.001)		(0.003)	
Health problems	-0.010***	-33%	-0.009	-30%
	(0.002)		(0.007)	
Log equivalized household income	-0.031***	-103%	-0.005	-17%
	(0.011)		(0.087)	
Constant (group membership)			0.080	267%
			(0.093)	

Decomposes the life satisfaction gap into the "explained" (characteristics) portion and the "unexplained" (coefficients) portions. The explained column shows the amount of the gap (value and percentage of the overall gap) explained by men and women's different characteristic – a negative value means the gap increases. For example, women have lower education levels, so they should have lower life satisfaction than men. The unexplained column shows the amount of the gap explained by how men and women differentially value those characteristics. For example, women get a smaller reduction from being unemployed than men do. The last line, group membership, is the remaining gap that is attributable only to gender.

Figure 1: Illustration of differential item functioning (DIF).



Adapted from a similar figure in Kapteyn et al. (2007).

Table 8: Benefit of the Doubt (Spearman) Rank Order Correlation Coefficient

	Vignette set A	**Vignette set B**
Median BDROCC	0.829	0.543
Mean BDROCC	0.650	0.463
% perfect	13.9%	5.4%
% near perfect	58.5%	32.8%

Near perfect means above a correlation value that allows at most one double-rank inversion (includes one single-rank inversion and two single-rank inversions).

Table 9: Vignette equivalence testing: fffect of gender on vignette ratings

	Both vignette sets				**Vignette set A only**			
Variables included	**Female & age only**		**All**		**Female & age only**		**All**	
	Coef.	**Std. err.**	**Coef.**	**Std. err.**	**Coef.**	**Std. err.**	**Coef.**	**Std. err.**
Vignette A1	-	-	-	-	-	-	-	-
Vignette A2	0.003	(0.013)	0.012	(0.016)	0.002	(0.015)	0.014	(0.017)
Vignette A3	0.021	(0.013)	0.005	(0.015)	0.024	(0.014)	0.006	(0.016)
Vignette A4	-0.001	(0.012)	-0.001	(0.014)	-0.002	(0.012)	-0.001	(0.015)
Vignette A5	0.004	(0.013)	0.000	(0.014)	0.005	(0.014)	0.001	(0.015)
Vignette A6	0.003	(0.015)	0.021	(0.017)	0.002	(0.016)	0.024	(0.018)
Vignette B1	0.002	(0.014)	0.009	(0.015)	-	-	-	-
Vignette B2	0.038**	(0.017)	0.031	(0.019)	-	-	-	-
Vignette B3	0.026*	(0.015)	0.006	(0.017)	-	-	-	-
Vignette B4	-0.032**	(0.015)	-0.043**	(0.017)	-	-	-	-
Vignette B5	-0.045***	(0.015)	-0.064***	(0.016)	-	-	-	-
Vignette B6	-0.014	(0.014)	-0.033**	(0.017)	-	-	-	-

Table 10: HOPIT regression

	Life satisfaction equation		**τ^1 equation**	
	(1)	(2)	(1)	(2)
Female	-0.019	0.043***	-0.024**	-0.032**
	(0.015)	(0.015)	(0.012)	(0.013)
Age	-0.006***	-0.012***	0.006***	0.005***
	(0.002)	(0.002)	(0.001)	(0.001)
Age^2	0.000	0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Urban		0.057***		0.020
		(0.020)		(0.020)
Married		0.103***		0.031
		(0.025)		(0.022)
Separated		-0.124***		0.035
		(0.042)		(0.054)
Divorced		-0.158***		0.062*
		(0.042)		(0.035)
Widowed		-0.107***		0.041
		(0.037)		(0.031)
Domestic partner		-0.015		0.046
		(0.039)		(0.031)
Employed full time for self		-0.001		0.006
		(0.022)		(0.024)
Employed part time do not want full time		0.017		-0.034
		(0.028)		(0.027)
Unemployed		-0.236***		0.014
		(0.032)		(0.024)
Employed part time want full time		-0.113***		-0.007
		(0.026)		(0.030)
Out of workforce		-0.014		-0.011
		(0.023)		(0.017)
Secondary education		0.158***		-0.042**
		(0.021)		(0.017)
Tertiary education		0.376***		-0.088***
		(0.034)		(0.025)
Health problems		-0.238***		0.075***
		(0.029)		(0.018)
Log equivalized household income		0.284***		0.015
		(0.017)		(0.012)
Observations	45,731	45,731	45,731	45,731

HOPIT regressions of life satisfaction on listed variables as well as country fixed effects (103 countries included). The left two columns show the coefficients on the listed variables in the life satisfaction equation, the right two columns show the coefficients on the listed variables in the equation for τ^1, the threshold between the lowest response category and the second-lowest. See appendix for the other threshold equations.

