

NBER WORKING PAPER SERIES

WHAT DO PEOPLE WANT?

Daniel J. Benjamin
Kristen B. Cooper
Ori Heffetz
Miles S. Kimball
Tushar Kundu

Working Paper 33846
<http://www.nber.org/papers/w33846>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2025, Revised June 2026

We acknowledge support from NIH/NIA grants R01-AG065364 to the Hebrew University of Jerusalem, and R01-AG051903 to the University of California Los Angeles. We are grateful to Dmitry Taubinsky and participants at the 2025 ASSA Annual Meeting, at Hebrew University, at SigmaCamp, at the International Society of Quality-of-Life Studies 2025 Annual Conference, and at the Workshop in the Honor of Richard Easterlin (Paris School of Economics) for helpful comments and discussion; and to Amir Adar, Tal Asif, Yehonatan Caspi, Colby Chambers, Arshia Hashemi, Dominic Kassirra, Mattar Klein, Dimitriy Leksanov, Rosie Li, Lev Maresca, Josef McCrum, Tuan Nguyen, Jeffrey Ohl, Shenhav Or, Ofri Piltz, Yonatan Rahimi, Becky Royer, Hannah Solheim, Pierre-Luc Vautrey, and Shira Zadik, for outstanding research assistance. This work utilized the Summit supercomputer and the Alpine high performance computing resource at the University of Colorado Boulder. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or other funding bodies. The authors received IRB approval from the relevant institutions and have no material financial interests that relate to the research described in this paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2025 by Daniel J. Benjamin, Kristen B. Cooper, Ori Heffetz, Miles S. Kimball, and Tushar Kundu. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

What Do People Want?

Daniel J. Benjamin, Kristen B. Cooper, Ori Heffetz, Miles S. Kimball, and Tushar Kundu

NBER Working Paper No. 33846

May 2025, Revised June 2026

JEL No. D12, D90, I31

ABSTRACT

Philosophical perspectives on human desires and values vary; economic theory-driven measurement techniques can provide relevant empirical evidence. We elicited over half a million stated preference choices over 126 dimensions or “aspects” of well-being from a sample of 896 online respondents. We also elicited, via self-reported well-being (SWB) questions, respondents’ current levels of the aspects. From the stated preference data, we estimate for each aspect its relative marginal utility per point on our 0-100 response scale. We validate these estimates by comparing them to alternative methods for estimating preferences, and we offer a range of estimates between those that take self-reports at face value and those that (over-)correct for potential social-desirability reporting bias. Our findings suggest that our respondents value, first and foremost, three basic things: family, money, and health (not necessarily in this order). While commonly studied concepts—such as happiness, life satisfaction, life’s ranking on a ladder, and meaning—are all important, respondents place the highest marginal utilities on aspects related to family well-being and health, and financial freedom and security. We document substantial heterogeneity in preferences across respondents within—but not between—demographic groups, with aspects’ current levels predicting preferences for the aspects.

Daniel J. Benjamin
University of California, Los Angeles
Anderson School of Management
Economics Department
and NBER
daniel.benjamin@anderson.ucla.edu

Kristen B. Cooper
Gordon College
kristen.cooper@gordon.edu

Ori Heffetz
Cornell University
Johnson Graduate School of Management
and The Hebrew University of Jerusalem
and also NBER
oh33@cornell.edu

Miles S. Kimball
University of Colorado Boulder
and NBER
miles.kimball@colorado.edu

Tushar Kundu
Columbia University
tk2859@columbia.edu

A data appendix is available at <http://www.nber.org/data-appendix/w33846>

Over many centuries, philosophers, psychologists, and others have offered varied answers to the fundamental questions of what people do or should want. However, the theories regarding what people do want are based on casual empiricism. In the meantime, economists have developed a sophisticated theoretical and empirical apparatus for studying people's wants—but usually focus exclusively on market goods, leaving out non-market aspects of well-being, such as quality of relationships and personal fulfilment. Here, we bring to bear tools from economics to study over half a million stated-preference choices that we elicited from 896 respondents on Amazon's Mechanical Turk (MTurk), focusing on tradeoffs over 126 aspects of well-being that include many non-market aspects. We find that family well-being, financial security, and health are among the most highly valued. Differences across people are explained both by differences in what they care about and by differences in constraints they face. Our results provide evidence relevant to evaluating theories of human wants, as well as a point of comparison for philosophical theories of what people should want. Furthermore, our results can help prioritize policies and interventions and prioritize what to ask about on national well-being surveys.

