2019-08-01

The sad truth about happiness scales

The Sad Truth about Happiness Scales

Timothy N. Bond

Purdue University and IZA

Kevin Lang

Boston University, NBER, and IZA

We are grateful to Christopher Barrington-Leigh, Larry Katz, Jeff Liebman, Jens Ludwig, Andy Oswald, Daniel Sacks, Miguel Sarzosa, Justin Wolfers, and participants in seminars and sessions at the American Economic Association, 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. We thank Brian Towell and Wanling Zou for their excellent research assistance. The usual caveat applies, perhaps more strongly than in most cases.

Abstract

Happiness is reported in ordered intervals (e.g. very, pretty, not too happy). We review and apply standard statistical results to determine when such data permit identification of two groups' relative average happiness. The necessary conditions for nonparametric identification are strong and unlikely to be ever satisfied. Standard parametric approaches cannot identify this ranking unless the variances are exactly equal. If not, ordered probit findings can be reversed by lognormal transformations. For nine prominent happiness research areas, conditions for nonparametric identification are rejected and standard parametric results are reversed using plausible transformations. Tests for a common reporting function consistently reject.

1 Introduction

A large literature has attempted to establish the determinants of happiness using ordered response data from questions such as "Taking all things together, how would you say things are these days – would you say that you are very happy, pretty happy, or not too happy?" [1] We review and synthesize how some well-known results from statistics and microeconomic theory apply to such data, and reach the striking conclusion that the results from the literature are essentially uninformative about how various factors affect average happiness.

The basic argument is as follows. There are a large (possibly infinite) number of states of happiness which are strictly ranked. In order to calculate a group's 'mean' happiness, these states must be cardinalized, but there are an infinite number of arbitrary cardinalizations, each producing a different set of means. The ranking of the means remains the same for all cardinalizations only if the distribution of happiness states for one group first order stochastically dominates that for the other. But, we do not observe the actual distribution of states. We instead observe their distribution in a small number of discrete categories, essentially a few intervals of their cumulative distribution functions.

Without additional assumptions we cannot rank the average happiness of two groups if each has responses in the highest and lowest category. Using observed covariates to achieve full nonparametric identification of the latent happiness distributions would require making assumptions that happiness researchers generally claim to reject. We are therefore forced to follow the standard approach and assume the latent distributions are from a common unbounded location-scale family (e.g., an ordered probit). If we do, it is (almost) impossible to get stochastic dominance, and the conclusion is therefore not robust to simple monotonic transformations of the scale. In the case of the normal, there are always distributions in the lognormal family that reverse the result. Even in the knife-edge case where it would be possible to get stochastic dominance, the result would still be subject to the assumption

[1] Our analysis applies, *mutatis mutandis*, to other common questions such as those with five, seven or ten response categories

that all individuals report happiness the same way (a common reporting function), something which we show empirically is unlikely to be the case.

We outline conditions under which we can identify the rank ordering of group happiness, and then apply them to nine prominent results from the happiness literature. Not a single one is satisfied in any case. We also describe a test for common reporting of happiness under the assumption that happiness follows a distribution from the normal family. Whenever the data allow us to perform such a test, we reject it.

2 Rank-Order Identification of Group Happiness

Suppose a researcher has two groups A and B, and she wishes to claim that members of group A are, on average, happier than members of group B. What assumptions are required to make such a claim?

Unfortunately, with happiness data we will never be able to simply compare the average group responses directly. That would require happiness to have been reported to the researcher on an interval scale, the impossibility of which should be uncontroversial.[2] Thus we assume we are presented with an ordinal ranking of individuals' happiness, and denote individual i's happiness in this order as H_i and the cdf of happiness for group k as F_k.[3] Provided this ranking is complete, transitive, and continuous, from Debreu (1954) there exists a cardinalization q such that, $H_i > H_j \Leftrightarrow q(H_i) > q(H_j)$. Thus, in principle, we can compare group means under this cardinalization. This approach is implicit in nearly the entire empirical happiness literature. However, we know from Pareto (1909, pp. 541-2), that q is not unique. There are infinite other cardinalizations, and changes in cardinalization can change the ranking of means unless, as is well-known, there is *first order stochastic*

[2]It would require some way of eliciting each individual's happiness on a scale where the difference between a "99" and a "98" is the same difference in happiness as between a "97" and a "96", or a "7" and a "6", and so on. See Stevens (1946).