Table 11: HOPIT vs. ordered probit marginal effects

		HOPIT (vignette-adjusted)			Ordered probit (unadjusted)		
Characteristic	**Comparison**	**LS**	**Change**	**x(gen. eff.)**	**LS**	**Change**	**x(gen. eff.)**
Income at 10th percentile	-	2.706	-	-	2.746	-	-
Income at 20th percentile	10th pctile	2.799	0.093	2.54	2.826	0.080	1.05
Income at 30th percentile	20th pctile	2.862	0.064	1.75	2.881	0.055	0.72
Income at 40th percentile	30th pctile	2.917	0.054	1.48	2.927	0.047	0.61
Income at 50th percentile	40th pctile	2.967	0.051	1.39	2.970	0.043	0.57
Income at 60th percentile	50th pctile	3.015	0.048	1.31	3.011	0.041	0.54
Income at 70th percentile	60th pctile	3.068	0.053	1.45	3.056	0.045	0.59
Income at 80th percentile	70th pctile	3.131	0.063	1.73	3.110	0.053	0.70
Income at 90th percentile	80th pctile	3.218	0.087	2.39	3.183	0.074	0.97
Low education	-	2.847	-	-	2.862	-	-
Medium education	Low	2.981	0.134	3.67	2.992	0.129	1.70
High education	Low	3.166	0.320	8.76	3.127	0.264	3.47
No health problems	-	3.008	-	-	3.017	-	-
Health problems	No health probs	2.807	-0.201	-5.51	2.802	-0.215	-2.83
Single	-	2.926	-	-	2.960	-	-
Married	Single	3.013	0.087	2.38	3.003	0.043	0.57
Separated	Single	2.822	-0.104	-2.86	2.855	-0.104	-1.37
Divorced	Single	2.794	-0.132	-3.63	2.793	-0.167	-2.19
Widowed	Single	2.837	-0.090	-2.45	2.849	-0.111	-1.45
Domestic partner	Single	2.913	-0.013	-0.35	2.923	-0.036	-0.48
Employed FT for employer	-	2.983	-	-	2.982	-	-
Employed FT for self	Employed FT for emp.	2.982	-0.001	-0.03	2.989	0.007	0.09
Employed PT, don't want FT	Employed FT for emp.	2.998	0.015	0.40	3.008	0.026	0.34
Unemployed	Employed FT for emp.	2.784	-0.199	-5.45	2.799	-0.183	-2.40
Employed PT, want FT	Employed FT for emp.	2.888	-0.095	-2.60	2.912	-0.070	-0.92
Out of the workforce	Employed FT for emp.	2.971	-0.012	-0.32	2.977	-0.005	-0.07
Rural	-	2.939	-	-	2.947	-	-
Urban	Rural	2.987	0.048	1.31	2.994	0.047	0.62
Male	-	2.939	-	-	2.925	-	-
Female	Male	2.976	0.036	1.00	3.001	0.076	1.00

Full sample (N = 45,731). Marginal effects of changing various characteristics, based on HOPIT regression coefficients in Table 10 column 2 and on ordered probit coefficients in Table 2 column 2. Income percentiles are from respondent's own country. "LS" shows estimated life satisfaction with that characteristic, all other characteristics held constant. "x(gen. eff.)" compares the magnitude of that marginal effect with the gender effect in the last row. Low education is elementary or less, medium education is secondary up to 3 years tertiary, high education is 4 years tertiary or more.

Table 12: Simulated life satisfaction with own scales and men's scales, global sample

	(1)	(2)
Own response thresholds		
Female average	2.972	2.972
Male average	2.945	2.945
Female - male gap	0.027***	0.027***
Female Pr(Life sat. = 1)	9.0%	9.1%
Female Pr(Life sat. = 2)	24.7%	24.4%
Female Pr(Life sat. = 3)	34.9%	35.0%
Female Pr(Life sat. = 4)	22.8%	23.3%
Female Pr(Life sat. = 5)	8.6%	8.2%
Male Pr(Life sat. = 1)	9.2%	9.3%
Male Pr(Life sat. = 2)	25.5%	25.2%
Male Pr(Life sat. = 3)	34.6%	34.8%
Male Pr(Life sat. = 4)	22.7%	23.2%
Male Pr(Life sat. = 5)	7.9%	7.5%
Men's response thresholds		
Female average	2.933	2.931
Male average	2.945	2.945
Female - male gap	-0.012***	-0.013***
Female Pr(Life sat. = 1)	9.3%	9.6%
Female Pr(Life sat. = 2)	25.7%	25.3%
Female Pr(Life sat. = 3)	35.0%	34.9%
Female Pr(Life sat. = 4)	22.5%	22.9%
Female Pr(Life sat. = 5)	7.6%	7.4%
% of gap explained	**145%**	**149%**

Simulated based on coefficients in a global HOPIT model. Specifications match the columns in Table 10. Significance stars indicate the results of t-tests of average life satisfaction by gender in each panel.

Table 13: Simulated life satisfaction with own thresholds and men's thresholds, by country

