

## **Happiness puzzles**

**Angus Deaton and Arthur Stone**

**December 29<sup>th</sup>, 2012**

Deaton: Princeton University, 361 Wallace Hall. Princeton, NJ 08544 ([deaton@princeton.edu](mailto:deaton@princeton.edu)). Stone: Stony Brook University, Putnam Hall, South Campus, Stony Brook, NY 11794-8790 ([Arthur.Stone@stonybrook.edu](mailto:Arthur.Stone@stonybrook.edu)).

Acknowledgements. We acknowledge support from National Institute on Aging through the National Bureau of Economic Research, Grants 5R01AG040629–02 and P01 AG05842–14, as well from the Gallup Organization. The authors are consulting Senior Scientists with Gallup.

Economists, as well as psychologists and philosophers, have become increasingly interested in self-reported measures of wellbeing, what they mean, and what they might be used for, especially in policymaking. The interest in wellbeing measures parallels a renewed awareness of the limitations of standard measures of GDP (and allied measures), as well as an interest in redirecting measurement away from GDP in an era when growth rates are diminishing across much of the rich world. Various self-reported wellbeing (SWB) measures have been used to provide new insights—for example that unemployment has effects well beyond the loss of income—and to capture a number of difficult-to-measure phenomena, such as the trade-off between inflation and unemployment, the costs of air pollution, or the values attached to environmental amenities.

In spite of those successes, the measures neither fully accepted nor fully understood. Traditional economics has been skeptical of measures whose content is unclear, is largely up to the respondent, and is sometimes deeply affected by the wording of the questions or by the context in which they are put. Although cognitive testing in the laboratory can provide insights about how people understand well-being questions, more often the interpretation is deduced by analyzing the correlates of the answers *ex post*. This poses a problem: we cannot use a correlation be-

tween a SWB measure and a variable  $X$  both to validate the measure *and* to claim that we have demonstrated the welfare consequences of  $X$ . One symptom is what we refer to as the “happiness fork:” that results that conform to standard welfare analysis (left fork)—that income, education and health are linked to SWB—are taken as support for SWB measures, while results that do not so conform (right fork)—that unemployment is much worse than the loss of income would imply, or that people hate to commute—are *also* taken as supportive and indeed demonstrative of the superiority of the SWB measures. Such interpretations are unlikely to persuade skeptics.

Rather than argue for one fork or the other, this paper looks at several SWB findings that are, on the face of it, internally contradictory, or that appear to contradict well-based if conventional judgments about what is or is not good for people. Those who see happiness as the *only* good, and who believe that SWB measures capture it, simply see these conventional judgments as mistaken. For those of us not so committed, we are likely to modify our interpretation of the measures, or the place that happiness plays in wellbeing, or both. From whatever perspective, internal contradictions need to be worked out, a process that is likely to be productive. Maintain a distinction throughout between evaluative and hedonic wellbeing.

### **I. A Relative Income and Wellbeing Puzzle**

There is a large literature, using data from the US, Canada, Europe, and developing countries in which researchers have regressed a wellbeing measure on both individual income and average income over some local area that is taken to include the people to whom the individual compares him or herself. In these regressions, individual income invariably has a positive effect, while local income has a negative effect, sometimes equal and opposite to the positive effect of own income. The United Nations *World Happiness Report* (pp. 60–66) documents these general findings summarizing dozens of studies using panel and cross-sectional data, only a few of which are

in contradiction. If these findings are correct, they imply that the cross-sectional response of SWB to income overstates a long-run time-series response in which incomes rise together, and that economic growth may fail to increase in SWB, as long claimed by Richard Easterlin. They also imply that individual incomes cause negative externalities to others, which could be used to justify Pigovian income taxes to discourage people from working hard and so harming others.

Our first puzzle is that, in our very large data sets, we can find no evidence for this apparently well-established conclusion. We investigate using two data sets, from the US, the Gallup Healthways Wellbeing Index survey, which has surveyed 1,000 randomly selected Americans every day since January 2008, and from the world as a whole, the Gallup World Poll, which has asked similar questions of 1,000 or more people in each of 160 countries beginning in 2006.