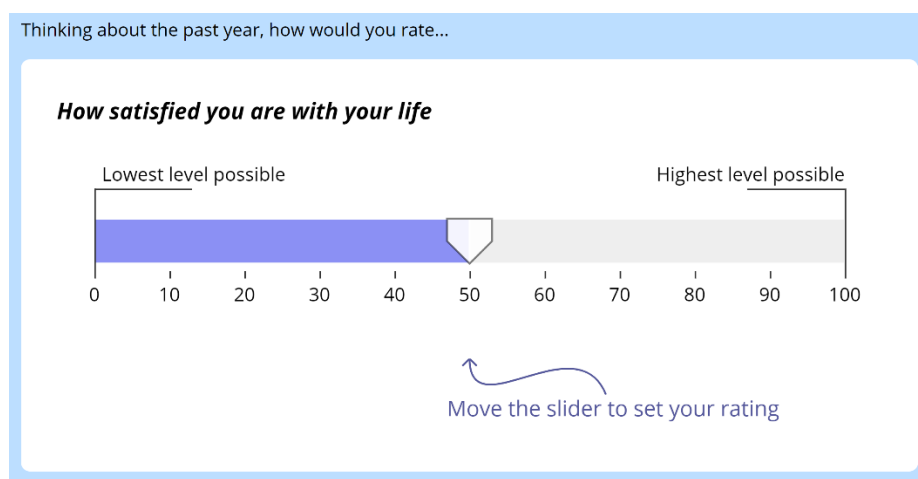
Our approach to quantifying people's wants draws on four insights from economics. First, whereas philosophers distinguish between the related concepts of value, desire, preference, well-being, and the good life, we focus on the concept of preference from economics, meaning what people would choose under idealized conditions when well-informed about the consequences. This choice-based concept is appealing because the main practical use of our results—informing policymakers about what to prioritize when they select among alternatives—could be viewed as supporting a social mechanism aimed at implementing would-be individual choices (Bernheim, 2016). Because the choices we elicit are hypothetical, we use the conventional term *stated preferences*. Our central assumption is that stated preferences are informative regarding “true” preferences. A strong version of this assumption is known to be false. Importantly, stated preferences can be biased by social desirability effects, as we demonstrate below. We therefore report two sets of estimates, *unadjusted* (raw) and *overadjusted* for social desirability, and view them as bounds. We also note that objections to stated preferences may have less force when, as here, stated preferences are used normatively to inform policymakers: to the extent that stated preferences reflect meta-preferences (what people *want* to

want) or “launder” unappealing features (e.g., racism), they may be more relevant for policy than actual choices.

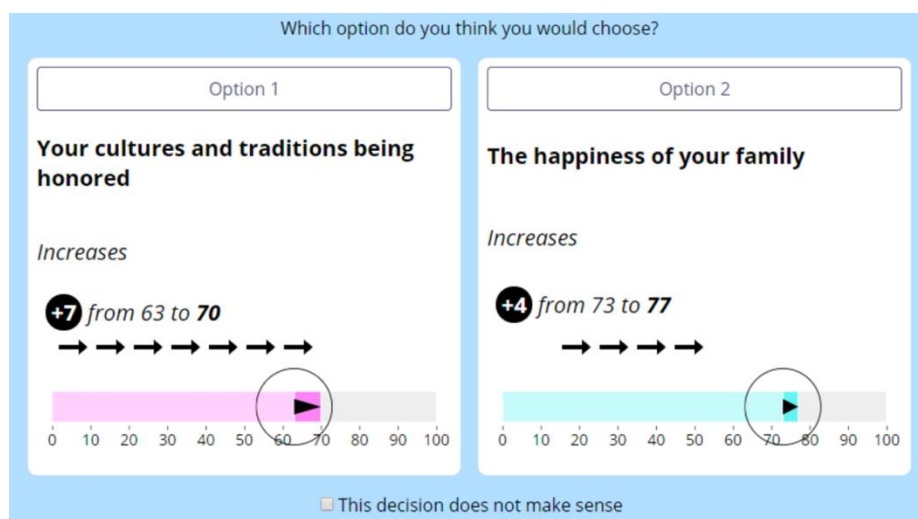
Second, what people want generally depends on what they currently have. For most decision-making and policy purposes, the relevant notion is *(relative) marginal utility*: the (relative) value of a small change from current levels. To estimate marginal utilities of different aspects of well-being, we first ask respondents about their current *levels* (i.e., how they currently rate themselves on the aspects of well-being) on a 0-100 scale (**Figure 1a**), then ask them what they would choose when making *tradeoffs* between improvements (or worsenings) of specific aspects’ current levels, holding other aspects constant (**Figure 1b**). The tradeoff questions enable us to estimate relative marginal utilities.

Figure 1. Screenshots of Stated-Preference Elicitation

(a) Levels Question



(b) Tradeoffs Question



Third, comparing across aspects of well-being requires a common unit. To illustrate with market goods, consider watermelons and grapes. Their relative marginal utilities depend on whether the unit is per piece of fruit or per pound. For market goods, a money metric—e.g., a dollar’s worth of grapes—is a natural unit, but there is no analog for non-market goods measured with subjective units, such as family relationships. We address this units problem by conducting our analysis with alternative, arguably reasonable units. In our main analysis, we adopt respondents’ own 0-100 scale as the common unit and compare marginal utilities of aspects of well-being per point. This choice of units relies on individuals’ judgments of how to translate the magnitude of a “point” across aspects of well-being. As an alternative, we examine marginal utilities in units of sample standard deviations. With those units, we are comparing marginal utilities between aspects by increasing an individual’s placement in the respondent distribution by the same amount. The most closely related prior work (Benjamin et al., 2014b) did not use a quantitative survey scale and did not address the units problem.

Fourth, variation in levels and marginal utilities across aspects of well-being can be understood as driven by a combination of “supply” (differences in opportunity costs) and “demand” (differences in tastes). Aspects are not market goods and rather must be “produced” using resources that may include money as well as time, cognitive and physical effort, emotional energy, and human capital (i.e., acquired skills). When the resource cost of producing an additional unit is high—for example, because of poor health—the aspect has low “supply.” In contrast, holding fixed the resource cost of producing an additional unit, an aspect has high “demand” if people are willing to sacrifice more for it—i.e., have greater taste for it. Economic theory predicts that aspects with low supply will have low levels accompanied by high marginal utilities, whereas aspects with low demand will have low levels accompanied by low marginal utilities. For example, consider ***You having good taste and being up on the latest things.*** Individuals who report high levels of this aspect also tend to have a high marginal utility for improvement. Thus, variation is largely about demand: some care a great deal about being stylish, while others do not. In contrast, for ***You not feeling anxious,*** those with high levels place little marginal utility on further improvement. Variation here is largely about supply, reflecting how easy or hard it is for people to avoid anxious feelings.

