

The Sad Truth About Happiness Scales*

Timothy N. Bond[†] Kevin Lang[‡]

December 15, 2014

Abstract

Nearly all happiness research assumes that happiness can be cardinalized to be distributed normally across different groups. We show that whenever this assumption is true, it is impossible to rank groups by average happiness. The CDFs will (almost) always cross when estimated using large samples. There are an infinite number of equally correct alternative cardinalizations that reverse this ranking. We provide several examples and a formal proof. Whether Moving-to-Opportunity increases happiness, men have become happier relative to women, and an Easterlin paradox exists depends on whether happiness is distributed normally or log-normally. We discuss restrictions that may permit such comparisons.

JEL Codes: D60, I31, O15

*We are grateful to Larry Katz, Jeff Liebman, Jens Ludwig, Andy Oswald, Justin Wolfers and participants in seminars at Boston University, the Federal Reserve Bank of New York, Georgetown University, the University of Michigan and the University of Waterloo, and an informal brownbag lunch at Purdue University for their helpful feedback and comments. The usual caveat applies, perhaps more strongly than in most cases.

[†]Department of Economics, Krannert School of Management, Purdue University, 100 S. Grant St., West Lafayette, IN 47907; email: tnbond@purdue.edu

[‡]Department of Economics, Boston University, 270 Bay State Road, Boston, MA, 02215; email: lang@bu.edu

1 Introduction

There is an extensive literature that relies on questions in which individuals are asked to report their happiness in a few ordered categories such as “very happy,” “pretty happy” or “not too happy.” We argue that with such scales it is essentially *never* possible to rank the overall happiness of two groups without auxiliary assumptions which must be made explicit and about which there has been no discussion on which to base a consensus. Yet, without such assumptions, it is impossible to use such data to make scientifically valid statements of the form “people in country A are, on average, happier than people in country B” or that “married men are happier than single men.” As a consequence, despite the obvious weaknesses of standard economic measures, it is premature to rely heavily on measures of subjective well-being to guide policy.

Our argument is simple. People perceive happiness as continuous, rather than discrete.¹ Thus, when asked to place their happiness into a small number of categories, they are placing their happiness in a range. For example, they describe themselves as “very happy” if their happiness exceeds some critical internal value. Oswald (2008) refers to this as the reporting function.

Suppose we have a scale with three categories (two cutoffs). Assuming all individuals use the same reporting function, we can, without apparent loss of generality, normalize the cutoffs to be 0 and 1. Given some belief about the full underlying distribution, such as that it is logistic or normal, we can estimate two parameters (e.g. the mean and variance) of the distribution from the distribution of the responses across categories.

Since we can calculate the mean, it might appear that we can compare average happiness. But, happiness is ordinal. This difference in mean is valid for just one of the infinite number of ways one can cardinalize happiness. Only if one group’s underlying happiness distribution stochastically dominates the other’s will each of these cardinalizations produce the same

¹More precisely, there are an infinite number of strictly ranked states of happiness an individual can be in.

ordering of the means. However, under conditions made precise later, we can establish first-order stochastic dominance of the underlying distributions only when the estimated variances are identical, which is an essentially zero-probability event. Moreover, even if our estimates of the variances are identical, since both are merely estimates subject to error, our posterior that they are identical must still be 0. All statements of ordering between the two groups are thus dependent on the researcher’s cardinalization of happiness, any of which is arbitrary and equally defensible from the data.²

Given that happiness research has been the frequent, perhaps even unfairly so, target of criticism, it is important to be clear what our critique is *not*. We do not argue that different individuals report their happiness in different ways. If, for example, women report on a survey that they are “very happy” when they are “quite happy” while men report being “very happy” only when they are “ecstatically happy,” there is, of course, no way for us to use survey data to compare the average happiness of men and women.³ We will show that even if *all* individuals report their happiness on a survey *in the same way*, one still cannot make a valid comparison of mean happiness.

We likewise do not argue that self-reported happiness is a poor measure of well-being.⁴ For example, Kahneman and Krueger (2006) attempt to show the validity of self-reported happiness by showing it is correlated with a large number of other objective and subjective indicators of well-being, while Krueger and Schadke (2008) find subjective well-being measures have a sufficiently high test-retest reliability, similar to the way test manufacturers

²It will be clear throughout that many of our notions of happiness are similar to the economic concept of utility. There is a literature (e.g. Frey and Stutzer, 2003) that seeks to distinguish between utility as measured by revealed preference and happiness as reported in surveys. We take no stance on this issue. Our point is merely that regardless of how utility or happiness is elicited, we cannot know more than the ranking of happiness.

³King et al. (2004) propose circumventing this issue through the use of ‘vignettes’ to anchor the scale on which people report their happiness. See also evaluations of this method by Peracchi and Rossetti (2013) and Ravallion et. al (2013).

⁴For a much more detailed and sophisticated discussion of measurement error in reported subjective well-being, see Bertrand and Mullainathan (2001). See also Olken (2009) for a discussion of perceived corruption, Waidmann, Bound and Schoenbaum (1995) on self-reported health and Bound, Brown and Mathiowetz (2001) for a broader discussion of measurement error in subjective survey measures.

assess the validity of standardized exams.⁵ Our argument holds *even if* surveys elicit *perfect* measures of happiness, so long as happiness itself is ordinal. That survey responses on happiness are correlated with positive outcomes is not evidence that they measure happiness on an interval scale. It is uncontroversial that test scores are ordinal measures of achievement, yet they have been frequently shown to be positively correlated with a range of outcomes.⁶

In fact it does not matter what happiness surveys measure for our criticism to be valid. Our critique is purely statistical and would hold equally for any variable that is measured on an ordinal scale with a small number of categories designed to capture a continuous underlying variable. While we could have focused on subjective pain at a doctor's office or cleanliness grades at restaurants, we choose happiness research because it appears to be the area where these variables are most widely used to make between-group comparisons.

In principle, the problems associated with ordinal scales can be solved if we are willing to anchor the happiness scale to some outcome measure. Bond and Lang (2014) develop interval measures of achievement by tying test scores to eventual completed education and to the associated expected wages.⁷ But as the parallel with their analysis of test scores shows, the conclusions we reach may depend on whether we relate the underlying happiness measure to the probability of committing suicide or some other outcome. Moreover it is not clear to us why in this case we would not prefer to measure the related outcomes directly. As discussed in section four, regardless of the concerns we raise about the measurement of happiness, the evidence is strong that Moving to Opportunity reduced symptoms of depression and improved other measures of psychological well-being.

The alternative approach is to place *ex ante* restrictions on plausible cardinalizations

⁵Interestingly, recent work by Boyd et. al. (2013) shows that conventional test-retest reliabilities dramatically understate the amount of measurement error in standardized tests.

⁶For discussion of the ordinality of test scores, see Stevens (1946), Thorndike (1966), Spencer (1983), and Bond and Lang (2013). Neal and Johnson (1996), Ritter and Taylor (2011), and Bond and Lang (2014), find positive correlations between test scores and wages, employment, and educational attainment, respectively. Chetty, Friedman and Rockoff (2014) find a positive correlation between adult outcomes and teacher value added, which is an ordinal measure of teacher quality.

⁷See also the work on human capital production by Cunha and Heckman (2008) and Cunha, Heckman, and Schennach (2010).

of happiness, which is equivalent to giving attention to only some population happiness distributions. This, too, raises difficulties. Our beliefs about what distributions are plausible are likely to depend on our beliefs about, among other things, the marginal utility of income. Yet, the relation between happiness and income is one of the key areas of debate in happiness research.

While our focus is on positive rather than normative economics, it is important to note that our results can be reinterpreted within the neo-utilitarian framework implicit in much happiness research. Fix the happiness distributions estimated using standard techniques; the welfare implications are (almost) never invariant to the choice of individualistic social welfare function.

However, it would be a mistake to interpret our main results as saying there is nothing researchers can learn from happiness surveys. As we will show, one cannot simultaneously argue both that Moving to Opportunity (MTO) decreased happiness and that economic growth in the United States has increased happiness, without also arguing that the skewness of the distribution of happiness amongst MTO recipients is dramatically different from and opposite to that of the United States as a whole.⁸ We are hopeful that researchers can come to a consensus about which of these distributions is more plausible.

In the next section, we present a series of simple examples. We show first that even if happiness is normally distributed, shifting respondents from “not too happy” to “pretty happy” can *lower* our estimate of average happiness. We then provide an example which appears to avoid this problem: the distribution of responses over the three categories is higher in the sense of first-order stochastic dominance, and estimated mean happiness for the group with more positive responses is higher. However, at one point in the happiness distribution, a substantial minority of the second group has higher happiness than the members in the first group. An alternative cardinalization that is a simple monotonic transformation of the happiness distribution fits the data equally well but reverses the comparison of mean

⁸Or, they could alternatively argue that we should measure the happiness effects of MTO on a different scale than we should measure the effects of GDP.

happiness. Finally, we further show that when mean happiness is estimated assuming happiness is normally distributed, a common implicit assumption, one of two simple exponential transformations can reverse *any* reported happiness gap.