We look at two different ways of measuring wellbeing, an evaluative measure, the Cantril ladder, and a hedonic measure, daily happiness. The former invites respondents to rate their lives on a “ladder” with eleven steps, marked from 0, which represents the worst possible life for you, to 10, which represents the best possible life for you. Answering such a question requires the respondent to think about their life and interpret the question. We also look at a dichotomous measure in response to the question “Did you experience a lot of happiness yesterday?” This is one of a number of hedonic questions, which should be distinguished from the evaluative question in the ladder. Hedonic questions do not require the cognitive effort required to answer evaluative questions, they refer to different aspects of experience, and often have different correlates. For example, hedonic measures are uncorrelated with education, vary over the days of the week, improve with age, and respond to income only up to a threshold. Evaluative measures remain correlated with income even at high levels of income, are strongly correlated with education, are often U-shaped in age, and do not vary over the days of the week, Stone et al (2010), Kahneman

and Deaton (2010). The important distinction between evaluative and hedonic wellbeing renders unhelpful the portmanteau use of the term “happiness.”

We explore the relationship between SWB and income at various levels of aggregation. If, for example, an individual’s relative income comparators live in the same county, the effects of income on SWB should be higher in the micro data than when we look at effects of *county* average income on *county* average wellbeing. More formally, if an individual’s outcome is positively and linearly related to individual log income and negatively to county average log income, we will obtain the coefficient on individual income from a regression of the outcome on log income and a county fixed effect. When we run a regression on averaged county-level data, the coefficient on county average log income will be the difference between the individual and group coefficients, potentially zero if only relative income matters.

The results at various levels of aggregation—persons, zip codes, counties, congressional districts, cities (Metropolitan Statistical Areas), and states—are shown in the top panel of Table 1. Consider first the ladder, our measure of evaluative wellbeing. In the micro data, the coefficient of log income on the ladder measure is 0.51; the regression controls for age groups, sex, and race, plus zipcode fixed effects. For all aggregations but zip codes, the coefficients are lower than in the micro data, and (arguably) decline modestly as the level of aggregation increases, though the state level coefficient is higher than the MSA level coefficient. If we interpret the coefficients as the difference in the effects of individual and average incomes, the maximum scope for the latter is quite small, less than a fifth of the effect of individual (log) income. There is no evidence at all for the idea that the positive effects of income on life evaluation are completely offset by negative effects of average income at any of the geographical areas considered here. Indeed, the zip code level analysis throws up an effect that is *higher* than in the micro re-

gression. One explanation is that living with rich people provides public goods—parks, schools, libraries, etc.—that enhance evaluative wellbeing. Another possibility is that life evaluation depends not on current but permanent income, and that the latter is better proxied by the zipcode average than by own income. There is no evidence that average zipcode income has a depressing effect on evaluative wellbeing and yet it is surely here, among people that one is most likely to know, that relative income theory has the greatest plausibility.

**Table 1: Self-reported well-being and relative income**

	Ladder			Happiness		
USA	$\beta$	s.e.	# obs	$\beta$	s.e.	# obs
<i>Person level data</i>	0.508	(0.002)	1,083,971	0.043	(0.0004)	1,084,890
<i>Averages by:</i>						
Zipcode	0.627	(0.011)	7,735	0.037	(0.002)	7,735
County	0.507	(0.018)	2,353	0.031	(0.003)	2,353
Congress District	0.491	(0.032)	470	0.023	(0.004)	470
MSA	0.401	(0.039)	365	0.012	(0.005)	365
State	0.419	(0.136)	51	-0.010	(0.015)	51
<b>World</b>						
<i>Person level data</i>	0.554	(0.003)	639,431	0.054	(0.0008)	504,543
<i>Country averages</i>	0.675	(0.078)	160	0.034	(0.034)	160

Notes: Regressions have either the ladder or happiness as the dependent variable, and we report the coefficient on log income. Regressions include controls for sex, age groups, and white/non-white (US only). Person level US regressions contain zipcode dummies, and person level world regressions contain country dummies. Counties and zipcodes are excluded if there are fewer than 50 people in the sample. The happiness question was not asked in all rounds of the World Poll, which accounts for the fewer number of observations in the world data.

The World results for the ladder are uncannily like the US results. The person level coefficient is 0.55, as opposed to 0.51 in the US, and the cross-country effect is 0.68, exactly what we would suppose if there are positive spillovers from having rich people nearby, and exactly the opposite of what would happen if rich people nearby create a negative externality.

