PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES OPEN ACCESS

Income and emotional well-being: A conflict resolved

Matthew A. Killingsworth[a,1], Daniel Kahneman[b,2], and Barbara Mellers[a]

Edited by Timothy Wilson, University of Virginia, Charlottesville, VA; received May 20, 2022; accepted November 29, 2022

Do larger incomes make people happier? Two authors of the present paper have published contradictory answers. Using dichotomous questions about the preceding day, [Kahneman and Deaton, Proc. Natl. Acad. Sci. U.S.A. 107, 16489–16493 (2010)] reported a flattening pattern: happiness increased steadily with log(income) up to a threshold and then plateaued. Using experience sampling with a continuous scale, [Killingsworth, Proc. Natl. Acad. Sci. U.S.A. 118, e2016976118 (2021)] reported a linear-log pattern in which average happiness rose consistently with log(income). We engaged in an adversarial collaboration to search for a coherent interpretation of both studies. A reanalysis of Killingsworth's experienced sampling data confirmed the flattening pattern only for the least happy people. Happiness increases steadily with log(income) among happier people, and even accelerates in the happiest group. Complementary nonlinearities contribute to the overall linear-log relationship. We then explain why Kahneman and Deaton overstated the flattening pattern and why Killingsworth failed to find it. We suggest that Kahneman and Deaton might have reached the correct conclusion if they had described their results in terms of unhappiness rather than happiness; their measures could not discriminate among degrees of happiness because of a ceiling effect. The authors of both studies failed to anticipate that increased income is associated with systematic changes in the shape of the happiness distribution. The mislabeling of the dependent variable and the incorrect assumption of homogeneity were consequences of practices that are standard in social science but should be questioned more often. We flag the benefits of adversarial collaboration.

Well-being | happiness | income | income satiation | experience sampling

Can money buy happiness? Two authors of this article have published contradictory claims about the relationship between emotional well-being and income. We later agreed that both studies produced valid results and that it was our responsibility to search for an interpretation that explains both findings. We engaged in an adversarial collaboration and asked Barbara Mellers to be the facilitator. This article reports the outcome of our work.

Kahneman and Deaton [(1); hereafter KD] reported data from "more than 450,000 responses to the Gallup-Healthways Well-Being Index, a daily survey of 1,000 US residents conducted by the Gallup Organization in 2008 to 9." The survey included several dichotomous questions about the emotional experience of the preceding day: "Did you experience the following feelings during a lot of the day yesterday? How about ______?" KD computed for each individual the average incidence (scored 1 or 0) of three happy states (happiness, enjoyment, and frequent smiling) and of two "blue" states (worry and sadness). A feature of this scoring method is that each measure of emotional well-being is a fraction, so the complement is a measure of unhappiness.

The article presented detailed analyses that linked emotional well-being (which we will call happiness) to various life circumstances. The two curves in Fig. 1*A* show the means of individual scores on happiness (positive and "not-blue" affect) for different household incomes. The main finding of the KD study is the flattening pattern: the average of happiness scores rises up to a threshold income and then levels off. The evidence for flattening is that, for both the positive and the not-blue measures, averages in the top two categories of income (90 to 120k and 120k or higher) are statistically undistinguishable, despite large numbers of observations in each.

KD concluded that "Emotional well-being [also] rises with log income, but there is no further progress beyond an annual income of ~$75,000." The threshold of $75,000, which has been frequently quoted, is simply the midpoint of the "60 to 90K" income category. A more precise statement would be that there is no further progress in average happiness beyond a threshold at or below 90K.

