The Sad Truth About Happiness Scales*

Timothy N. Bond† Kevin Lang‡

August 31, 2016

Abstract

Happiness research typically assumes happiness can be cardinalized to be distributed normally in each group studied. We show that under this assumption, we can never rank groups by average happiness. The CDFs will (almost) always cross when estimated using large samples. Moreover, there is always a cardinalization in the lognormal family that reverses the result. We summarize an extensive online appendix: many surprising results in this literature can be reversed by assuming a moderately left-skewed lognormal distribution; we can reject the joint assumption of a normal distribution and a common reporting function for the disabled and non-disabled.

JEL Codes: D60, I31, O15

*We are grateful to Christopher Barrington-Leigh, Larry Katz, Jeff Liebman, Jens Ludwig, Andy Oswald, Justin Wolfers, Miguel Sarzosa, and participants in seminars at Boston University, Indiana University-Purdue University Indianapolis, the Federal Reserve Bank of New York, Georgetown University, McGill University, the University of Michigan and the University of Waterloo, an informal brownbag lunch at Purdue University and the referees and editor 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 reporting their happiness on a scale with few categories, they place it 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.

Consider 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 underlying distribution (e.g. 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 means 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 these cardinalizations all produce the same ordering of the means. However, under conditions made precise later, establishing

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

first-order stochastic dominance requires that the estimated variances be identical, which is an essentially zero-probability event. Moreover, even if the estimated variances are identical, both are estimates subject to error, so our posterior that they are identical must still be 0. Thus, all statements of ordering between the two groups depend on the researcher's cardinalization of happiness, any of which is arbitrary and equally defensible from the data.

Our critique is separate from concerns about the reporting function. Even if *all* individuals report their happiness on a survey *in the same way*, we still cannot compare mean happiness. We are sympathetic to concerns about reporting differences, and the online appendix provides evidence of their importance, but our principal message does not rely on their existence.

We recognize that there are strong conditions under which the distribution of happiness would be nonparametrically identified. As we will discuss briefly, a key requirement is that there is a variable that is known to affect *either* happiness *or* the reporting of happiness and is orthogonal to unobservable factors which influence happiness. We know of no strong candidate for a variable satisfying this restriction, and it is not an approach that we are aware of any happiness researcher pursuing.

We also do not argue that self-reported happiness is a poor measure of well-being.[2] 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.[3]

[2] For a much more detailed and sophisticated discussion of measurement error in reported subjective well-being, see Bertrand and Mullainathan (2001). See Bound, Brown and Mathiowetz (2001) for a broader discussion of measurement error in subjective survey measures.

[3] For 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.

In fact it does not matter what happiness surveys measure for our criticism to be valid.[4] 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 focus on subjective pain or cleanliness grades at restaurants, we choose happiness measures because they are widely used to make between-group comparisons intended to influence policy.

In principle, the problems associated with ordinal scales can be solved if we are willing to anchor the happiness scale to some outcome measure. But as the parallel with the analysis of test scores shows, our conclusions may depend on whether we relate the underlying happiness measure to the probability of committing suicide or some other outcome.[5] Moreover it is not clear to us why we would not prefer to measure the related outcomes directly. As discussed in the online appendix, 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 is to place *ex ante* restrictions on plausible cardinalizations 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 a key area of debate in happiness research.

While our focus is on positive not normative economics, 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.

Our point is not merely theoretical. In the online appendix, summarized in section 3, we replicate a large number of results from the happiness literature. In almost every case,

[4]Therefore, we use happiness, life satisfaction and utility interchangably. There is a literature (e.g. Frey and Stutzer, 2003) that distinguishes among them. We take no stance on this issue.

[5]See, for example, Cunha and Heckman (2008), Cunha, Heckman, and Schennach (2010), and Bond and Lang (2014)

we can reverse the finding by assuming happiness is distributed as either a standard or left-skewed lognormal. Three major anomalies, the Easterlin paradox, the female happiness paradox and that children reduce happiness, can be eliminated by the assumption of a plausible left-skewed lognormal. Some other results are also not robust. When comparing the life satisfaction of the disabled and non-disabled, we can reject the joint hypothesis of a common reporting function and a distribution of happiness in the normal family (including lognormal).

2 The Inevitable Failure of Stochastic Dominance

We assume initially that individuals use a common reporting function. We do not believe that this assumption is plausible, only that the purposes of this section do not require allowing for stochastic and/or systematic differences in the reporting function. We return to the reporting function later in the section.

2.1 Categories v. the Latent Variable

Suppose we know the distribution of happiness on some ordinal scale for two distinct groups, say men and women. It is by now well known that if we use only the ordinal properties of the scale, we can rank their average happiness if and only if the distribution of one group is higher in the sense of first order stochastic dominance (FOSD).

There is only a limited sense in which this result carries over when responses represent intervals rather than a single point. Suppose, for example, respondents are asked to describe themselves as "very happy," "somewhat happy" or "not very happy." Most people will recognize that, even ignoring differences in reporting functions, it is unlikely that all "not too happy" people are equally unhappy. The three responses should be understood as representing intervals rather than points on a scale. We have been told several times that while this point is correct, experience shows that techniques appropriate for ordered data such as ordered probit or ordered logit almost always give essentially the same result as the naive approach of valuing the responses as 0, 1 and 2. This is not surprising. If we estimate a single ordered probit (or logit) with a dummy variable for male, stochastic dominance in the categories is sufficient to ensure that the coefficient on the male dummy has the same sign regardless of our choice of scale and whether we use ordered probit or linear regression.

However, the standard application of ordered probit assumes that the effect of being male is solely to shift the mean and not any other aspect of the distribution. But this is a strong and, at least *a priori,* untenable assumption. Consider the following extreme case portrayed in example 1. 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." At first blush group A appears happier, on average. But, normalizing the cutoffs to be 0 and 1 and assuming a normal distribution, only if the variance is infinite can no one have happiness between 0 and 1. With more observations to the left than to the right of 0, 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 1

	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.

2.2 Reversing Rankings: The Normal Distribution

Even if estimated mean happiness changes in the same direction as the movement among categories, the distributions of happiness will rarely be ranked in the sense of FOSD. Consider example 2. Again group B appears happier than group A. But let us assume that happiness is normally distributed and normalize the cutoffs to 0 and 1 as before. Now estimated mean happiness for group B (.60) is indeed above the estimated mean for group A (.50), but the standard deviation is also larger (.68 v .59) so that the happiness distributions cross at the 13th percentile.[6] The results are similar if we assume happiness is logistically distributed.

Example 2

	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 the normal or logistic distribution, that group B is happier on average. Unfortunately, this conclusion necessarily holds only for specific cardinalizations of happiness, in this case the one that produces a normal (or logistic) distribution of happiness for both groups.[7] The data could be just as well represented by any other cardinalization produced by monotonically transforming the

[6] Using an LR test, we would reject equality of the variances at rhe .1 level with a sample of somewhat fewer than 350/group, at the .05 level with somewhat fewer than 500/group and at the .01 level with a sample of somewhat fewer than 800/group.

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

underlying happiness data. And given that the distributions cross, there are an infinite number of other equally correct alternative cardinalizations which reverse the ordering of mean happiness.

Suppose we instead estimate a lognormal distribution, by transforming the utilities by e^X. Our new mean is

$$e^{\mu+.5\sigma^2}. \tag{1}$$

If we compare two groups, one with a higher mean when estimated normally and the other a higher variance, this transformation alone could reverse the ranking obtained using the normal distribution. If not, we can multiply our latent happiness variable by a positive constant c before exponentiating. The mean under the lognormal transformation would thus become

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

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 lognormal distribution. The mean of happiness becomes

$$\overline{\mu} = -e^{-c\mu+.5c^2\sigma^2} \tag{3}$$

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.[8]

[8] Our analysis is related to that of Hammond, Liberini and Proto (2013) who suggest ordering polices 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.

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 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 on the dispersion of happiness. We can generate plausible examples in which different assumptions about the distribution of happiness change our conclusions about relative variances, but we do not claim there is always an easy transformation that generates a reversal.

2.3 Reversing Rankings: Two-Parameter Distributions

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.[9]

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

[9] Here 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.

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 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.[10] 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 this 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.[11] Most importantly, it includes the normal and logistic distribution, the distributions underlying ordered probit and 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 differs between the groups. Then there also exists a cardinalization $q' \in Q$ under which the ranking of group means is the opposite of that under q^*.*

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

[11] We 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.

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

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

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 equal. This means we cannot strictly rank groups by mean happiness. Any ranking of happiness depends on the cardinalization, of which there are infinitely many which are equally defensible from the data.

How often is there 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.

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* $L = \Sigma_{j \in G_i} \Sigma_c d_c \ln F_j^c(m_i, s_i)$ *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 $\widehat{m}_i, \widehat{s}_i$ be the parameter estimates that maximize the estimated likelihood, then

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

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.[12]

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.*

We note that such a restriction might take the form of assuming that the distribution does not fit our definition of a standard distribution, but that this will not necessarily solve the problem.

2.4 Further Complications: The Reporting Function

Our focus is on the limits imposed by ordinal scales reported in categories. Nevertheless, the fact that we rely on self-reports rather than observing "true happiness" adds complexity to the problem. It is a corollary of our earlier discussion that if true happiness is normally distributed with different variances between groups, then there is always a reporting function

[12] The intuition behind this is relatively straightforward. The variances are estimated by maximum likelihood and are thus asymptotically normal as is their difference which means that the probability of the estimate falling in a small range around 0 is small provided that range shrinks appropriately as sample size grows. More formally, 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 $\widehat{s}_A \rightarrow^d N\left(s, \sigma_A^2\right)$ and $\widehat{s}_B \rightarrow^d N\left(s, \sigma_B^2\right)$. Define $\alpha = \widehat{s}_A - \widehat{s}_B$. Then since $\widehat{s}_A$ and $\widehat{s}_B$ are asymptotically independent normals, $\alpha \rightarrow^d N\left(0, \sigma^2\right)$ where $\sigma \equiv \left(\sigma_A + \sigma_B\right)^{.5}$. 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 $2\varepsilon\sigma\left(2\pi\sigma^2\right)^{-.5} = (2/\pi)^{.5}\varepsilon$ 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.

that transforms true happiness into reported happiness such that the difference in mean reported happiness has the opposite sign from the true difference.

The problem becomes more severe when reporting functions differ across individuals. To see this, let the reported happiness of individual i in group j be

$$h_i^j = \overline{h}^j + \varepsilon_i^j + \nu_i \tag{4}$$

where $\overline{h}^j$ is the group mean, ε_i^j is the deviation from the group mean in true happiness and ν is the deviation from the standard report. Note that while the distribution of true happiness may differ between groups, reporting error is not group dependent. Nevertheless, even if the resulting true and reported distributions can be cardinalized by a standard two parameter distribution, we will generally estimate the points where the CDFs cross incorrectly.

In the special case where there is a cardinalization under which the happiness of two groups are each normal and reported with a mean-zero normally distributed 'error' with a common variance, our estimate of the c (shape of the lognormal) that reverses the ranking of the means remains consistent. However, if the variances of the reporting function error are group specific, this is no longer true. Thus even if at a given level of happiness, all groups report the same happiness on average, reporting function differences are problematic.

Of course, the problem is even more severe if the reporting function differs across groups, and there is ample reason to be concerned that it does (see apendix section A.3.9). King et al. (2004) propose circumventing reporting function differences by using 'vignettes' to anchor the scale on which people report their happiness. Applications include Peracchi and Rossetti (2013) and Ravallion et. al (2013). This approach requires that (1) individuals evaluate the vignettes on the same scale as they evaluate their own well-being ('response consistency'), and (2) each individual perceives the same value from the vignettes ('vignette equivalence'). These assumptions are strong; given that preferences are heterogenous, the second assumption is particularly unlikely to hold. Moreover, this approach does not address the issue we raise. Even if we were able to place all individuals' happiness onto a common scale, this is just one of an infinite number of equivalent scales, and other scales will be able to reverse group rankings.

If we are willing to impose a functional form for happiness, with adequate data it is possible to test for a common reporting function. Suppose, for example, we believe that happiness is normally distributed. Our data consist of happiness reported in $N \geq 4$ categories. We normalize the first two cutoffs to be 0 and 1. This is sufficient to allow us to identify the mean and variance of the distributions and leaves $N - 3$ cutoffs free. Under the maintained assumption that there is a cardinalization consistent with two normal distributions and common reporting function, the free cutoffs should be equal across groups.

This is easily tested using standard techniques. In the appendix, we show that it is rejected for a sample of disabled and non-disabled respondents. Of course, the researcher is free to conclude either that her initial conjecture about the cardinalization was wrong[13] or that the reporting functions differ, but we believe that testing the joint hypothesis should become standard procedure in this literature.

2.5 Conditions for Nonparametric Identification

It is impossible to have stochastic dominance, and thus rank two groups in mean happiness, when assuming what we have called a standard two-parameter distribution. It is possible, at least in theory, that we could obtain a stochastic dominance result if we could avoid such assumptions. In this subsection we briefly review the conditions under which the underlying distribution can be identified nonparametrically. The main point is that, at least at the current state of knowledge, the conditions should be understood either as excessively strong or as an arbitrary approach to choosing a specific cardinalization.

Several papers in the literature on latent variables, for example Manski (1988), Lewbel (2000) and Carneiro, Heckman, and Hansen (2003), have shown that one can use variation

[13] Note that other cardinalizations with cutoffs that are monotonic transformations of the normal cutoffs are also rejected.

in a continuous covariate to nonparametrically identify the distribution of a latent variable which is observed only in discrete, ordered categories. In the simplest case outlined by Lewbel, given the existence of a continuous covariates ν and x, the distribution of some variable u reported in categories can be nonparametrically identified if $u_i = v_i + x_i\beta + e_i$ (or $-v_i + x_i\beta + e_i$), the distribution of e is independent of v, and the support of $-x\beta - e$ is a subset of the support of the conditional distribution of v given x. These exact conditions can be modified and relaxed in different contexts, but in general to identify a latent variable distribution using variation in a continuous covariate, one must find a v that is orthogonal to unobservables and has a sufficiently large support, and place a restriction on the relation between v and u.[14]

We make a number of minor remarks:

1. While imposing a coefficient of 1 or -1 on v is simply a normalization, assuming that the effect of v is linear (or any other restriction) is not. In our context, it amounts to choosing a particular cardinalization of the happiness function. Thus, unless we had a strong prior on the true form of the relation between u and v, we would still require stochastic dominance to rank two groups in mean happiness.

2. The support condition in the happiness context is strong. It says that some observed variable must have an effect that spans the effect of all the unobservables. So if there are unmeasured determinants of happiness such as marital quality which range from newlywed bliss to intense fear of spousal abuse, there must be a single measured determinant that affects happiness sufficiently to outweigh the full range of potential effects of the unmeasured determinants. We are unaware of a strong candidate to fill this role.

3. In principle, it does not matter whether v influences the true underlying latent variable or its reporting. In the latter case, we would then be interested in identifying the

[14] For example, Matzkin (1992) shows that one can identify the distribution of a latent variable nonparametrially while imposing only that the relationship between v and u is a member of a functional familiy which satsifies a set of conditions, all of which are satisfied by linear functions.

distribution of e or perhaps $x\beta + e$, depending on what is contained in x. Of course, our estimate of the difference in means and the distributions will depend crucially on whether v (and x) affects the reporting function or the true latent variable, something for which it is unlikely we can obtain guidance from the data.[15]

4. It is similarly difficult to imagine a variable which affects either reporting or happiness and which is orthogonal to all the unmeasured factors that determine happiness.