Country	$\mathbf{N_m}$	$\mathbf{N_f}$	$\mathbf{Gap_{own}}$	$\mathbf{Gap_m}$	**Avg. adj.**	**Marg. eff.**
Albania	189	229	0.181***	-0.033	-0.213	0.043
Argentina	131	281	-0.128***	-0.131***	-0.003	-0.080
Armenia	152	251	0.069*	0.102**	0.033	0.111
Australia	194	224	0.104***	0.054**	-0.050	0.042
Austria	222	207	-0.047*	-0.008	0.040	0.036
Azerbaijan	196	168	0.086**	-0.013	-0.099	0.096
Bahrain	339	175	-0.065***	-0.171***	-0.106	-0.160
Bangladesh	222	233	0.169***	0.157***	-0.012	0.243
Belarus	120	267	-0.329***	-0.308***	0.020	-0.150
Benin	217	167	0.160***	0.051*	-0.109	0.132
Bolivia	184	267	-0.020	-0.070***	-0.051	0.069
Bosnia & Herzegovina	169	197	0.003	-0.049	-0.052	0.057
Botswana	175	184	-0.209***	-0.275***	-0.067	-0.074
Brazil	176	299	-0.082***	-0.089***	-0.007	-0.038
Bulgaria	146	280	0.070**	0.119***	0.049	0.092
Cambodia	150	329	-0.218***	-0.191***	0.027	-0.101
Cameroon	247	173	0.150***	-0.019	-0.168	0.024
Canada	213	192	0.146***	0.024	-0.122	0.063
Central Afr. Republic	209	260	-0.027	0.026	0.053	0.056
Chile	203	227	-0.114***	-0.191***	-0.077	-0.114
China	804	1141	-0.093***	-0.037**	0.056	0.011
Colombia	129	294	-0.166***	-0.213***	-0.048	-0.143
Costa Rica	153	186	0.208**	0.097***	-0.110	0.155
Croatia	181	162	-0.003***	0.001***	0.004	-0.142
Czech Republic	173	189	-0.098	-0.157	-0.059	-0.205
Dem. Rep. of the Congo	220	137	-0.062**	-0.090***	-0.028	0.032
Dominican Republic	163	281	0.030	-0.014	-0.044	-0.012
El Salvador	176	217	-0.060*	-0.044	0.015	-0.011
Ethiopia	146	340	0.072**	0.075**	0.003	0.069
France	157	206	-0.077**	-0.174***	-0.098	-0.065
Germany	335	345	-0.024	-0.210***	-0.186	-0.114
Ghana	91	109	-0.057	0.095*	0.153	0.218
Greece	164	184	0.133***	0.104**	-0.029	0.174
Guatemala	222	207	-0.198***	-0.191***	0.008	-0.053
Haiti	70	67	-0.179***	-0.275***	-0.095	-0.233
Honduras	216	246	0.093***	0.011	-0.082	0.037
Hungary	147	225	-0.168***	-0.294***	-0.127	-0.048
India	1182	1161	0.032**	-0.080***	-0.111	0.051
Indonesia	221	204	-0.030	0.024	0.054	0.021
Iran	203	194	0.246***	0.260***	0.015	0.161
Iraq	258	193	0.146***	0.145***	0.000	0.208
Israel	206	202	-0.258***	-0.344***	-0.085	-0.132

Italy	86	98	-0.082	-0.076	0.005	-0.050
Japan	173	168	0.040	-0.110***	-0.150	-0.069
Jordan	206	281	0.329***	0.083***	-0.247	0.226
Kazakhstan	164	212	-0.040*	0.067**	0.107	0.028
Kenya	208	242	0.046*	0.083***	0.038	0.215
Kuwait	312	197	0.320***	-0.005	-0.325	0.058
Kyrgyzstan	202	244	-0.036*	-0.021	0.014	-0.012
Laos	139	221	-0.109***	-0.116***	-0.006	-0.017
Liberia	82	76	0.007	0.093	0.086	0.015
Macedonia	213	212	0.154***	0.175***	0.020	0.198
Madagascar	168	330	-0.019	-0.018	0.002	-0.035
Malawi	157	335	-0.053*	-0.180***	-0.128	-0.176
Malaysia	279	198	0.197***	0.134***	-0.063	0.227
Mauritania	207	146	-0.163***	-0.192***	-0.029	-0.030
Mexico	202	207	-0.080**	-0.096***	-0.015	-0.078
Moldova	174	228	-0.080***	0.016	0.096	0.074
Mongolia	226	248	0.229***	0.131***	-0.098	0.161
Morocco	110	108	0.097*	0.107**	0.010	0.285
Myanmar	201	276	0.025	-0.143***	-0.168	-0.075
Namibia	130	186	0.109**	-0.007	-0.116	0.220
Nepal	170	232	0.097***	0.103***	0.005	0.040
New Zealand	166	242	-0.041	-0.130***	-0.090	-0.068
Nicaragua	231	184	-0.161***	-0.155***	0.006	-0.081
Nigeria	223	141	0.002	0.038	0.036	0.127
Pakistan	278	157	0.438***	0.707***	0.270	0.592
Palestine	123	334	0.331***	0.223***	-0.108	0.242
Panama	202	263	0.083**	0.119***	0.036	0.193
Paraguay	171	271	0.083**	0.058*	-0.025	0.123
Peru	197	255	-0.092***	-0.085***	0.007	-0.014
Philippines	188	276	0.219***	0.109***	-0.110	-0.015
Poland	148	193	0.018	-0.029	-0.047	0.091
Portugal	188	236	0.004	-0.191***	-0.195	-0.138
Russia	160	362	0.076**	0.026	-0.049	0.074
Rwanda	190	236	-0.043	-0.018	0.025	0.080
Saudi Arabia	319	140	0.355***	-0.050**	-0.405	-0.059
Slovakia	161	240	0.116**	0.049	-0.066	0.077
Slovenia	209	260	0.066*	-0.005	-0.071	0.084
South Africa	233	259	0.050*	0.118***	0.069	0.169
South Korea	180	223	-0.083**	-0.224***	-0.141	-0.097
Spain	177	233	0.062**	-0.051**	-0.113	0.057
Sri Lanka	191	278	0.144***	0.017	-0.127	-0.033
Sudan	202	119	0.176***	0.045	-0.130	0.144
Syria	219	241	-0.092***	-0.069***	0.023	-0.095
Taiwan	173	215	0.045	0.041	-0.004	0.178
Tajikistan	189	246	-0.046*	-0.060**	-0.014	0.037
Tanzania	210	280	-0.031	0.012	0.044	0.057