Study Design

Aspects of well-being. While our stated-preference methodology can be applied regardless of what is being traded off, we aim to study aspects of well-being that capture human desires for “goods” more fundamental (and abstract) than market goods: ends (e.g., sense of purpose, happiness) rather than means (e.g., consumption of particular goods). Rather than relying on existing sets of aspects believed to be important, we aimed to cover what might be important as comprehensively and agnostically as possible. We compiled an initial list of over 2000 aspects for a pilot study. We focus here on a subset of 126, chosen because they: (i) are commonly studied in the subjective well-being literature and/or collected by statistical agencies (e.g., *How satisfied you are with your life*); (ii) appeared to have high importance (although estimated without units) in Benjamin et al. (2014b) (e.g., *The overall well-being of you and your family*); (iii) contributed most to the average R^2 of predicting all 2000+ aspects in the pilot data (e.g., *The people who have been most important to you in your life still being alive*); or (iv) are regularly traded off in common decisions (e.g., *You getting enough sleep*); we included five additional aspects for idiosyncratic other reasons (e.g., *You and your family having enough to eat*).

Surveys. We conducted four surveys, fielded to the same pool of respondents. We use the first primarily for demographic information and quality-control screening. The second is our main instrument, focusing on two types of questions: *self-reported well-being (SWB) questions (Figure 1a)* and *stated-preference (tradeoff) questions (Figure 1b)*. Every respondent in our main sample answered SWB questions for all 126 aspects of well-being. Every respondent also faced 648 stated-preference questions, in which the starting point for each aspect’s change is the level that the respondent had reported earlier in the SWB question. We use two supplemental surveys to correct for social desirability bias and, more generally, to better understand our results. In one, respondents were asked to rate aspects (using a 9-point Likert scale, see **Extended Data Figure 1**) on 28 characteristics including importance, breadth, susceptibility to social desirability biases, and reasons why people might value the aspect differently. The other survey included socio-demographic and behavioral questions and standard psychology measures including the Marlowe-Crowne Short Form scale for social desirability bias (Reynolds, 1982). We collected data on Amazon’s Mechanical Turk (MTurk) platform, with strict quality control

filters and prior to the popularity of LLM agents. Our main analysis sample consists of 896 respondents.

Results

Estimated marginal utilities of aspects of well-being: unadjusted. Using a Bayesian hierarchical probit model with weakly informative priors for the hyperparameters, for each aspect j we estimate the population mean of log marginal utilities across individuals, μ_j , normalized such that the mean across aspects is zero (i.e., μ_j is mean log marginal utility relative to the average aspect). **Table 2** column 2 shows our estimates (with 95% credible intervals) of the population mean marginal utility, $M_j = e^{\mu_j}$, for each of the 126 aspects. Note that for 5 aspects involving children (colored orange in the table) and 4 aspects involving spouse/partner (colored green), we only include data from individuals with children and spouse/partner, respectively.