In the third section we prove our main result: it is (almost) never possible to rank the mean happiness of two groups when the data are reported on a discrete ordinal scale. We apply this result, in section four, to three findings from the happiness literature: the effect of Moving to Opportunity on happiness (Ludwig et. al, 2013), the decline in the relative happiness of women despite the dramatic progress they have made economically and socially since the 1970s (Stevenson and Wolfers, 2009), and the Easterlin paradox (Easterlin, 1973). We also investigate the impact of different distributional assumptions on comparisons more generally, by looking at the rank order of mean happiness by country. In the fifth and final section we discuss whether it is possible to weaken our result. We conclude that we can do so only under (perhaps overly) strong assumptions although we hold out some hope for a consensus on plausible restrictions on the happiness distribution which would permit strong conclusions in some cases.

If we are willing to accept that happiness is strongly left-skewed (or in a neo-utilitarian framework adopt a highly inequality averse social welfare function), then economic growth increased average happiness (or social welfare), U.S. women became happier relative to men and the effects of MTO were even more positive than previously reported.

2 Some Simple Examples

Suppose we ask a large number of people belonging to two groups to assess their happiness on a 3-point scale, and they respond as shown in example 1.

Example 1

	Group A	Group B
Very happy	20	15
Pretty happy	25	30
Not too happy	55	55

The responses in group A are higher than those in group B in the sense of first-order stochastic dominance so that regardless of the values assigned to the three categories, two of which are in any event mere normalizations, group A will have higher average happiness than group B does.⁹ However, increasingly researchers recognize that the three categories capture a continuum. Therefore they are likely to estimate underlying happiness using ordered logit or probit. For a normal distribution of happiness with mean μ and standard deviation σ , textbook ordered probit will estimate μ/σ . Different computer packages use somewhat different normalizations to identify the model. We will use Stata which sets the constant term equal to zero and the variance to 1. Stata informs us that group B is .07 standard deviations less happy than group A if we use ordered probit and about .08 units less happy if we use ordered logit.

But this conclusion is problematic because it assumes that the distribution of happiness differs between the two groups only through a shift in the mean. It is highly unlikely that a shift in the mean would induce only a shift between the top two categories and not one between the bottom two categories. Indeed this cannot happen with either the normal or logistic distribution. If there were roughly 400 observations in each group, a

⁹We focus on what we view as the more sophisticated approach in this literature which views these categories as capturing three parts of a continuum. We note, however, that is common for researchers to assign the values 0, 1 and 2 to the three categories, in which case, group A would have mean happiness .65 while group B would have mean happiness of only .6. Alternatively, they may perform a linear transformation by subtracting by the mean and dividing by the standard deviation. This has no substantive impact on the results. These approaches assume that the three points on the scale represent known points on an interval scale, which, it will be clear, we view as incorrect. There are even cases where such scales have been treated as ratio scales: "... the data revealed that those making \$55,000 were just 9 percent more satisfied than those making \$25,000." (Dunn and Norton, 2013, p. xiv)

maximum likelihood estimator for either a normal or logistic distribution would reject the null hypothesis that the distributions differ only due to a shift in their mean.

Of course, we could estimate the ordered probit or logit separately for the two groups, but this makes it difficult to interpret the difference. When estimated on a single group with no explanatory variables, normalizing the constant to 0, as in Stata, sets the mean equal to 0. Therefore, we cannot find a difference in mean happiness between the two groups. Instead, we would conclude that for some unfathomable reason, members of group B declare themselves very happy only when their happiness exceeds 1.04 standard deviations above the mean while members of group A are very happy as long as their happiness exceeds .84 standard deviations above the mean although both groups declare themselves not too happy if their happiness is less than .13 deviations above the mean.

Needless to say, this is an unsatisfactory conclusion. The normalizations rule out differences in the true distributions of happiness, the very phenomenon we are trying to investigate. A more reasonable assumption is that the members of groups A and B define the categories of happiness similarly but have both different means and standard deviations of happiness. Without loss of generality, we set the cutoff between “not too happy” and “pretty happy” to 0 and the cutoff between the latter category and “very happy” to 1.

Given normality, we solve

$$\Phi\left(\frac{-\mu}{\sigma}\right) = .55 \tag{1}$$

$$\Phi\left(\frac{1-\mu}{\sigma}\right) = .80 \tag{2}$$

for group A and similarly for group B except that we replace .80 with .85.

We find that average happiness is actually lower for group A at $-.18$ than for group B which has average happiness $-.14$.¹⁰

¹⁰For the logistic distribution the means are $-.17$ and $-.13$.

To gain some intuition into this seemingly perverse result, consider a more extreme situation portrayed in example 2. In this case, in both groups 55% are “not too happy” but the remaining 45% of group A are “very happy” whereas their counterparts in group B are only “pretty happy.” Given a normal distribution, the only way for no one to have happiness between 0 and 1 is for the variance to be infinite. With more observations to the left of 0 than to the right of it, as variance goes to infinity, mean happiness goes to minus infinity. So, on average, group A is infinitely unhappy. In contrast, when nobody reports being “very happy,” the variance must be near zero. As the variance goes to zero, all observations are clustered very close to zero. Even though somewhat more people find themselves with happiness just below zero than just above it, they are all so close to zero that mean happiness among group B is also very close to zero.

Example 2

	Group A	Group B
Very happy	45	0
Pretty happy	0	45
Not too happy	55	55

As the example may suggest, and it is straightforward to show, with the normal and logistic distributions, perverse examples arise when the median response lies at one of the extremes. In the happiness data for the United States, the median generally lies in the middle category. However, the normal and logistic distributions are both symmetric distributions. Asymmetric distributions can produce different results.

Even if estimated mean happiness changes in the same direction as the movement among categories, it will rarely be the case that the distributions of happiness can be ranked in the sense of first-order stochastic dominance. Consider example 3. Again group B appears to be happier than group A. But let us assume that happiness is logistically distributed and normalize the cutoffs to 0 and 1 as before. Now our estimate of mean happiness for group

B (.61) is indeed above the estimated mean for group A (.50), but the spread coefficient is also larger (.42 v .36) so that the happiness distributions cross at the 14th percentile. The results if we instead assume that happiness is normally distributed are similar.

Example 3

	Group A	Group B
Very happy	.2	.28
Pretty happy	.6	.53
Not too happy	.2	.19

At first blush this may not seem problematic. Although neither group is happier in the sense of first-order stochastic dominance, we can still say, using either distribution, that group B is happier on average. Unfortunately, this conclusion necessarily holds only for one specific cardinalization of happiness, the one which produces a normal distribution of happiness for both groups.¹¹ The data could be just as well represented by any other cardinalization produced by monotonically transforming the underlying happiness data. And given that the distributions cross, there are an infinite number of other equally correct alternative cardinalizations for which the conclusion about mean happiness is reversed. In example 3, starting from the normal distribution, we can redefine all values of happiness below -.163 to be

$$u^* = c(u + .163) - .163. \tag{3}$$

For c sufficiently positive, the estimates of average happiness will be reversed.

In fact, perverse examples can even come from standard distributions. Suppose that we used ordinal data to estimate an underlying happiness distribution assuming normality. If we normalize the cut-points to 0 and 1, we will obtain a parameter for the mean, μ , and standard deviation, σ . Suppose we instead estimated a log-normal distribution, by transforming the

¹¹This is also assuming such a cardinalization exists, which we have no way of knowing.

utilities by e^X . Our new mean is

$$e^{\mu+.5\sigma^2}.$$

If we are comparing two groups, one of which has a higher mean and the other a higher variance, this transformation alone could reverse the ranking obtained by the normal distribution. If not, we could simply multiply our latent happiness variable by a positive constant c before exponentiating. The mean under the log-normal transformation would thus become

$$\bar{\mu} = e^{c\mu+.5c^2\sigma^2}. \tag{4}$$

There then will always be a c large enough to reverse the ranking.

What if one group has both a higher mean and higher variance when estimated normally? We can then transform the data by $-e^{-cX}$ to generate a left-skewed log-normal distribution. The mean of happiness becomes

$$\bar{\mu} = -e^{-c\mu+.5c^2\sigma^2} \tag{5}$$

which is decreasing in σ . Thus there must be some c that will reverse the gap. It should be noted that in both cases these are just simple monotonic transformations of the latent happiness variable. Since happiness is ordinal, these transformations represent the responses equally well.¹²

There is a risk that our criticism will be confused with one that is trite. It is, of course, possible to argue that even though a lower proportion of group A than of group B is very happy, the As in this group are much happier than the Bs or that the unhappy Bs are much more unhappy than the unhappy As. In essence, since the top and bottom categories are unbounded, the Manski/Tamer (2002) bounds on both means are plus and minus infinity

¹²Our analysis is related to that of Hammond, Liberini and Proto (2013) who suggest ordering policies using Suppes-Sen dominance. That is, policy A is greater than policy B if and only if the distribution of happiness under policy A stochastically dominates the distribution of happiness under policy B. Our views contrast in that they view stochastic dominance in happiness as reported in categories as a necessary and sufficient condition for Suppes-Sen dominance, while we require stochastic dominance in the continuous latent well-being variable that underlies these categories.

so that either group could be infinitely happier than the other. But our argument is different. The allocation of the responses across the three categories strongly suggests that the variance of happiness differs between the two groups. Therefore, there must be a simple transformation that reverses the ranking of the two groups.