Matthew Killingsworth [(2); hereafter MK] recruited a large number of participants for a study in which he obtained "1,725,994 experience-sampling reports from 33,391 employed US adults." The participants were prompted on their smartphones to report their current happiness, typically three times per day for several weeks. They answered the question, "How do you feel right now?" on a continuous response scale with end points

Significance

Measures of well-being have often been found to rise with log (income). Kahneman and Deaton [*Proc. Natl. Acad. Sci. U.S.A.* **107**, 16489–93 (2010)] reported an exception; a measure of emotional well-being (happiness) increased but then flattened somewhere between $60,000 and $90,000. In contrast, Killingsworth [*Proc. Natl. Acad. Sci. U.S.A.* **118**, e2016976118 (2021)] observed a linear relation between happiness and log(income) in an experience-sampling study. We discovered in a joint reanalysis of the experience sampling data that the flattening pattern exists but is restricted to the least happy 20% of the population, and that complementary nonlinearities contribute to the overall linear-log relationship between happiness and income. We trace the discrepant results to the authors' reliance on standard practices and assumptions of data analysis that should be questioned more often, although they are standard in social science.

Author affiliations: [a]University of Pennsylvania, Philadelphia, PA 19104; and [b]Princeton University, Princeton, NJ 08544

Author contributions: M.A.K., D.K., and B.M. performed research; M.A.K. analyzed data; and M.A.K., D.K., and B.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Note: Angus Deaton did not participate in this collaboration and should not be taken as endorsing its conclusions.

[1]To whom correspondence may be addressed. Email: mattkil@upenn.edu.

[2]Retired.

Published March 1, 2023.