[3]To draw anything at all from the happiness literature, it is absolutely essential that it is possible to make interpersonal utility comparisons, an assumption we maintain throughout. But this is, itself, a controversial topic. For a brief review, see Binmore (2009).

dominance (FOSD).[4] Therefore, to conclude that group A is happier than group B, we must observe that F_A first order stochastically dominates F_B.

In practice, happiness researchers never observe F_k directly. Instead they observe subjective responses in a small number of categories, such as "not too happy," "pretty happy," or "very happy." Suppose we have three categories $S = \{0, 1, 2\}$, and let r_S^k be the fraction of respondents in group k who report happiness S. It is plausible that individuals follow a coherent introspective reporting rule, so that they report 0 if $H_i \leq H_i^1$ and 1 if $H_i^1 \leq H_i \leq H_i^2$.[5] This will be uninformative about F_k if H_i^1 and/or H_i^2 vary across individuals; a large number of "very happy" responses in group A may indicate a high F_A or low H_i^2s. For the moment we assume $H_i^1 = H^1, H_i^2 = H^2 \forall i$, that is a *common reporting function*, so the subjective responses inform us of two points on each cdf: $F_k(H^1) = r_0^k$ and $F_k(H^2) = 1 - r_2^k$.

Unfortunately, $F_A(H^1) \leq F_B(H^1)$ and $F_A(H^2) \leq F_B(H^2)$ (i.e., "stochastic dominance" in the categories) does not directly imply F_A FOSD F_B. It simply narrows the set of possible group distributions of happiness (see, for example, Manski 1988) to the set of all combinations of cdfs for which $F_A(H^1) = r_0^A, F_B(H^1) = r_0^B, F_A(H^2) = 1 - r_2^A, F_B(H^2) = 1 - r_2^B$. In other words, any two distributions that produce the observed happiness distribution in categories, by group, describe the data equally well.

Following this, we say the *rank order* of A and B is *identified* only if for every pair of distributions consistent with the data we have F_A FOSD F_B or F_B FOSD F_A.[6] That is, if A and B are rank order identified, the rankings of $\bar{q}_A$ and $\bar{q}_B$ will be the same for any cardinalization of any distribution that can describe the data. Requiring rank-order identification should again be uncontroversial. Rejecting this, requires making definitive statements about rankings that hold for some arbitrary cardinalizations or equivalently some arbitrary distributions, but not others satisfying the same *a priori* restrictions and describing

[4]See Lehmann (1955). This result entered the economics literature in the late 1960s (e.g, Hader and Russell 1969; Hanoch and Levy 1969).

[5]We will use a 3-point happiness scale going forward for illustrative purposes, but the discussion easily extends to 4 or more categories.

[6]For a formal definition of rank order identification, see the Online Technical Appendix.

the data equally well.

We now explore the conditions under which group happiness is rank-order identified. We begin with methods that do not rely on distributional assumptions, then consider restricting the set of permissible distributions (i.e. parametric assumptions such as ordered probit).

2.1 Non-Parametric Identification

To simplify exposition, we normalize the cutoffs between the categories to $H^1 = 0$ and $H^2 = 1$. Put differently, we restrict ourselves to the set of cardinalizations with this property.