Thailand	160	330	0.166***	0.142***	-0.023	0.040
Turkey	183	271	-0.056**	-0.019	0.036	-0.038
Uganda	134	160	0.004	0.012	0.008	-0.006
United Arab Emirates	191	208	0.054***	-0.193***	-0.247	-0.238
United Kingdom	171	206	0.123***	0.077**	-0.046	0.085
United States	219	205	0.205***	0.116***	-0.089	0.224
Uruguay	135	239	0.085***	0.074**	-0.011	0.070
Uzbekistan	183	242	0.027	-0.088***	-0.115	0.063
Venezuela	123	219	0.157***	0.160***	0.003	0.250
Vietnam	122	165	-0.026	-0.034	-0.008	0.042
Yemen	214	191	0.212***	-0.276***	-0.488	-0.292
Zambia	205	242	0.039	0.093**	0.053	0.186
Zimbabwe	183	239	0.127***	0.143***	0.016	0.265
Overall	**20,431**	**24,346**	**0.020*** **	**-0.026*** **	**-0.046**	**0.038**

Simulated based on coefficients in HOPIT models run separately for each country, with the same specification as in Table 10. N_m and N_f are observation counts for men and women respectively. Gap_{own} is the gender life satisfaction gap when men and women use their own thresholds, Gap_m is the gender life satisfaction gap when both men and women use men's thresholds. Avg. Adj. is the average adjustment in women's life satisfaction when women in that country move from their own thresholds to men's thresholds. Marg. eff. is the marginal gender effect in that country.

Table 14: Gap_{own} and Gap_{m} by region

	N	**Gap_{own}**	**Gap_{m}**	**Avg. adj.**
Latin America & the Caribbean	4,791	-0.027**	-0.053***	-0.027***
East Asia	3,551	-0.026**	-0.035**	-0.009***
Sub-Saharan Africa	8,198	-0.013*	-0.023**	-0.009***
Western Europe	3,215	0.000	-0.095***	-0.095***
Transition countries	7,683	0.000	-0.029***	-0.029***
Southeast Asia	3,459	0.005	-0.035**	-0.040***
Central America	2,503	0.018	-0.001	-0.019***
Middle East & North Africa	5,618	0.022	-0.120***	-0.142***
South Asia	4,104	0.080***	0.020*	-0.060***
Australia/New Zealand/Canada/USA	1,655	0.108***	0.022	-0.086***

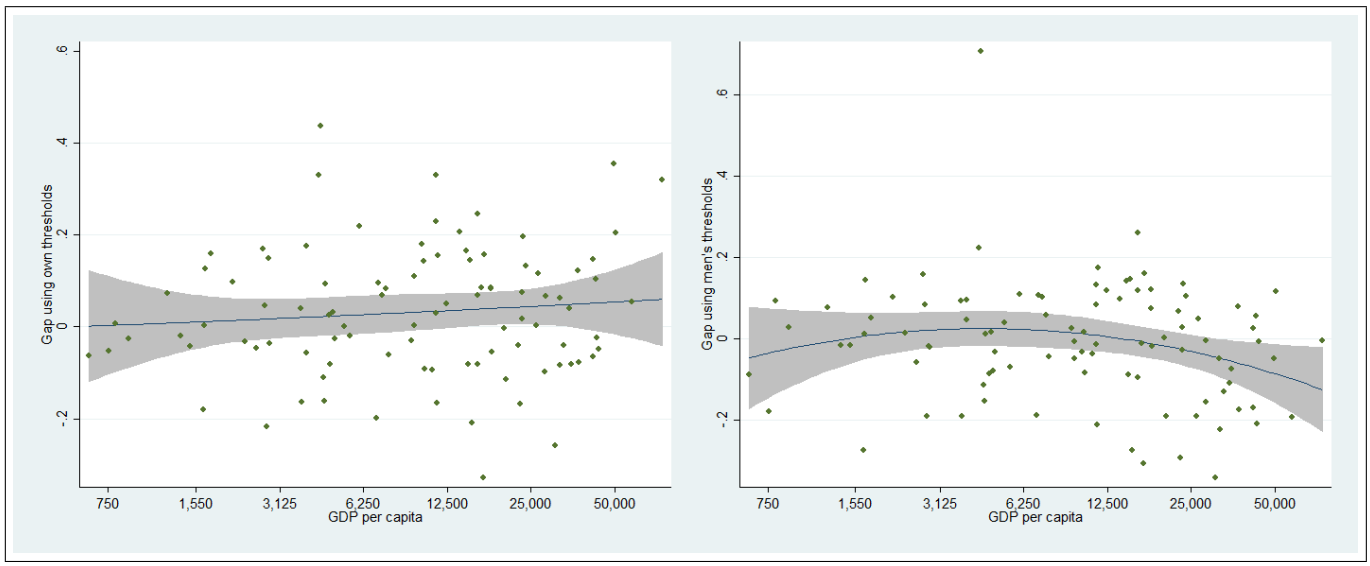
Significance stars indicate the results of t-tests of Gap_{own} and Gap_{m} by gender.