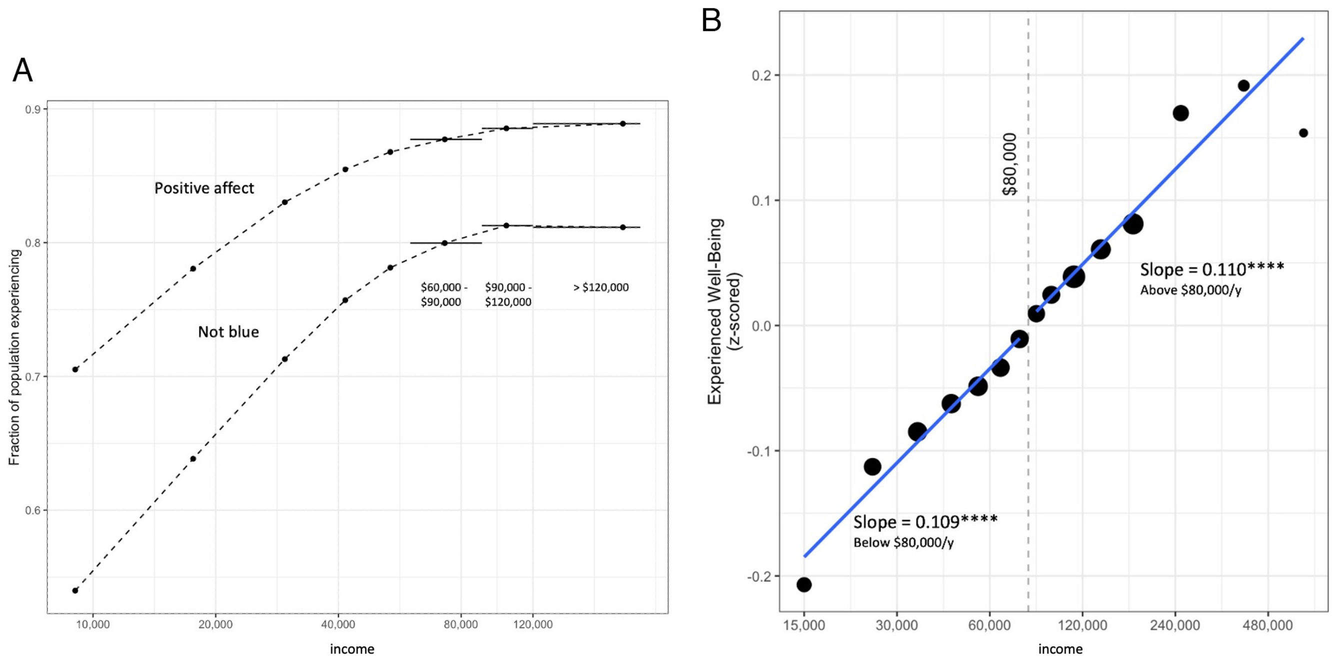

Fig. 1. (*A*) Redrawn from KD. Average fraction of population reporting positive affect (happiness, joy, frequent smiling) and average fraction not reporting negative affect (sadness, worry). The lines connecting the midpoints of the income categories are reproduced from KD. (*B*) Average experienced (emotional) well-being in experience sampling in MK. (Note: Fig. 1*B* differs slightly from Fig. 1 in MK. Fig. 1 in MK compared the income trend of experienced well-being to life satisfaction in participants who provided data for both measures. Fig. 1*B* also includes people who were not asked the life satisfaction question. Fig. 1*B* corresponds to the "Main Results" described in MK, Table 1.)

labeled "Very bad" and "Very good." MK analyzed the means of his measure of happiness as a function of income. His conclusion is described in the title of his article, "Experienced well-being rises with income, even above $75,000." As shown in Fig. 1*B*, MK found a linear relationship between average experienced happiness and log(income) which extended well beyond $200,000. We will refer to the steady increase in average happiness with log(income) as the linear-log pattern.

We agreed that KD and MK had attempted to measure the same construct of emotional well-being. We also agreed that experience sampling was the gold standard for the measurement of emotional well-being, and that MK's continuous measure of happiness was more sensitive than the average of dichotomous Gallup questions. We therefore accepted MK's conclusion about the consistent linear relationship between average happiness and log income. We also agreed that KD had demonstrated the flattening pattern as a robust feature of emotional well-being, which should be replicated in an adequate study.

How could the two patterns be present in the same data? A line of reasoning that eventually proved flawed led us to a hypothesis that proved to be correct. The hypothesis consists of two propositions: 1) There is an unhappy minority, whose unhappiness diminishes with rising income up to a threshold, then shows no further progress; 2) In the happier majority, happiness continues to rise with income even in the high range of incomes. We investigated these propositions in an analysis of MK's experience sampling data.

This article consists of two sections. The first confirms that the linear-log pattern and the flattening pattern are both present in MK's data. The second section explores the question of why the correct solution was not reached earlier. Why did KD overstate the scope of the flattening pattern? And why did MK fail to observe flattening in his original analysis? The answers point to reasonable procedures and standard methodological norms that were applied in the original studies. We conclude that these common practices should be questioned more often.

Results

The goal of our analysis was to search for the flattening pattern in MK's data, but we first had to specify precisely where it would be found. As illustrated in Fig. 1*A*, flattening is characterized by a threshold income, beyond which further increases are not associated with improved happiness. A replication of KD's results in MK's data should not only demonstrate the flattening; it should also reproduce its coordinates.

For the two studies to be comparable, the threshold income beyond which happiness is expected to flatten requires an adjustment because KD and MK used different income categories and collected data at different times: KD gathered data in 2008 to 2009, and MK from 2009 to 2015. Adjusting for inflation, the threshold of <= 90K observed in KD becomes <=97K. The category that contains 97K in MK's study is 90 to 100K. We, therefore, identify two ranges of income, below and above 100k. We use the logarithm of income in our analyses as both KD and MK did.

Fig. 2 presents a comprehensive analysis of the relationship between happiness and log(income) in the experience sampling study. It shows the happiness scores that correspond to five different quantiles of the distribution. For example, the 50th percentile plots the medians of happiness in the different income categories. We used quantile regression to investigate the trends shown in Fig. 2, computing separate slopes in the low range of incomes (less than $100,000) and in the higher range (above $100,000) for each of the five quantiles.

Starting from the low end of the happiness distribution, we find flattening for the 15th percentile, consistent with KD. The happiness of the least happy 15% rises quickly in the lower range of income, leveling off abruptly at $100,000 to a near-zero, statistically nonsignificant slope in the higher range of incomes.