The results for our hedonic measure, daily happiness, conform more closely to the predictions of the relative income story. The coefficient on log income of 0.043 (remember the scale for happy is 0 to 1, not 0 to 11 as for the ladder), and declines steadily with the level of aggregation becoming negative and insignificant at the level of states. Another story is that hedonic happiness, unlike life evaluation, is more closely related to transitory than to permanent income. Aggregation over larger units annihilates an increasingly large share of transitory income and drives down the coefficient on income. Either explanation is consistent with the evidence, as well as with the sharp decline in the effects of income on daily happiness in the global data as we move from within to across countries. Graphical examination of the global data shows that the cross-country coefficient is as high as it is because of the presence of a few very unhappy countries at the bottom of the world income distribution together with one very rich, reportedly very happy country (the US). Over a broad range of per capita GDP levels, aggregate happiness is unrelated to aggregate income; the countries of the former Soviet Union are among the unhappiest in the world, unhappier than either Congo, Benin, or Chad, for example, and Italy and Denmark are unhappier than Mozambique, Sudan, and Rwanda. These results cast serious doubt on using (the hedonic version of) happiness to provide an overall assessment of human wellbeing. While it makes sense for SWB measures to paint a different picture than GDP, it is hard to credit a measure that says that Denmark is worse off than Rwanda; being happy is a good thing, but other things surely outweigh it in any credible overall assessment of life. Life evaluation measures, such as the ladder measure analyzed here, do not share this particular problem.

The reconciliation of the results in Table 1 with the large number of contradictory results in the literature is beyond the scope of this short paper. More substantively, the standard insights from permanent income theory seem promising for interpreting the wellbeing data, and offer an

alternative to the relative income theory, just as was the case in consumption theory half a century ago. Yet our results should highlight that the relative income story is inconsistent with at least these data, and cannot be used to support the view that income generates externalities for others, or that universal growth will leave wellbeing unaffected.

## **II. Is religion good for you? An aggregation puzzle**

Many previous studies have found that more religious individuals have higher wellbeing; our second puzzle is that this finding does not carry through to comparison of more or less religious places. Idler and Kasl (1997) discuss the interpretative, regulative, and integrative functions of religion. Religion provides meaning to life, especially in times of difficulty, religion regulates conduct and promotes behaviors that are good for health and wellbeing, but which might be difficult to adhere to on one's own, and religion may provide direct help, in the form of healthcare or education, particularly under circumstances where provision by the state is absent or limited. Religion provides social capital through churchgoing, and Chaeyoon Lim and Robert Putnam (2010) have argued that friends from church promote happiness more effectively than friends in general.

We use the Gallup data to examine the correlates of “religiosity,” defined by the yes/no answer to the question of whether religion is an important part of your daily life. In the US, 66 percent of people answer this question positively; across US states, the fraction varies from 41 percent in Vermont to 87 percent in Mississippi. Across the world, 65 percent of people answer positively, with a range from less than 20 percent in Sweden, China, Estonia and Denmark, to more than 98 percent in Niger, Mauritania, Egypt, Malawi, Sri Lanka, Bangladesh and Somalia.

The US, although close to the world average overall, is much more religious than most other rich countries.

**Table 2: Self-reported well-being and religiosity**

	Ladder			Happiness		
<b>World: within countries</b>	$\beta$	s.e.	# obs	$\beta$	s.e.	# obs
<i>No income control</i>	0.039	(0.006)	796,121	0.044	(0.002)	560,396
<i>Log income control</i>	0.073	(0.007)	595,630	0.049	(0.002)	469,846
<b>Between countries</b>						
<i>No income control</i>	-0.550	(0.428)	159	0.125	(0.072)	152
<i>Income control (1)</i>	0.264	(0.329)	159	0.186	(0.071)	152
<i>Income control (2)</i>	-0.025	(0.334)	155	0.165	(0.071)	149
<b>USA</b>						
<i>Person level data</i>	0.248	(0.004)	1,388,404	0.041	(0.0006)	1,391,165
<i>Log income control</i>	0.263	(0.004)	1,080,485	0.042	(0.0007)	1,081,420
<i>State averages</i>	-0.059	(0.202)	51	0.022	(0.024)	51
<i>Log income control</i>	0.368	(0.188)	51	0.062	(0.025)	51

Notes: Regressions have either the ladder or happiness as the dependent variable, and we report the coefficient on religiosity. Regressions include controls for sex, age groups, and white/non-white (US only). Micro US regressions contain zipcode dummies, person level world regressions contain country dummies. In the bottom panel we show results for state averages: similar results were obtained for other levels of aggregation, with small or even negative effects without income controls, and positive, often significant, effects with income controls. Income control (1) uses the average of log income from the survey, while income control (2) uses the log of real chained GDP per capita in 2010 from the Penn World Table, Version 7.1.