We thus know of no obvious candidate for v in the happiness literature. Therefore we conclude that, at best, comparisons of happiness distributions across groups will be dependent on our ability to find acceptable restrictions on the distribution of happiness.

3 Empirical Applications

The online appendix contains detailed assessments of the robustness of important results in the happiness literature. We briefly summarize these results and refer the reader to the appendix for more detail. In each case, we begin by estimating the distribution of happiness using an ordered probit and thus assuming that happiness is normally distributed allowing both the mean and variance to differ among the groups we are comparing. We then determine the minimal c that using a transformation to a lognormal or left-skewed lognormal would reverse our conclusions. We view values of $c \leq 2$ as plausible. At this value, happiness is no more skewed than the income distribution. Somewhat higher values may also be plausible. We leave that judgment to the reader although we find any $c > 3$ quite implausible.

Note that we allow *only* the mean and variance to differ among groups. It is quite possible that skewness differs as well. In the context of the lognormal distribution, this would require

[15] Cunha, Heckman, and Navarro (2007) show a set of conditions under which one can nonparametrically identify a latent variable model when there is unobservable differences in individuals' reporting functions. However, their approach requires imposing that there are no unobservable individual differences in the latent variable. In their context, in which they are interested in the return to observable variables on an economic parameter of known scale (e.g. years of education), this is simply a normalization. In our context, this is an unreasonably strong assumption that differences in responses to a happiness survey by two observably identical people can only be due to differences in reporting functions.

allowing c to differ across groups, which is isomorphic to permitting the reporting functions to differ. The data for one of our replications is sufficient to allow us to test and reject the joint hypothesis that the two groups we study share a common reporting function and have happiness distributions drawn from the normal family (including lognormal). No such test is available in the remaining cases. We find the frequency with which we can reverse standard results even without relying on reporting function differences to be striking.

Given the restriction we put on a plausible c, only one regularity we study is fully robust: we find (in the General Social Survey [GSS]) that married individuals are happier than those who never married. We can only reverse this result with an implausibly left-skewed lognormal. (See section A.3.6)

As we discuss in the conclusion, we find left-skewness more plausible than right-skewness, both on *a priori* grounds and because it eliminates several surprising results in the happiness literature. If we rule out right-skewness, the finding in the Moving to Opportunity experiment, that individuals who move out of poor neighborhoods are substantially happier than those who do not (Ludwig et. al 2012) is also robust. (See section A.3.5)

The remaining regularities can be eliminated, and in some cases reversed, if we assume a plausible left-skewed lognormal distribution.

We begin with the Easterlin Paradox, the observation by Easterlin (1973, 1974) that higher levels of national income are, if anything, associated with lower national happiness. Using data from the GSS, we confirm that for the United States from 1973 to 2006, per capita income and mean happiness are negatively correlated when we assume that happiness is distributed normally. But there is also a negative correlation between the variance of happiness and per capita income, which means that the relation can be reversed by a left-skewed lognormal distribution. The requisite distribution to reverse the sign is mild, $c = 0.7$, although generating a positive relation that is significant at the 5% level requires $c = 2.6$; this results in a happiness distribution that is still less skewed than the wealth distribution in the United States, although in the opposite direction. (See section A.3.1)

We obtain a similar result regarding the robustness of cross-country comparisons. Using the World Values Survey (WVS), we rank 57 countries by mean happiness. Under normality, we find that Mexico, Trinidad and Tobago, Great Britain, Ghana and Colombia are the five happiest countries in the world. When happiness is right-skewed ($c = 2.0$), Ghana becomes the happiest country in the world, Guatemala moves from 12 to 2, and Great Britain drops to 8. When happiness is left-skewed ($c = 2.0$). Mexico falls from first to 20th, Trinidad and Tobago falls to 36th, and Ghana falls all the way to 55th. The top 5 countries become New Zealand, Sweden, Canada, Norway, and Great Britain, all highly economically developed countries. When happiness is left-skewed there is a strong positive relation between GDP per capita and mean happiness, but when it is right-skewed there is a strong negative relation. (See section A.3.4)

We also study the 'paradox' discovered by Stevenson and Wolfers (2009), that women's relative happiness has declined over the last 40 years despite their great economic progress. Using the Stevenson/Wolfers data, we confirm that women's happiness decreased relative to men's in the GSS, but also find evidence that their relative variance declined, although the difference is not statistically significant. As a consequence, we can eliminate the statistical significance of the 'paradox' with a mildly left-skewed lognormal ($c = .75$); however, we require $c = 3.75$ to reverse the point estimate. (see section A.3.7)

We also reexamine attempts to use happiness data to correct the "misery index," society's weighting of the inflation and unemployment rates. Using the Eurobarometer and data on unemployment and inflation from the OECD, we find that a one percentage point increase in unemployment has a negative impact on average happiness 1.91 times as large as a one percentage point increase in inflation, a slightly larger effect than in Di Tella, MacCulloch, and Oswald's (2001) seminal paper. However, we also find that unemployment has a statistically significant and positive impact on the variance of happiness. Thus when we skew happiness to the right, the importance of unemployment dissipates. At $c = 1.125$, the misery index is optimal. At $c = 2.0$, the estimated impact of unemployment on happiness is approximately zero. At $c = 2.5$, our point estimate suggests the unemployment rate *increases* mean happiness. Neither inflation nor unemployment is ever statistically significant under right-skewed transformations. More importantly, the effect of inflation also becomes insignificant when we skew left. Thus the adverse effect of inflation on happiness is not robust. At the same time, the effect of the unemployment rate increases. At $c = 2.0$, a one percentage point increase in unemployment has the same negative effect on happiness as a three percentage point increase in inflation. (See section A.3.3)

We also explore the relation between life events and happiness. Blanchflower and Oswald (2008) find a robust, cross-country U-shaped relation between happiness and age. Using data on 16 countries from the Eurobarometer, we find fairly supportive evidence for this conclusion when happiness is assumed to be normally distributed. The pattern is only obviously violated in 3 cases. However, only 2 of the 16 countries continue to show U-shape under both right and left skewness ($c = 2.0$). These transformations produce both strictly increasing and strictly decreasing age-happiness profiles for some countries. In at least two countries, our transformations allow for both a U-shape and its inverse. With the left-skewness assumption, at most four countries display the posited U-shape. (See section A.3.2)

We also find, again consistent with the literature, that people who live with children are less happy than those who do not. Without controls, this is difficult to reverse. However, once we include a full set of demographic controls, we can reverse the result with a mild left-skewed transformation ($c = 1.20$). (See section A.3.6)

In our final exercises, we explore issues related to the reporting function. We begin by asking whether there is robust evidence that individuals adapt to disabilities, as suggested by Brickman et al (1978). Using the British Household Panel Survey (BHPS), we find only weak evidence for adaptation when happiness is distributed normally. Comparing the distribution of happiness for the same set of individuals both N and $N + 1$ years after they became disabled, we can always easily reverse the results for any N for which we estimate happiness to have increased when we assume normality. More strikingly, the BHPS measures life satisfaction on a 7 point scale, which allows us to test whether, under the maintained assumption that the underlying distribution is normal or lognormal, individuals with and without disabilities used different reporting functions. We find strong evidence for a heterogeneous reporting function. In particular, it appears that relative to the rest of the population, the disabled require a much lower level of life satisfaction to report being in the top categories of life satisfaction. (See section A.3.8)

Building on this result, we discuss evidence on the stability of the reporting function and present our own evidence using data on subjective income. Using the WVS, which asked individuals to report, on a 10-point scale, to which income group in their country they belonged, we find substantial heterogeneity in the responses across countries. Even when we compare just Australia and New Zealand, which have similar income inequality across a broad spectrum of measures, the distribution of responses differs dramatically, suggesting that individuals use a different reporting function in these two countries. (See section A.3.9)

4 Discussion and Conclusions

As we have demonstrated, key conclusions of happiness studies depend on the chosen cardinalization of happiness, something about which the data give little or no guidance. Our review of the literature on nonparametric identification of ordered response models suggests that this route is not likely to be productive for happiness research. Since, using standard parametric assumptions, the estimated CDFs (almost) always cross, there is always some transformation that preserves the rank order of individuals and changes the direction of the estimated gap in mean happiness.

We believe that our results create a compelling case against rapidly adopting happiness measures as a basis for policy. But, despite accusations to the contrary, we do not seek to be nihilists. We have shown that some puzzling results can be eliminated or even reversed if the underlying distribution of happiness is a plausibly left-skewed lognormal. And in no case did this assumption generate a, to us, implausible result. It is possible the profession could achieve near consensus about reasonable restrictions on the happiness distribution that would allow us to reach strong conclusions about the ranking of mean happiness in some cases. One possibility is what we call the "Tolstoy assumption," that there is far greater variation in unhappiness than in happiness.[16] In other words, happiness is left-skewed.

Even if we agreed that, for example, the happiness distribution can be well-represented by a left-skewed lognormal, equal skewness in all groups would be a very strong assumption and more or less isomorphic to assuming a common reporting function.[17] With only three categories, we cannot test for differences in reporting functions directly. Good practice should therefore rely on scales with several categories and test whether a common reporting function is consistent with the researcher's assumption about the underlying happiness distribution. Assuming reasonable power, a finding that failed to reject a common reporting function and was robust to reasonable modifications of the assumption about the underlying distribution of happiness should stand a good chance of broad acceptance.

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. Under 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. But we begin to venture into moral philosophy, an area in which we have no expertise.

[16]We 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."

[17]The skewness of the lognormal is determined by the values of the cutoff points. But allowing different cutoffs is equivalent to allowing different reporting functions.

References

[1] Bertrand, Marianne and Sendhil Mullainathan. 2001 "Do People Mean What They Say? Implications for Subjective Survey Data," *American Economic Review,* 91 (2): 67-72.

[2] 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.

[3] Bond, Timothy N. and Kevin Lang. 2014, "The Black-White Education-Scaled Test-Score Gap in Grades K-7," 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] Brickman, Phillip, Dan Coates, and Ronnie Janoff-Bulman. 1978. "Lottery Winners and Accident Victims: Is Happiness Relative?" *Journal of Personality and Social Psychology*, 36 (8): 917-927.

[6] Cantril, Hadley. 1965. *The Pattern of Human Concerns*. New Brunswick, NJ: Rutgers University Press.

[7] Carneiro, Pedro, Karsten T. Hansen, and James J. Heckman. 2003. "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice." *International Economic Review*, 44 (2): 361-422.

[8] 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.

[9] 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.

[10] Cunha, Flavio, James J. Heckman, and Salvador Navarro. 2007. "The Identification and Economic Content of Ordered Choice Models with Stochastic Thresholds." *International Economic Review*, 48 (4): 1273-1309.

[11] Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica,* 78 (3): 883-931.

[12] Di Tella, Rafael, Robert J. MacCulloch, and Andrew J. Oswald. 2001. "Preferences over Inflation and Unemployment: Evidence from Surveys of Happiness." *American Economic Review*, 91 (1): 335-341.

[13] Easterlin, Richard A. 1973. "Does Money Buy Happiness?" *The Public Interest*, 30 (3): 3-10.

[14] 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.

[15] Frey, Bruno S. and Alois Stutzer. 2003. "Maximising Happiness?" *German Economic Review*, 1 (2): 145-67.

[16] 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.

[17] 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.

[18] Layard, Richard, Guy Mayraz, and Stephen Nickell. 2010. "Does Relative Income Matter? Are the Critics Right?" in Ed Diener, Daniel Kahneman, and John Helliwell (eds.) *International Differences in Well-Being*. Oxford University Press: Oxford. 139-165.

[19] Lewbel, Arthur. 2000. "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables." *Journal of Econometrics*, 97: 145-177.

[20] 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.

[21] 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.

[22] Manski, Charles F. 1988. "Identification of Binary Response Models." *Journal of the American Statistical Association*, 83 (403): 729-738.

[23] Manski, Charles F., and Elie Tamer. 2002. "Inference on Regressions with Interval Data on a Regressor or Outcome." *Econometrica*, 70 (2): 519-546.

[24] Matzkin, Rosa L. 1992. "Nonparametric and Distribution-Free estimation of the Binary Threshold Crossing and the Binary Choice Models." *Econometrica*, 60 (2): 239-270.

[25] 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.

[26] Oswald, Andrew J. 2008, "On the Curvature of the Reporting Function from Objective Reality to Subjective Feelings," *Economics Letters*, 100 (3): 369–372.

[27] 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.

[28] Ravallion, Martin, Kristen Himelein, and Kathleen Beegle. 2013. "Can Subjective Questions on Economic Welfare be Trusted? Evidence for Three Developing Countries," unpublished.

[29] Ritter, Joseph A. and Lowell J. Taylor. 2011. "Racial Disparity in Unemployment," *Review of Economics and Statistics*, 93 (1): 30-42.

[30] Spencer, Bruce D. 1983. "On Interpreting Test Scores as Social Indicators: Statistical Considerations," *Journal of Educational Measurement*, 20 (4): 317-333.

[31] Stevens, S.S. 1946. "On the Theory of Scales and Measurement," *Science*, 103 (2684): 677-680.

[32] Stevenson, Betsey and Justin Wolfers. 2009. "The Paradox of Declining Female Happiness." *American Economic Journal: Economic Policy*, 1(2): 190-225.

[33] Thorndike, Robert L. 1966. "Intellectual Status and Intellectual Growth." *Journal of Educational Psychology*, 57 (3): 121-127.

A Online Empirical Appendix

In the main text, we proved that if there exists a cardinalization under which the happiness of each of two groups is distributed normally with a different mean and variance for each group, and group A has a higher mean than group B, then there also exists a cardinalization of happiness using the family of log-normal distributions under which group B has a higher mean than group A.

This appendix explores empirically the types of cardinalizations that are required to reverse the major results in the happiness literature. While as demonstrated theoretically there is no single result that is irreversible, some results require more dramatic deviations from normality than others. Additionally, some results are only reversible by skewing the distribution to the left, while others are only reversible by skewing the distribution to the right. Our exercises indicate that the set of results one can claim from the happiness literature is highly dependent on what one's beliefs are on the underlying distribution of happiness in society, or the social welfare function one chooses to adopt.

This appendix is organized as follows. In section A.1, we discuss the methods we will use to implement our assessment. Section A.2 reviews our data sources. Our main results lie in section A.3, where we determine the cardinalizations (within the log-normal family) that reverse nine key results from the happiness literature: the Easterlin Paradox, the 'U-shaped' relation between happiness and age, the happiness trade-off between inflation and unemployment, cross-country comparisons of happiness, the impact of the Moving to Opportunity program on happiness, the impact of marriage and children on happiness, the 'paradox' of declining female happiness, and the effect of disability on happiness. The final subsection of A.3 discusses issues related to the reporting function, and presents evidence that the reporting function for subjective income differs across countries. Section A.4 concludes.