Our focus is on reversals in the estimation of means. However, there is a small but growing literature discussed below that analyzes the dispersion of happiness. It is straightforward to generate plausible examples in which different assumptions about the distribution of happiness lead to different conclusions about relative variances. However, we make no claim that there is always an easy transformation that generates such a reversal.

3 The General Argument

To be clear on our concept of happiness, define V as a vector of factors that influence one's well-being. These could be things like marriage stability, income, the weather, etc. Each vector V is associated with a happiness state H . It follows then that while there can be many V 's which yield the same H , each H must be strictly ranked.¹³

We will assume that there is an infinite number of happiness states, so that the set of all states \mathcal{H} is a well-ordered infinite set. We will further assume the set of happiness states in the population with positive mass is of measure zero. These assumptions allow us to represent happiness by a continuous distribution.

Our assumptions are plausible if we can always imagine being just a little bit more or less happy and if there are a large number of factors affecting happiness, many of which are themselves continuous. We also note that continuity and no mass points are sufficient, not necessary for our results. The CDFs of discrete functions with mass points can, of course, also cross.

We describe happiness as ordinal. Our argument is correct *mutatis mutandis* if there is a

¹³Here we are implicitly assuming that happiness is transitive. If this is not true, it is unclear how one could use any statistical tools to analyze differences in group level happiness.

true underlying interval happiness scale which is unknown. We can define a function q to be a cardinalization of H if for every $H_i > H_j$, $q(H_i) > q(H_j)$. Each of these cardinalizations measures happiness on a different scale.¹⁴ Without selecting a cardinalization, happiness has no numeric value through which we can calculate group means. But, it should be clear that the set of all cardinalizations Q is infinite and our choice of q will influence the magnitudes of differences in happiness between individuals and groups.

We now introduce a series of results, based on the above definition of happiness.

Definition 1 *A “standard” two-parameter distribution is an unbounded probability distribution whose cumulative distribution function can be written as a function of $(u - m)/s$ where u is the level of happiness, m is (typically) a measure of central tendency, and s is (typically) a measure of spread*

Our concept of “standard” encompasses a large range of distributions including Cauchy, Laplace, and extreme value.¹⁵ Most importantly, it includes the normal and logistic distribution, which are the distributions of happiness assumed by any estimation using ordered probit or ordered logit.

Theorem 1 *Suppose there exists a cardinalization q^* under which the distribution of happiness of two groups can be represented by the same standard two-parameter distribution except for the values of m and s , and s is not identical for the two groups. Then there also exists a cardinalization $q' \in Q$ under which the ranking of group means is the opposite of that under q^**

Proof. We can write the cumulative distribution functions for the two groups for standard distributions as $F(\frac{u-m_1}{s_1})$ and $F(\frac{u-m_2}{s_2})$. We can solve for the crossing point by basic algebra

¹⁴For example, one cardinalization could equate each state to a level of earnings that a typical family would require to be in such a state.

¹⁵We note that the idea that happiness is “bounded” or “unbounded” has no meaning for ordinal happiness, so long as the distribution of happiness has no mass points. For any cardinalization where happiness is bounded, we can take the cumulative distribution function $F(u)$ and transform happiness by $u' = \ln(F(u)/(1 - F(u)))$ to get to an unbounded cardinalization.

as

$$u^* = \frac{m_2 s_1 - m_1 s_2}{s_1 - s_2}$$

which is finite for $s_1 \neq s_2$. Therefore, there cannot be first-order stochastic dominance, which is a necessary condition for the the mean of two distributions to be invariant to monotonic transformations (e.g. Spencer, 1983). ■

The proof shows that if we can represent the happiness of two groups through the same standard two-parameter distribution with differing parameters, their cdfs will always cross, except for a knife-edge case where the variances are the same. This means we cannot strictly rank groups by mean happiness. Any ranking of happiness will be dependent on choice of cardinalization, of which there are infinitely many which are equally defensible from the data.

How often does there exist a cardinalization under which the happiness of two groups can be represented by a standard two-parameter distribution? There is always a cardinalization under which one group's happiness follows a normal distribution (or any other standard distribution). This simply involves assigning arbitrary values to happiness states to get the correct shape. What is unclear is whether it is likely that another group's distribution would also follow a normal distribution under that same cardinalization. However, all happiness research that uses ordered probit or ordered logit to estimate differences in mean happiness *implicitly assumes* the existence of such a cardinalization. In these situations then, it must either be the case that the results can be reversed, or the model is invalid.

Further, we lack other statistical methods to analyze ordinal data with a small number of categories. In particular, with the ubiquitous three-point scale from the General Social Survey, which we will focus on in the following section, our model for ordered probit or logit is just-identified, so we have no way of testing the validity of the cardinalization/distributional assumption. As we will discuss later, we have some hope for scales with more than 3 points.

Finally, we note that since the ordinal responses are reported in categories, in finite samples, depending on the number of observations from each group, there can be a positive

probability that the estimated s will be the same for two independent samples. However, as the sample gets large, this probability gets small.

Remark 1 *Let $L = \sum_{j \in G_i} \sum_c d_c \ln F_j^c(m_i, s_i)$ be the log-likelihood function of distribution $F(m, s)$ for group G_i from J independent observations of data d with C categories and let \hat{m}_i, \hat{s}_i be the parameter estimates that maximize the estimated likelihood, then*

$$N^{.5}(\hat{s}_i - s_i) \rightarrow^d N(0, \sigma_s^2)$$

This remark follows from the standard properties of maximum-likelihood estimators. It follows directly that for large but finite samples, the estimated measures of spread will almost never be equal.¹⁶

This, in turn, leads to our main result.

Conclusion 1 *If happiness is reported using a discrete ordinal scale, in large samples it will (almost) never be possible to rank the mean happiness of two groups without additional restrictions on the nature of the happiness distribution.*

4 Empirical Applications

4.1 Moving to Opportunity

Economists have long postulated that living in a poor neighborhood may make it more difficult to escape poverty. Motivated by this idea and the positive results of the Gautreaux

¹⁶The intuition behind this is relatively straightforward. Needless to say, if the true variances are not equal, then asymptotically the probability that the estimates differ by less than ε goes to 0 as the samples become large. Suppose, however, that $s_A = s_B = s$. Then $\hat{s}_A \rightarrow^d N(s, \sigma_A^2)$ and $\hat{s}_B \rightarrow^d N(s, \sigma_B^2)$. Define $\alpha = \hat{s}_A - \hat{s}_B$. Then since \hat{s}_A and \hat{s}_B are asymptotically independent normals, $\alpha \rightarrow^d N(0, \sigma^2)$ where $\sigma \equiv (\sigma_A + \sigma_B)^{.5}$. The asymptotic density at $\alpha = 0$ is $(2\pi)^{-.5} \sigma^{-1}$. Since the density is maximized at 0, the probability that α falls in the range $-\varepsilon\sigma < \alpha < \varepsilon\sigma$ is less than $2\varepsilon\sigma (2\pi\sigma^2)^{-.5} = (2/\pi)^{.5} \varepsilon$ which can be made arbitrarily small for any sequence of σ approaching 0. We are grateful to Zhongjun Qu for providing us with this argument.

desegregation program in Chicago,¹⁷ the Moving-to-Opportunity experiment targeted families living in public housing in high poverty areas. Eligible families were invited to apply for the chance to receive a Section 8 housing (rental assistance) voucher. Applicants were randomly assigned into three groups: no voucher (Control group), Section 8 voucher that could only be used in an area with a poverty rate below 10% (Experimental group), and a standard Section 8 voucher (Section 8 group).

The program has been assessed at multiple stages.¹⁸ A long-term follow-up (Ludwig et al, 2012, 2013) emphasizes that subjects in the experimental group were substantially happier than those in the control group. We reexamine the evidence for this conclusion.

The participants in the long-term MTO evaluation study were asked “Taken all together, how would you say things are these days – would you say that you are very happy, pretty happy, or not too happy?” The authors focus on the effect on the distribution of responses across categories. Nothing we write below can or will contradict the finding that MTO increases the proportion of individuals who report that they are “very happy” and reduces the proportion who say they are “not too happy.” If these are the socially relevant categories, then there is no need to estimate a mean, or any other single summary measure of happiness. Thus, for example, we might believe that people are “not depressed,” “mildly depressed” or “severely depressed” and view variation in depression within these categories as unimportant. We take no position on the accuracy of this view of depression, but if we accept it, an intervention that reduces the proportion of severely and mildly depressed individuals reduces depression since variation in depression within categories is unimportant. We return to this point in our conclusions.