Table 2 documents the evaluative and hedonic advantage for individuals who report that they are religious. Again, we begin with the ladder. The first rows of the table show that, within countries, there is a modest advantage in life evaluation among those who are more religious—the effect is equivalent to less than a ten percent increase in income, though about 50 percent in the US—and that the effect is larger when income is controlled; within countries income is negatively correlated with religiosity. Looking across countries, the most religious countries have *lower* ladder values, which comes entirely from the fact that poorer countries are more religious and that income is strongly positively correlated with the ladder. Once log income is controlled for, there is an insignificant positive partial correlation between religion and the ladder if we use the average of log income from the survey itself, or an insignificant negative partial correlation if



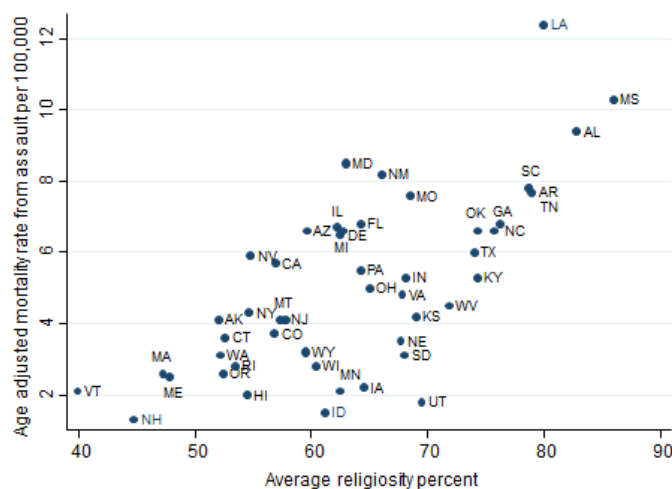
we use log per capita GDP from the Penn World Table 7.1. In the US, religious states have lower average ladder values than non-religious states. Once again, the more religious states are sharply poorer, and once income is controlled, the positive effect of religiosity returns.

The sharpest puzzle here is why the religiosity advantage for individuals does not carry through to the country averages. There is *something* associated with religion in aggregate that offsets the effects of religiosity on individuals; this could be religion itself—if religion favors the in-group at the expense of the out-group—but there are many other possibilities, as we shall see. This puzzle parallels a similar puzzle with health; within countries, poor health is strongly negatively correlated with life evaluation yet, conditional on income, life expectancy adds nothing to the prediction of country average ladder scores, see Deaton (2008).

For daily happiness, these paradoxes do not exist. Religious people experience more happiness each day, and religious countries also experience more happiness on average. Religious Americans are happier, and religious states in America are happier. Controlling for income makes little difference to these results. In this case, judged by the correlations, daily happiness does better than evaluative wellbeing, but without understanding why, we have no basis for choosing between them.

The individual versus aggregate religion puzzle appears in another form in the American data. Religious people may have higher life evaluation and experience more daily happiness, and there is a large literature demonstrating that they are healthier, but more religious *states* experience more social pathologies, including crime, smoking, divorce rates, teen birth rates, venereal disease, and poor health, see David Myers (2012). Figure 1 plots state level data on age-adjusted murder rates (technically, deaths from assault, from the Centers for Disease Control for years

2008–9) against the average religiosity data from the Gallup poll (excluding Washington, DC, as well as North Dakota where the two year murder rate is too small to be reliably estimated.) Note that murders are rare, only 5.6 per 100,000 people per year, and that while the correlation is high, 0.39, it is far from perfect, and there are very religious states (e.g. Utah) that enjoy the lowest murder rates. The correlation in the Figure is even stronger for all-cause mortality rates (0.72), for mortality from cardiovascular disease (0.68), and is about the same as murder for cancer mortality (0.38). The citizens of more religious states are generally unhealthier, not just more at risk of dying from assault.



**Figure 1: Religiosity and deaths by assault, 2008-09**

With fifty observations, it is easy enough to find spurious correlations, so we have replicated the results using county level data, including all counties for which the Gallup data have at least 50 observations; we have 3,094 matched counties with both Gallup and CDC data for the years 2008 and 2009.

A county-level regression of the age-adjusted all-cause mortality rate on religiosity has a coefficient of 4.36 and a  $t$ -value of 24. A move from San Francisco County, California, where 31 percent of respondents said that religion was important in their daily lives, to one of the several counties in Georgia, Kansas, or Texas, where everyone said so, would increase the age-adjusted mortality rate by 3.01 deaths per 100, compared with a US average of 8.31, with a standard deviation over counties of only 1.45. The introduction of obvious possible confounders, like log income, fraction white, and even average

daily happiness and average ladder values, reduce the effect to 2.84 (the coefficients on income, white, happiness and ladder are all negative; conditional on religion, mortality rates are negatively correlated with happiness, life evaluation, and fraction of the population that is white.)