Table 15: Gap_{own} Gap_{m}, and GDP per capita

	(1)	(2)
Log GDP per capita	0.003	0.322*
	(0.186)	(0.191)
Log GDP per capita squared	0.001	-0.019*
	(0.010)	(0.011)
Constant	-0.039	-1.328
	(0.821)	(0.844)
Observations	96	96
R-squared	0.010	0.055

OLS regressions of Gap_{own} (column 1) and Gap_{m} (column 2).

Figure 2: Relationship between life satisfaction gender gaps and GDP per capita



Quadratic regressions of Gap_{own} (left) and Gap_{m} on log GDP per capita. See Table 15.

Table 16: Gap_{own} and Gap_{m} by various measures of gender equality

		$\textbf{Gap}_{\textbf{own}}$	$\textbf{Gap}_{\textbf{m}}$
	N	(1)	(2)
Percent of parliament made up of women	99	-0.002*	0.000
		(0.001)	(0.001)
SIGI: Disciminatory family code	99	0.155***	0.108*
		(0.056)	(0.06)
SIGI: Restricted physical integrity	79	0.041	0.015
		(0.07)	(0.076)
SIGI: Son bias	82	0.211***	0.211***
		(0.057)	(0.064)
SIGI: Restricted resources and assets	100	0.109**	0.113**
		(0.051)	(0.054)
SIGI: Restricted civil liberties	100	0.070	-0.024
		(0.054)	(0.058)
SIGI: Overall	72	0.260**	0.123
		(0.111)	(0.124)

Linear regressions of Gap_{own} Gap_{m} on measures of gender equality; see Section 6.1.

Table 17: Gap_{own} and Gap_{m} by respondent's religion

	N	$\textbf{Gap}_{\textbf{own}}$	$\textbf{Gap}_{\textbf{m}}$	**Avg. adj.**
Judaism	373	-0.272***	-0.355***	-0.082
Blank	2,927	-0.142***	-0.113***	0.029
Buddhism	2,877	0.022	-0.038**	-0.060
Christianity	21,755	0.023***	-0.008	-0.030
Hinduism	2,433	0.030*	-0.057***	-0.087
Other	2,901	0.040**	-0.033**	-0.072
Islam	10,849	0.067***	-0.014*	-0.082
Secular	662	0.105**	0.070*	-0.035

Significance stars indicate the results of t-tests of Gap_{own} and Gap_{m} by gender.

Table 18: Scale adjustment by country's main religion

	N	**Gap_{own}**	**Gap_{m}**	**Avg. adj.**
Judaism	408	-0.258***	-0.344***	-0.085
Unaffiliated	2,648	-0.103***	-0.085***	0.018
Christianity	24,645	0.008	-0.024***	-0.032
Hinduism	2,745	0.028*	-0.064***	-0.092
NoMain	1,442	0.045**	0.006	-0.038
Islam	10,140	0.071***	-0.014*	-0.086
Buddhism	2,749	0.078***	0.015	-0.063

A religion in a country is the "main" religion if more than 50 % of the country identifies with that religion. Significance stars indicate the results of t-tests of Gap_{own} and Gap_{m} by gender.

Appendix

Table A1: Benefit of the doubt Spearman rank order correlation coefficients (BDROCCs), by country, for each set of vignettes. See section 5.3.1.

Table A2: Coefficients in the remaining threshold equations not reporting in Table 10.

Tables A3 and A4: Replicate Table 10 (HOPIT model coefficients) and 12 (simulated life satisfaction with own and men's scales) with all non-representative samples removed, as discussed in the Data section.