However, we believe happiness is well-approximated by a continuous distribution. Therefore, statements about mean happiness and not just the frequency of responses within categories are potentially relevant. Ludwig et al (2012, table S4) report intent-to-treat estimates

¹⁷The Gautreaux program came out of a court-ordered desegregation program in Chicago in the 1970s. See Rosenbaum (1995) for a detailed analysis.

¹⁸For the earliest evaluation, see Katz, Kling, and Liebman (2001). For an intermediate-term evaluation, see Kling, Liebman, and Katz (2007).

on the experimental group using intervals of 1 unit between the categories, as is common in the literature, but also ordered probit and logit. In all three cases, they find positive effects on average happiness that fall just short of significance at the .05 level.

If we cardinalize happiness to be normally distributed, normalizing the cutoffs to 0 and 1, we find that the control group does have a lower mean (.44 v. .60). But, the control group also has a higher variance (.79 vs .63). The cdfs cross at the 83rd percentile, which is 1.20 units of happiness (and also in the extreme left tail of the distributions). Thus if we simply transform the underlying happiness data to increase the values above 1.20 we can reverse the mean happiness. This cardinalization would explain the data equally well.

Alternatively, we can perform an exponential transformation to get a log-normal distribution of happiness. Keeping our underlying cut-points fixed at 0 and 1, the exponential transformation will still show that the experimental group (2.22) is happier than the control group (2.14). But, as discussed in section 2, since the control group has a higher variance of happiness, we can raise their mean relative to the experimental group by multiplying the underlying latent (log) happiness variable by a constant. If we multiply each individual's happiness by 1.33 before performing the exponential transformation, the mean happiness of the two groups are equal. As we show in figure 1, this amounts to a somewhat right-skewed distribution of happiness, meaning the differences among the happiest individuals are greater than the differences among the least happy. This is just a monotonic transformation of the cardinalization underlying the normal distribution and thus fits the data equally well. Therefore, we cannot determine whether the causal effect of MTO on happiness is positive on average. One plausible interpretation of the data is that moving to a low poverty area increased happiness for most people, but that there is a group of people who were extremely happy in their old environment who could not match positive aspects of their former social environment in their new community.

One solution to this indeterminacy is to tie our assessment of (un)happiness to some other outcome variable. This is the approach we use in Bond and Lang (2014) where we

scale test scores in a given grade by the eventual educational attainment of students with those test scores. The limited number of points on the happiness scale makes this difficult. This discreteness may be missing variation within the categories that represents important distinctions in happiness. But, compared with variation at the high end of the scale, variation in happiness at the low end of the scale might prove to be more closely correlated with other signs of psychological distress, which were also shown in Ludwig et al to be beneficially influenced by moving to a neighborhood with a lower poverty rate.

Thus, in settings where we do not have direct measures of psychological well-being, it *may* be possible, we are agnostic on this point, to use data from other settings such as MTO to scale happiness in a more compelling way. For MTO, the strongest evidence of positive psychological benefits comes from direct measures of the prevalence of psychological problems. Provided that these conditions are discrete rather than continuous, our concerns about happiness scales do not apply to such things as measures of depression.¹⁹

4.2 The Paradox of Declining Female Happiness

One surprising result from the happiness literature, documented by Stevenson and Wolfers (2009), is that in the United States women’s happiness appears to have fallen relative to men’s from 1972-2006 despite the great social and economic progress women made during this period.²⁰ Again, this result is easily reversed.

We use the publicly available file created by Stevenson and Wolfers from the General Social Survey (GSS), a nationally representative survey of social attitudes conducted annually or biennially since 1972. The GSS assesses subjective well-being on the same three point scale later adopted in the MTO study. While the question remains constant over time, its position in the survey does not, which could lead to biases in responses in different years.²¹

¹⁹Luttmer (2005) is another example where the author is careful to ensure that the results from other measures are consistent with those from subjective measures of well-being.

²⁰By labeling this “surprising,” we do not mean to imply that it could not be true. In a related area, Black et al (2009) suggest that apparent black-white earnings convergence was accompanied by black mobility to higher cost localities, suggesting much less convergence in real incomes.

²¹For example, Stevenson and Wolfers (2009) note that in every year but 1972, the question followed a

Stevenson and Wolfers use split-ballot experiments to modify the data to account for these differences.²²

To simplify the analysis and ease exposition, we create two subgroups: those from the first five surveys (1972-1976) and those from the last five surveys (1998-2006) but can obtain similar results using the full time series. We display the distribution of happiness in these groups in Table 2. Using ordered probit, Stevenson and Wolfers found that women lost ground to men at the rate of .376 standard deviations per century. We confirm this result between the two subgroups; ordered probit estimates that women were .09 standard deviations less happy relative to men in the later sample than the early.

However, as discussed previously, ordered probit assumes that we can cardinalize happiness so that the variance is constant across sex and over time, an assumption we can easily reject. When we allow the variances to differ, we find women’s happiness has more variance than men’s and that the variance of happiness has declined over time. This is what one would expect from looking at the data. Most of the differences between the sexes and over time are due to there being more “very happy” women in the early years. When more people are “very happy” but there is no difference in the number “not too happy,” the distribution must have higher variance to fit the data. Relaxing the constant variance assumption lowers the growth of the gap to .07.

Now, we transform the latent happiness variable by

$$\tilde{u} = -e^{-Cu} \tag{6}$$

so that the distribution of happiness is given by the left-skewed log normal distribution. Since their happiness distribution has the highest variance under the normality assumption, choosing a C sufficiently large lowers the mean happiness of women in the early period by more than it does men’s. As we show in in Figure 2, for $C \geq 3.9$ women become happier over

question on marital happiness, which may cause differences in the impact of one’s marriage on his or her response to the general happiness assessment. See Smith (1990).

²²For details of this process, see appendix A of Stevenson and Wolfers (2008b).

time relative to men, as one might expect given their social progress in the period. Large values of C will show large increases in relative female happiness.

Admittedly, when $C = 3.9$, the distribution is fairly skewed. This implies that differences among the unhappiest people are far greater than differences among the happiest. All happy people have happiness between 0.02 and 0, while 5% of the distribution has happiness below -250 . Of course, this distribution fits the data just as well as the normal. From the data alone it is difficult to argue that one happiness distribution is clearly more plausible.

Further, this is just one scale and distribution based on a simple transformation under which women gain happiness relative to men. There are an infinite number of others, and more complex transformations may create distributions that are more intuitively appealing. Ultimately if we can gain consensus about plausible restrictions on the happiness distribution (e.g. skewness, kurtosis) or at least a reasonable loss function involving these moments, it may be possible in some cases to conclude that no plausible transformation will reverse a particular finding.

4.3 Easterlin Paradox

No question in the happiness literature has received more attention than the “Easterlin Paradox,” the observation that in some settings higher incomes do not appear correlated with higher levels of happiness. Easterlin (1973, 1974) found that income and subjective well-being assessments were strongly and positively correlated within a country in a given year, but not over time and across countries. This, and subsequent studies, led Easterlin (1994) to conclude, “Will raising the incomes of all increase the happiness of all? The answer to this question can now be given with somewhat greater assurance than twenty years ago. It is ‘no’.” Easterlin instead concludes that the weight of the evidence supports the idea that individuals judge their happiness relative to their peers and not on an absolute scale.

The paradox was recently called into question in a comprehensive study by Stevenson

and Wolfers (2008a).²³ They use ordered probit both across countries and over time within countries and find a strong relation between happiness and economic development. However, they find that the United States is an exception. Happiness has not increased despite substantial growth in per capita incomes. They attribute this to the substantial rise in income inequality over the last 30 years which occurred simultaneously with the rise in real GDP.

We match the GSS data from Stevenson and Wolfers (2009) with U.S. per capita real GDP data from the Federal Reserve Bank of St. Louis to get a time-series of national happiness and income data. Fixing the cut-points to 0 and 1, we estimate the two parameters of a normal distribution for each year using the GSS and regress the means on the log of real GDP per capita. As we show in Figure 3, we do indeed find an Easterlin Paradox. Ordinary Least Squares estimates imply that a 10% increase in GDP per capita is actually associated with a decrease in average happiness in the United States of .46 units, although, with a p -value of only .18, it is not statistically significant.

However, figure 4 shows that we also estimate a strong negative relation between real GDP per capita and the variance of happiness. A 10% increase in GDP per capita is associated with a statistically significant 1.06 unit decrease in the standard deviation of happiness. This may be somewhat surprising given the increase in income inequality over the time period, but is what one would expect from the data and has been demonstrated previously by Stevenson and Wolfers (2008b) and Dutta and Foster (2013).²⁴ As real GDP has increased, fewer people report being very happy, but there is a zero to slightly negative change in the number of people who report being not too happy.

Since high-GDP periods have a lower mean and variance than low-GDP periods, we know that a left-skewed log normal distribution will reverse the trend. In fact, we do not need to skew the distribution that much. For values of $C \geq .70$ we find the expected positive relation between income and happiness. In Figure 5, we show the distribution of happiness

²³See also Deaton (2008) who finds similar results from the Gallup World Poll using OLS on a basic 10-point scale.