Why might there be this sharp contradiction between religious people being happy and healthy, and religious places being anything but? As with the other puzzles in this paper, we can only suggest possibilities. One story is that religiosity is a response to a hostile environment, not the cause of it. People are religious in countries where incomes are low, disease is prevalent, personal insecurity is high, and where the state has little capacity or control, so that there are low quality and/or ineffective public services. Religion provides a partial refuge, from disease and distress when it can, and a promise of relief in the next world when it cannot. As states gain capacity to provide for the economic and personal security of their citizens, religiosity declines, and this is one version of the secularization hypothesis, see in particular Ronald Inglehart and Pippa Norris (2001). This theory does relatively little to explain why the US is so religious compared with other rich countries, but does a much better job of explaining variations within the US, as we have seen.

### **III. Other puzzles and conclusions**

We have discussed two unresolved puzzles that we believe deserve further attention. There are many others. The most famous is the Easterlin paradox that in spite of the positive effect of income on life evaluation and on happiness, there appears to have been little effect of economic growth. That there is a paradox at all has been robustly challenged by Daniel Sacks, Betsey Stevenson, and Justin Wolfers (2012), and Easterlin's counter-evidence rests heavily on long-run

Chinese data of dubious comparability. This literature typically does not make the distinction between evaluative and hedonic measures that is so important here.

Other puzzles concern conflicts between human agency and the evaluation of that agency, as when greater autonomy for women is associated with a decline in their self-reported well-being, Stevenson and Wolfers (2009), or when people are less satisfied with their lives when they have children, the vast majority of whom were willingly and even joyously conceived, see Thomas Hansen (2012) for a review. These findings are a good deal less puzzling to psychologists than to economists. According to the former, people suffer from focusing illusions: if women are asked about their increased freedom, and so focus on it, they will say that it makes them happy, but report lower happiness when they are not thinking about it: if parents are asked about their children, they will report that their parenthood is the most important and satisfying thing in their lives, but report lower life evaluation when asked about their lives in a questionnaire in which children are not the focus. Time spent with children is remembered fondly, but not rated very highly at the time. Yet, these explanations do not help us with welfare assessments, nor with the role of SWB measures in the kind of welfare that we seek to promote. To many of us, the capacities that come with greater freedom are as or more important than what we report about our feelings having been granted that freedom. In which case, self-reported wellbeing measures are at best one of many measures that we should be taking into account.

## REFERENCES

- Deaton, Angus, 2008. "Income, health, and wellbeing around the world: evidence from the Gallup World Poll," *Journal of Economic Perspectives*, 22(2), 53–72.
- Easterlin, Richard A. 2013. "Happiness, growth, and public policy," *Economic Inquiry*, 51(1), 1–15.

- Hansen, Thomas, 2012. "Parenthood and happiness: a review of folk theories versus empirical evidence," *Social Indicators Research*, 108, 29–64.
- Helliwell, John, Richard Layard, and Jeffrey Sachs (eds). 2012. *The World Happiness Report*.  
<http://www.earth.columbia.edu/sitefiles/file/Sachs%20Writing/2012/World%20Happiness%20Report.pdf>
- Idler, Ellen L. and Stanislav V. Kasl, 1997. "Religion among disabled and non-disabled persons I: cross-sectional patterns in health practices, social activities, and wellbeing," *Journal of Gerontology Social Sciences*, 52B:6, S294–305.
- Kahneman, Daniel and Angus Deaton. 2010. "High income improves evaluation of life but not emotional well being," *Proceedings of the National Academies of Science*, 107(38), 16489–93.
- Lim, Chaeyoon, and Robert D. Putnam, 2010. "Religion, social networks, and life satisfaction," *American Sociological Review*, 75, 914–33.
- Myers, David G, 2012, "Reflections on religious belief and prosociality," *Psychological Bulletin*, 138(5), 913–7.
- Norris, Pippa and Ronald F. Inglehart, 2004. *Sacred and secular: religion and politics worldwide*. New York. Cambridge University Press.
- Sacks, Daniel W, Betsey Stevenson, and Justin Wolfers, 2012, "Subjective wellbeing, income, economic development, and growth," in Philip Booth, ed., ...*and the pursuit of happiness*, IEA, London.
- Stevenson, Betsey, and Justin Wolfers, 2009, "The paradox of declining female happiness," *American Economic Journal: Economic Policy*, 1(2), 190–225
- Stone, Arthur A., Joseph E. Schwarz, and Joan E. Broderick. 2010. "A snapshot of the age distribution of psychological wellbeing in the United States," *Proceedings of the National Academies of Science*, 107(22), 9985–90.