Table 2 – Relative Marginal Utility Estimates

(1) Aspect	(2) M_j (LB - UB)	(3) Rank (M_j)	(4) e^{α_j}	(5) M_j Adjusted (LB - UB)	(6) Rank (M_j Adjusted)
Your children's health	66.9 (42.4 - 115.5)	1	25.02	10.6 (6.7 - 18.3)	1
Your children being able to live happy lives	18.6 (13.7 - 26.3)	2	10.94	3.4 (2.5 - 4.8)	13
Your spouse/partner's health	16.2 (11.5 - 23.8)	3	17.03	3.5 (2.5 - 5.2)	12
You having a strong bond with your children	10.8 (7.8 - 15.8)	4	12.40	3.9 (2.8 - 5.6)	11
The health of your parents and siblings	8.6 (6.8 - 11.0)	5	13.26	3.4 (2.7 - 4.3)	14
The people who have been most important to you in your life still being alive	8.0 (6.4 - 10.1)	6	12.31	3.2 (2.5 - 4.0)	18
The overall well-being of your family	7.2 (5.9 - 8.6)	7	6.31	2.4 (2.0 - 2.9)	31
Your family living a wonderful life	5.8 (5.0 - 6.8)	8	5.01	2.8 (2.4 - 3.2)	26
The happiness of you and your family	5.8 (5.0 - 6.8)	9	4.36	1.9 (1.6 - 2.2)	39
Your ability to protect your loved ones	5.6 (4.6 - 6.8)	10	7.81	2.2 (1.8 - 2.7)	36
The overall well-being of you and your family	5.6 (4.8 - 6.7)	11	5.33	3.3 (2.8 - 3.9)	17
Your health	5.0 (4.3 - 5.9)	12	5.65	3.1 (2.6 - 3.6)	20
The happiness of your family	4.9 (4.4 - 5.5)	13	4.52	3.4 (3.1 - 3.8)	15
You being able to support your family financially	4.6 (4.1 - 5.1)	14	4.40	4.3 (3.9 - 4.8)	8
Your financial security	4.3 (3.7 - 5.2)	15	5.33	4.0 (3.4 - 4.8)	9
You having enough money to pay for healthcare and/or medicines that you or your family need	4.2 (3.5 - 5.1)	16	7.83	4.8 (4.0 - 5.8)	7
Your ability to take care of your family	4.0 (3.5 - 4.6)	17	4.79	2.5 (2.2 - 2.9)	29
You not worrying about money	3.8 (3.3 - 4.5)	18	4.50	9.3 (8.0 - 10.8)	2
You having enough money to buy food that you or your family need	3.8 (3.2 - 4.6)	19	7.99	5.2 (4.4 - 6.3)	5
Your physical health	3.7 (3.3 - 4.1)	20	4.18	2.4 (2.2 - 2.7)	32
Your children being kind and caring about other people	3.6 (2.9 - 4.6)	21	7.35	1.1 (0.8 - 1.4)	61
You having enough financial resources for your retirement years	3.3 (2.8 - 3.8)	22	5.19	3.9 (3.4 - 4.6)	10
Your mental health	3.2 (3.0 - 3.5)	23	3.17	4.9 (4.5 - 5.3)	6
Your relationship with your spouse, partner, or closest friend being stronger than ever	3.2 (2.7 - 3.9)	24	4.70	2.5 (2.1 - 3.0)	28
You having enough material assets and resources for the future	2.9 (2.6 - 3.3)	25	4.03	3.0 (2.6 - 3.5)	21
You and your family having enough to eat	2.9 (2.3 - 3.6)	26	14.65	2.0 (1.6 - 2.5)	37
You having enough money to do everything you want to do	2.7 (2.3 - 3.1)	27	4.73	6.4 (5.5 - 7.3)	4
You not having to worry about where your family's next meal will come from	2.6 (2.2 - 3.1)	28	8.05	2.0 (1.7 - 2.4)	38
Your children growing up to follow the golden rule— treating others as they would like to be treated	2.5 (2.0 - 3.2)	29	7.34	0.6 (0.5 - 0.8)	81
The quality of your spouse/partner as a mate	2.5 (2.1 - 3.0)	30	5.18	2.5 (2.1 - 3.0)	30
You not having depression	2.3 (2.0 - 2.6)	31	4.26	3.1 (2.7 - 3.6)	19
You feeling that you have enough money for the things that are most important to you	2.3 (2.0 - 2.6)	32	3.69	3.4 (2.9 - 3.8)	16
Your freedom from physical pain	2.1 (1.8 - 2.4)	33	4.98	1.8 (1.5 - 2.0)	43
The absence of worry in your life	2.1 (1.9 - 2.3)	34	3.11	2.4 (2.1 - 2.7)	33
How happy you feel	2.0 (1.9 - 2.1)	35	2.13	2.3 (2.2 - 2.5)	35
How satisfied you are with your life	2.0 (1.9 - 2.1)	36	1.84	3.0 (2.8 - 3.1)	22
How much you enjoy your life	2.0 (1.9 - 2.1)	37	1.91	2.4 (2.2 - 2.5)	34
You not feeling downhearted and depressed	1.9 (1.7 - 2.1)	38	2.56	2.9 (2.6 - 3.3)	23
The absence of stress in your life	1.9 (1.6 - 2.1)	39	3.38	1.7 (1.5 - 2.0)	44
Your rating of your life on a ladder where the lowest rung is "worst possible life for you" and the highest rung is "best possible life for you"	1.8 (1.7 - 2.0)	40	2.59	2.8 (2.6 - 3.0)	25
The quality of your romantic relationships, marriage, love life or sex life	1.8 (1.5 - 2.1)	41	8.71	2.8 (2.3 - 3.4)	24
You enjoying every day	1.7 (1.5 - 1.8)	42	2.16	1.8 (1.7 - 2.0)	42
You getting the health care you need	1.7 (1.4 - 1.9)	43	5.22	0.7 (0.6 - 0.8)	78
How high your income is compared to the income of other people around you	1.6 (1.4 - 2.0)	44	7.65	6.8 (5.7 - 8.2)	3
You not feeling anxious	1.6 (1.5 - 1.7)	45	3.52	1.8 (1.7 - 2.0)	41
Your physical safety and security	1.6 (1.4 - 1.9)	46	5.41	0.8 (0.7 - 1.0)	70
The absence of sadness in your life	1.5 (1.4 - 1.6)	47	2.54	1.5 (1.4 - 1.6)	47
You feeling calm and peaceful	1.4 (1.3 - 1.6)	48	2.64	1.2 (1.1 - 1.3)	53
Your sense of control over your life	1.3 (1.2 - 1.3)	49	2.29	1.7 (1.6 - 1.8)	45
You having hope	1.3 (1.1 - 1.4)	50	3.24	0.9 (0.8 - 1.1)	66