When is the ranking of A over B identified without assuming a distribution for F_A and F_B? If we use just the ordered responses, from Manski and Tamer (2002) we must have $r_0^A = 0, r_2^B = 0$, and $r_2^A \geq r_1^B$.[7] If both groups have responses in the lowest category, the data can be represented by distributions where nearly all the unhappy As are close to $-\infty$ and all the unhappy Bs close to $-\varepsilon$, and vice versa. Similar logic applies when both groups have responses in the highest category. Even if the lowest category for A and the highest category for B are empty, all As might be clustered at the bottom of their categories (i.e., ε and $1 + \varepsilon$) while Bs might be clustered at their tops (i.e., $-\varepsilon$ and $1 - \varepsilon$); $r_2^A > r_1^B$ ensures that the distribution of A still dominates B in these situations. In practice, such a strong set of conditions will never be satisfied (see Table 1 below) for a sample of any reasonable size. Thus if happiness is reported using a discrete ordered scale, in practice we cannot rank the mean happiness of two groups without additional information or restrictions on the happiness distribution.

Now, suppose we have a vector of observable determinants of happiness X, and assume we can partition i's latent happiness

$$h_i^* = \psi(X_i) + u_i \tag{1}$$

[7] If there are $S+1$ categories, we require that $r_0^A = 0$, $r_S^B = 0$ and $\Sigma_0^s r_j^A \geq \Sigma_0^{s-1} r_j^B \ \forall s = 1, ...S$. We follow the literature on focusing on comparing means. Conditions for comparing, for example, medians are weaker and sometimes satisfied in practice.

into an observable component X_i with distribution F_k^X in group k, and an additively separable unobservable component u_i with distribution F_k^u. The function ψ transforms the observable components from their reported scale (e.g. dollars of income) to happiness under the chosen cardinalization. Note that in addition to a common reporting function, we assume ψ does not vary across groups. This assumption is implausible and also not strictly necessary for identification, but we simplify the environment to achieve rank-order identification as easily as possible.[8]

As Manski (1988) shows for a binary response, and Cameron and Heckman (1998) extend to a general ordered response model, we can identify ψ, F^u, and F^h through assumptions on the relations among X, u, and h, thereby avoiding assumptions about the distribution of h or u.[9] In our context, the key condition for nonparametric identification (Carneiro, Hansen, and Heckman 2003, Cunha, Heckman, and Navarro 2007) is that the support of u is contained in the support of $\psi(X)$. As Manski discusses, this condition is critical;[10] yet it is problematic for happiness studies. The major observable determinants such as income and marital status are bounded in practice, while factors such as physical and psychological health are, at best, observed only in discrete ordered categories and at worst unobservable and unbounded.[11]

To illustrate this problem, consider some authors' claim that there is a "satiation point" beyond which income does not increase happiness.[12] If they are correct, then as $X \to \infty, \psi(X) \to \bar{\psi}$ where X is income. Above $-\bar{\psi}$ we can identify the distribution of u based on differences between reported happiness and that which would be predicted by $\psi(X)$. But

[8] In order to place both groups on the same cardinalization, it is necessary that either both groups follow the same reporting rule, or both groups have some common X that has the same effect on the true latent variable. See Urzua (2008).

[9] See the Online Technical Appendix for the full list of conditions.

[10] See Manski's (1988) discussion surrounding Proposition 2, Corollary 1.

[11] While the condition is not formally testable, since we do not observe all possible draws of X from F^X, we note that the individual with the lowest family income (as calculated by Stevenson and Wolfers 2009) in the General Social Survey reports being "pretty happy."

[12] Note that our appendix implicitly refutes such claims. Stevenson and Wolfers (2013) reviews a number of papers which argue for a satiation point with respect to income, and find no support for this claim in any dataset using methods conventional within the happiness literature. Nonetheless, this has not stopped such claims from being made. See, for example, Jebb et al. (2018).

we cannot learn anything about the distribution of u below $-\bar{\psi}$ since all values in this range result in a report of "not very happy" for any X. Within the set of admissible distributions, u could be concentrated just below $-\bar{\psi}$ for one group and concentrated near $-\infty$ for another. Even if A stochastically dominates B above $-\bar{\psi}$, the distributions could cross below $-\bar{\psi}$, thus we cannot have rank-order identification. A similar problem arises when X is bounded.