A.1 Methods

A.1.1 Estimation

We begin in each case by assuming happiness is normally distributed within each group, but allow its mean and variance to differ across groups. There is, of course, no guarantee that happiness can be cardinalized so that this assumption holds; as we examine more groups, this assumption becomes increasingly fragile. However, almost all papers in the happiness literature that recognize the ordinal nature of the data assume normality. If normality is violated, the vast majority of happiness results would automatically be suspect.

When happiness is recorded on a three-point scale, it is straightforward to calculate the means and variances under normality as the model is just identified (see section 2 of the main text). In many of our applications, we will be interested in estimating the distribution of happiness net of factors such as income or employment status that vary across groups. To do this, we estimate a heteroskedastic probit model using the oglm command in STATA created by Williams (2010). Denote $h \in \{0, 1, 2\}$ as an individual's answer to a 3-point subjective well-being survey. The model assumes that h is determined by a latent variable h^*,

$$\begin{aligned} h &= 0 \text{ if } h^* < k_0 \\ h &= 1 \text{ if } k_0 \leq h^* \leq k_1 \\ h &= 2 \text{ if } k_1 < h^* \end{aligned} \tag{1}$$

and that

$$h_i^* = \alpha_m D_i + \beta_m X_i + \varepsilon_i \tag{2}$$

where k_0 and k_1 are cut-off values of the latent variable that determine the observed response, D_i is an indicator for the group we are studying, X_i is a vector of individual specific controls and ε_i is a normally distributed error term with mean 0 and variance σ_i with

$$\sigma_i = \bar{\sigma} \exp(\alpha_s D_i + \beta_s X_i) \tag{3}$$

In other words, the model varies from the classic ordered probit in that it allows the observable characteristics to influence the variance of the error term in the latent variable.

Just as with a textbook ordered probit, one cannot separately identify the cut points from the variance. The oglm routine normalizes $\bar{\sigma} = 1$ and estimates $\alpha_s, \beta_s, a_m \equiv (\alpha_m/\bar{\sigma})$, and $b_m \equiv (\beta_m/\bar{\sigma})$. It easy to transform the oglm estimates into their equivalents under our preferred normalization, where $k_0 = 0$ and $k_1 = 1$. The estimated mean and variance for our control group $D_i = 0$ will be

$$\hat{\mu}_0 = -\hat{k}_0(\hat{\sigma}_o) \tag{4}$$

$$\hat{\sigma}_0 = \frac{1}{\hat{k}_1 - \hat{k}_0} \tag{5}$$

and for $D_i = 1$,

$$\hat{\mu}_1 = \hat{\sigma}_0 \hat{a}_m + \hat{\mu}_0 \tag{6}$$

$$\hat{\sigma}_1 = \hat{\sigma}_0 \exp(\hat{a}_s) \tag{7}$$

These produce estimates for the mean and variance of the distribution only at a specific set of values for controls, namely $X = 0$. In each of our exercises we will de-mean the controls, so our estimates could be thought of as characterizing the distribution of happiness for individuals who differ in their group membership, but otherwise possess the mean characteristics throughout the entire population (regardless of group membership).

This method extends easily to more groups.

A.1.2 Applying Other Cardinalizations

Once we have estimated μ and σ under the normality assumption, we can re-cardinalize happiness. We limit attention to cardinalizations that transform the distribution from normal to left-skewed and right-skewed log-normal distributions. Recall from the main text that for a given constant c, the mean of happiness once re-cardinalized to a right-skewed log-normal distribution will be

$$\mu^{\tau} = \exp(c\mu + \frac{1}{2}c^2\sigma^2) \tag{8}$$

while for a left-skewed log-normal

$$\mu^{\tau} = -\exp(-c\mu + \frac{1}{2}c^2\sigma^2) \tag{9}$$

As the transformed mean of the right-skewed log-normal distribution is increasing in the variance of the normal cardinalization, and the mean of left-skewed log-normal distribution is decreasing in the variance of the normal cardinalization, there will always be a c such that one of these transformations reverses the original ordering of two groups.

A.1.3 Assessing Robustness

We ask whether the choice of c that reverses the conclusion produces a 'reasonable' distribution of happiness. Increasing c increases the skewness of the distribution, which grows rapidly as c gets large. Our assessment of whether the transformed distribution is reasonable depends on our tolerance for skewness in the happiness distribution.

Of course there has been little or no discussion about what features a reasonable happiness distribution would possess. Hence any threshold we choose is *ad hoc.* While recognizing this, we find that log-normal distributions with $c \leq 2$ look reasonable to us and become less comfortable when c is noticeably greater than 2. Using the

General Social Survey data (described below), we estimate that if the distribution of happiness were log-normal with $c = 2$ the mean level of happiness would lie at the 73rd percentile. This compares favorably to the income distribution in the United States where the mean is at the 74th percentile (Diaz-Gimenez, Glover, and Rios-Rull, 2011). Of course others may be willing to tolerate more skewness. The mean of the wealth distribution in the United States is at the 80th percentile; we would not reach that level of skewness until $c = 2.8$. It is conceivable that a policy maker who cared only about helping the worst-off in society would adopt a social welfare function that is even more skewed, but towards the left.

It is important to note that as the c required to reverse a happiness result depends on parameters that are estimated with error, it, too, is estimated with error. Even when we require a c larger than 2, it may not be possible to rule out that the result is reversible by a log-normal transformation that we find palatable. As we showed in the text, when comparing two groups, the c required to reverse the means is

$$c = \left| \frac{2(\mu_1 - \mu_2)}{\sigma_2^2 - \sigma_1^2} \right| \tag{10}$$

A negative term within the absolute value indicates a left-skewed log normal is required. Constructing a confidence interval for $\widehat{c}$ is problematic. When the estimated variances are similar, we may not be able to rule out $\sigma_1^2 = \sigma_2^2$, in which case c is undefined. In such cases, the required transformation may be right- or left-skewed.

To side-step this problem, we use the delta method to test

$$2(\mu_1 - \mu_2) - c^*(\sigma_2^2 - \sigma_1^2) = 0$$

where c^* is our boundary for reasonableness, generally equal to two in our applications. This transformation of the hypothesis ensures that the test-statistic is always defined. We expect this test to be conservative; it should always reject in the problematic cases where the estimated variances are close given there is a significant difference in μ's. Note that when there is not a significant difference in the μ's, we cannot rejected either ordering of the means when the distribution is normal, so the result is already not robust. Restating the test in terms of our parameters estimates, we have

$$2\hat{a}_m - \frac{c^*}{\hat{k}_2 - \hat{k}_1}[1 - \exp(2\hat{\alpha}_s)] = 0$$

which can be estimated using the testnl command in STATA. When a left-skewed log-normal is required, the test becomes

$$2\hat{a}_m + \frac{c^*}{\hat{k}_2 - \hat{k}_1}[1 - \exp(2\hat{\alpha}_s)] = 0.$$

Moreover, any conclusions we draw about the robustness of results in the happiness literature will be conservative. By limiting ourselves to log-normal transformations, we are not necessarily presenting the *most* 'reasonable' distributional assumption that reverses a key result in the literature. Instead we ask whether a palatable, to us, transformation from a very restrictive class of distributions can reverse the result. It is certainly possible that results we have difficulty reversing using a log-normal, can be reversed by some other reasonable distribution.

A.2 Data

We use and describe here data common in the happiness literature.

A.2.1 General Social Survey

The General Social Survey (GSS) is the most widely used data to study happiness in the United States. It has surveyed a nationally representative sample of Americans on a variety of social attitudes annually or biennially since 1972. It asks, "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?" This language and its 3-point scale has been commonly adopted by other studies, including the assessments of the Moving to Opportunities project (MTO) which we discuss in the next section. While the question remains constant over time, its position in the survey does not, which could lead to biases in responses in different years.[1] We therefore use the publicly available replication file provided by Stevenson and Wolfers (2009), who use split-ballot experiments to modify the data to account for these differences.[2]

A.2.2 Eurobarometer Trend File

The Mannheim Eurobarometer Trend File 1970-2002 combines and harmonizes several different annual surveys of the European Community, thus enabling within- and cross-country comparisons over time. The surveys included a question on life satisfaction in 1973 and then continuously from 1975-2002. There were some slight

[1]For example, Stevenson and Wolfers (2009) note that in every year but 1972, the question followed a 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 also Smith (1990).

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

differences in question wording in some years, but in general it asked, "On the whole, are you very satisfied, fairly satisfied, not very satisfied, or not at all satisfied with the life you lead?" While the question measures "life satisfaction" on a four-point scale, very few individuals reported being "not at all satisfied." We therefore combine the lowest two categories, which eases estimation and provides consistency with our analyses in other sections using the standard three-point question from the General Social Survey. The survey included Austria, Belgium, Denmark, Spain, Finland, France, West Germany, the United Kingdom, East Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, and Sweden, but typically only in years when these countries were members of the European Economic Area.

A.2.3 World Values Survey

The 2006 World Values Survey (WVS) is a comprehensive global survey on prevailing beliefs and social attitudes across a large number of nations. The fifth wave was conducted in 2005-2008 and included the following question on happiness, "Taking all things together would you say you are: very happy, rather happy, not very happy, or not at all happy." As with the Eurobarometer, very few people respond in the lowest category, and so we combine the bottom two categories to improve comparability with the GSS.

The survey also ask respondents about their relative income. We discuss this question in section A.3.9.

A.2.4 British Household Panel Survey[3]

The British Household Panel Survey (BHPS) is a panel survey which began in 1991 with a representative sample of 10,300 individuals. The BHPS included a question on life satisfaction in the waves from 1996 to 2008, with the exception of 2001.[4] The survey asked "Here are some questions about how you feel about your life. Please tick the number which you feel best describes how dissatisfied or satisfied you are with the following aspects of your current situation." (Box 1 is marked "Not satisfied at all" while box 7 is marked "completely satisfied.") After questions about particular aspects of life, the survey continues "Using the same scale how dissatisfied or satisfied are you with your life overall?"

A.3 Empirical Results

In this section, we report c without taking its absolute value. Therefore when we report a negative c, we are referring to a left-skewed log normal with c equal to the absolute value of the c to which we refer. Similarly $c = 0$, refers to a standard normal. While there is some risk of confusion, it greatly simplifies presentation to refer to considering values of $c = -2, -.5, 0, .5$ and 2 rather than explaining that the first two are left-skewed log normals, the third is standard normal and the last two right-skewed log normals.

[3] University of Essex. Institute for Social and Economic Research, British Household Panel Survey: Waves 1-18, 1991-2009 [computer file]. 7th Edition. Colchester, Essex: UK Data Archive [distributor], July 2010. SN: 5151.

[4] The 2001 wave surveyed life satisfaction with a question that was worded slightly differently than the other years, and represented the scale with faces rather than simply by boxes. Most researchers have felt these differences were sufficiently minor to ignore. We are less sanguine and did not even download the data from that year.

A.3.1 Easterlin Paradox

No question in the happiness literature has received more attention than the "Easterlin Paradox," the observation that in some settings higher incomes are not associated 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 (1995) 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 individuals judge their happiness relative to their peers and not on an absolute scale.

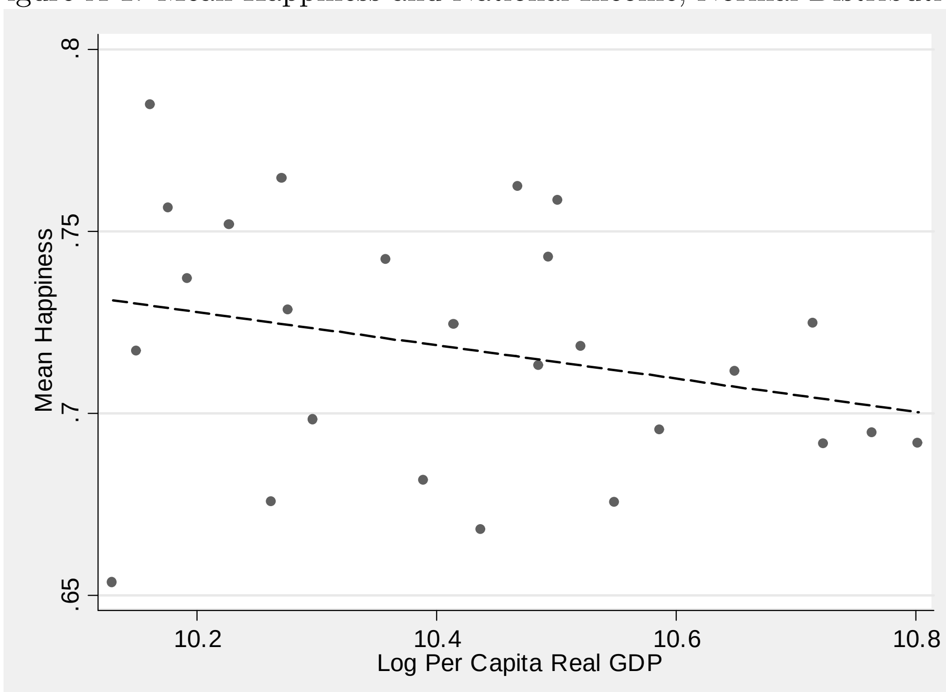
The paradox was recently called into question in a comprehensive study by Stevenson and Wolfers (2008a).[5] They use ordered probit both across countries and over time within countries and find a strong relation between happiness and economic development. However, they find that the United States is an exception. Happiness has not increased despite substantial growth in per capita income. They attribute this to the substantial rise in income inequality over the last 30 years which occurred simultaneously with the rise in real GDP.

To assess this paradox, we first use the GSS to estimate the mean and variance of happiness in the United States assuming normality for each available year from 1973 to 2006. We then match our estimates with U.S. per capita real GDP data from the Federal Reserve Bank of St. Louis and regress mean happiness on the log

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

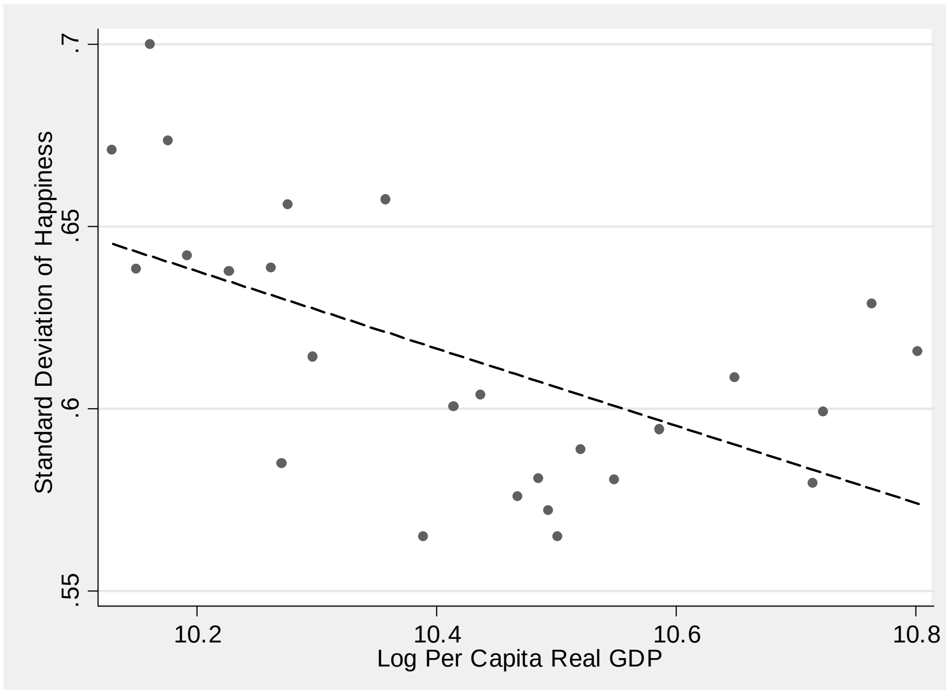
of real GDP per capita. As we show in Figure A-1, we do indeed find an Easterlin Paradox. OLS estimates imply that a 10% increase in GDP per capita is associated with a decrease in average happiness in the United States of .46 units, although, with a p-value of only .18, it is not statistically significant.