You feeling well-rested	1.2 (1.0 - 1.3)	51	3.95	1.1 (1.0 - 1.2)	59
You liking what you do every day	1.1 (1.0 - 1.3)	52	2.22	1.5 (1.4 - 1.7)	48
You having people you can turn to in time of need	1.1 (1.0 - 1.3)	53	4.98	1.2 (1.0 - 1.4)	54
You reaching most of your goals over the past 12 months	1.1 (1.0 - 1.3)	54	3.69	1.3 (1.1 - 1.5)	51
You having many options and possibilities in your life and the freedom to choose among them	1.1 (1.0 - 1.2)	55	3.47	1.2 (1.0 - 1.3)	56
You not having to worry about being unemployed	1.1 (0.9 - 1.3)	56	5.94	1.2 (1.0 - 1.4)	52
You feeling active and productive every day	1.1 (1.0 - 1.2)	57	2.50	1.0 (0.9 - 1.1)	64
You having a lot of energy	1.1 (0.9 - 1.2)	58	3.31	1.1 (1.0 - 1.3)	57
Women being treated fairly in your nation	1.1 (0.9 - 1.3)	59	12.70	0.7 (0.5 - 0.8)	80
Your life being more than just getting things done	1.0 (0.9 - 1.2)	60	3.01	1.1 (1.0 - 1.3)	58
The extent to which you feel the things you do in your life are worthwhile	1.0 (1.0 - 1.1)	61	1.89	1.0 (1.0 - 1.1)	62
Your knowledge and skills	1.0 (0.9 - 1.2)	62	4.30	1.0 (0.9 - 1.2)	63
Your sense of purpose	1.0 (1.0 - 1.1)	63	2.42	1.1 (1.0 - 1.1)	60
You not feeling tired all the time	1.0 (0.9 - 1.1)	64	3.95	1.2 (1.0 - 1.3)	55
Your living environment not being spoiled by crime and violence	1.0 (0.8 - 1.2)	65	6.31	0.5 (0.4 - 0.6)	88
You getting enough sleep	1.0 (0.8 - 1.1)	66	5.30	0.7 (0.6 - 0.8)	77
How much of the time the things you do seem meaningful	1.0 (0.9 - 1.1)	67	2.71	1.3 (1.2 - 1.5)	50
You being a good person	0.9 (0.9 - 1.0)	68	3.52	0.8 (0.7 - 0.9)	68
You feeling that you have enough time for the things that are most important to you	0.9 (0.9 - 1.0)	69	2.34	0.8 (0.7 - 0.8)	72
You being able to turn your good intentions into actions	0.9 (0.8 - 1.0)	70	3.40	0.7 (0.6 - 0.8)	76
How satisfied you are with your standard of living compared to the people you spend time with	0.8 (0.8 - 1.0)	71	3.19	2.6 (2.3 - 2.9)	27
How often you smile or laugh	0.8 (0.7 - 1.0)	72	3.86	0.8 (0.7 - 0.9)	71
You being a winner in life	0.8 (0.7 - 1.0)	73	4.33	1.9 (1.6 - 2.2)	40
Your spouse/partner taking on their fair share of responsibility for the household	0.8 (0.7 - 1.0)	74	5.21	0.8 (0.7 - 1.0)	69
Your home being comfortable	0.8 (0.7 - 0.9)	75	3.87	0.7 (0.6 - 0.8)	79
The wisdom you have been able to accumulate	0.8 (0.7 - 0.9)	76	3.65	0.9 (0.8 - 1.1)	65
You not being lonely	0.8 (0.7 - 0.9)	77	3.65	1.5 (1.4 - 1.7)	46
Your friends and family giving you positive energy every day	0.8 (0.7 - 0.9)	78	3.73	0.7 (0.7 - 0.8)	74
You taking care of others instead of just yourself	0.8 (0.6 - 0.9)	79	4.62	0.6 (0.5 - 0.7)	82
The absence of anger in your life	0.8 (0.7 - 0.9)	80	4.20	0.9 (0.7 - 1.0)	67
You having people in your life who see the good in you	0.7 (0.6 - 0.8)	81	3.60	0.8 (0.7 - 0.9)	73
Your ability to climb a flight of stairs	0.7 (0.6 - 0.9)	82	8.69	0.4 (0.3 - 0.5)	98
You not being bothered by having little interest or pleasure in doing things	0.7 (0.6 - 0.8)	83	3.14	1.4 (1.2 - 1.5)	49
You being a good friend and doing your duty by your friends	0.7 (0.6 - 0.8)	84	4.20	0.5 (0.4 - 0.6)	91
You learning or doing something interesting every day	0.6 (0.6 - 0.7)	85	3.82	0.6 (0.5 - 0.7)	83
You not having things stolen from you	0.6 (0.5 - 0.8)	86	5.75	0.3 (0.3 - 0.4)	105
You feeling like you have work-life balance	0.6 (0.5 - 0.7)	87	3.08	0.6 (0.5 - 0.6)	87
You always making time for regular trips or vacations with friends and family	0.6 (0.5 - 0.7)	88	5.72	0.6 (0.5 - 0.7)	86
How much fun you have when you get together with friends	0.6 (0.5 - 0.7)	89	3.84	0.5 (0.4 - 0.6)	90
You finding meaning in the world	0.6 (0.5 - 0.7)	90	3.22	0.4 (0.4 - 0.5)	95
You being satisfied with your job	0.6 (0.5 - 0.7)	91	4.05	0.7 (0.6 - 0.8)	75
You getting to use your strengths to do what you do best every day	0.5 (0.5 - 0.6)	92	3.19	0.4 (0.4 - 0.5)	94
You having a life outside of work	0.5 (0.5 - 0.6)	93	3.88	0.5 (0.4 - 0.6)	89
How much you like your home (house or apartment)	0.5 (0.4 - 0.6)	94	4.41	0.6 (0.5 - 0.7)	84
You showing people the right way to live by your words and actions	0.5 (0.4 - 0.6)	95	4.53	0.5 (0.4 - 0.5)	92
The ability of ordinary citizens to influence your national government	0.5 (0.4 - 0.6)	96	11.33	0.3 (0.2 - 0.4)	110
The overall quality of your experience at work	0.5 (0.4 - 0.6)	97	4.18	0.6 (0.5 - 0.7)	85
You being able to spend a lot of time with your friends	0.5 (0.4 - 0.5)	98	5.11	0.4 (0.4 - 0.5)	93
The air in your area not being polluted	0.5 (0.4 - 0.5)	99	8.23	0.2 (0.2 - 0.2)	117
You having the frequent opportunity to see things of beauty	0.4 (0.4 - 0.5)	100	3.91	0.4 (0.3 - 0.4)	103