²⁴Clark, Fleche and Senik (2014, forthcoming) argue that this is a standard pattern – growth reduces happiness inequality.

under this set of parameters in 2006. Here the cut-point values of happiness would be -1 to go from not too happy to pretty happy and $-.5$ to go from pretty happy to very happy. There is variation among the happiest and least happy individuals, although more so among the latter given the skewness of the distribution.

For $C \geq 2.60$, this positive relation becomes statistically significant. We plot the $C = 2.60$ case in Figure 6. Here, a 10% increase in real GDP per capita is associated with a 2.04 unit increase in average happiness. If we are willing to accept this amount of skewness in the happiness distribution, then raising the incomes of all does not raise the happiness of all but does raise average happiness. There are other distributions and transformations that replicate this result; there is no way to determine from the data which cardinalization is correct.

As in the case of MTO, if we are convinced that the response categories in the survey are the ones that are relevant for policy purposes, we can avoid this problem. However, unlike the case of the female happiness paradox where it *might* be possible to conclude that no plausible happiness cardinalization would reverse the result, it is evident that plausible (at least to us) distributions can reverse the basic finding.

4.4 Cross-Country Comparisons

In the previous sections, we found that three conclusions based on normally-distributed happiness assumptions could be reversed by simple log-normal transformations. In this subsection we explore the sensitivity of happiness comparisons to such transformations in general. Using data from the World Values Survey, we estimate mean happiness at the country level for a normal distribution, as well as a log-normal distribution with $C = 2, .5, -.5$, and -2 .²⁵

The ordering of countries in Table 3 represents their happiness ranking when happiness is distributed normally, and the columns list their ranking under the different log-normal

²⁵In contrast with the other data we use, the World Values Survey elicits happiness responses on a 4-value scale: “not at all happy,” “not very happy,” “quite happy,” and “very happy.” Because the fraction of “not at all happy” responses is almost universally trivial, we combine these responses with those of not very happy to get a 3-point scale. This allows us to follow the same approach as in the previous three subsections.

transformations. Although the actual degree of skewness varies across countries due to differences in the variance of the underlying normal distribution, moving from left to right in the columns represents moving from a relatively right-skewed to a relatively left-skewed distribution. Doing so has dramatic effects on the rank-ordering of happiness. The five happiest countries when happiness is right-skewed are Ghana, Guatemala, Mexico, Trinidad and Tobago, and South Africa. Three of these countries rank in the *bottom* ten when happiness is left-skewed, and only one (Mexico) ranks in the upper half. The top five under the extreme left-skewed distribution of happiness (New Zealand, Sweden, Canada, Norway, and Great Britain) fare relatively better under right-skewed happiness, though only Great Britain remains in the top ten. The rank-correlation between the log-normal transformations with $C = 2$ and $C = -2$ is .156.

There are some countries whose rank remains fairly stable throughout the transformations. Great Britain is the third happiest country under a normal distribution and has its rank vary between 2 and 8 under the skewed distributions. Moldova, the world's least happy country under the normal distribution, is never able to rise above 4th worst in the skewed transformations. These cases are counterbalanced by countries like Ghana and Ethiopia. Ghana ranges from the world's happiest to the world's 3rd least happy depending on whether happiness is right- or left-skewed. Ethiopia, the 10th least happy under the normal distribution, is able to rise as high as 7th when happiness is right-skewed, placing it above the United States, Australia, and Great Britain, among others.

The wide variation in ranking suggests that in most cases the amount of skewness allowed in the distribution can have substantial impacts on cross-group comparisons. Even the most skewed-distributions we explored here are not, to us, implausible. They involve a smaller exponential transformation than required to have a significantly positive relation between average happiness and per capita GDP over time in the United States (see figure 5). We do find the ranking under the left-skewed distribution to be more in-line with our priors than the right-skew or the normal, though we stress there is nothing in the happiness data

itself that would allow us to choose among the distributions. Interestingly, the right-skewed distributions would imply a strong negative correlation between per capita GDP and mean happiness, while the left-skewed implies a strong positive relation.²⁶

5 Discussion and Conclusions

As we have demonstrated, key conclusions of happiness studies depend on the chosen cardinalization of happiness, something about which the data can give us little or no guidance. Since the estimated cdfs (almost) always cross when we assume a standard distribution of happiness, there is always some transformation that preserves the rank order of individuals and changes the direction of the estimated gap in mean happiness.

Is there any way to create compelling cross-group comparisons?

5.1 Assume the Policy-Relevant Distribution

Perhaps the simplest assumption one could make is to assume that we know the policy-relevant distribution of happiness. For instance, if we believe happiness is distributed normally, we can fix the cut-points between “not too happy”, “pretty happy”, and “very happy” and estimate the means and variances of each group through ordered probit.

It should be clear that the choice of a particular distribution almost inevitably implies taking a stand on some of the very issues that have been the focus of the happiness literature. Thus since wealth and income are highly skewed, a normal distribution of happiness would almost necessarily require the marginal utility of wealth or income to be sharply diminishing. We do not, however, rule out the possibility that the profession could achieve near consensus on some reasonable restrictions on the happiness distribution and that these restrictions would be adequate to allow us to reach strong conclusions about the ranking of

²⁶Using 2005 data from the World Bank on purchasing power parity equivalent per capita GDP, the coefficient on a regression of estimated mean happiness and the natural logarithm of per capita GDP is -5.35 for the right-skewed ($C = 2$) distribution and .61 for the left-skewed distribution ($C = -2$).

mean happiness in some cases.

Of course, there are difficulties even if we can rank means. Unless we are very traditional utilitarians who wish to maximize the sum of utilities, we will still encounter problems for policy purposes. We may, many philosophers would argue should, care more about increases in happiness at some parts of the distribution than at others. In this case, the Bond and Lang (2013) criticism of test scores applies directly.

A similar approach would be to impose a standard distribution and qualify results with a discussion of the implied cdfs. For example, if we found that when both groups' happiness is estimated at the normal distribution, the cdfs cross only at the 95th percentile, we could conclude that one group was happier on average provided that the happiness differences amongst the bottom 95 percent are greater than that among the top 5 percent. In some circumstances, the crossing points may be so extreme that reasonable people will likely agree that we can draw a conclusion. Of course this will also frequently not be the case. Still, this approach brings the advantage of ensuring the researcher specifies her underlying assumptions clearly.

An alternative solution is to declare the ordinal scale on which people report their happiness to be the policy-relevant one. Group A is happier than group B if its members' responses "stochastically dominate" B's using the categories provided in the question about self-assessed happiness. This approach has a great deal of intuitive appeal, and we confess that in some cases we are inclined to accept it. However, it is trivial to find examples where, using what appears to be a sensible partition of the data into three categories, groups appear to be ordered in the sense of stochastic dominance, but for which the means have the reverse order when the full underlying distribution is examined. For example, there are many occupations (e.g. actors) in which mean income is relatively high but most people in the occupation have very low incomes. Other occupations have high variance but also relatively low mean wages. When stochastic dominance fails using the full underlying distribution, it is possible to group wages (or other variables) so that using the grouped data, stochastic

dominance appears to hold.

5.2 More Categorical Responses

When happiness is surveyed using three discrete categories, it is impossible to estimate a distribution with more than two parameters. As we showed in Theorem 1, there can never be stochastic dominance when comparing two groups whose happiness follows the same two parameter distribution with differing spreads and locations. Would increasing the number of categories allow us to estimate more parameters and thus find stochastic dominance?

When looking at only one group, more categories does not help. Adding an additional response adds an additional interval which will allow us to estimate one more parameter in a maximum likelihood framework. However, we must also identify the cutoff value of the latent happiness variable that causes individuals to report in this new category. Unlike the first two cutoff values, the third cannot be assigned as a normalization. Thus, even with a ten-point Cantril (1965) scale, we are not able to identify additional parameters of the distribution – each added interval adds one additional auxiliary parameter to estimate.

However, assuming a common reporting function, these auxiliary parameters will be the same across different groups. Thus when comparing two groups, if we estimate the distributions jointly, each additional category gives us one new moment. Thus with four categorical responses, we would have one overidentifying restriction to test the distributional assumptions of our model. However, a failure to reject only suggests that it is possible to cardinalize happiness so that both groups' distribution is normal. It says nothing about whether this cardinalization is “correct” for calculating happiness means for policy purposes. With five categories, it *might* be possible to estimate an additional common parameter of a three-parameter distribution, but this is by no means obvious.²⁷ And it is not clear that adding a few more points to the scale would frequently, if ever, lead to stochastic dominance.

²⁷Given that the choice of distribution for one group is in some sense a normalization, it may be more logical with 5 categories to estimate a 2-parameter distribution for one group and a more general 4-parameter distribution for the other.