Table 1 provides a more detailed view of the relationship between happiness and log(income) in the lower range (5 to 35%) of the happiness distribution in the experience sampling data. We see flattening in the three lowest quantiles. Note that the range of

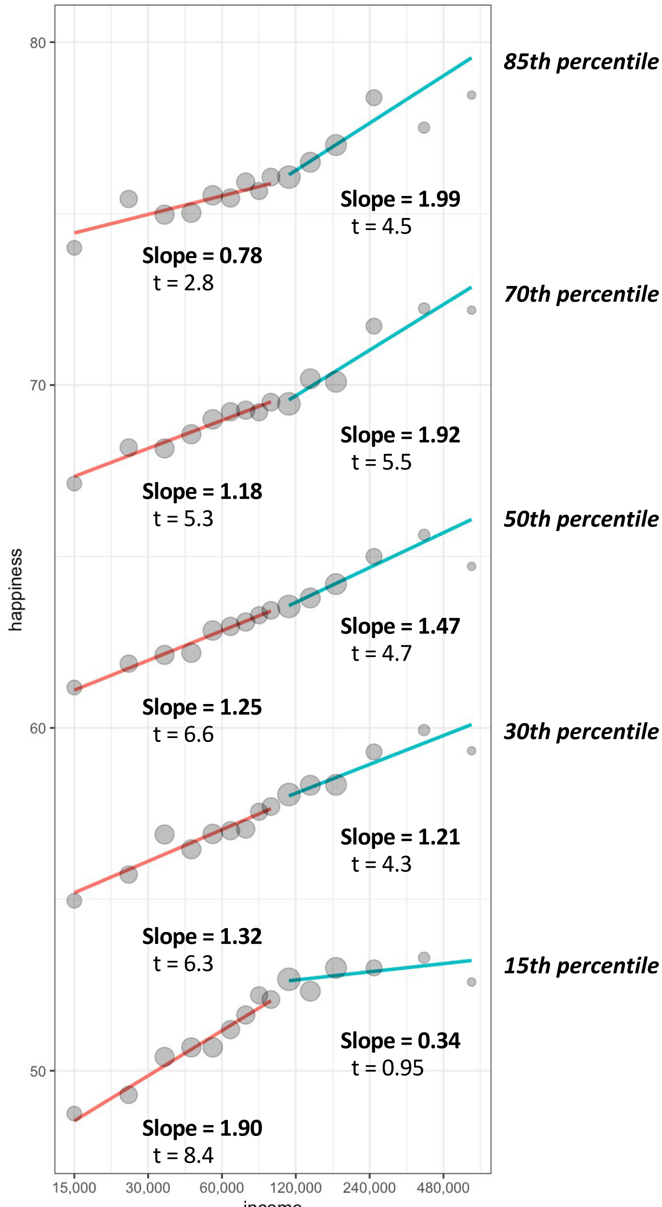

Fig. 2. Emotional well-being of the 15th, 30th, 50th, 70th, and 85th percentiles of the person-level happiness distribution in MK, calculated within each income category. Slopes were calculated below and above 100k, using quantile regression.

happiness values for which flattening occurs in Table 1. (up to 15 to 20%) is quite similar to KD's observations: In the flat region of Fig. 1*A*, the fraction reporting negative affect was 19%, and the fraction not reporting positive affect was 11%.

Also of interest, Fig. 2 shows that the 15th percentile of the happiness distribution is close to the midpoint of MK's continuous scale. Approximately 15 to 20% of people frequently experience negative affect, and the relationship between happiness and income is different in that group and in the happier majority. The suffering of the unhappy group diminishes as income increases up to 100k but very little beyond that. This income threshold may represent the point beyond which the miseries that remain are not alleviated by high income. Heartbreak, bereavement, and clinical depression may be examples of such miseries.

Fig. 2 shows that a different pattern appears in the 30th and 50th percentiles, which matches the trend of the overall average. A third pattern, which we had not anticipated, appears in the 70th and 85th percentiles. For the happiest 30% of people in the various income categories, happiness rises with log(income) at an accelerated rate beyond 100k.