You living in a city or area that is perfect for you	0.4 (0.4 - 0.5)	101	4.52	0.4 (0.4 - 0.5)	97
You having someone in your life who always encourages you to be healthy	0.4 (0.4 - 0.5)	102	5.52	0.4 (0.3 - 0.5)	100
The pleasure you get from helping people	0.4 (0.3 - 0.5)	103	4.78	0.3 (0.3 - 0.4)	108
How much you can trust most people in your nation	0.4 (0.3 - 0.5)	104	6.28	0.3 (0.3 - 0.4)	104
You doing your duty	0.4 (0.3 - 0.4)	105	4.60	0.3 (0.3 - 0.3)	111
At work, you getting to use your strengths to do what you do best every day	0.3 (0.3 - 0.4)	106	4.04	0.3 (0.3 - 0.4)	106
You feeling generous	0.3 (0.3 - 0.4)	107	4.95	0.4 (0.3 - 0.5)	99
You having an impact on the world	0.3 (0.2 - 0.3)	108	7.00	0.4 (0.3 - 0.5)	101
How much you trust the police in your nation	0.3 (0.2 - 0.3)	109	12.54	0.3 (0.2 - 0.3)	113
You not being discriminated against for any reason	0.3 (0.2 - 0.3)	110	7.80	0.3 (0.2 - 0.3)	114
How much you trust the courts in your nation	0.3 (0.2 - 0.3)	111	9.60	0.2 (0.2 - 0.2)	119
You having the courage to talk to people you don't already know	0.3 (0.2 - 0.3)	112	7.42	0.3 (0.3 - 0.4)	109
You being satisfied with your city	0.2 (0.2 - 0.3)	113	4.45	0.2 (0.2 - 0.2)	118
How easy it is for you to express who you are through your lifestyle, culture, and beliefs	0.2 (0.2 - 0.3)	114	4.49	0.3 (0.2 - 0.3)	115
You having a leader in your life who makes you enthusiastic about the future	0.2 (0.2 - 0.2)	115	7.10	0.2 (0.2 - 0.3)	116
You having others remember your accomplishments long after your death	0.2 (0.1 - 0.2)	116	8.54	0.4 (0.3 - 0.5)	96
Your understanding of the society you live in	0.2 (0.1 - 0.2)	117	5.98	0.1 (0.1 - 0.2)	121
How easy it is for you to get from home to work and back	0.2 (0.1 - 0.2)	118	8.28	0.1 (0.1 - 0.1)	122
People doing what you tell them to do	0.2 (0.1 - 0.2)	119	9.65	0.4 (0.3 - 0.5)	102
Your stature in your community	0.2 (0.1 - 0.2)	120	6.75	0.3 (0.2 - 0.4)	112
You being part of a grand story	0.1 (0.1 - 0.2)	121	7.63	0.3 (0.3 - 0.4)	107
Your freedom from government interference in your dealings with people who are citizens of other countries	0.1 (0.1 - 0.2)	122	10.22	0.1 (0.1 - 0.1)	123
Your cultures and traditions being honored	0.1 (0.1 - 0.1)	123	14.95	0.1 (0.1 - 0.1)	125
You having good taste and being up on the latest things	0.1 (0.1 - 0.1)	124	8.15	0.2 (0.1 - 0.2)	120
How much you trust the media in your nation	0.1 (0.1 - 0.1)	125	16.97	0.1 (0.1 - 0.1)	126
Your enjoyment of risky situations	0.0 (0.0 - 0.1)	126	21.12	0.1 (0.1 - 0.1)	124

Notes: Bayesian hierarchical model estimates using Hamiltonian Monte Carlo, based on 574,282 tradeoff responses from 896 respondents who passed quality control and completed the first nine blocks of our Multi-Block Survey. Aspects related to standard self-reported well-being questions in purple. Estimates for aspects related to children (in orange), and those related to a spouse/partner (in green), only use data from individuals with children or a romantic partner, respectively. Sample sizes: 423 respondents have at least one child and a romantic partner, 213 have neither, 79 have children but no partners, and 181 have partners but no children. Columns: (1) Aspect names; (2) Relative marginal utility per point $M_j = e^{\mu_j}$ with 95% credible intervals in parentheses; (3) Rank based on M_j ; (4) e^{σ_j} , where σ_j is the estimated standard deviation of individual-level log marginal utilities for aspect j ; (5) Relative marginal utility per point, adjusted for social desirability and sample shares of respondents with children/partners; (6) Rank based on the adjusted M_j . Estimates are normalized so that the average log marginal utility across aspects equals zero; thus M_j in columns (2) and (5) can be interpreted as the marginal utility of aspect j relative to the geometric mean across all aspects.