Figure A-1: Mean Happiness and National Income, Normal Distribution



However, Figure A-2 shows a strong negative relation between real GDP per capita and the estimated 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

Figure A-2: Standard Deviation of Happiness and National Income, Normal Distribution

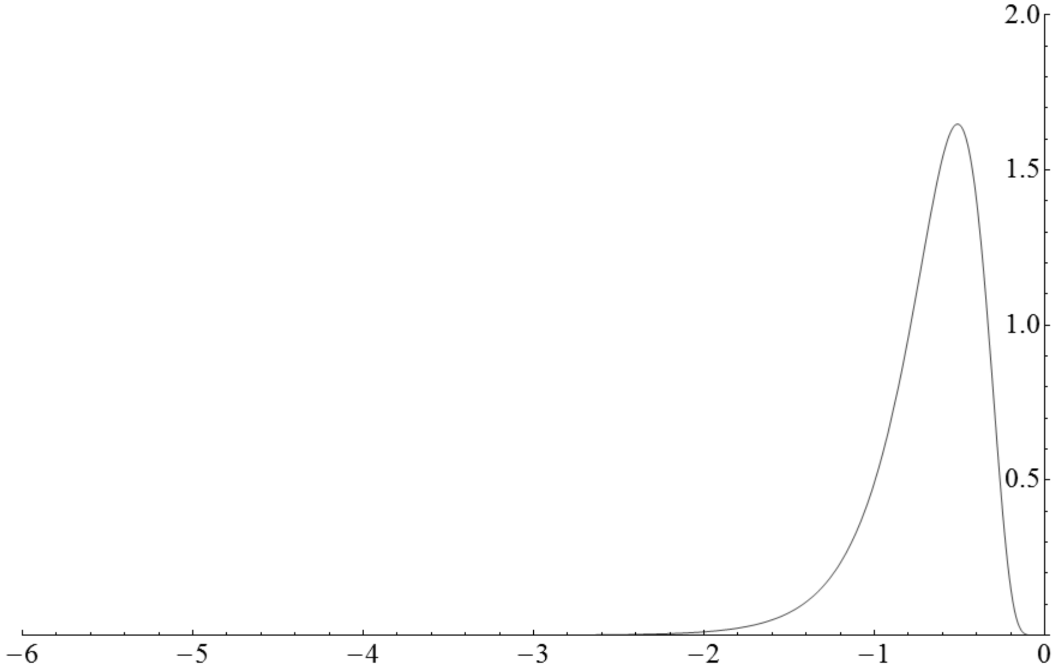


Foster (2013).[6] As real GDP has increased, fewer people report being very happy, but there is a zero to slightly negative change in the number of people who report being not too happy.

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, only mild skewness is required. For values of $C \geq .70$ we find the expected positive relation between income and happiness. Figure A-3 shows the distribution of happiness in 2006 with these parameters. Under our transformation, individuals with happiness

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

Figure A-3: 2006 Log-Normal Distribution of Happiness with no Easterlin Paradox



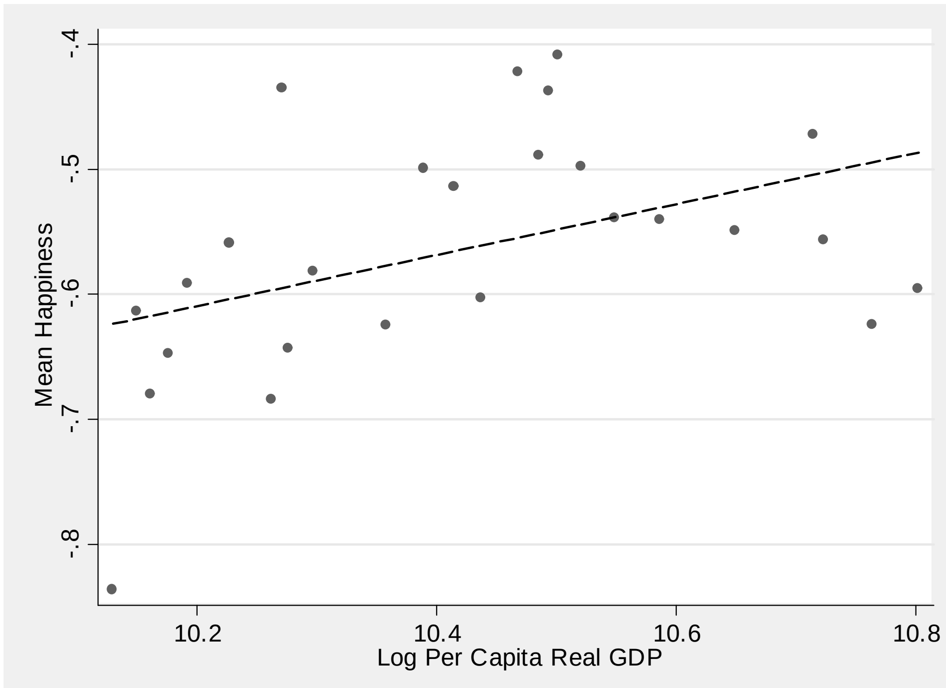
below -1 report being not too happy and those above $-.5$ report being very happy.[7] There is variation among the happiest and least happy individuals, although more so among the latter given the skewness of the distribution. The mean lies at the 41st percentile.

For $C \geq 2.60$, this positive relation becomes statistically significant at the 5% level. We plot the $C = 2.60$ case in Figure A-4. Here, a 10% increase in real GDP per capita is associated with a 2.04 unit increase in average happiness. If the happiness distribution is sufficiently skewed, then the results imply that average income and average happiness are positively associated. There are other distributions and transformations that also produce this result; there is no way to determine from

[7] Our normalized cut-points are 0 and 1, and $-\exp(0) = -1$ and $-\exp(-.7) = -.5$.

the data which cardinalization is correct.

Figure A-4: Mean Happiness and National Income, Left-Skewed Log-Normal Distribution



A.3.2 Happiness over the Lifecycle

There is a substantial literature that finds happiness is U-shaped over the lifecycle.[8] Individuals begin their adulthood fairly happy, see a decrease during much of their working life, and then rebound in happiness as they reach retirement. Blanchflower and Oswald (2008) obtain this result across over 70 countries, and there is even some evidence that it holds in apes (Weiss et. al, 2012). This claim is, however, not

[8] For some recent reviews, see Frijters and Beatton (2012), and Steptoe, Deaton, and Stone (2015)

without controversy. In some data sets, the shape depends on the choice of control variables, and whether one uses fixed effects or a pooled regression (e.g., Glenn, 2009; Kassenboehmer and Hasiken-DeNew, 2012).

In this section, we show that even when a conventional approach confirms the customary U-shape, this conclusion is highly sensitive to the choice of cardinalization of happiness. Using the Eurobarometer we find that with a standard set of controls and using ordered probit but allowing for heteroskedasticity, happiness is U-shaped with respect to age among men in most EU countries. However, there are also plausible cardinalizations in which happiness is strictly increasing or strictly decreasing; and linear, convex, or concave. In at least two countries the relation can be U-shaped or hump-shaped depending on the distributional assumption.

Data and Methods We use the Eurobarometer and restrict attention to men in the sixteen members of the European Union (excluding West Germany) as of 1986, as these countries have the most years of data. The literature is divided on whether the relation between age and happiness should be estimated with controls, and which controls should be included. For example, in the United States, the relation between happiness and age is U-shaped with a standard set of time-varying controls, but hump-shaped without (Easterlin 2006; Blanchflower and Oswald, 2009). In particular, since year - age = birth year, the effect of age in the presence of year and cohort controls is only identified through the arbitrary functional form restrictions.[9] While we appreciate this debate, we have nothing to contribute. Since our focus

[9] In the happiness literature, this problem has been discussed and an alternative method proposed by de Ree and Alessie (2011) and Cheng et al. (forthcoming). The age-period-cohort problem is well known in economics and dates to at least Heckman and Robb (1985).

is on the effect of choice of cardinalization on the result, we adopt a similar set of controls to those used by Blanchflower and Oswald (2008) which have been shown to be consistent with finding a U-shape. Controls include cohort (in 5-year groups) and year fixed effects, as well as controls for marital status, work status, and education.[10]

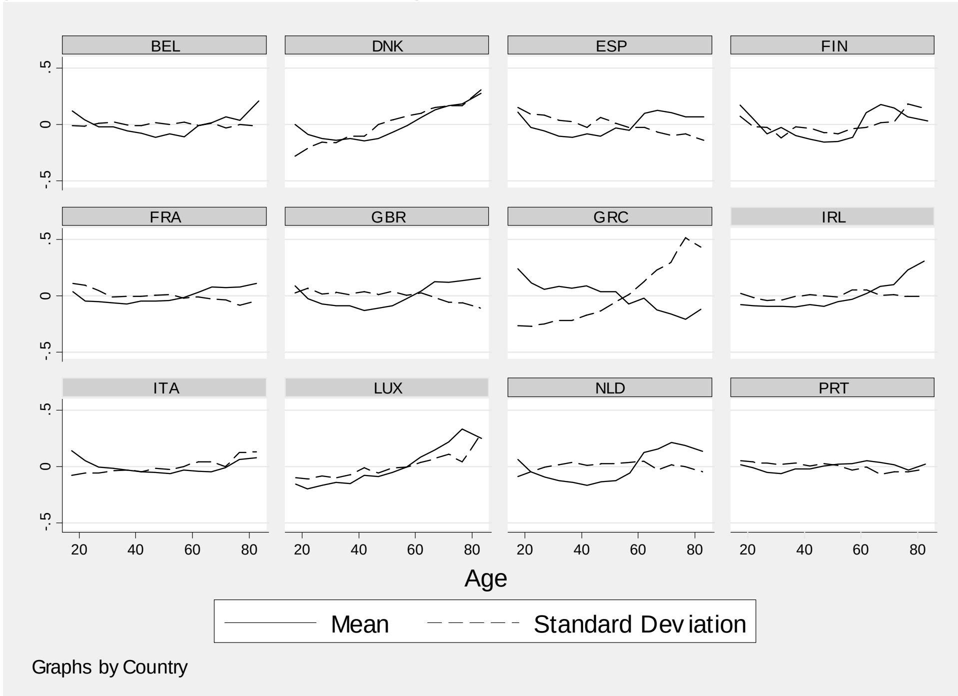
Results To ensure consistency with the literature, we follow Blanchflower and Oswald (2008) and group individuals by 5-year age category.[11] We group together all individuals age 80 and above due to the small number of individuals who are this old. We then estimate our model using the oglm command in STATA, assuming that the conditional distribution of happiness is normal and that our age categories and other controls may influence both the mean and the variance. We obtain this estimate for each country separately. We then use our estimates to calculate the mean and variance of each age-group for each country under normality at the (country-specific) mean of our controls.

Figure A-5 displays these estimated means and standard deviations for each of the 16 countries. We demean each country's estimates to allow for a better comparison of the shapes across countries. Consistent with Blanchflower and Oswald (2008), the estimates suggest a U-shape relation between happiness and age in the vast majority of European countries. While the patterns are generally noisy, only in Greece, Luxembourg, and Portugal is the U-shape obviously violated. However, there

[10] Our main departure from Blanchflower and Oswald (2008) is that we do not include a control for income. Income data from the Eurobarometer Trend file is problematic. It is measured in discrete categories which are specific to each country, and the boundaries on these categories change over time. It is also measured in nominal values of currencies which no longer exist, although in one year it was measured in U.S. dollars.

[11] The survey only includes individuals above age 16, so our youngest category covers only four years, 16-19.

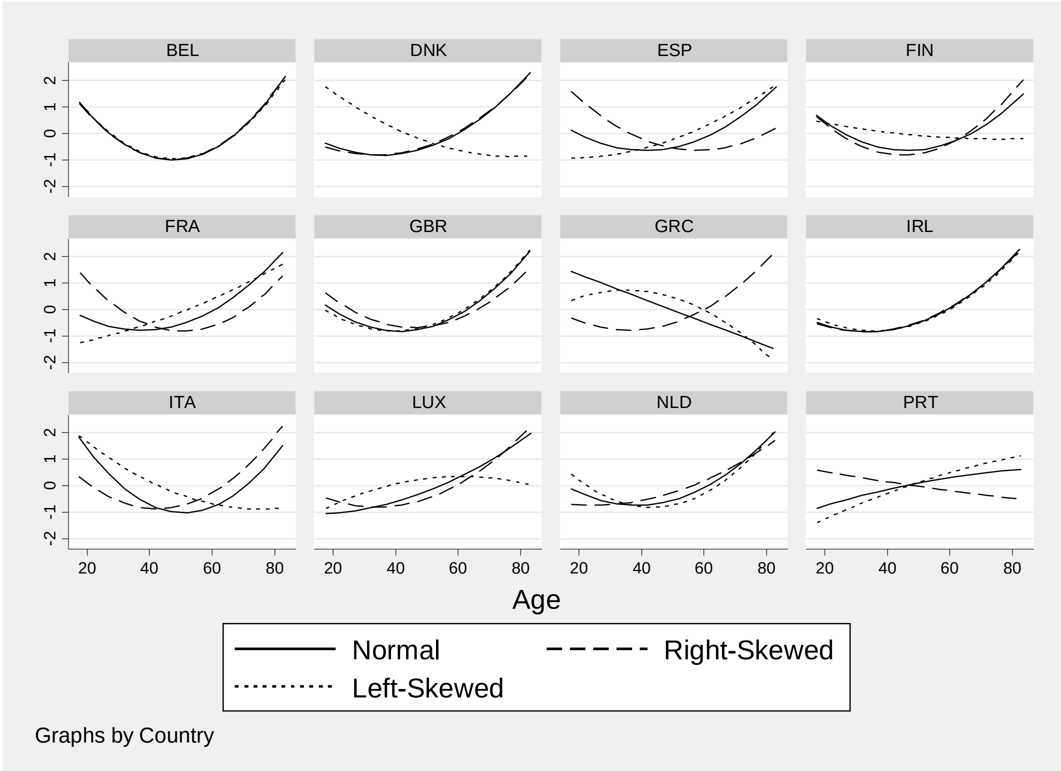
Figure A-5: Relation between Age when Happiness is Distributed Normally



are substantial differences in the relation between the variance of happiness and age. The variance appears to increase with age in Denmark, Greece, and Luxembourg but decrease in Britain and Spain and is U-shaped in Finland.