In sum, while theoretically possible, in practice we are unlikely to have the observables necessary to non-parametrically identify the tails of the happiness distribution, without which the means of two groups are never rank-order identified.

2.2 Parametric Identification

Absent nonparametric identification, we must either conclude that identification is impossible or rely on parametric identification. Happiness researchers almost universally assume either that the ordered responses are measured on a discrete interval scale, or that each group's latent happiness distribution is normal (i.e., ordered probit) or logistic (i.e., ordered logit). These different approaches often yield similar results, leading some researchers to conclude falsely that such assumptions are innocuous (e.g., Ferrer-i-Carbonell and Frijters 2004).[13]

Estimating ordered probit is equivalent to assuming the existence of a single cardinalization under which *both* groups' happiness is distributed normally, which is in and of itself very strong.[14] Further, the standard ordered probit happiness researchers use also assumes that under this hypothesized cardinalization the variance of happiness is equal for all groups.[15] The assumption is implausible and unnecessary for estimation, but *necessary* to obtain rank-

[13]This conclusion may be problematic even sidestepping the issues raised by this paper. Heckman and Singer (1984) show that despite three common distributional assumptions showing negative unemployment duration dependence (and two strongly so), using an estimator that does not impose a parametric assumption produces the expected positive sign.

[14]Note that imposing that a *single* group's happiness is distributed normally is simply selecting an arbitrary cardinalization.

[15]The normalization in many statistical packages (such as STATA) imposes that the variance of the normally distributed unobservables is 1 and that the constant term is 0. If the ordered probit were applied separately without additional covariates, the program would report different cutpoints, implying a different reporting function but the same mean and variance for all groups. In the online appendix, we normalize the first two cutpoints to 0 and 1 and estimate separate means and variances for each group.

order identification through ordered probit. It is well known that for unbounded distributions from the location-scale family, one distribution is greater than the other in the sense of FOSD if and only if their location parameters differ and their scale parameters are equal.[16]

Moreover, in practice, as the sample size gets large, we will never estimate identical scale parameters even if the true scale parameters are equal.[17] Of course, we may be unable to reject equality, but we hardly need remind the reader that failure to reject and accepting the null are not equivalent. Thus, for large sample sizes, these techniques will never be able to identify which group is happier without further restrictions. Interestingly, this is true even if the condition in section 2.1 is satisfied.

What types of cardinalizations reverse the conclusion drawn from the normal? Suppose we estimate $\mu_A > \mu_B$, but $\sigma_A < \sigma_B$. The rank-ordering is preserved by all concave but not all convex transformations (Hanoch and Levy 1969).[18] One simple convex transformation is e^{ch^*}, where c is a positive constant. As is well known, the resulting distribution is log normal with mean

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

which is increasing in σ. For c sufficiently large this transformation reverses the ranking. When $\mu_A > \mu_B$ and $\sigma_A > \sigma_B$, the ranking is preserved by convex transformations, but not concave ones, such as $-e^{-ch^*}$. This generates a left-skewed lognormal distribution, with mean

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

which is decreasing in σ. Thus a sufficiently large c reverses the gap. These are both simple monotonic transformations of the latent happiness variable. Since happiness is ordinal, these

[16]The proof of this is a simple algebra exercise dating at least to Bawa (1975). A close corollary serves as an exercise in a popular graduate textbook (Casella and Berger 2002, p. 407). An alternative proof for the multivariate normal is in Müller (2001).

[17]Assuming estimation by maximum likelihood, the estimates will be asymptotically normal and their difference will also have a normal limiting distribution. See the working paper for more discussion.

[18]Hanoch and Levy show this result for any distribution which can be characterized by two parameters that are independent functions of the mean and variance of the distribution, a subset of the location-scale family.

transformations represent the responses equally well. It *is* possible to achieve stochastic dominance using bounded distributions from the location-scale family, such as the uniform. But, there is little or no reason to prefer a bounded distribution over unbounded ones, from the location-scale family or otherwise, that do not exhibit FOSD.