Whereas the estimates in column 2 are in units of points on the 0-100 scale, **Extended Data Table 2** (column 4) shows analogous estimates in units of standard deviations of aspect levels in the survey population. The correlation between the two sets of estimates is 0.992 (rank correlation: 0.992). The correlation is nearly perfect because, measuring aspect levels in points, the variance across aspects in log marginal utility (1.47) is two orders of magnitude larger than the variance across aspects of their log standard deviation (0.023). That is, the variation in aspects' marginal utilities (in points) swamps the variation in aspects' standard deviations (in points). Thus, while either choice of units could be justified, it makes little difference to the results.

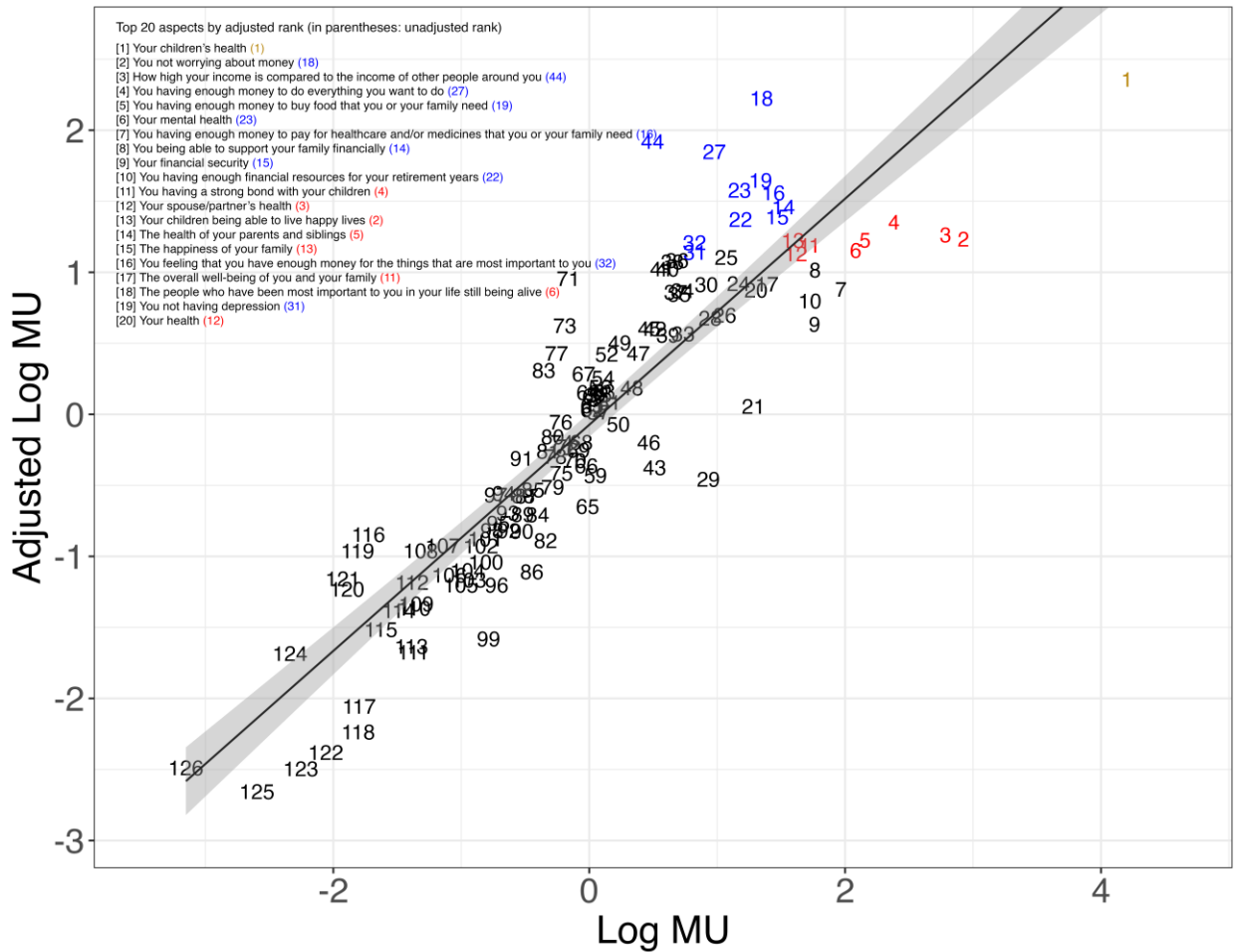
Estimates adjusted for social-desirability bias and child/partner shares. To the extent that respondents want to give a good impression of themselves (Edwards, 1957), their responses to the stated-preference questions would exaggerate the relative marginal utility of prosocial aspects. To quantify and correct for this bias, we asked respondents (in a separate survey) to rate each aspect (on a 9-point Likert scale) on whether “someone would downplay how much they want this aspect because openly wanting it might make them look bad.” Consistent with social-desirability bias, an aspect’s mean rating correlates -0.40 with its marginal utility estimate ($t(124) = -4.83, P = 3.93 \times 10^{-6}$). However, this correlation might exaggerate the magnitude of bias if, on average, aspects with estimates biased downward (upward) also have lower (higher) true marginal utilities—as is likely the case because, on average, things that people consider more desirable individually (e.g., being healthy, not feeling anxious) come to also be considered more desirable socially.