To test the robustness of the U-shape to distributional assumptions, we calculate these age-specific means under a right-skewed log-normal distribution with $c = 2$ and a left-skewed log-normal distribution with $c = -2$. Note that our right-skewed (left-skewed) log-normal transformations will mechanically increase (decrease) the mean at every age group. Therefore, we normalize happiness under each distribution within country by subtracting the average of the happiness mean across age groups and dividing by the standard deviation of these averages. This facilitates comparison

Figure A-6: Age-Happiness Profile Under Different Cardinalizations of Happiness



of the shape of the relation across distributions. Following the literature, to provide a smooth estimate of the relation between happiness and age, for each country we then regress the mean happiness estimates for the 14 age intervals on a quadratic in age.[12]

Figure A-6 displays these results graphically for each of the three distributional assumptions. Consistent with figure A-5, nearly every country shows a U-shaped relation between happiness and age when we assume that happiness is distributed

[12] We calculate the mean age of the respondents in the Eurobarometer in each category for each country seperately, and use that value to convert our age categories into continuous data.

normally. However, in nearly every country we arrive at a different pattern under one of our alternative distributional assumptions. While France and Spain reveal a U-shape under our normal and right-skewed distributions, the relation becomes almost linearly upward sloping if happiness is assumed to be left-skewed. Italy and Denmark, in contrast, go from U-shaped to strictly downward sloping under a left-skewed distribution. Greece and Luxembourg both lack a U-shape under normally distributed happiness. Greece is strictly downward sloping while Luxembourg is strictly upward sloping. Remarkably, both are U-shaped under the right-skewed distribution and *hump*-shaped under the left-skewed distribution. In contrast, Belgium shows a pronounced U-shape under all three distributional assumptions.

There are many equally plausible conclusions one could draw from figure A-6. If we had a strong prior that happiness is always normally distributed within age group and country, then we would believe that most countries have U-shaped happiness/age profiles but that Luxembourg and Greece violate this pattern. If we had a strong belief that happiness follows a U-shape across the lifecycle (and that the skewness of the distribution does not change with age,) then we would have to conclude that happiness is right-skewed in Luxembourg and Greece, but must not be in the Netherlands. In the absence of any prior, we might conclude that Belgium and Ireland have a well-determined happiness/age relation, but that we lack sufficient evidence to say anything with confidence about the other EU countries.

A.3.3 The Unemployment-Inflation Trade-off

Many happiness researchers have advocated evaluating policy based on its ability to raise the 'average' response on measures of subjective well-being. The application that has perhaps received the most widespread interest is correcting the misery index. While the misery index was developed as a political slogan, the idea that both unemployment and inflation are costly is intuitive. But it is by no means obvious that the proper weights, even assuming linearity, are equal.

Di Tella, MacCulloch, and Oswald (2001, hereafter DMO) provide one prominent attempt to use subjective well-being data to determine the appropriate trade-off between unemployment and inflation. They match estimated national well-being from happiness surveys with time-series data on inflation and unemployment across countries. They find that both unemployment and inflation are negatively related to national happiness but that the cost of unemployment is 1.7 times that of inflation. Thus the politically-derived "misery index" (inflation plus unemployment) biases policy towards too much unemployment relative to inflation.

We explore the robustness of this result. We find that the estimated social costs of inflation and, particularly, unemployment are remarkably fragile. Under reasonable transformations (or alternative social welfare functions), optimal policy should place three and a half times the value on a one percentage point decrease unemployment relative to a one percentage point decrease in inflation. Under other transformations, optimal policy should ignore unemployment altogether. At slightly higher right-skewness than we personally find plausible, society becomes *happier* when unemployment increases, consistent with arguments that recessions are "good for your health" (Ruhm 2000). The distribution necessary for the "misery index" to approximate the social welfare function is only slightly more right-skewed than a log normal.

Even if we ignore other concerns about using subjective wellbeing data, unless we are willing to take a strong stand about the underlying distribution of happiness, we can only conclude that the correct weight on unemployment is somewhere between roughly 0 and much higher than the weight on inflation.

Data and Methods Following DMO, we use happiness data from the Eurobarometer Trend File, and national unemployment and inflation data from the Organisation for Economic Co-operation and Development (OECD). DMO study the time period 1975-1991. However, the OECD currently only offers harmonized unemployment data for European nations beginning in 1983. We therefore focus on 1983-2002, which is slightly later than DMO but of similar duration.

We estimate the effect of inflation and unemployment on national happiness in two steps. First we estimate country-specific heteroskedastic ordered probits using individual-level data on life satisfaction from the Eurobarometer Trend File.[13] In these regressions we control for marital status, education, a quadratic in age, and a set of year dummies.[14] We then calculate the mean and standard deviation of

[13]We use all countries in the Eurobarometer file with the exception of Germany due to its reunification during the time frame. We thus estimate 15 separate models; for Austria, Belgium, Denmark, Spain, Finland, France, the United Kingdom, Greece, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, and Sweden.

[14]Unlike DMO, we do *not* control for unemployment status in these first stage regressions. This would cause our second stage to understate the true welfare cost of national unemployment. DMO recognize this problem and adjust their results using the estimated effect of unemployment on happiness in the first stage. However, changes in the distribution of happiness will also change the estimated effect of unemployment on individual-level happiness, so performing such an adjustment

happiness in each country in each year using the estimated coefficients on the year dummies, using the method discussed in section A.1.

In our second stage, we estimate a pooled regression of mean happiness on annual unemployment, inflation, a set of year fixed effects, country fixed effects, and country-specific time trends.[15] This represents DMO's preferred specification. We then estimate this regression under alternative right-skewed and left-skewed transformations of happiness. To provide comparability across estimates, we normalize each transformation so that the distribution of happiness is mean zero with standard deviation one across country-years.

Results In the first row of Panel A of Table A-1, we estimate the relation between national happiness and unemployment and inflation assuming normality (i.e. a heteroskedastic ordered probit). Inflation and unemployment are scaled in decimals, so that an inflation rate of 100% would correspond to a value of 1. Despite the different time horizon, our result is remarkably similar to that in DMO. Both inflation and unemployment have large and statistically significant negative effects on national happiness. The effect of unemployment is larger, suggesting that a 1 percentage point increase in unemployment would have the same negative impact on welfare as a 1.92 percentage point increase in inflation, only slightly larger than that estimated by DMO. However, as we show in the second row, unemployment also has a statistically significant and positive impact on the variance of happiness. Therefore

wold be inappropriate in our context.

[15] DMO use 3-year moving averages of inflation and unemployment rather than the annual values. It is not clear whether this is due to a data limitation or a preference for smoothing year-to-year variation in these variables. We use the annual figures as our results more closely resemble DMO's under a normal distribution than when using 3-year moving averages.

transformations that skew happiness to the right will decrease the estimated impact of unemployment on national happiness, while those which skew left will increase its estimated impact.

In Panel B, we first explore right-skewed distributions using the standard log-normal transformation used throughout the appendix. Beginning with our most modest transformation, we can see that $c = 1.5$ is already sufficient to reverse society's relative preferences. Under this transformation, inflation is more costly to social welfare, and society would be willing to raise unemployment by 1 percentage point in order to reduce the inflation rate by .62 percentage points. Remarkably, at $c = 2.0$, which we have chosen as our guideline for 'reasonable' transformations, society no longer incurs any discernible cost from unemployment. A 1 percentage point increase in the unemployment rate would only decrease mean happiness by .0004 standard deviations, compared to a .018 standard deviation cost from a 1 percentage point increase in inflation. When we skew the distribution only slightly more at $c = 2.5$, we estimate that society *gains* .012 standard deviations in happiness from a 1 percentage point increase in unemployment.

What about the "misery index," where society places equal weight on inflation and unemployment? It is easy to achieve this result. When $c = 1.125$ (not shown), the effect of inflation and unemployment are identical; a one percentage point increase in either leads to a .023 standard deviation decrease in national happiness.

Our conclusion that the importance of unemployment relative to inflation declines as we choose an increasingly right-skewed log-normal distribution must be tempered by the fact that, in all of the distributions in this class for which we have obtained

Table A-1: Effect of Inflation and Unemployment on Happiness under Various Distributional Assumptions

	Unemployment (1)	Inflation (2)	Trade-off (3)
Panel A: Normal Distribution			
μ	-5.185***	-2.706*	1.916
	(1.403)	(1.535)	
σ	0.465*	-0.169	
	(0.260)	(0.210)	
Panel B: Right-Skewed Log-Normal			
c=1.5	-1.300	-2.104	0.618
	(1.693)	(1.894)	
c=2.0	-0.041	-1.804	0.022
	(1.994)	(2.324)	
c=2.5	1.116	-1.390	-0.803
	(2.332)	(2.882)	
Panel C: Left-Skewed Log-Normal			
c=-1.5	-8.511***	-3.109	2.738
	(2.336)	(2.425)	
c=-2.0	-9.195***	-3.041	3.023
	(2.734)	(2.763)	
c=-2.5	-9.407***	-2.730	3.446
	(3.081)	(3.006)	

Each row represents the results of a seperate regression on mean happiness. All regressions include year and country fixed effects and country specific time-trends. "Trade-off" represents the implied trade-off in the social welfare function between unemployment and inflation and is computed as the ratio of the point estimate for unemployment to the point estimate for inflation. Robust standard errors in parenthesis.

$^{*}p < 0.10$ $^{**}p < 0.05$ $^{***}p < 0.01$

estimates, both the coefficient on inflation individually and the coefficients on inflation and unemployment jointly are insignificant at conventional levels of significance. Thus an alternative interpretation of the results using these distributions is that neither unemployment nor inflation affects happiness.

In Panel C, we instead use left-skewed log-normal transformations. As expected given the relation between the variance of happiness and unemployment, each of these transformations increases the impact of unemployment both relative to inflation and absolutely. A transformation of $c = -2.0$ is sufficient to raise the unemployment/inflation trade-off by 50% relative to that estimated under a normal distribution. In our most skewed transformation, $c = -2.5$, we estimate that a 1 percentage point increase in unemployment causes a .094 standard deviation decrease in happiness, and is equivalent to a 3.45 percentage point increase in inflation. In contrast to our right-skewed transformations, the estimated impact of unemployment is statistically significant under each distribution.

Strikingly, while our point estimate of the adverse effect of inflation is reasonably robust and largest when happiness is left-skewed, the estimated effect is statistically significant at conventional levels only when we assume normality.

Conclusions On a somewhat optimistic note, our results make intuitive sense. We would expect that unemployment would generally make people less happy, a result consistent with a large literature using self-reported happiness (e.g. Clark and Oswald, 1994; Blanchflower, 2001; Blanchflower and Oswald, 2004). It is at least plausible that those who are most directly affected by unemployment are located in the left-tail of the happiness distribution. Increasing the left-skewness of happiness is equivalent to using a social welfare function that places more weight on the least happy individuals relative to the happiest. When unemployment increases, there are more unhappy unemployed individuals, and the more weight we place on these individuals, the larger the social cost will appear. In contrast, a right-skewed transformation increases the weight on the happiest individuals relative to the least happy. As the happiest people will disproportionately hold stable jobs, increases in unemployment are unlikely to bother them. Even a positive effect of unemployment is plausible for a highly right-skewed social welfare function if, as some have suggested, individual's report happiness based on their relative circumstances.[16] When many are without a job, those still employed may report particularly high levels of happiness.[17]

Our results suggest that, even ignoring concerns about the reporting function, policy prescriptions that target maximizing happiness will vary wildly depending on what one believes the shape of the happiness distribution is, or the social welfare function one chooses to adopt. Minimizing the "misery index" would be appropriate if happiness is modestly right-skewed or if we believe that happiness is normally distributed and society places modestly more weight on the happiness of the happiest than the unhappy. If happiness were left-skewed, or society were more concerned with improving the welfare of the least happy than making the happiest even happier, a policy which placed substantially more weight on unemployment could be optimal.

[16] See Clark, Frijters, and Shields (2008) for a review of the empirical evidence that social comparisons are important for happiness.

[17] While Clark (2003) presents evidence that macro-level unemployment has a negative impact on the average well-being of the employed, this does not necessarily mean that macro-level unemployment has a negative impact on the well-being of the happiest employed individuals, who would be weighted heaviest by a right-skewed social welfare function.

Perhaps equally important is the lack of robustness of any adverse effect of inflation on happiness. As we note, while the estimated coefficient on inflation is reasonably robust, its significance is not. Thus while under normality we conclude that both inflation and unemployment are important although only at the .1 level for the former, under plausible levels of right-skewed log-normality, we conclude that neither unemployment nor inflation affects reported well-being while under plausible levels of left-skewed log-normality only unemployment matters.

A.3.4 Cross-Country Comparisons

In previous sections we found that the ranking of happiness across groups is highly sensitive to distributional assumptions. To explore this sensitivity in a larger context, we use the WVS to rank nations based on mean happiness. For each country, we estimate mean happiness under a normal distribution, as well as log-normal distributions with $c = 2, .5, -.5$, and -2.

The ordering of countries in Table A-2 represents their happiness ranking when happiness is assumed to be distributed normally; the columns give their ranking under the different log-normal transformations. Although the implied degree of skewness varies across countries, moving from left to right in the columns represents moving from a more right-skewed to a more 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.

Table A-2: 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

Table A-2 Continued

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
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

Table A-2 Continued

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
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 trasformations. Countries listed in order of

Table A-2 Continued

estimated mean happiness under normal distribution.

Source: World Values Survey 2005.

There are some countries whose rank remains fairly stable throughout the transformations. Great Britain is the third happiest country under a normal distribution and its rank varies 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. 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.[18]

[18]Using 2005 data from the World Bank on purchasing power parity equivalent per capita GDP,

A.3.5 Moving to Opportunity

Happiness data have also been used to evaluate micro-level policies, as in the case of the Moving to Opportunity (MTO) program. Motivated by the positive results of the Gautreaux desegregation program in Chicago,[19] the MTO 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 to 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.[20] 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 the standard GSS happiness question, "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?" Ludwig et al (2012, table S4) report the distribution of responses across the experimental and control group, which we reproduce in Table A-3. They calculate intent-to-treat estimates using intervals of 1 unit between the categories, as is common in the

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 ($C = -2$).

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

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

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.