Table A1: Benefit of the doubt Spearman rank order correlation coefficients

A vignette set				B vignette set			
Country	**Median BDROCC**	**% perfect**	**% near perfect**	**Country**	**Median BDROCC**	**% perfect**	**% near perfect**
Chad	-0.371	8.5%	29.4%	Palestinian Terr.	-0.086	0.4%	2.7%
Czech Republic	0.257	15.4%	66.6%	UAE	0.029	0.2%	7.7%
Austria	0.429	20.7%	79.1%	Japan	0.086	0.8%	8.4%
Cameroon	0.486	14.5%	49.8%	Chad	0.314	4.3%	23.1%
Japan	0.543	6.6%	61.2%	Syria	0.314	1.8%	14.6%
Vietnam	0.543	7.2%	32.6%	Bangladesh	0.371	2.8%	21.0%
Jordan	0.543	13.9%	46.5%	El Salvador	0.371	4.3%	24.1%
Colombia	0.543	21.6%	80.7%	Kuwait	0.371	4.5%	22.3%
New Zealand	0.600	10.1%	68.2%	Liberia	0.371	6.3%	26.3%
Central Afr. Rep.	0.600	9.6%	43.3%	South Korea	0.371	3.2%	24.2%
Myanmar	0.600	24.7%	75.3%	Vietnam	0.371	6.0%	23.9%
Liberia	0.600	11.7%	39.0%	Afghanistan	0.429	1.3%	17.9%
Bolivia	0.600	11.6%	67.5%	Bahrain	0.429	3.9%	24.4%
India	0.600	19.4%	50.4%	Canada	0.429	4.3%	23.0%
Mongolia	0.600	10.9%	59.7%	Central Afr. Rep.	0.429	1.8%	27.9%
Nicaragua	0.600	10.3%	46.8%	Iraq	0.429	4.7%	23.7%
Germany	0.657	1.8%	69.5%	Israel	0.429	2.0%	20.5%
Bangladesh	0.657	7.9%	39.7%	Mexico	0.429	5.7%	28.3%
Brazil	0.657	12.2%	59.8%	Morocco	0.429	4.0%	24.0%
Greece	0.657	22.8%	76.8%	Saudi Arabia	0.429	5.5%	25.1%
Palestinian Terr.	0.657	2.5%	10.4%	Spain	0.429	3.7%	26.2%
Honduras	0.657	10.8%	44.0%	Thailand	0.429	3.4%	24.2%
France	0.657	21.1%	79.1%	United Kingdom	0.429	3.1%	24.6%
Indonesia	0.657	16.0%	65.9%	Yemen	0.429	3.0%	24.1%
Thailand	0.714	11.2%	57.3%	Uganda	0.457	4.3%	24.7%
Malawi	0.714	19.0%	60.0%	Australia	0.486	3.1%	25.1%
Uganda	0.714	9.5%	43.5%	Germany	0.486	1.9%	23.2%
Cambodia	0.714	12.6%	54.2%	Ghana	0.486	3.5%	28.8%
Croatia	0.714	15.9%	68.0%	Haiti	0.486	7.4%	29.7%
UAE	0.771	1.0%	7.2%	Kenya	0.486	5.0%	31.0%
Australia	0.771	9.6%	69.3%	Macedonia	0.486	5.0%	28.0%
United Kingdom	0.771	9.8%	72.3%	Panama	0.486	5.4%	27.8%
United States	0.771	15.0%	79.1%	United States	0.486	4.5%	27.8%
Italy	0.771	17.4%	79.8%	Zambia	0.486	3.2%	26.5%
Belarus	0.771	19.8%	71.0%	Armenia	0.543	2.9%	26.4%
Kenya	0.771	11.8%	40.4%	Botswana	0.543	3.2%	29.6%
Iran	0.771	14.8%	61.9%	Brazil	0.543	3.7%	30.0%
Uzbekistan	0.771	14.1%	43.4%	Czech Republic	0.543	2.7%	32.1%
Ethiopia	0.771	14.5%	60.8%	Dominican Republic	0.543	5.4%	32.6%
Malaysia	0.771	14.1%	64.0%	Guatemala	0.543	5.1%	35.8%

Kuwait	0.771	8.1%	51.7%	Honduras	0.543	5.9%	30.8%
Afghanistan	0.771	11.0%	47.2%	Iran	0.543	5.5%	30.0%
Mexico	0.771	7.3%	41.3%	Jordan	0.543	4.0%	26.3%
Saudi Arabia	0.771	13.1%	52.2%	Mauritania	0.543	6.5%	33.3%
Mauritania	0.771	11.8%	40.7%	Nepal	0.543	2.2%	27.4%
Slovakia	0.771	16.4%	70.8%	New Zealand	0.543	2.9%	32.6%
Congo (Kinshasa)	0.771	22.9%	61.0%	Nigeria	0.543	2.6%	27.8%
Kyrgyzstan	0.771	17.6%	66.1%	Pakistan	0.543	2.9%	26.3%
Madagascar	0.771	13.8%	56.3%	Philippines	0.543	8.2%	38.6%
Canada	0.829	16.7%	75.1%	Portugal	0.543	7.2%	34.2%
Spain	0.829	16.6%	79.9%	Senegal	0.543	6.8%	31.0%
Russia	0.829	12.8%	58.8%	Slovenia	0.543	5.1%	34.5%
Syria	0.829	2.2%	20.2%	Sri Lanka	0.543	6.8%	31.9%
Yemen	0.829	13.1%	51.8%	Taiwan	0.543	7.1%	34.4%
Panama	0.829	6.6%	38.4%	Tanzania	0.543	6.2%	33.7%
Kazakhstan	0.829	13.5%	55.9%	Uzbekistan	0.543	5.0%	33.3%
Zimbabwe	0.829	19.6%	72.9%	Venezuela	0.543	6.7%	33.7%
Azerbaijan	0.829	22.1%	62.3%	Albania	0.600	6.6%	36.2%
Bulgaria	0.829	25.0%	81.4%	Argentina	0.600	4.9%	37.4%
South Africa	0.829	19.8%	62.2%	Bolivia	0.600	6.3%	38.0%
Iraq	0.829	11.4%	47.5%	Cambodia	0.600	3.7%	32.2%
Israel	0.829	2.9%	37.8%	Cameroon	0.600	5.1%	36.8%
Morocco	0.829	6.8%	37.1%	Costa Rica	0.600	6.2%	35.0%
Haiti	0.829	13.2%	35.1%	Ethiopia	0.600	5.8%	32.7%
Armenia	0.829	15.8%	54.4%	France	0.600	6.3%	34.8%
Botswana	0.829	9.9%	46.5%	India	0.600	9.3%	36.9%
Dominican Republic	0.829	14.3%	65.4%	Indonesia	0.600	8.0%	35.5%
Pakistan	0.829	6.7%	40.9%	Italy	0.600	6.7%	35.8%
Philippines	0.829	14.3%	60.1%	Kazakhstan	0.600	3.4%	32.8%
Portugal	0.829	13.2%	74.6%	Malawi	0.600	5.4%	36.4%
Senegal	0.829	11.2%	49.0%	Moldova	0.600	6.0%	35.0%
Sri Lanka	0.829	10.4%	54.2%	Mongolia	0.600	5.5%	40.0%
Albania	0.829	20.2%	72.5%	Namibia	0.600	6.5%	36.0%
Costa Rica	0.829	21.3%	72.0%	Nicaragua	0.600	3.1%	36.1%
Namibia	0.829	10.1%	51.3%	Poland	0.600	7.9%	36.9%
Poland	0.829	16.1%	73.7%	Rwanda	0.600	9.0%	40.4%
Rwanda	0.829	15.8%	40.9%	Slovakia	0.600	5.0%	35.3%
Sudan	0.829	15.2%	62.0%	Sudan	0.600	3.8%	30.4%
Tajikistan	0.829	10.2%	34.9%	Tajikistan	0.600	6.5%	35.6%
Uruguay	0.829	15.5%	70.9%	Uruguay	0.600	5.7%	37.6%
Benin	0.829	22.9%	63.7%	Zimbabwe	0.600	1.4%	33.9%
Bosnia Herzegovina	0.829	18.8%	75.9%	Benin	0.629	6.1%	38.0%
Paraguay	0.829	19.0%	60.0%	Austria	0.657	7.9%	40.3%
Macedonia	0.857	10.2%	53.2%	Azerbaijan	0.657	6.8%	35.5%
South Korea	0.886	5.9%	48.9%	Belarus	0.657	10.9%	43.0%
Slovenia	0.886	8.8%	73.0%	Bosnia Herzegovina	0.657	4.4%	42.3%