In sum, it is impossible to rank order groups using only the parametric assumptions made in the literature or any of the location/scale distributions typically used by empirical researchers. Identification *might* be achieved by applying additional restrictions to a standard distribution (i.e. placing restrictions on the set of permissable cardinalizations). Addressing whether a researcher could make a compelling case for such restrictions or the profession could reach a consensus on them would take us into the philosophy and sociology of science and beyond the scope of this paper. In the empirical section below we show that standard results can almost always be reversed using what we are confident are plausible transformations within the lognormal.

2.3 Reporting Function

It follows from our discussion that if true happiness is normally distributed with different variances between groups, there is always a reporting function such that the difference in mean reported happiness has the opposite sign from the true difference. But even this result assumes that all individuals report their happiness the same way. Up to now, it has been convenient to assume a common reporting function, but as we show in the empirical section, this assumption is unlikely to be correct.[19]

King et al. (2004) propose circumventing differences in reporting functions by using 'vignettes' to anchor the scale on which people report happiness. This 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

[19]See also Ravallion, Himelein, and Beegle (2016), who find evidence of substantial differences in the reporting function ("scale heterogeneity" in their terminology) for subjective poverty status both within and across countries.

("vignette equivalence"). Both assumptions are strong, the second particularly so, given heterogenous preferences. Moreover, as discussed above, even if we could place all individuals' happiness on a common scale, monotonic transformations of this scale would reverse group rankings.

Still, it is often easy to test for a common reporting function under a given parametric assumption. Suppose, for example, happiness is reported in $N \geq 4$ categories, and we impose normality, setting the first two cutoffs to be 0 and 1. This 'identifies' the means and variances of the distributions and leaves $N - 3$ cutoffs free. It is straightforward to test whether the free cutoffs are the same for all groups. Of course, faced with a rejection, the researcher is free to conclude either that her initial conjecture about the cardinalization was wrong or that the reporting functions differ, but we believe that, at the very least, testing the joint hypothesis should be standard procedure.[20]

3 Empirical Tests of the Happiness Literature

In the previous section we outlined the conditions under which the rank order of happiness for groups can be identified using categorical data on subjective well-being. We now put these into practice for nine key results from the happiness literature: the Easterlin (1973, 1974) Paradox for the United States, whether happiness is U-shaped in age, the optimal policy trade-off between inflation and unemployment, rankings of countries by happiness, whether the Moving to Opportunity program increased happiness, whether marriage increases happiness, whether children decrease happiness, the relative decline of female happiness in the United States, and whether disabilities decrease happiness.[21] For each of these, we first ask if we can draw any conclusions without assuming a parametric distribution by applying the criterion in section 2.1. We then test whether we can reject equal variances under an

[20]Note that we can also reject for any cardinalization in which the latent variable and cutpoints can be represented by the same monotonic transformation of the normal.

[21]For details of the estimation procedures and literature review of these results, see the Online Empirical Appendix.

assumed normal distribution, and determine whether the conclusions can be reversed using a left-skewed or right-skewed lognormal transformation, as outlined in section 2.2, and the degree of skewness required. Finally, when we have sufficient happiness categories, we also test for equal reporting functions assuming the existence of a cardinalization from the normal family as discussed in section 2.3.

Table 1 summarizes the results. None of these results are identified non-parametrically. Moreover, in the eight cases for which we can test for equality of variances under a parametric normal assumption, we reject equality. Thus we never have rank order identification, and can always reverse the standard conclusion by instead assuming a left-skewed or right-skewed lognormal. Further, in the seven cases where we are able to test for stability of the reporting function, we reject every time.