Table A-3: Distribution of Happiness - Moving to Opportunities

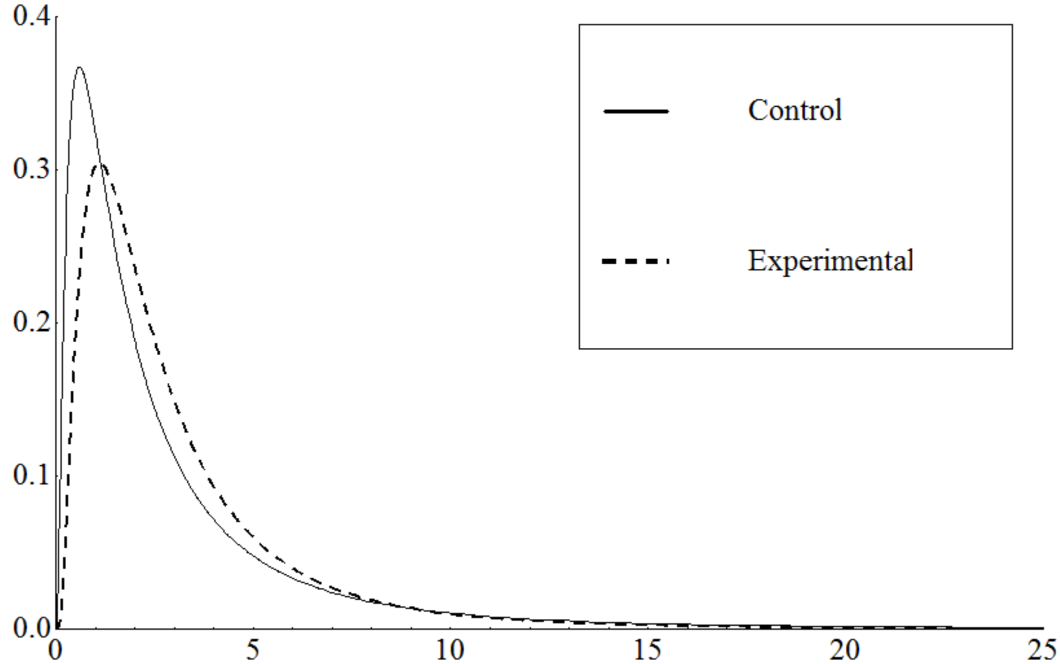
	Control Compliers	Experimental Compliers
Very Happy	0.242	0.262
Pretty Happy	0.470	0.564
Not Too Happy	0.288	0.174

Source: Ludwig et al (2013), Appendix Table 7.
Experimental estimates are TOT.

These results do not take into account that MTO may have had an impact on the *variance* of happiness. When we allow for this, while assuming the distribution is normal, we find that the control group has a lower mean (.44 v. .60), but also a higher variance (.79 v. .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 underling happiness data to increases the values above 1.20 we can reverse the mean happiness. This cardinalization would explain the data equally well.

Since the control group has a higher variance of happiness, it will be possible to reverse this result by applying a right-skewed distribution. From our point estimates, the required transformation is only slightly more skewed than a standard log-normal; we need only apply a $c = 1.33$. We plot the resulting distributions in figure A-7. The mean of the control group lies at the 70th percentile, while the mean of the experimental group likes at the 72 percentile. 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

Figure A-7: MTO Log-Normal Happiness Distribution with Equal Means



who could not match positive aspects of their former social environment in their new community

A.3.6 Marriage and Children

One of the most robust results in the happiness literature is that married individuals are happier than non-married individuals. This phenomenon has been observed both across countries and across time (e.g., Stack and Eshelman, 1998; Diener et. al, 2000;

Blanchflower and Oswald, 2004.)[21] In contrast, happiness researchers generally find that individuals with children are less happy than those without (e.g., Alesina et. al, 2004; Blanchflower, 2009; Stanca, 2012; Deaton and Stone, 2015).[22] These conflicting results present a bit of a conundrum. As both getting married and having children are actions that individuals (generally) take voluntarily, revealed preference suggests they should both raise happiness although as Deaton and Stone (2015) point out, if only people who expect children to make them happier become parents, in a cross-section having children and happiness might be unrelated. And, most parents are unwilling to express publicly that their children are a source of unhappiness.

In this section we explore the robustness of cross-section comparisons between those who are married and those who are not, and those who live with and without children. We find estimates consistent with the prevailing literature when we assume that happiness is distributed normally: married individuals are happier than those who never marry and people who live with children are less happy than people who do not. We find that even when we include a large set of controls, it is difficult to reverse the gap in happiness associated with marriage. In contrast, adding in a set of standard economic and demographic controls allows us to reverse the gap in happiness associated with children with only a very modest transformation.

[21] Note this is a distinct question from whether marriage *causes* individuals to become happier, which is a much more controversial claim. Lucas et. al (2003), for instance, present evidence consistent with the idea that becoming married causes a short term increase in one's well-being, but that the individual eventually 'adapts' and returns to his or her pre-marital happiness. We will discuss the evidence for adaptation in a different context in section A.3.8.

[22] There are important exceptions however. For example, Angeles (2010) finds positive effects of children on the happiness of married individuals in Britain.

Data We use the GSS both with and without a basic set of demographic controls which include birth cohort (in 10-year bins), family income, a quadratic in age, gender, race, region, work status, highest degree completed, and year fixed effects. When looking at marriage, we also include an indicator for whether the individual lives with children. These controls are similar to those used by Blanchflower and Oswald (2004).

Results In panel A of table A-4, we estimate the difference in happiness between married and never married individuals under the assumption that the distribution of each is normal. In the absence of controls, we find unsurprisingly that married individuals appear much happier than individuals who never married. We also find that they have a slightly higher variance of happiness, which means that this gap can be reversed by a left-skewed log normal. However our numerical estimates for the difference in variance parameters (.587 v. .552) is much smaller than our estimates for the difference in mean parameters (.864 v. .606). Given that our transformation reverses the gap by exploiting differences in the estimated variances, we require an incredibly skewed and probably implausible transformation with $c = -12.82$ in order to conclude that never married individuals were happier than married individuals. Reasonable transformations ($c > -2$) can be rejected at the 1 percent level.

There are large differences in the observable characteristics of married individuals and never married individuals. In particular, married individuals have substantially higher family incomes. In columns (3) and (4) of panel A, we re-estimate our model including our full set of controls. We again see that married individuals are happier than unmarried individuals, although that gap is slightly narrower. We also again

Table A-4: Estimated Happiness Distributions for Family Events

	No Controls		Controls	
	(1)	(2)	(3)	(4)
	Panel A: Marriage			
	Married	Never Married	Married	Never Married
Normal Mean	0.864	0.606	0.856	0.627
Normal Variance	0.587	0.552	0.569	0.525
Required c	-12.82		-9.33	
	Panel B: Children			
	Child	No Child	Child	No Child
Normal Mean	0.821	0.911	0.851	0.875
Normal Variance	0.572	0.603	0.550	0.584
Required c	-5.10		-1.20	

Source: General Social Survey Stevenson-Wolfers file.

see that married individuals have a higher variance of happiness than those who never married, but we can only reverse this happiness gap with a highly left-skewed log normal where $c = -9.33$. We can again reject $c > -2$. Thus the cross-sectional difference in happiness appears to be highly robust.

Turning attention to children, we divide the sample into those who live with at least one child under the age of 18 and those who do not. We focus only on those who are currently married to avoid comparing the happiness of, for instance, single mothers to childless families. In the first two columns of Panel B, we estimate the raw means and variances of happiness, assuming normality. Consistent with the majority of the literature we observe that individuals who live with children are less happy than those who do not. We also observe that those without children have a slightly higher variance in happiness than those with. This means that we can reverse the gap with a left-skewed log normal. We estimate that the required transformation would have $c = -5.1$ which, while less extreme than that required to reverse the marriage gap is still more extreme than we are comfortable advocating for. We can again reject that $c > -2$ at the 1% level.

Unlike with marriage, however, the addition of controls has a marked impact on our confidence in this result. As can be seen in columns (3) and (4), while adding controls does not change the rank ordering of groups, it does reduce the estimated gap in happiness between those with children and those without. This primarily comes from the fact that married individuals without children have higher family incomes than those with children. We note that the difference between the two groups remains statistically significant under normality. However, the reduced gap in means makes the rank ordering of happiness easy to reverse. We require only a left-skewed log normal with $c = -1.20$. This suggests that, if happiness is left-skewed, the difference in cross-sectional estimates between families with children and families without is primarily due to differences in income, and that otherwise families with children may be happier.

A.3.7 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.[23] In this section we explore whether this result is

[23] By 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.

robust to other cardinalizations. When happiness is distributed normally, we confirm that women's estimated mean happiness has declined relative to men's. However, the variance of female happiness has also declined relative to men's. This means that we can reverse the result using a left-skewed log-normal. We find, though, that the required transformation is quite extreme.

Data and Methods Like Stevenson and Wolfers, we use the GSS and focus on the years 1972-2006. Again we first estimate the mean and variance of happiness for both men and women in each year using a heteroskedastic probit. We then regress these estimates on a female dummy, a time trend, and the interaction between the female dummy and the time trend. Finally, we repeat the exercise under left-skewed log-normal distributions.

Results Figure A-8 plots our estimates of male and female happiness in each year. Consistent with Stevenson and Wolfers, we observe a clear decline over time in women's happiness relative to men. Using OLS, we find that women lost about .03 units of happiness per decade to men. However, figure A-9 shows some evidence that the gap in the variances of men and women's happiness also changed over the time period. Using OLS, relative to men, women's standard deviation of happiness decreased by roughly .008 units per decade, although this difference is not statistically significant. Given that both the mean and the variance of women's happiness is decreases relative to men's, we can reverse the result using a left-skewed log-normal.

In figure A-9 we plot estimated difference in trends for left-skewed log-normal transformations, varying c in .05 unit intervals. We normalize our means under

Figure A-8: Mean Happiness over Time under Normality

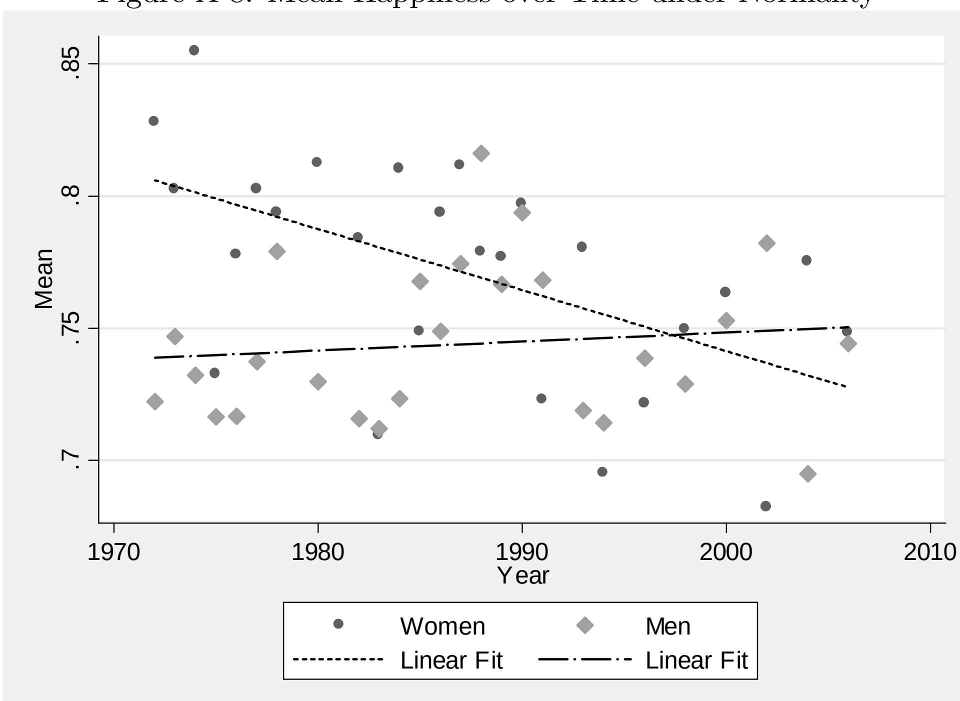


Figure A-9: Standard Deviation of Happiness over Time under Normality

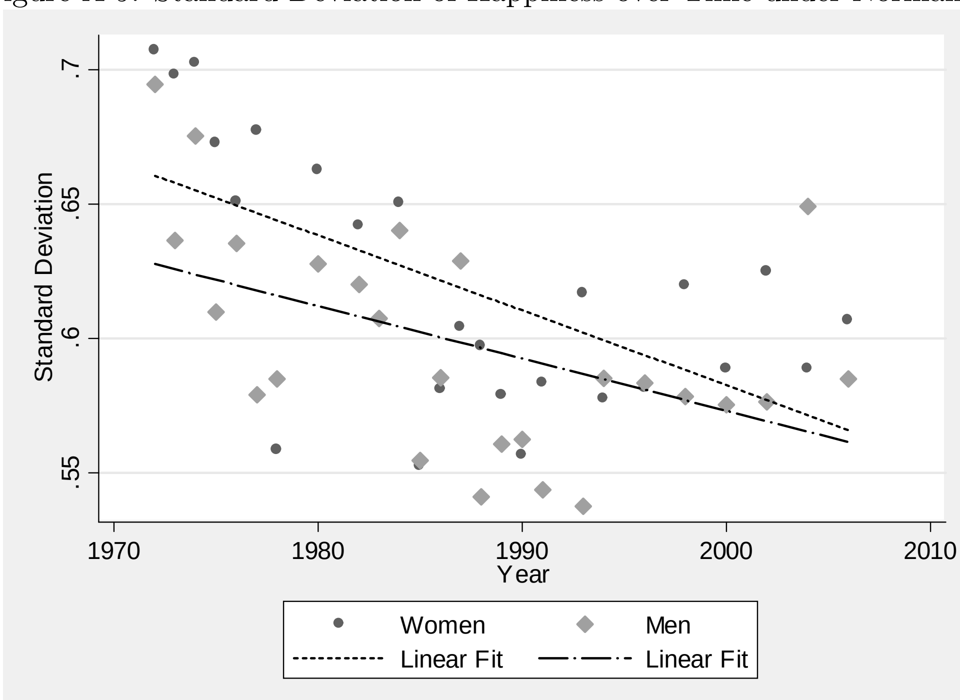
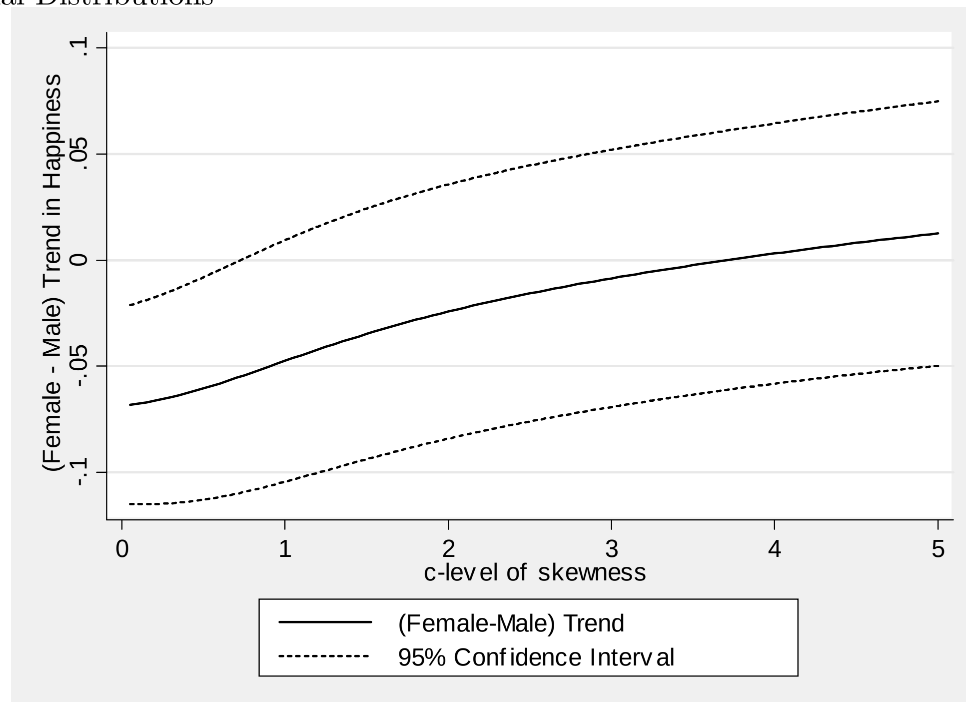


Figure A-10: Difference in Female-Male Happiness Trend for Left-Skewed Log-Normal Distributions



each transformation to be mean zero, standard deviation one across gender-year observations. That women's relative happiness did not increase is robust; that it decreased is not. While a left-skewed log normal with $c = -.75$ is enough to eliminate the statistical significance of the result, reversing the estimated sign requires $c \leq -3.75$, substantially more skewed than we feel comfortable with. We note, however, that the assumption that the shape of the distribution is constant over time may be particularly problematic in this case. If happiness has become more skewed (as income and wealth have) the required transformation may be less severe.