Bahrain	0.886	9.3%	54.2%	Bulgaria	0.657	6.2%	45.9%
Hungary	0.886	20.6%	74.5%	China	0.657	7.2%	39.0%
El Salvador	0.886	8.9%	38.1%	Colombia	0.657	7.3%	39.9%
Ghana	0.886	13.8%	51.1%	Congo (Kinshasa)	0.657	6.4%	40.7%
Zambia	0.886	11.0%	49.3%	Greece	0.657	4.9%	35.9%
Taiwan	0.886	14.8%	73.8%	Kyrgyzstan	0.657	5.7%	43.6%
China	0.886	19.5%	60.3%	Madagascar	0.657	4.4%	37.3%
Nepal	0.886	13.2%	54.6%	Malaysia	0.657	8.7%	37.8%
Nigeria	0.886	12.0%	52.1%	Myanmar	0.657	8.0%	44.3%
Tanzania	0.886	15.6%	49.5%	Paraguay	0.657	10.3%	41.6%
Venezuela	0.886	14.4%	62.3%	Peru	0.657	6.8%	44.5%
Argentina	0.886	16.1%	73.9%	Russia	0.657	6.8%	35.3%
Moldova	0.886	11.7%	57.3%	South Africa	0.657	4.6%	43.0%
Peru	0.886	18.1%	81.7%	Turkey	0.657	8.7%	42.9%
Turkey	0.886	12.4%	53.4%	Chile	0.714	6.4%	45.0%
Laos	0.886	21.3%	74.8%	Croatia	0.714	7.0%	46.9%
Guatemala	0.943	12.5%	60.1%	Hungary	0.714	6.8%	46.6%
Chile	0.943	20.4%	74.3%	Laos	0.771	8.3%	51.1%

Near perfect means above a correlation value that allows at most one double-rank inversion (includes one single-rank inversion and two single-rank inversions).

Table A2: Coefficients in equations for thresholds past τ^1 in HOPIT specifications

	$ln(\tau^2 - \tau^1)$		$ln(\tau^3 - \tau^2)$		$ln(\tau^4 - \tau^3)$	
	(1)	(2)	(1)	(2)	(1)	(2)
Female	-0.016**	-0.014	-0.003	0.002	-0.032***	-0.022**
	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.009)
Age	-0.000	-0.000	0.000	0.000	-0.002	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age^2	-0.000	0.000	-0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Urban		-0.017		-0.005		0.013
		(0.017)		(0.011)		(0.014)
Married		0.015		0.012		-0.006
		(0.012)		(0.012)		(0.012)
Separated		-0.022		-0.007		-0.053
		(0.032)		(0.023)		(0.034)
Divorced		0.008		-0.028		-0.065***
		(0.023)		(0.018)		(0.023)
Widowed		-0.016		-0.002		-0.019
		(0.021)		(0.018)		(0.022)
Domestic partner		0.004		-0.011		-0.045*
		(0.019)		(0.024)		(0.023)
Employed full time for self		0.000		-0.022*		-0.010
		(0.015)		(0.012)		(0.017)
Employed part time do not want full time		0.029		-0.019		0.017
		(0.020)		(0.015)		(0.018)
Unemployed		-0.016		-0.013		-0.031
		(0.021)		(0.016)		(0.021)
Employed part time want full time		0.002		-0.026*		-0.030
		(0.019)		(0.015)		(0.021)
Out of workforce		0.017		-0.009		0.008
		(0.017)		(0.009)		(0.012)
Secondary education		0.014		0.046***		0.072***
		(0.012)		(0.010)		(0.015)
Tertiary education		0.075***		0.056***		0.165***
		(0.017)		(0.014)		(0.023)
Health problems		-0.045***		-0.028***		-0.035***
		(0.010)		(0.011)		(0.010)
Log equivalized household income		0.007		0.003		0.054***
		(0.006)		(0.006)		(0.008)

Effects of individual characteristics on the remaining thresholds. Note that while the coefficients in the right columns of Table 10 are for the first threshold, τ^1, the coefficients here are for $ln(\tau^j - \tau^{j-1})$. See Section 5.