Of course, as the number of response categories becomes very large, the distinction between analyzing the categorical responses as points or intervals becomes less important. In an interesting experiment, Oswald (2008) asked participants to report their height on a continuous scale from 0 to 10. Assuming that there are no responses at the extreme, if all respondents use the same reporting function, then stochastic dominance is sufficient to ensure that the ranking of the means is independent of scale.²⁸

5.3 Closing Remarks

One solution when working with ordinal scales is to relate them to some measurable outcome. In the traditional economics literature, we measure the utility of a good or outcome by willingness to pay, imperfectly captured by the equivalent or compensating variation. The happiness literature has called this approach into question and with it some basic assumptions, such as positive marginal utility of money. Invoking a monetary scale thus brings us to a Catch-22. We cannot answer the main questions of the happiness literature using the most obvious tool because the literature seeks to invalidate that very tool.

We note that our examples require different transformations to reverse the results in the literature. We showed that moving to a low poverty area reduces mean happiness if happiness is log normally distributed and strongly right skewed. But the Easterlin paradox is resolved for the United States if happiness follows a sufficiently left-skewed log normal distribution. It is not obvious that there is an assumption about the distribution that would reverse both results. It may be that we can reach sufficient consensus about which distributions are acceptable that we can make definitive statements in some cases. One possibility is what we call the “Tolstoy assumption,” that there is far greater variation in unhappiness than in

²⁸For each of these suggestions it is critical the reporting function is common across individuals and time. We take no stance on whether this is the case, but we note that in Oswald’s experiment he found that men and women appeared to report their height differently. Likewise, in repeated cross-sections, immigrants show no improvement in their host-country language skills, but they report improvement when asked to compare their current and earlier language skills, suggestion that reporting functions over time are unlikely to be stable (Berman, Lang and Siniver, 2003).

happiness.²⁹ In other words, happiness is left skewed. In this case, it is very likely that MTO did raise happiness for those who moved to less impoverished areas. This is consistent with the standard assumption that expanding the choice set of rational agents should never *ex ante* decrease utility.

The underlying distribution of happiness elicited from categorical survey data is almost unknowable. Many natural occurring variables are symmetric and normal, but many are also skewed. As we have shown, using ordinal happiness data it is (almost) never possible to draw conclusions which are robust to all scales and distributions. Researchers must take great care in examining the robustness of their results to plausible alternative cardinalizations of happiness, and be explicit about what distributional assumptions are underlying their conclusions.

As mentioned briefly in the introduction, our conclusions have important implications for attempts to guide policy using happiness research. If one takes the neo-utilitarian position that we should maximize an individualistic social welfare function, the finding that the happiness distributions (almost) always cross means that the preferred policy is never invariant to the choice of social welfare function. A highly left-skewed happiness distribution can be viewed as a generalization of a Rawlsian utility function that puts considerably more weight on improvements at the bottom of the happiness distribution. With this interpretation, economic growth increases social welfare, the welfare of U.S. women may not have decreased relative to that of men, and the benefits of MTO were even greater than previously estimated.

²⁹We apologize to lovers of Russian literature for this deliberate misinterpretation of Anna Karenina – “All happy families are alike; each unhappy family is unhappy in its own way.”

References

- [1] Berman, Eli, Kevin Lang and Erez Siniver, 2003. "Language-Skill Complementarity: Returns to Immigrant Language Acquisition,' *Labour Economics*, 10 (3): 265-90.
- [2] Bertrand, Marianne and Sendhil Mullainathan. 2001 "Do People Mean What They Say? Implications for Subjective Survey Data," *American Economic Review*, 91 (2): 67-72.
- [3] Black, Dan, Natalia Kolesnikova, Seth Sanders, Lowell Taylor and Mark Wessel. 2009. "A Divergent View on Black-White Earnings Convergence," unpublished.
- [4] Bound, John, Charles Brown and Nancy Mathiowetz. 2001. "Measurement Error in Survey Data," in James J. Heckman and Edward Leamer, Editor(s), *Handbook of Econometrics*, New York and Amsterdam: Elsevier, 3705-3843.
- [5] Clarke, Andrew E., Sarah Fleche and Claudia Senik, forthcoming. "The Great Happiness Moderation," in A.E. Clark and C. Senik, eds., *Happiness and Economic Growth: Lessons from Developing Countries*, Oxford: Oxford University Press.
- [6] Clarke, Andrew E., Sarah Fleche and Claudia Senik, 2014. "Economic Growth Evens Out Happiness: Evidence from Six Surveys,' Paris School of Economics Working Paper No. 2014-03.
- [7] Cunha, Flavio and James J. Heckman. 2008. "Formulating, Identifying, and Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Journal of Human Resources*, 43 (4): 738-782.
- [8] Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica*, 78 (3): 883-931.

- [9] Bond, Timothy N. and Kevin Lang. 2013. "The Evolution of the Black-White Test Gap in Grades K-3: The Fragility of Results." *Review of Economics and Statistics*, 95 (5): 1468-79.
- [10] Bond, Timothy N. and Kevin Lang. 2014, "The Black-White Education-Scaled Test-Score Gap in Grades K-7," unpublished.
- [11] Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2013. "Measuring Test Measurement Error: A General Approach." *Journal of Educational and Behavioral Statistics*, 38 (6): 629-663.
- [12] Cantril, Hadley. 1965. *The Pattern of Human Concerns*. New Brunswick, NJ: Rutgers University Press.
- [13] Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impact of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review*, 104 (9): 2633-2680.
- [14] Deaton, Angus. 2008. "Income, Health, Wellbeing Around the World: Evidence from the Gallup World Poll." *Journal of Economic Perspectives*, 22 (2): 53-72.
- [15] Dunn, Elizabeth and Michael Norton, *Happy Money: The Science of Smarter Spending*, New York: Simon & Schuster, 2013.
- [16] Dutta, I. and J. Foster, 2013. "Inequality of Happiness in the United States: 1972-2010," *Review of Income and Wealth*, 59: 393-415.
- [17] Easterlin, Richard A. 1973. "Does Money Buy Happiness?" *The Public Interest*, 30 (3): 3-10.
- [18] Easterlin, Richard A. 1974. "Does Economic Growth Improve the Human Lot?" In *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*, ed. Paul A. David and Melvin W. Reder, 89-125. New York: Academic Press.

- [19] Easterlin, Richard A. 1994. "Will Raising the Incomes of All Increase the Happiness of All?" *Journal of Economic Behavior and Organization*, 27 (1): 35-47.
- [20] Frey, Bruno S. and Alois Stutzer. 2003. "Maximising Happiness?" *German Economic Review*, 1 (2): 145-67.
- [21] Hammond, Peter J., Federica Liberini and Eugenio Proto. 2013. "Individual Welfare and Subjective Well-Being: Commentary Inspired by Sacks, Stevenson and Wolfers." in Claudia Sepulveda, Ann Harrison, and Justin Yifu Lin (eds.) *Annual World Bank Conferences on Development Economics 2011: Development Challenges in a Postcrisis World*. Washington, DC: World Bank Publications: Washington, DC. 339-353.
- [22] Kahneman, Daniel and Alan B. Krueger. 2006. "Developments in the Measurement of Subjective Well-Being." *Journal of Economic Perspectives*, 20 (1): 3-24.
- [23] Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman. 2001. "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment." *Quarterly Journal of Economics*, 116 (2): 607-654.
- [24] King, Gary, Christopher J. L. Murray, Joshua A. Salomon, and Ajay Tandon. 2004. "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research." *American Political Science Review*, 98 (1): 191-207.
- [25] Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Labor Market Effects." *Econometrica*, 75 (1): 83-119.
- [26] Krueger, Alan B. and David . Schkade. 2008. "The Reliability of Subjective Well-Being Measures." *Journal of Public Economics*, 92 (8-9): 1833-1845.
- [27] Ludwig, Jens, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, Lisa Sanbonmatsu. 2012. "Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults," *Science*, 337: 1505-1510.

- [28] Ludwig, Jens, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, Lisa Sanbonmatsu. 2013. "Long-Term Neighborhood Effects on Working Families: Evidence from Moving to Opportunity." NBER Working Paper No. 18772.
- [29] Luttmer, Erzo F. P. 2005. "Neighbors as Negatives: Relative Earnings and Well-Being," *Quarterly Journal of Economics*, 120 (3): 963-1002.
- [30] Neal, Derek A. and William R. Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy*, 104 (5): 869-895.
- [31] Olken, Benjamin A. 2009. "Corruption Perceptions vs. Corruption Reality," *Journal of Public Economics*, 93 (7-8): 950-964.
- [32] Oswald, Andrew J. 2008, "On the Curvature of the Reporting Function from Objective Reality to Subjective Feelings," *Economics Letters*, 100 (3): 369-372.
- [33] Peracchi, Franco and Claudio Rossetti. 2013. "The Heterogeneous Thresholds Ordered Response Model: Identification and Inference," *Journal of the Royal Statistical Society*, 176 (3): 703-722.
- [34] Ravallion, Martin, Kristen Himelein, and Kathleen Beegle. 2013. "Can Subjective Questions on Economic Welfare be Trusted? Evidence for Three Developing Countries," unpublished.
- [35] Ritter, Joseph A. and Lowell J. Taylor. 2011. "Racial Disparity in Unemployment," *Review of Economics and Statistics*, 93 (1): 30-42.
- [36] Rosenbaum, James E. 1995. "Changing the Geography of Opportunity by Expanding Residential Choice: Lessons from the Gautreaux Program." *Housing Policy Debate*, 6 (1): 231-269.
- [37] Spencer, Bruce D. 1983. "On Interpreting Test Scores as Social Indicators: Statistical Considerations," *Journal of Educational Measurement*, 20 (4): 317-333.