We compared the slopes below and above 100k for each quantile by including an interaction term in quantile regressions. For the 15th percentile, the slope flattens above 100k ($P = 0.0002$). For the 85th percentile, the slope accelerates above 100k ($P = 0.023$). The slopes do not differ significantly in the remaining quantiles (30th, 50th, 70th), although the 70th percentile approaches significance ($P = 0.075$). The acceleration at the 85% percentile and the flattening at the 15th percentile are contrary to the trend that would be expected from floor or ceiling effects (e.g., a ceiling effect would predict a diminishing slope at the high end of the happiness distribution, not an acceleration). The simple linear-log relationship shown in Fig. 1*B* turns out not to be simple after all: it is produced, in part, by complementary nonlinearities.

The correlation between income and well-being is much discussed, both by the public and by social scientists and has been

Table 1. Happiness at different percentiles in MK

Percentile of happiness	Slope up to $100k	Slope above $100k
5% (least happy)	2.34 (t = 4.7)	0.25 (t = 0.5)
10%	1.75 (t = 6.6)	0.52 (t = 1.4)
15%	1.90 (t = 8.4)	0.34 (t = 1.0)
20%	1.84 (t = 8.5)	0.62 (t = 1.8)
25%	1.52 (t = 6.8)	1.12 (t = 3.3)
30%	1.33 (t = 6.3)	1.21 (t = 4.3)
35%	1.26 (t = 6.8)	1.21 (t = 4.1)

The table displays piecewise quantile regression slopes at the low end of the happiness distribution (5 to 35% percentiles). Slopes above $100k were not statistically significant (i.e., flat) for the lowest three percentiles (5%, 10%, and 15%) and marginally positive for the fourth (20%, $P = 0.08$). The remaining slopes above $100k and all slopes below $100k were significantly positive.

the focus of considerable research (3–13). Yet is important to note that the relationship is weak, even if statistically robust. The correlation between average happiness and log(income) is 0.09 in the experience sampling data, for example, and the difference between the medians of happiness at household incomes of $15,000 and $250,000 is about five points on a 100-point scale. The flattening and accelerating patterns are even smaller modulations of a small effect. However, the emotional effects of other circumstances are also small. KD reported that the effect of an approximately fourfold difference in income is about equal to the effect of being a caregiver, twice as large as the effect of being married, about equal to the effect of a weekend, and less than a third as large as the effect of a headache.

Errors and Their Origins. We have noted that KD claimed that the flattening pattern applies to the happiness of the entire population when in fact it is restricted to the lowest 15 to 20% of the distribution. For his part, MK did not discover that the linear-log relationship he observed is due in part to offsetting nonlinearities. We think we now know how these errors occurred.

How did KD come to overstate the scope of the flattening pattern that they discovered? The answer is that they quite reasonably believed that the Gallup questions on which they relied provided a measure of happiness in general, when in fact these questions were only useful as a measure of unhappiness in particular. We now turn to an explanation of this surprising claim.

KD analyzed the relationship between happiness (positive and not-blue affect) and income. The orientation of the variable was the obvious choice because KD were investigating happiness, not misery, just as scholars who study intelligence have tests of intelligence and not of stupidity. But there is an argument against that choice.

The critical observation is that Fig. 1*A* shows the distribution of happiness to be markedly lopsided. In the range of high incomes, in particular, the average reported positive affect is 89% of a perfect score (equivalent to 2.67 on a 0 to 3 scale) and the average of two not-blue items is 81% of the maximum. For example, on average, each dichotomous positive affect item distinguishes between a happier 89% of people and a less happy 11% of people, while the composite report combines three such items. We know, of course, that happy people are not all equally happy. In MK's experience sampling data, for example, the distribution of happiness was approximately normal. The high density of maximum scores in the KD items indicates that the items do not adequately discriminate among degrees of happiness—there is a ceiling effect.