Thus if researchers wanted to draw any conclusions from these data, they would have to eschew rank order identification. In other words, they would have to argue that it is appropriate to inform policy based on one arbitrary cardinalization of happiness but not another, or equivalently that some cardinalizations are "less arbitrary" than others. It is unclear from where such an argument would come, or why we should apply a different standard for happiness research than other branches of economics.

Even if someone were to make this case, we cannot see how such a standard would say that distributions that resemble objective economic variables would be implausible. In the Online Empirical Appendix we further show that nearly every result can be reversed by a lognormal transformation that is no more skewed than the wealth distribution of the United States.[22] Even within this class of distributional assumptions, we cannot draw conclusions stronger than "Nigeria is somewhere between the happiest and least happy country in the world," or "the effect of the unemployment rate on average happiness is somewhere between very positive and very negative." To be clear, we are not proposing that satisfying this minimal criterion would make a result convincingly robust. It is plausible that happiness

[22]The exceptions are that the disabled are less happy than the non-disabled and that married women are happier than unmarried women. In both cases, we reject a common reporting function.

is more skewed than wealth is. And it is certainly not self-evident that happiness must be normal or lognormal. Happiness could be left-skewed for men and right-skewed for women, and their distributions might come from different families. And any claims of robustness would have to address our consistent finding of different reporting functions across groups. While we cannot rule out the possibility that happiness researchers will be able to find cardinalizations that are both consistent with a common reporting function and are robust within some parametric restrictions that most economists will find compelling, we are not sanguine about this prospect.

4 Conclusion

It is essentially impossible to rank two groups based on their mean happiness using the types of survey questions prevalent in the literature. Because happiness is ordinal, two groups can be ranked only if the distribution of one group first order stochastically dominates the distribution of the other. We can infer FOSD from the ordered response data alone only in extreme cases that do not occur in practice. The conditions for full identification through variation in observables are also violated. The parametric assumptions in the literature are incapable of producing FOSD without untenably strong assumptions about the happiness distribution. All of these conclusions are direct implications of well-established and uncontroversial results.

What then can we learn from such data? The regression estimates from ordered probit or logit are only accurate for one particular, arbitrary (and possibly nonexistent) cardinalization of happiness. This does not discount the actual self-reports, themselves. If we are only concerned about the number of people who subjectively consider their emotional state "not too happy," we can estimate effects using conventional binary response models. But, it is important to recognize that such an interpretation is much narrower than proponents of the use of average happiness measures currently claim for them. Subjective perceptions are

subjective and introspective, and generalizations drawn from such analysis are particularly sensitive to differences in the reporting function.

Researchers who wish to continue to interpret such questions more broadly need to be explicit about the assumptions underlying their conclusions, and justify their particular cardinalization or parametric assumption relative to other plausible alternatives. It is not clear what this justification could be, and our empirical work finds little evidence that even very strong restrictions will yield interpretable results. At a bare minimum, we would require a functional form assumption that survived the joint test of the parametric functional form and common reporting function across groups. Certainly calls to replace GDP with measures of national happiness are premature.

References

Bawa, Vijay S. 1975. "Optimal Rules for Ordering Uncertain Prospects." *Journal of Financial Economics* 2 (1): 95-121.

Binmore, Ken. 2009. "Interpersonal Comparison of Utility" In *The Oxford Handbook of Philosophy of Economics*, edited by Harold Kincaid and Don Ross, 540-59. New York: Oxford University Press.

Cameron, Stephen V., and James J. Heckman. 1998. "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males." *Journal of Political Economy* 106 (2): 262-333.

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): 361-421.

Casella, George, and Roger L. Berger. 2002. *Statistical Inference*. Pacific Grove, CA: Duxbury.

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): 1273-309.

Debreu, Gerard. 1954. "Representation of a Preference Ordering by a Numerical Function." in *Decision Processes*, edited by R.M Thrall, C.H. Coombs, and R.L Davis, 159-64. Oxford, England: John Wiley & Sons.