- [38] Stevens, S.S. 1946. "On the Theory of Scales and Measurement," *Science*, 103 (2684): 677-680.
- [39] Stevenson, Betsey and Justin Wolfers. 2008a. "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox." *Brookings Papers on Economic Activity*, 1-87.
- [40] Stevenson, Betsey and Justin Wolfers. 2008b. "Happiness Inequality in the United States." *Journal of Legal Studies*, 37(2): S33-79.
- [41] Stevenson, Betsey and Justin Wolfers. 2009. "The Paradox of Declining Female Happiness." *American Economic Journal: Economic Policy*, 1(2): 190-225.
- [42] Thorndike, Robert L. 1966. "Intellectual Status and Intellectual Growth." *Journal of Educational Psychology*, 57 (3): 121-127.
- [43] Waidmann, Timothy, John Bound and Michael Schoenbaum. 1995. "The Illusion of Failure: Trends in the Self-Reported Health of the U S Elderly," *Milbank Quarterly* 73 (2):253-287.

Figure 1: MTO Log-Normal Happiness Distribution with Equal Means

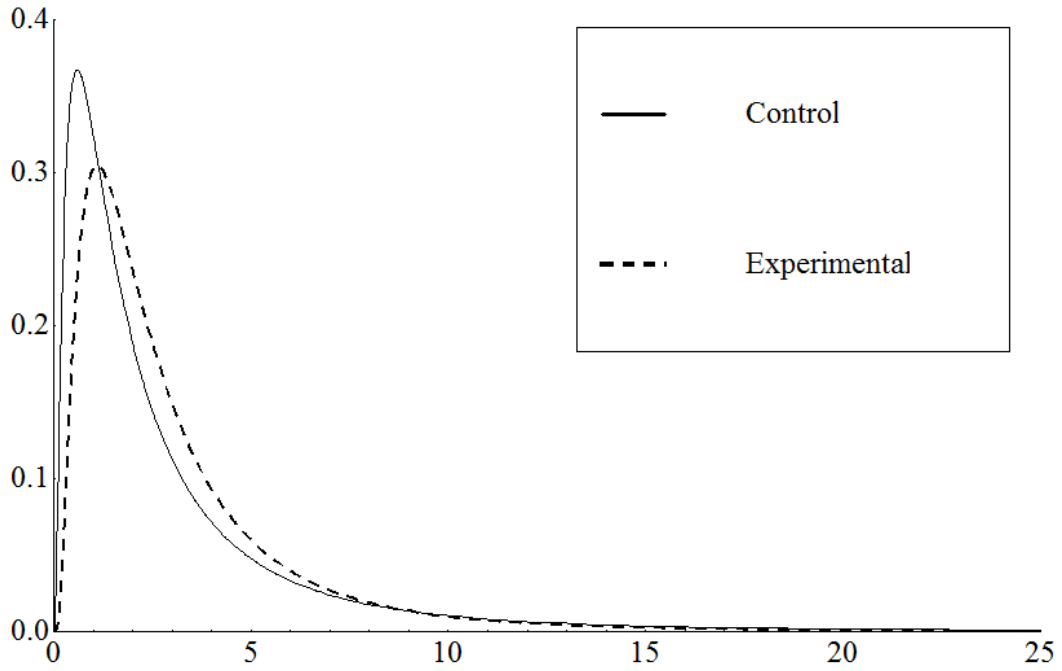


Figure 2: Trend in Female-Male Happiness Gap for Log-Normal Distributions

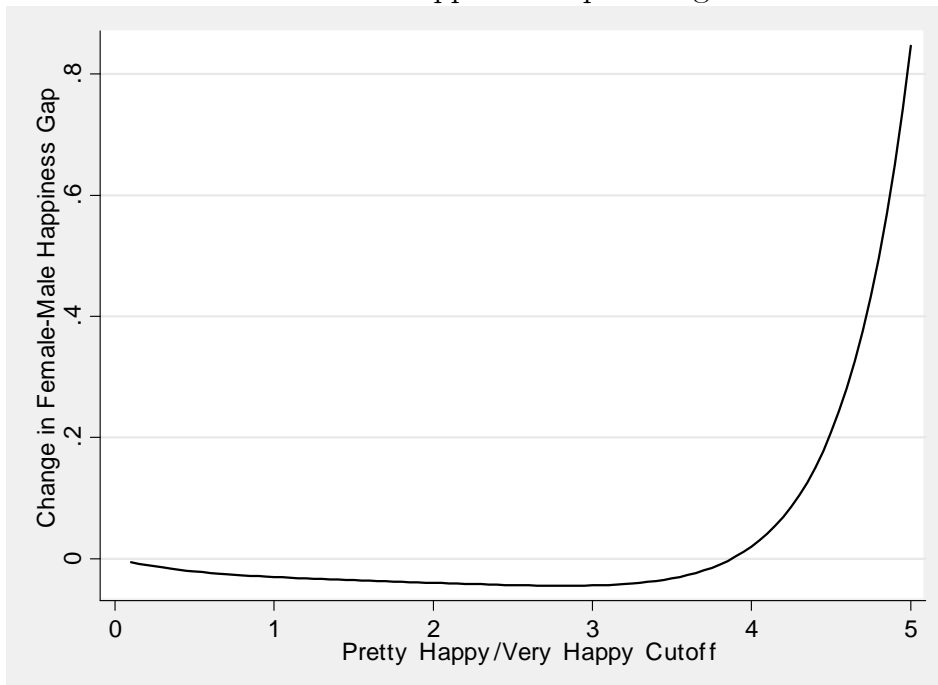


Figure 3: Mean Happiness and National Income, Normal Distribution

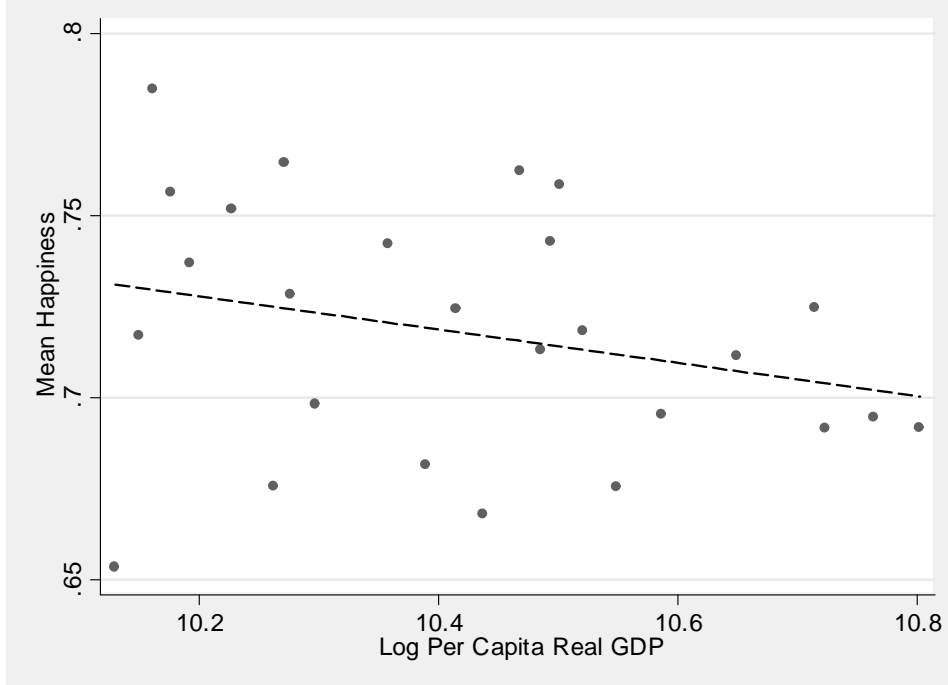


Figure 4: Standard Deviation of Happiness and National Income, Normal Distribution