An example illustrates the relevance of a ceiling effect to the naming of a variable. Imagine a test of cognitive functioning which consists of items that most elderly patients pass easily, with a few exceptions due to inattention or momentary confusion. Such a test would rightly be considered a measure of dementia: the number of items failed is an indication of the severity of dementia, but the scale does not discriminate among levels of normal cognitive functioning because most normal people get the same perfect score. A similar argument implies that KD's affect items are best interpreted as measures of unhappiness.

The label of the scale is important because it determines the interpretation of results. Consider two possible summaries of KD's data. With "emotional well-being" replaced by "happiness," KD write: "Happiness rises with income, but there is no further progress beyond ~$75,000." With the scale flipped, the natural summary would be "Unhappiness diminishes with increasing income, but there is no further progress beyond ~$75,000." The two statements describe the same results and initially appear interchangeable. When read carefully, however, they do not have the same meaning.

The difference is in the scope of the assertion. Without further qualification, a statement about wealth, intelligence, or happiness is taken to apply to the average of the whole population, as in the following example: "In this region, there is an East–West gradient in (wealth/intelligence/happiness). "With reversed labels, statements about poverty, dementia, or misery are taken to describe the poor, the demented, and the unhappy. They do not support inferences about other cases.

We believe that if KD had labeled their scale "unhappiness," they could have concluded with confidence that the flattening pattern applies to a category of unhappy people. We also believe that they would have had no grounds to infer that the happiness of happier people flattens in the same way. The narrower statement of the flattening pattern is entirely compatible with a linear-log pattern in average happiness. Indeed, both patterns were confirmed in our analysis of MK's data. In summary, we suggest that KD overstated the scope of flattening because they followed conventional practice in labeling their dependent variable. We are surely not the first social scientists to note that ceiling effects and labels matter.

The main finding of our reanalysis of MK's study is that the shape of the distribution of happiness changes—slightly, but systematically—as income rises. The same increases of income have different effects on the happy and on the unhappy regions of the distribution. In the low range of incomes, unhappy people gain more from increased income than happier people do. In other words, the bottom of the happiness distribution rises much faster than the top in that range of incomes. The trend is reversed for higher incomes, where very happy people gain much more from increased income than unhappy people do. The upper part of the happiness distribution rises with log(income) at an accelerated rate in that range, while the lower 20% is almost completely flat. The middle of the happiness distribution shows approximately linear gains in happiness with rising log(income). We use terms such as "increase" and "gain" for ease of exposition but, to be clear, we are simply describing cross-sectional associations between happiness and income (just as KD and MK did).

The results of this analysis violate a homogeneity assumption that is routinely made—and rarely checked—in the study of bivariate relationships. The assumption is more restrictive than the familiar condition of homoscedasticity. It holds if the conditional distributions of the predicted variable retain the same shape over the entire range of the predictor. In the present case, homogeneity requires the entire distribution of happiness to move in unison as income changes: If the 44th percentile rises by three points, then every quantile of the distribution also rises by three points.

Neither KD nor MK investigated the possibility that the distribution of happiness might change shape as income rises. The two panels of Fig. 1 both describe the happiness/income relationship by plotting conditional means. The choice of display implicitly invokes the assumption that these means provide a sufficient description. For the means to be sufficient, however, homogeneity must hold. MK would have discovered the flattening pattern if he had studied the joint distributions in detail, but the standard practice of social science did not require that step.

We have traced KD's overstatement of the flattening pattern and MK's failure to find it in his data to routine methodological practices which should perhaps be questioned more often. The choice of the orientation of a variable (in this case, happiness or unhappiness) is not always obvious, nor is it inconsequential. In particular, researchers should not take it for granted that their measures function as intended. In KD's case, a scale that was interpreted as a measure of happiness turned out to be a measure of unhappiness, with much consequent confusion. The other source of error in this episode was the norm that allows bivariate relationships to be described without detailed inspection of the joint distribution. The conflict between the two studies would have been avoided by following analytical procedures that are only slightly stricter than current standards.