Table A3: HOPIT regression with non-representative sample removed, all equations

	Life sat.		τ^1		$ln(\tau^2-\tau^1)$		$ln(\tau^3-\tau^2)$		$ln(\tau^4-\tau^3)$	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Female	0.047***	-0.025	-0.020	-0.015	-0.022**	-0.026***	-0.000	-0.005	-0.025**	-0.034***
	(0.017)	(0.016)	(0.015)	(0.014)	(0.010)	(0.010)	(0.009)	(0.009)	(0.010)	(0.010)
Age	-0.013***	-0.006***	0.005***	0.008***	-0.001	-0.001	-0.000	-0.000	-0.001	-0.002
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age^2	0.000***	-0.000	-0.000***	-0.000***	0.000	0.000	0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Urban	0.033		0.013		-0.003		-0.016		0.008	
	(0.025)		(0.025)		(0.016)		(0.013)		(0.013)	
Married	0.121***		0.034		0.002		0.020		-0.007	
	(0.022)		(0.023)		(0.013)		(0.014)		(0.014)	
Separated	-0.085**		0.050		-0.032		-0.009		-0.049	
	(0.037)		(0.060)		(0.036)		(0.025)		(0.036)	
Divorced	-0.158***		0.058		-0.017		-0.016		-0.059**	
	(0.046)		(0.040)		(0.025)		(0.020)		(0.025)	
Widowed	-0.076*		0.035		-0.019		0.020		-0.009	
	(0.039)		(0.031)		(0.024)		(0.020)		(0.021)	
Domestic partner	-0.019		0.031		0.008		-0.009		-0.039	
	(0.040)		(0.032)		(0.022)		(0.027)		(0.025)	
Employed full time for self	-0.032		0.024		-0.026		-0.021		-0.002	
	(0.026)		(0.029)		(0.020)		(0.014)		(0.020)	
Employed part time do not want full time	0.017		-0.004		0.004		-0.032*		0.032*	
	(0.030)		(0.031)		(0.018)		(0.017)		(0.017)	
Unemployed	-0.255***		0.025		-0.034*		-0.021		-0.016	
	(0.038)		(0.028)		(0.019)		(0.018)		(0.023)	
Employed part time want full time	-0.156***		0.005		-0.013		-0.031		-0.011	
	(0.029)		(0.031)		(0.017)		(0.019)		(0.022)	
Out of workforce	-0.046*		-0.012		0.000		-0.015		0.017	
	(0.025)		(0.022)		(0.013)		(0.011)		(0.014)	
Secondary education	0.140***		-0.051**		0.016		0.042***		0.080***	
	(0.024)		(0.021)		(0.011)		(0.013)		(0.018)	
Tertiary education	0.359***		-0.106***		0.084***		0.046***		0.178***	
	(0.036)		(0.028)		(0.018)		(0.017)		(0.024)	
Health problems	-0.257***		0.084***		-0.046***		-0.033**		-0.035***	
	(0.027)		(0.021)		(0.012)		(0.014)		(0.012)	
Log equivalized household income	0.300***		0.027**		0.008		0.001		0.053***	
	(0.014)		(0.013)		(0.008)		(0.008)		(0.008)	
Observations	33,991	33,991	33,991	33,991	33,991	33,991	33,991	33,991	33,991	33,991

Duplication of Table 10 with non-representative samples removed. The left two columns show the coefficients on the listed variables in the life satisfaction equation, the remaining columns show the coefficients on the listed variables in the threshold equations. The first of these is τ^1; the following are for $ln(\tau^j - \tau^{j-1})$. See Section 5

Table A4: Simulated life satisfaction with own scales and men's scales, non-representative samples removed

	(1)	(2)
Own response thresholds		
Female average	3.008	3.007
Male average	2.968	2.968
Female - male gap	0.04***	0.039***
Female Pr(Life sat. = 1)	8.8%	9.0%
Female Pr(Life sat. = 2)	23.6%	23.1%
Female Pr(Life sat. = 3)	34.7%	34.8%
Female Pr(Life sat. = 4)	23.8%	24.5%
Female Pr(Life sat. =5)	9.1%	8.7%
Male Pr(Life sat. = 1)	9.1%	9.1%
Male Pr(Life sat. = 2)	24.8%	24.4%
Male Pr(Life sat. = 3)	34.4%	34.6%
Male Pr(Life sat. = 4)	23.6%	24.1%
Male Pr(Life sat. = 5)	8.1%	7.7%
Men's response thresholds		
Female average	2.967	2.970
Male average	2.968	2.968
Female - male gap	-0.001	0.002
Female Pr(Life sat. = 1)	9.0%	9.2%
Female Pr(Life sat. = 2)	24.7%	24.1%
Female Pr(Life sat. = 3)	34.8%	34.8%
Female Pr(Life sat. = 4)	23.6%	24.1%
Female Pr(Life sat. = 5)	7.9%	7.7%
% of gap explained	**103%**	**96%**

Duplication of Table 12 with all non-representative samples removed, as discussed in the Data section. Significance stars indicate the results of t-tests of average life satisfaction by gender in each panel.