A.3.8 The Reporting Function and Adaptation to Disability

One of the most striking results in the happiness literature is the finding that people adapt to disability. Kahneman (2011) suggests that the experienced utility of paraplegics and nonparaplegics is fairly similar after a period of adaptation. Whether adaptation is complete is controversial, but partial adaptation is widely accepted. For example, Oswald and Powdthavee (2008) argue that the widely cited Brickman et al (1978) study has been misinterpreted and also find notable but only partial recovery of happiness from what they define as moderate and severe disability.

Yet the question of adaptation to disability and other adverse events is, perhaps, unusually sensitive to concerns about the reporting function. Individuals may reconcile themselves to such events either by becoming less unhappy about them or by reducing the standard for reporting themselves as unhappy.

In this section, we begin by providing evidence that is weakly consistent with adaptation although, consistent with much of this appendix, we find the result to be sensitive to how the distribution of happiness is modeled. Our primary focus, however, is on evidence that disabled and other individuals use different reporting functions suggesting that it is the reporting function rather than true underlying happiness that is adjusting.

Data We approach this question using the BHPS, which includes a measure on life satisfaction discussed before as well as asks the respondents about their disabilities. Unfortunately, there are some inconsistencies in the disability question across waves that we are forced to ignore in the interest of having an adequate time series. From

1996 to 2000, we use the question "Can I check, are you registered as a disabled person, either with Social Services or with a green card?" For 2002 and 2003, we use "Can I check, are you registered as a disabled person?" Finally, from 2004 on, we use "Can I check, do you consider yourself to be a disabled person?" On the face of things, we would expect the largest difference to be between 2003 and 2004 when the question shifts from being purely factual to one of self-image. In fact the difference between 2002-3 and 2004-8 is negligible. There is a large jump between 2000 and 2002, but this seems largely to reflect an upward trend between 1996 and 2002.

We limit the sample to individuals ages 16 to 60. If the information on disability was missing in any given year, we used the response from the previous year (11346 cases) or, if that was missing either because of non-response or because the missing year was the respondent's first in the sample, we use the response from the following year (669 cases). Finally, the remaining 237 missing cases were recoded as not disabled. For most of our estimates, we delete anyone who was disabled upon entering the sample (4713 cases). The final sample consists of 125,945 person-years of which 3,281 are disabled.

Results

The Basic Relation between Disability and Life Satisfaction We begin by examining the effect of becoming disabled. We limit ourselves to individuals whom we observe in both the wave before the individual became disabled and the wave in which they became disabled and limit the observations to these two waves.[24] We

[24] Waves are approximately one-year apart except for the comparison between 1990 and 1992.

estimate a heteroskedastic ordered probit model with the life satisfaction question capturing the latent variable and disabled as the sole explanatory variable.[25] This is effectively a fixed effect estimator since all observations come in pairs in which the first year the individual is not disabled but is in the second year.

Consistent with the literature and what we would expect, individuals are less satisfied with life in the year that they become disabled (the coefficient is -0.13 with a standard error of 0.04). The point estimate suggests that they also have a higher variance of life satisfaction although the difference is not statistically significant at any conventional level. Given the point estimates, reversing the mean utilities would require a left-skewed log-normal with $c = -2.9$, somewhat higher than we feel confident using. But the large standard errors on the relative variances leave plausible transformations within the confidence interval.

As an imperfect check on the importance of combining different disability questions, we experimented with including year dummies and cannot reject at any conventional level of significance that the year dummies are jointly insignificant.

In addition, anything else that varies systematically with the transition from not disabled to disabled will be captured by the coefficient on disabled. Some such variables change endogenously in response to the disability or the cause of the disability and can reasonably be ignored. However, by design each disabled individual is approximately one-year older when disabled than when not disabled. This is not necessarily a problem because prior researchers have found happiness or life satisfaction

[25] There are 1427 pairs involving 1239 individuals because some individuals report not being disabled after becoming disabled and then report being disabled again. Standard errors are clustered at the level of the individual.

to be quadratic in age (although see the discussion in section A.3.2). Consistent with this when we add age to the model, the coefficients are tiny and jointly insignificant at the .1 level. Consequently for the remainder of this section, we limit ourselves to simple analysis of the relation between disability and life satisfaction without further controls.

Adaptation To test for adaptation, we create samples of individuals in proximate years of disability. Thus the first estimate limits the sample to people whom we observe both in their first wave of disability and in their second wave and to those two waves. Sample sizes become too small to be useful after six waves.[26]

As can be seen in table A-5, there is at best weak evidence for adaptation. There are four positive coefficients, consistent with adaptation, and two negative ones. The sum of the coefficients over the period is consistent with complete adaptation, but we simply do not have the power to conclude that adaptation occurs. More importantly from our perspective, however, even if we ignored power, the conclusion of adaptation would not be robust to simple transformations of the underlying distribution. The strongest evidence for adaptation comes from the change between the fourth and fifth waves of disability. The coefficient is large albeit statistically insignificant. Yet given the estimated difference in the variance of the distributions, we need transform the normal into only a very modestly left-skewed log normal to reverse the conclusion of adaptation.

[26] Clustered standard errors tended to be slightly smaller than the traditional estimates. Therefore, we show traditional standard errors.

Table A-5: Changes in Mean Life Satisfaction with Disability Duration

Wave	Mean (1)	Log of Relative σ (2)
1 to 2	0.029	-0.033
	(0.057)	(0.047)
2 to 3	-0.005	-0.077
	(0.078)	(0.065)
3 to 4	-0.044	-0.016
	(0.098)	(0.080)
4 to 5	0.123	-0.049
	(0.112)	(0.094)
5 to 6	0.010	0.091
	(0.156)	(0.123)
6 to 7	0.043	-0.072
	(0.171)	(0.143)

Source: British Household Panel Survey. Life satsifaction is measured on a scale from 1 to 7. See text for explanation of the disability measure.

The Reporting Function Perhaps our most striking results regarding the relation between reported life satisfaction and disability address the relation between the reporting function and disability. Since ordered probit requires only two normalizations, we can test for the equality of the reporting functions under the maintained hypothesis that the distribution of happiness belongs to the normal family (including the log-normal and left-skewed log-normal, all of which we have shown would fit the data equally well). Of course, if we reject equality of the reporting functions, we may instead be rejecting the use of ordered probit, but in that case it is equally unclear what we should conclude from the ordered probits.

Table A-6 reports the estimated cutoffs for disabled and other individuals where we have normalized the cutoff between the lowest and second lowest reports to be 0 and the cutoff between the next two lowest to be 1. We, of course, allow the means and variances to differ. We can reject equality of the four free cutoffs at any conventional level of significance ($\chi^2(4) = 185.96$). Moreover, while the cutoffs in the middle of the life satisfaction range are similar, the cutoffs for the two highest responses are much lower for the disabled. This suggests that the disabled lower their standard for expressing complete satisfaction with their lives, which in turn leads us to underestimate the adverse effect of disability on happiness.

Table A-7 shows the distribution of life satisfaction in our data for the not disabled, disabled using the cutoffs estimated for the disabled, and disabled using the cutoffs estimated for the not disabled. We see that applying the standards of the not disabled, we get a much smaller fraction of disabled respondents who report being completely satisfied with their lives.

Table A-6: Estimated Cutpoints under Normality

Cutpoint	Disabled (1)	Not Disabled (2)
1 to 2	0	0
2 to 3	1	1
3 to 4	2.23	2.22
4 to 5	3.44	3.16
5 to 6	4.93	5.58
6 to 7	6.74	8.15

Source: British Household Panel Survey

Table A-7: Distribution of Reported Life Satisfaction (percentages)

	Not Disabled	Disabled	
	(1)	(2)	(3)
Cutpoints Used	Not Disabled	Disabled	Not Disabled
Not satisfied at all	1.00	8.15	8.15
2	1.96	8.41	8.41
3	5.87	16.08	15.86
4	14.18	19.88	22.93
5	32.03	23.03	28.00
6	34.51	17.25	14.66
Completely Satisfied	10.44	7.21	0.82

Source: British Household Panel Survey.

We emphasize that these results depend on one particular choice of cardinalization. Nevertheless, we do believe that these calculations reinforce the potential importance of differences in reporting functions for the interpretation of happiness data.

A.3.9 Heterogeneous Reporting Functions

An enduring critique of the happiness literature, touched on above, is that each individual reports his or her happiness on a different scale.[27] It should be clear that should this also be the case, the problems for the happiness literature are even more severe than we have heretofore asserted. Consider our discussion of marriage in section A.3.6. Assuming that both married and single individuals report their happiness the same way, the distribution of happiness in the population would have to be, in our view, unreasonably skewed to support the notion that never married individuals are more happy than those who are married. However, the level of skewness required would be substantially less if we *also* believed that married individuals have a lower threshold for reporting that they are "pretty happy" and "very happy."

Testing this hypothesis using data on individuals' subjective well-being is problematic. For example, In the previous section, we showed that a common reporting function that is independent of disability status is inconsistent with the normality assumption underlying ordered probit. But this result is consistent either with in-

[27] We have referred throughout this article to the way in which an individual transforms subjective feelings into a numerical value or category reported on a survey as the "reporting function," and that if two individuals transform their feelings into numerical values differently they lack a "common reporting function." This follows the language used by Oswald (2008). The problem has been alternatively referred to as "differential item functioning" by King et. al (2004), "scale recalibration" by Adler (2013), and "heterogeneous standards" by Fleurbaey and Blanchet (2013).

dividuals' reporting functions depending on their disability status or with failure of the normality assumption. We cannot exclude the possibility that some other cardinalization of happiness is consistent with a common reporting function.

A substantial body of work across multiple disciplines, however, has found that individuals assess the same circumstances differently on a discrete subjective scale. Salomon, Tandon, and Murray (2004) briefly described several hypothetical individuals with a mobility impairment (i.e., vignettes) to a cross-section of individuals in 6 countries, and then asked them to assess how 'mobile' the hypothetical individual was on a discrete 5-point scale. They found substantial heterogeneity in how individuals rated the same vignette both across countries and across age groups. King et. al (2004) finds that Chinese individuals consistently describe individuals described in the same vignettes as having more political say than do Mexicans. Kapteyn, Smith, and van Soest (2007) find that when evaluating disabilities, Americans are more likely to label vignettes with minor disabilities as fully able to work than are similar Dutch respondents. Beegle, Himelein, and Ravallion (2012) find that wealthy individuals in Tajikistan have higher thresholds for placing vignettes into the 'richest' category of an economic ladder, but also a lower threshold to place vignettes above the 'poorest'.[28] Ubel et al (2005) ask representative samples of older Americans to rate their health on a scale of 100. The sample was split into three with one-third having no further qualification, one-third being asked for someone of their age and

[28]Most of these studies use the vignettes to adjust individual responses to their own subjective assessment of their circumstances using methodology developed by King et. al (2004). This will only be valid however if these differences in evaluation of the vignettes reflect *only* differences in the reporting function *and* respondents use the same reporting function on the vignettes as they do for evaluating their own circumstances. Peracchi and Rossetti (2013) develop a formal test of these assumptions and find that they can generally be rejected on self-reported health data.

the last third asked for someone 20-years old. The first two groups reported similar health but the last reported poorer health, consistent with the view that people answer relative to some expectation. Oswald (2008) asked respondents to report their height *relative to their gender* on a vertical line with the bottom endpoint labeled "Very Short" and the top, "Very Tall." He then measured and recorded each response with a ruler, and found that, while men reported a lower subjective height than women conditional on objective height, they did not do so sufficiently to reflect the true gender difference in height. Consequently, the difference between the mean subjective height of men and women was a full point on a scale of 0 to 10.

In this section, we explore this issue further using ordinal reporting of income, an object with an underlying cardinal scale. We use data from the World Values Survey which asks people to place themselves in groups 1 through 10 of the income distribution in their country. It is difficult to reconcile the pattern of the answers with any reporting distribution that is common across countries.

Data We use the World Values Survey, Wave 6, conducted between 2010 and 2014. There are 87,226 respondents in 60 nations.[29] We merge these data with World Bank data on national gini coefficients. When we do so, we lose twelve countries, leaving 74,705 observations. For Australia and New Zealand, which we study in more detail, there are 1477 and 841 respondents, of whom 1422 and 770 answer the income question.

The survey asks respondents "On this card is an income scale on which 1 indicates

[29] Some of the 60 are, like Hong Kong, political entities which are not independent but are nevertheless distinct from the country to which they belong, or are not universally recognized as countries.

the lowest income group and 10 the highest income group in your country. We would like to know in what group your household is. Please, specify the appropriate number, counting all wages, salaries, pensions and other incomes that come in."

The question is unfortunately ambiguous about what constitutes an income group. The most obvious, to us, interpretation is that it refers to income deciles. However, the raw data strongly indicate that either respondents do not share this interpretation or are badly misinformed about where they stand in the income distribution in their country. Across all individuals in all countries (not weighting by country population), only 15 percent of respondents are in the bottom two categories and only 11 percent are in the top three categories.

Results Since respondents do not answer as if they were interpreting the question in terms of deciles of the income distribution, we begin by asking if there is some other equally simple meaning that is common to all countries. Perhaps it is always understood that the top income group refers to the top 1.5% of the income distribution.

Again this is decidedly not the case. A simple test of the equality of the distributions across countries has a χ^2 test-statistic of roughly 17,000 with 531 degrees of freedom which rejects the null well beyond any conventional level of significance.

Of course, since respondents apparently do not interpret the question as referring to deciles, the distribution of responses might reflect the distribution of income in a country. If the income distribution is very unequal, more people may view themselves as much richer or much poorer than the mean or median. In this case, we would expect the top category to be most highly populated in countries with considerable income inequality. In fact, the three countries with the largest fraction of people in the top group are, in order, New Zealand, Qatar and Japan. In Spain, none of the respondents reports being in the top category and in Morocco, none reports being in either of the top two categories. Using data on country-level income inequality from the World Bank, we find that there is no correlation between a country's gini coefficient and the fraction of respondents who answer in the top 2, bottom 2, or middle 6 subjective income categories.

Aussies and Kiwis The most direct comparison we can make is between respondents in Australia and New Zealand. Language differences between the two countries are trivial so that there is no concern that some fine point of translation might account for differential responses. Moreover, the income distributions in the two countries are similar. According to the OECD income distribution database,[30] in 2012 Australia and New Zealand had the following metrics: Gini for disposable income (.326, .323), Gini for market income (.463, .461), Gini for gross income (.368, .360), Palma ratio (1.24, 1.30), P90/P10 ratio for disposable income (4.4, 4.2), P90/P50 ratio for disposable income (2.2, 2.1), S90/S10 disposable income share (8.8, 8.2). These data show that the income distribution is probably marginally more unequal in Australia than in New Zealand, but the difference is small. In contrast, real income levels are much higher in Australia than in New Zealand using any standard measure and even correcting for the noticeably higher cost of living in Australia.