Concluding Remarks

We wish to flag the increasingly popular process of adversarial collaboration, in which researchers with different views attempt to resolve their disagreement by doing joint research with the help of a friendly arbiter (14–19). The surprising results of Fig. 2 would not have come to light had we not attempted to understand the basis of our conflicting conclusions.

The argument that we have presented is fairly straightforward, and it may suggest that we formulated it quickly and only then turned to the MK data for confirmation. This impression would be misleading. Although we developed the hypothesis that the flattening pattern is restricted to unhappy people quite early, the details of the current treatment took shape over a long time and many versions. Our difficulties were similar to those that affected KD and MK. Routine assumptions have a powerful hold, and precisely formulating how they fail is not easy.

Finally, we must also note that we benefited from luck. The close correspondence that we found between the KD and MK studies is a better result than social scientists can reasonably expect, even when their hypotheses are correct.

Materials and Methods

Sample Information. Participants were 33,391 employed adults living in the United States. The median age was 33, the median household income was \$85,000/year (25th percentile = \$45,000; 75th percentile = \$137,500; mean = \$106,548; SD = \$95,393), 36% were male, and 37% were married. To reduce confounding effects on the association between income and well-being such as unemployment, retirement, and family income transfers, the participants were restricted to employed adults living in the United States of working age (18 to 65) who reported household incomes of at least \$10,000/year.

Experience Sampling Procedure. After providing informed consent, participants completed an intake survey, which included demographic questions as well as two measures of life satisfaction, as detailed below. The participants were next asked to indicate the times at which they typically woke up and went to sleep, and how many times during the day they wished to report on their experiences (default = 3). A computer algorithm then divided each participant's day into a number of intervals equal to the number of desired reports, and a random time was chosen within each interval. New random times were generated each day, and the times were independently randomized for each participant. At each of these times, the participants were signaled via a notification on their smartphone, asking them to respond to a variety of questions about their experiences at the moment just before the signal. The experienced well-being question was asked in every survey. Other questions unrelated to the present investigation were also asked. The participants received notifications requesting a report until they chose to discontinue participation. If 50 samples had been collected, reporting stopped for 6 mo or until the participant requested that it be restarted.

Compliance rate was calculated by dividing the number of actual reports by the number of notifications sent during a participant's "active period," which was defined as the interval between a participant's first and last responses. For example, if a participant received 50 notifications but only completed 25 reports, their compliance rate would have been 50%. The median compliance rate observed was 72%.

Income Measure. Income was measured on an intake survey that occurred prior to and on a different occasion from any of the experienced well-being measures. Thus, income was not made salient by the study design to participants when they were reporting experienced well-being.

Income was measured by asking people, "What is your total annual household income before taxes?" with response options in \$10,000 increments up to \$100,000/year, followed by "\$100,001 to \$125,000, \$125,001 to \$150,000, 150,001 to \$200,000, and over \$200,000.

If a person selected "over \$200,000," then an expanded income range was offered including \$200,001 to \$300,000, \$300,001 to \$500,000, \$500,001 to \$750,000, \$750,001 to \$1,000,000, \$1,000,001 to \$2,000,000, \$2,000,001 to \$4,000,000, \$4,000,001 to \$7,000,000, \$7,000,001 to \$10,000,000, \$10,000,001 to \$20,000,000, \$20,000,001 to \$50,000,000, \$50,000,001 to \$100,000,000, and more than \$100,000,000.

For analysis and visualization, income values were set to the midpoint of the income range selected, e.g., the income value for the income band \$100,001 to \$125,000 was set to \$112,500. In practice, 90.96% of people indicated incomes below \$200,000/year. Incomes over \$500,000 were quite rare, collectively comprising just 1.2% of the sample, and were pooled together and set to a value of \$625,000/year for visualization and analysis (the midpoint of the income band was above \$500,000/year).