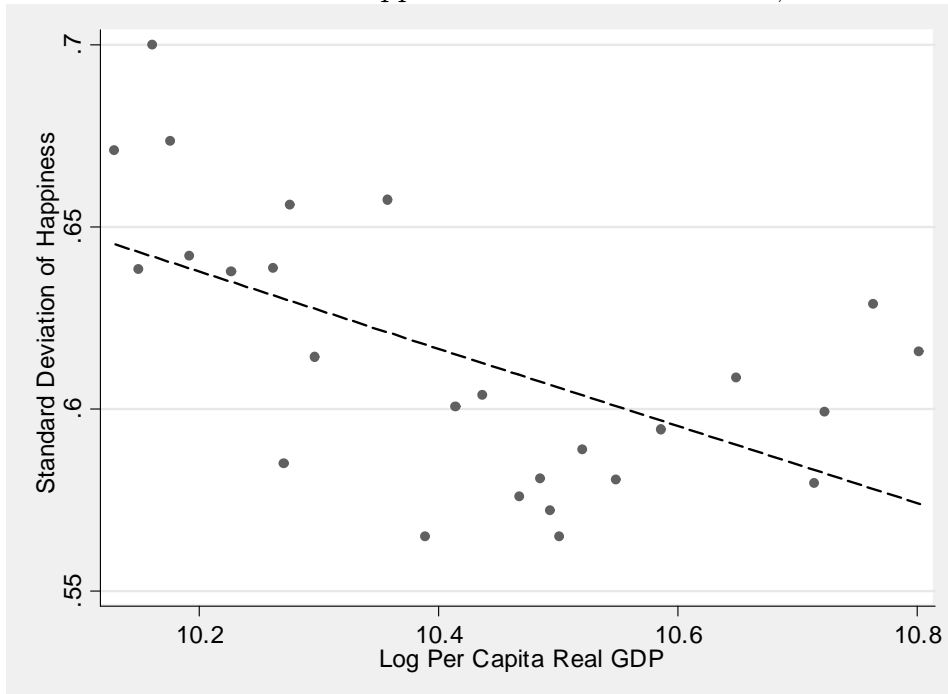


Figure 5: 2006 Log-Normal Distribution of Happiness with no Easterlin Paradox

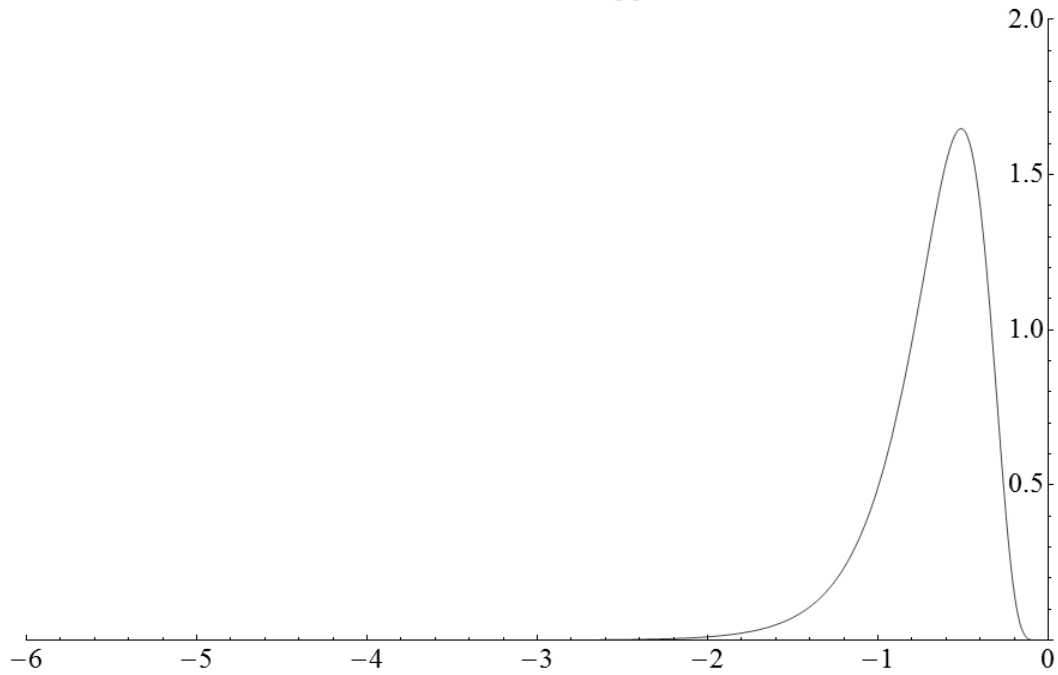


Figure 6: Mean Happiness and National Income, Log-Normal Distribution

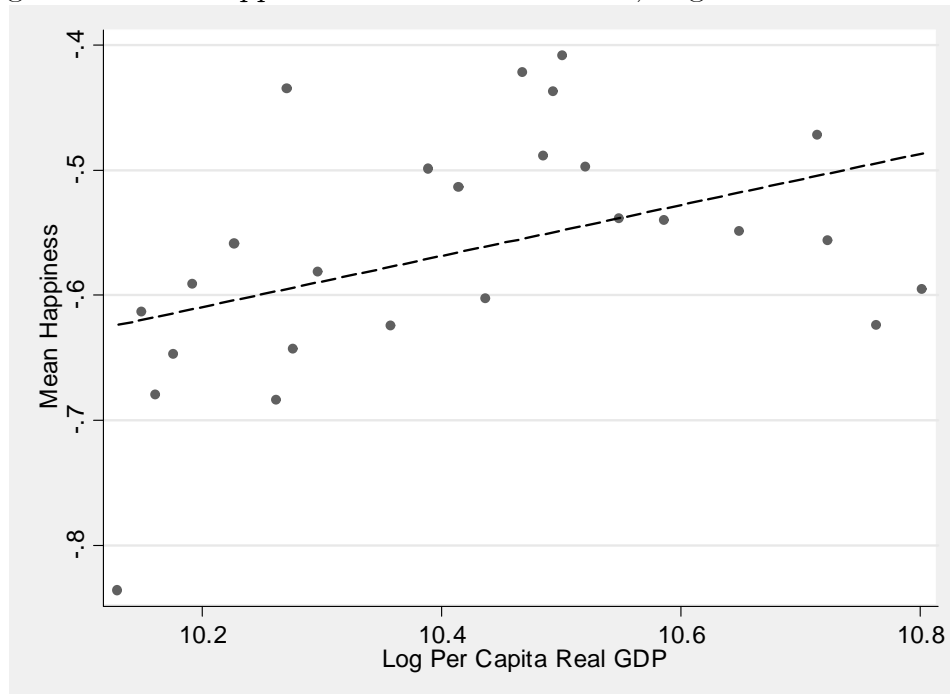


Table 1: Distribution of Happiness - Moving to Opportunities

	Control Compliers	Experimental Compliers
Very Happy	0.242	0.262
Pretty Happy	0.470	0.564
Not Too Happy	0.288	0.174

Source: Ludwig et al (2013), Appendix Table 7.

Experimental estimates are TOT.

Table 2: Distribution of Happiness - General Social Survey

	Male	Female
Panel A: 1972-1976		
Very Happy	0.337	0.384
Pretty Happy	0.530	0.493
Not Too Happy	0.132	0.122
Normal Mean	0.727	0.798
Normal Variance	0.424	0.471
Panel B: 1998-2006		
Very Happy	0.330	0.339
Pretty Happy	0.566	0.553
Not Too Happy	0.104	0.109
Normal Mean	0.742	0.748
Normal Variance	0.346	0.367

Source: General Social Survey Stevenson-Wolfers file.

Normal means and variances calculated from answers under assumption that happiness follows a normal distribution with separate means and variances

Table 3: Country Rankings of Mean Happiness under Log-Normal Distributions

	C=2	C=0.5	C=-0.5	C=-2.0
Mexico	3	2	1	20
Trinidad and Tobago	4	3	5	36
Great Britain	8	6	2	5
Ghana	1	1	26	55
Colombia	6	4	9	33
Canada	12	8	3	3
Sweden	19	9	4	2
Switzerland	14	10	7	8
Netherlands	15	11	8	6
New Zealand	27	14	6	1
Thailand	16	13	11	9
Guatemala	2	5	30	49
Norway	29	16	10	4
Malaysia	25	17	12	7
South Africa	5	7	31	48
France	20	19	15	17
Australia	22	20	14	13
United States	28	21	13	10
Mali	9	12	23	39
Turkey	11	15	20	29
Cyprus	13	18	19	26
Brazil	23	22	16	16
Argentina	24	23	22	23

Table 3 Continued

Finland	32	24	18	14
Andorra	35	26	17	11
Japan	31	27	24	21
Indonesia	36	30	21	12
Uruguay	26	25	28	27
Jordan	30	28	27	24
Viet Nam	39	33	25	15
Poland	40	34	29	18
Chile	18	29	35	42
Italy	44	39	32	22
Taiwan	38	38	34	30
Spain	45	41	33	19
Morocco	33	36	37	38
India	17	32	43	45
Burkina Faso	34	37	39	40
Germany	41	40	36	31
South Korea	46	45	38	25
Slovenia	43	43	41	35
Iran	42	44	42	37
China	37	42	45	43
Rwanda	47	46	40	28
Peru	10	35	48	53
Egypt	52	47	44	34
Hong Kong	55	49	46	32
Ethiopia	7	31	52	57
Ukraine	49	48	47	41

Table 3 Continued

Russian Federation	51	52	49	44
Georgia	48	51	50	46
Serbia	53	53	51	47
Zambia	21	50	55	56
Bulgaria	50	54	53	52
Romania	56	55	54	50
Iraq	57	57	56	51
Moldova	54	56	57	54

Rank of estimated country mean happiness under various log-normal transformations. Countries listed in order of estimated mean happiness under normal distribution.

Source: World Values Survey 2005.