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

Easterlin, Richard A. 1974. "Does Economic Growth Improve the Human Lot?" In *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*, edited by Paul A. David and Melvin W. Reder, 89-125. New York: Academic Press.

Ferrer-i-Carbonell, Ada, and Paul Frijters. 2004. "How Important is Methodology for the

Estimates of the Determinants of Happiness." *Economic Journal* 114 (497): 641-59.

Hader, Joseph, and William R. Russell. 1969. "Rules for Ordering Uncertain Prospects." *American Economic Review* 59 (10): 25-34.

Hanoch, G. and H. Levy. 1969. "Efficiency Analysis of Choices Involving Risk." *Review of Economic Studies* 36 (3): 335-46.

Heckman, James J., and Burton Singer. 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica* 52 (2): 271-320.

Jebb, Andrew T., Louis Tay, Ed Diener, and Shigehiro Oishi. 2018. "Happiness, Income Satiation, and Turning Points around the World." *Nature Human Behavior* 2: 33-8.

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.

Lehmann, E. L. 1955. "Ordered Families of Distributions." *Annals of Mathematical Statistics* 26 (3): 399-419.

Manski, Charles F. 1988. "Identification of Binary Response Models." *Journal of the American Statistical Association* 83 (403): 729-38.

Manski, Charles F., and Elie Tamer. 2002. "Inference on Regressions with Interval Data on a Regressor or Outcome." *Econometrica* 70 (2): 519-46.

Müller, Alfred. 2001. "Stochastic Ordering of Multivariate Normal Distributions." *Annals of the Institute of Statistical Mathematics* 53 (3): 567–75.

Pareto, Vilfredo. 1909. *Manuel d'économie politique.* Translated by Alfred Bonnet. (Paris: V. Giard et E. Brière).

Ravallion, Martin, Kristen Himelein, and Kathleen Beegle. 2016. "Can Subjective Questions on Economic Welfare be Trusted? Evidence for Three Developing Countries." *Economic Development and Cultural Change* 64 (4): 697-726.

Stevens, S.S. 1946. "On the Theory of Scales and Measurement." *Science* 103 (2684): 677-80.

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

Stevenson, Betsey and Justin Wolfers. 2013. "Subjective Well-Being and Income: Is there any Evidence of Satiation?" *American Economic Review* 103 (3): 598-604.

Urzua, Sergio. 2008. "Racial Labor Market Gaps: The Role of Abilities and Schooling Choices." *Journal of Human Resources* 43 (4): 919-74.

Table 1: Rank Order-Identification in the Happiness Literature: A Summary of Results

	Non-Parametric (1)	Equal Variances (2)	Transformation to Reverse (3)	Equal Reporting (4)
Easterlin Paradox (U.S.): Happines does not increase with per capital income	No	Reject	Left-Skewed	[e]
Happiness is U-shaped with respect to age	No	Reject	[b]	Reject
Inflation/Unemployment Tradeoff	No	Reject	[c]	Reject
Moving to Opportunity increasd subjective well-being	No	[a]	Right-skewed	Reject
Country Rankings of Happiness	No	Reject	[c]	Reject
Marriage increases happiness	No	Reject	[d]	Reject
Children at home reduce happiness	No	Reject	Left-skewed	Reject
Relative female happiness has declined (U.S.)	No	Reject	Left-skewed	[e]
Disability reduces happiness	No	Reject	Right-skewed	Reject

Notes - For details of the procedures for each case, see the online appendix. Column (1) checks the condition for non-parametric identification discussed in section 2.1. Column (2) reports result of test of equal reporting functions conditional on the existance of a common cardinalization in the normal family. Column (3) reports result of test of equal variances under a normal cardinalization. Column (4) reports whether the results from the normal cardinalization are reversed by a left-skewed or right-skewed log normal.

[a] Not testable because calculated from published results.

[b] Depends on country.

[c] Left-skewed makes unemployment more important.

[d] Right-skewed for men, left-skewed for women.

[e] Not testable as only three happiness categories used.