Informed Consent. At initial sign-up, participants completed an informed consent form electronically. This research was approved by the UC Berkeley Committee for Protection of Human Subjects and the Harvard University Committee on the Use of Human Subjects.

Data, Materials, and Software Availability. Person-level data are available on OSF at https://osf.io/qye4a/. For privacy, person-level well-being values are slightly rounded (20).

ACKNOWLEDGMENTS. We thank Ville Satopaa and Paul Tetlock for helpful feedback.

1. D. Kahneman, A. Deaton, High income improves evaluation of life but not emotional well-being. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 16489–16493 (2010).
2. M. A. Killingsworth, Experienced well-being rises with income, even above \$75,000 per year. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2016976118 (2021).
3. R. Easterlin, Will raising the incomes of all increase the happiness of all? *J. Econ. Behav. Organ.* **27**, 35–47 (1995).
4. E. Diener, S. Oishi, "Money and happiness: Income and subjective well-being across nations" in *Culture and Subjective Well-Being* (The MIT Press, 2000), pp. 185–218.
5. A. T. Jebb, L. Tay, E. Diener, S. Oishi, Happiness, income satiation and turning points around the world. *Nat. Hum. Behav.* **2**, 33–38 (2018).
6. E. Diener, R. Biswas-Diener, Will money increase subjective well-being? *Soc. Indic. Res.* **57**, 119–169 (2002).

7. J. Gardner, A. J. Oswald, Money and mental wellbeing: A longitudinal study of medium-sized lottery wins. *J. Health Econ.* **26**, 49–60 (2007).
8. B. Stevenson, J. Wolfers, "Economic growth and subjective well-being: Reassessing the Easterlin Paradox" in *Brookings Papers on Economic Activity* (Spring, 2008), 1–87.
9. A. Deaton, Income, health, and well-being around the World: Evidence from the gallup World poll. *J. Econ. Perspect.* **22**, 53–72 (2008).
10. R. E. Lucas, U. Schimmack, Income and well-being: How big is the gap between the rich and the poor? *J. Res. Pers.* **43**, 75–78 (2009).
11. B. Stevenson, J. Wolfers, Subjective well-being and income: Is there any evidence of satiation? *Am. Econ. Rev.* **103**, 598–604 (2013).
12. K. Kushlev, E. W. Dunn, R. E. Lucas, Higher income is associated with less daily sadness but not more daily happiness. *Soc. Psychol. Pers. Sci.* **6**, 483–489 (2015).
13. N. W. Hudson, R. E. Lucas, M. B. Donnellan, K. Kushlev, Income reliably predicts daily sadness, but not happiness: A replication and extension of Kushlev, Dunn, and Lucas (2015). *Soc. Psychol. Pers. Sci.* **7**, 828–836 (2016).
14. G. P. Latham, M. Erez, E. A. Locke, Resolving scientific disputes by the joint design of crucial experiments by the antagonists: Application to the Erez-Latham dispute regarding participation in goal setting. *J. Appl. Psychol.* **73**, 753–772 (1988).
15. B. Mellers, R. Hertwig, D. Kahneman, Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychol. Sci.* **12**, 269–275 (2001).
16. T. Gilovich, V. H. Medvec, D. Kahneman, Varieties of regret: A debate and partial resolution. *Psychol. Rev.* **105**, 602–605 (1998).
17. D. Kahneman, G. Klein, Conditions for intuitive expertise: A failure to disagree. *Am. Psychol.* **64**, 515–526 (2009).
18. N. Cowan *et al.*, How do scientific views change? Notes from an extended adversarial collaboration. *Perspect. Psychol. Sci.* **15**, 1011–1025 (2020).
19. N. Ellemers, S. T. Fiske, A. E. Abele, A. Koch, V. Yzerbyt, Adversarial alignment enables competing models to engage in cooperative theory building toward cumulative science. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 7561–7567 (2020).
20. M. A. Killingsworth, Income and emotional well-being: A conflict resolved. Open Science Framework. https://osf.io/qye4a/. Deposited 1 February 2023.