So if people are accurately portraying their income group relative to the country as the question requires, answers should be similar in the two countries. If contrary to

[30] http://stats.oecd.org/index.aspx?queryid=66670, accessed June 28, 2016

the wording of the question, people take into account their absolute income, answers in New Zealand should be lower than in Australia.

What do we actually observe? Table A-8 shows the distribution of responses. It is evident that respondents in New Zealand report themselves as being higher in the income distribution with some possible departure from this near the bottom of the distribution. The proportions placing themselves in the bottom three groups are similar, but far fewer Australians (9 percent) than New Zealanders (32 percent) say they are in the top three groups.

Table A-8: Reported Income Groups: Australia and New Zealand

	Australia (1)	New Zealand (2)
Lowest	7.0	7.0
2	6.4	10.3
3	13.8	9.2
4	12.4	11.3
5	18.6	9.4
6	16.6	9.2
7	16.1	11.2
8	7.3	10.6
9	1.3	11.8
Highest	0.4	10.0

Source World Values Survey Wave 6. Each cell represents the proportion of respondents representing themselves as being at that level in their country's income distribution. Sampling weights have been applied.

Conclusions We certainly should not conclude that people in New Zealand are, on average, richer than those in Australia. The survey question was not designed to address mean income differences across countries. But neither should we conclude that the income distribution is more unequal in New Zealand; this is demonstrably false from objective data. For some reason, far fewer people in Australia than in new Zealand report themselves as being in the upper end of the income distribution. Roughly speaking to report being in the top three tiers in Australia, one must be in the top decile while being in the top decile in New Zealand is adequate to report being in the top tier. While it may be true that people in the top 1 percent but not the top .5 percent of the income distribution genuinely perceive their relative status as noticeably worse in Australia than in New Zealand, it is hard to reconcile this conclusion with the objective situation. It seems much more plausible that Australians use a different reporting function.

Similarly, there are substantial differences in the distribution of reported income groups across countries that are not explained by differences in income distribution statistics as gathered by the World Bank. Again, the simplest explanation is that residents of different countries use different criteria for reporting their relative income.

A.4 Closing Remarks

This appendix demonstrates clearly that the problem of ordinality is empirically relevant. Even when we restricted ourselves to log-normal transformation, and imposed that the transformation must be constant across time and group, only two of the nine results we studied were robust to 'reasonable' transformations. There is no reason to think that happiness must be normal or log-normal. There is also no reason to think that the distribution must be the same across different groups, or that it is stable over time. Relaxing these restrictions would certainly allow one to find even more 'reasonable' distributions that would reverse the main results of the happiness literature, possibly including the decline in female relative happiness and the happiness gap between married and never married individuals.

Further, we also imposed for most of our analysis that the reporting function is constant across groups and time. There is no *a priori* reason to expect this to be true; as we and others have now demonstrated, it is not true for a variety of subjective responses. Without any outside reference for determining the reporting function, the model is simply not identified. There always exists a group-specific reporting function that rationalizes a distributional assumption, and always exists a distributional assumption that would rationalize a proposed reporting function.

That is not to say we have learned nothing at all about happiness from these exercises. Concerns about the reporting function aside, under a left-skewed distribution, happiness can be rising in per capita income and people with children may be happier than those without. In contrast, inflation has a higher impact on happiness than the unemployment rate only for a very right-skewed distribution. It is difficult to simultaneously argue that there exists a Easterlin paradox, and that children increase happiness; or that inflation is more important for national happiness than the unemployment rate, but that economic growth increases average happiness.

Likewise, our exercises take nothing away from studies addressing similar questions using objective data. While we cannot rule out that MTO had no impact on the average happiness of voucher recipients, we still know from Ludwig et. al (2013) that it reduced the prevalence of diabetes among adults and mental health problems among young girls. Likewise, while the pattern of a ‘U-shaped’ relation between age and happiness depends heavily on the researcher’s cardinalization choice of the researcher, we know that anti-depressant usage peaks at mid-life across 27 European nations (Blanchflower and Oswald 2016) , and that the age distribution of admittance to psychiatric hospitals is ‘hump-shaped’ (Le Bon and Le Bon 2014).

This presents the possibility that one could use objective measures to calibrate cardinalizations of happiness, as Bond and Lang (2014) do with test scores. However, this would not address the fundamental problems associated with the use of discrete categories. Moreover, it is unclear why using the subjective well-being data would be better than using objective data, if these objective outcomes are what we care about.

References

[1] Adler, Matthew D. 2013. "Happiness Surveys and Public Policy: What’s the Use?" *Duke Law Journal*, 62: 1509-1601.

[2] Alesina, Alberto, Rafael Di Tella, and Robert MacCulloch. 2004. "Inequality and happiness: are Europeans and Americans different?" *Journal of Public Economics*, 88 (9-10): 2009-2042.

[3] Angeles, Luis. 2010. "Children and Life Satisfaction." *Journal of Happiness Studies*, 11 (4): 523-538.

[4] Beegle, Kathleen, Kristen Himelein and Martin Ravallion. 2012. "Frame-of-reference bias in subjective welfare." *Journal of Economic Behavior & Organization*, 81 (2): 556-570.

[5] Black, Dan, Natalia Kolesnikova, Seth Sanders, Lowell Taylor and Mark Wessel. 2009. "A Divergent View on Black-White Earnings Convergence," unpublished.

[6] Blanchflower, David G. 2001. "Unemployment, Well-Being, and Wage Curves in Eastern and Central Europe." *Journal of the Japanese and International Economies*, 15 (4): 364-402.

[7] Blanchflower, David G. 2009. "International Evidence on Well-Being." in Krueger, Alan B. (ed.) *Measuring the Subjective Well-Being of Nations: National Accounts of Time Use and Well-Being*. University of Chicago Press: Chicago: 155-226.

[8] Blanchflower, David G. and Andrew J. Oswald. 2004. "Well-Being over time in Britain and the USA." *Journal of Public Economics*, 88 (7-8): 1359-1386.

[9] Blanchflower, David G. and Andrew J. Oswald. 2008. "Is well-being U-shaped over the life cycle?" *Social Science & Medicine*, 66 (8): 1733-1749.

[10] Blanchflower, David G. and Andrew J. Oswald. 2009. "The U-shape without controls: A response to Glenn." *Social Science & Medicine*, 69 (4): 486-488.

[11] Blanchflower, David G. and Andrew J. Oswald. 2016. "Antidepressants and age: A new form of evidence for U-shaped well-being through life." *Journal of Economic Behavior & Organization*, 127: 46-58.

[12] Bond, Timothy N. and Kevin Lang. 2014, "The Black-White Education-Scaled Test-Score Gap in Grades K-7," unpublished.

[13] Brickman, Phillip, Dan Coates, and Ronnie Janoff-Bulman. 1978. "Lottery Winners and Accident Victims: Is Happiness Relative?" *Journal of Personality and Social Psychology.* 36 (8): 917-927.

[14] Cheng, Terence C., Nattavudh Powdthavee and Andrew J. Oswald. forthcoming. "Longitudinal Evidence for a Midlife Nadir in Human Well-Being: Results from Four Data Sets." *Economic Journal.*

[15] Clark, Andrew E. 2003. "Unemployment as a Social Norm: Psychological Evidence from Panel Data." *Journal of Labor Economics*, 21 (2): 323-351.

[16] Clark, Andrew E., Sarah Fleche and Claudia Senik. 2014. "The Great Happiness Moderation," in A.E. Clark and C. Senik, eds., *Happiness and Economic Growth: Lessons from Developing Countries*, Oxford: Oxford University Press.

[17] Clark, Andrew E., Sarah Fleche and Claudia Senik, forthcoming. "Economic Growth Evens Out Happiness: Evidence from Six Surveys," *Review of Income and Wealth.*

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

[19] Clark, Andrew E. and Andrew J. Oswald. 1994. "Unhappiness and Unemployment." *Economic Journal*, 104 (424): 648-659.

[20] Deaton, Angus. 2008. “Income, Health, Wellbeing Around the World: Evidence from the Gallup World Poll.” *Journal of Economic Perspectives*, 22 (2): 53-72.

[21] Deaton, Angus and Arthur A. Stone. 2014. "Evaluative and hedonic wellbeing among those with and without children at home." *Proceedings of the National Academy of Sciences*, 111 (4): 1328-1333.

[22] Dias-Gimenez, Javier, Andy Glover and Jose-Victor Rios-Rull. 2011. "Facts on the Distribution of Earnings, Income, and Wealth in the United States: 2007 Update." *Federal Reserve Bank of Minneapolis Quarterly Review,* 34 (1): 2-31.

[23] Diener, Ed, Carol L. Gohm, Eunkook Suh, and Shigehiro Oishi. "Similarity of the Relations between Marital Status and Subjective Well-Being across Cultures." *Journal of Cross-Cultural Psychology*, 31 (4): 419-436.

[24] Di Tella, Rafael, Robert J. MacCulloch, and Andrew J. Oswald. 2001. “Preferences over Inflation and Unemployment: Evidence from Surveys of Happiness.” *American Economic Review*, 91 (1): 335-341.

[25] Dutta, Indranil and James Foster, 2013. “Inequality of Happiness in the U.S.: 1972-2010,” *Review of Income and Wealth*, 59 (3): 393-415.

[26] Easterlin, Richard A. 1973. “Does Money Buy Happiness?” *The Public Interest*, 30 (3): 3-10.

[27] 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.

[28] Easterlin, Richard A. 1995. "Will Raising the Incomes of All Increase the Happiness of All?" *Journal of Economic Behavior and Organization*, 27 (1): 35-47.

[29] Easterlin, Richard A. 2006. "Life cycle happiness and its sources Intersections of psychology, economics, and demography." *Journal of Economic Psychology*, 27 (4): 463-482

[30] Fleurbaey, Marc and Didier Blanchet. 2013. *Beyond GDP: Measuring Welfare and Assessing Sustainability*. Oxford University Press: Oxford.

[31] Frijters, Paul and Tony Beatton. 2012. "The mystery of the U-shaped relationship between happiness and age." *Journal of Economic Behavior & Organization*. 82 (2): 525-542.

[32] Glenn, Norval. 2009. "Is the apparent U-shape of well-being over the life course a result of inappropriate use of controls? A commentary on Blanchflower and Oswald (66:8, 1733-1749)." *Social Science & Medicine*, 69 (4): 481-485.

[33] Heckman, James and Richard Robb. 1985. "Using Longitudinal Data to Estimate Age, Period, and Cohort Effects in Earnings Equations." in Mason, William M. and Stephen E. Feinberg (eds.) *Cohort Analysis in Social Science Research*. New York: Springer-Verlag: 137-150.

[34] Kahneman, Daniel. 2011. *Thinking Fast and Slow*. New York: Farrar, Strauss and Giroux, 2011.

[35] Kassenboehmer, Sonja C. and John P. Haisken-DeNew. 2012. "Heresy or enlightenment? The well-being age U-shape effect is flat." *Economics Letters*, 117: 235-238.

[36] Kapteyn, Arie, James P. Smith, and Arthur van Soest. 2007. "Vignettes and Self-Reports of Work Disability in the United States and the Netherlands." *American Economic Review*, 97 (1): 461-473.

[37] 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.

[38] 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.

[39] Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Labor Market Effects." *Econometrica*, 75 (1): 83-119.

[40] Lucas, Richard E., Andrew E. Clark, Yannis Georgellis, and Ed Diener. 2003. "Reexamining adaptation and the set point model of happiness: Reactions to changes in marital status." *Journal of Personality and Social Psychology*, 84 (3): 527-539.

[41] 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.

[42] 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 Low-Income Families: Evidence from Moving to Opportunity." NBER Working Paper No. 18772.

[43] Le Bon, Olivier and Serge-Daniel Le Bon. 2014. "Age distribution curve in psychiatric admissions inversely correlates with Life Satisfaction." *Psychiatry Research*. 219 (1): 232-234.

[44] Oswald, Andrew J. 2008. "On the curvature of the reporting function from objective reality to subjective feelings." *Economics Letters*, 100 (3): 369-372.

[45] Oswald, Andrew J. and Nattavudh Powdthavee. 2008. "Does happiness adapt? A longitudinal study of disability with implications for economists and judges." *Journal of Public Economics*. 92 (5-6): 1061-1077.

[46] Perrachi, Franco and Claudio Rossetti. 2013. "The heterogeneous thresholds ordered response model: identification and inference." *Journal of the Royal Statistical Society*, 176 (3): 703-722.

[47] de Ree, Joppe and Rob Alessie. 2011. "Life satisfaction and age: Dealing with underidentification in age-period-cohort models." *Social Science & Medicine*, 73 (1): 177-182.

[48] 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.

[49] Ruhm, Christopher. 2000. "Are Recessions Good for Your Health?" *Quarterly Journal of Economics*, 115 (1): 617-50.

[50] Salomon, Joshua A., Ajay Tandon and Christopher J. L. Murray. 2004. "Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes." *British Medical Journal*, 328 (7434): 258-261.

[51] Smith, Tom W. 1990. *Timely Artifacts: A review of measurement variation in the 1972-1989 GSS*. GSS Methodological Report No. 56. Chicago: NORC.

[52] Stack, Steven and J. Ross Eshelman. 1998. "Marital Status and Happiness: A 17-Nation Study." *Journal of Marriage and Family*, 60 (2): 527-536.

[53] Stanca, Luca. "Suffer the little children: Measuring the effects of parenthood on well-being worldwide." *Journal of Economic Behavior & Organization*, 81 (3): 742-750.

[54] Steptoe, Andrew, Angus Deaton, and Arthur A Stone. 2015. "Subjective well-being, health, and ageing." *Lancet*. 385: 640-648.

[55] Stevenson, Betsey and Justin Wolfers. 2008a. "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox." *Brookings Papers on Economic Activity*, 1-87.

[56] Stevenson, Betsey and Justin Wolfers. 2008b. "Happiness Inequality in the United States." *Journal of Legal Studies*, 37(2): S33-79.

[57] Stevenson, Betsey and Justin Wolfers. 2009. "The Paradox of Declining Female Happiness." *American Economic Journal: Economic Policy*, 1(2): 190-225.

[58] Ubel, Peter A., Aleksandra Jankovic, Dylan Smith, Kenneth Langa and Angela Fagerlin. 2005. "What Is Perfect Health to an 85-Year-Old?: Evidence for Scale Recalibration in Subjective Health Ratings." *Medical Care*, 43(10): 1054-7.

[59] Weiss, Alexander, James E. King, Miho Inoue-Murayama, Tetsuro Matsuzawa, and Andrew J. Oswald. 2012. "Evidence for a midlife crisis in great apes consistent with the U-shape in human well-being." *Proceedings of the National Academy of Sciences*, 109 (49): 19949-19952.

[60] Williams, Richard. 2010. "Fitting heterogeneous choice models with oglm." *Stata Journal*, 10 (4): 540-567.