In column 5 of **Table 2** we implement two adjustments. First, to address social desirability, we regress the μ_j point estimates (from column 2) on these survey-based aspect-level measures of social desirability; we then use the exponentiated residuals as social-desirability-corrected estimates of relative marginal utility. These likely *overcorrect* for social desirability if, as noted above, an aspect’s social desirability is correlated with its true marginal utility. Second, for the 5 aspects related to children and 4 related to a spouse/partner, we multiply their exponentiated residuals by the share of our sample with children (0.56) and with a romantic partner (0.67), respectively, effectively treating the marginal utility of these aspects as 0 for individuals without children or without a partner—also likely an overcorrection. We thus view our estimates as bounds: the unadjusted estimates in column 2 are undercorrected whereas our adjusted estimates in column 5 are overcorrected.

Figure 2 compares the unadjusted with the adjusted estimates. Aspects are indicated by their unadjusted rank (column 3 in **Table 2**). The rank correlation is reassuringly high: 0.925 (SE 0.018). However, the figure also reveals some notable effects of the adjustment. For example, as the list of top 20 aspects by adjusted rank (top left of **Figure 2**, based on column 6 in **Table 2**) shows, our adjustment moves up (in blue) many aspects related to financial security (roughly speaking, from top twenty and top thirty to top ten), while moving down (in red) aspects related to family (roughly, from top ten to top twenty). Specifically, while the very top aspect—*Your children’s health*—which combines both family and health, remains ranked #1 with and without

adjustment, the rest of the unadjusted top ten (unadjusted #2-10) drop to the next ten and below (adjusted #11 and below). At the same time, *You not worrying about money* (unadjusted #18), *How high your income is compared to the income of other people around you* (unadjusted #44), and *You having enough money to do everything you want to do* (unadjusted #27) are ranked, after adjustment, #2, #3, and #4. We also note that after adjustment, *Your mental health* (unadjusted #23) climbs to #6—while *Your physical health* (unadjusted #20) drops to #32 (see Table 2).

Figure 2. Log Marginal Utility (MU) versus Adjusted Log MU



Notes: Comparison of log marginal utility estimates before and after (over-)adjustment for social desirability and child/partner shares. Scatterplot points are numbers showing unadjusted log MU rank (column 3 in Table 2). Horizontal axis is unadjusted log MU (column 2 in Table 2). Vertical axis is adjusted log MU (column 5 in Table 2). OLS regression line shown with 95% confidence interval. Top 20 aspects by adjusted log MU are listed in upper-left quadrant. Within this set, aspects that move up in rank after adjustment are indicated in blue; aspects that move down, in red; and a single aspect ranked the same, in gold.

To further validate our social desirability bias adjustment, we tested the hypothesis that respondents who were more susceptible to the bias would have lower estimated marginal utilities

for less socially desirable aspects. We use the Marlowe-Crowne Short Form (Reynolds, 1982) to get a respondent-level measure of social desirability susceptibility, split the sample at median social desirability, and re-estimate our model on each subsample. For each aspect, we calculate the difference in log marginal utility between subsamples. We regress this difference in log marginal utility on our (9-point Likert scale) aspect-level “want to downplay” measure (see **Extended Data Figure 4**). As hypothesized, we observe a steep negative slope (-0.22 , $t(124) = -5.72$, $P = 7.44 \times 10^{-8}$).

Unadjusted and adjusted marginal utility estimates: summary. The very top aspect (**Table 2** and **Figure 2**), ***Your children’s health***, stands distinctly apart from the rest, with an estimated post-adjustment relative marginal utility approximately 11 times larger than the average aspect (and 67 times larger pre-adjustment). Aspects related to social taboos, such as money and mental health, move up dramatically after adjustment for social-desirability bias. At the same time, roughly the top third of our 126 aspects, as well as roughly the bottom third, consist of many of the same aspects before and after adjustment (see **Table 2** and **Figure 2**). Looking at the top third, those with the highest marginal utilities before and after adjustment cluster around three key domains: family well-being, health (both mental and physical), and financial security. The relative importance of the three depends on the extent to which our adjustment over-corrects.

While the top family-related aspects have post-adjustment marginal utilities at least 3 times higher than the average aspect (10-20 times higher pre-adjustment) and money-related aspects at least 4 times higher, standard SWB measures (purple in **Table 2**), such as ***How happy you feel*** and ***How satisfied you are with your life***, have marginal utilities at *most* 3 times the average. Interestingly, their estimated geometric standard deviation (column 4) in marginal utility across respondents is low, ranging from 1.84 to 3.52. This makes them—along with several other often-discussed aspects of well-being such as enjoying life, avoiding depression and sadness, and having a sense of purpose, control, and meaning—among the aspects valued most consistently across respondents.

Coherence checks. We conduct two additional analyses to check the coherence of our results. First, we examine subset/superset pairs (e.g., ***Your health*** vs. ***The health of you and your family***) and find that the ranking of aspects largely follows the expected pattern: for 25 out of 27 pairs, the superset aspect receives higher marginal utility. For the two exceptions, we

speculate that salience or mental averaging may explain the reversal (see **Supplementary Discussion section 3.1**). Second, we were concerned that if respondents differ in how they weigh numeric changes versus aspect text in stated-preference questions, this could result in an inflated variance of marginal utility estimates for text-focused respondents, distinct from true differences in the intensity of preferences. (An extreme example would be innumerate respondents appearing to have lexicographic preferences.) We compared the variance of log marginal utility means for high- and low-numeracy respondents (based on seven math questions) and found similar spreads, suggesting our results are not driven by this type of differential attention (see **Supplementary Discussion section 3.2**).

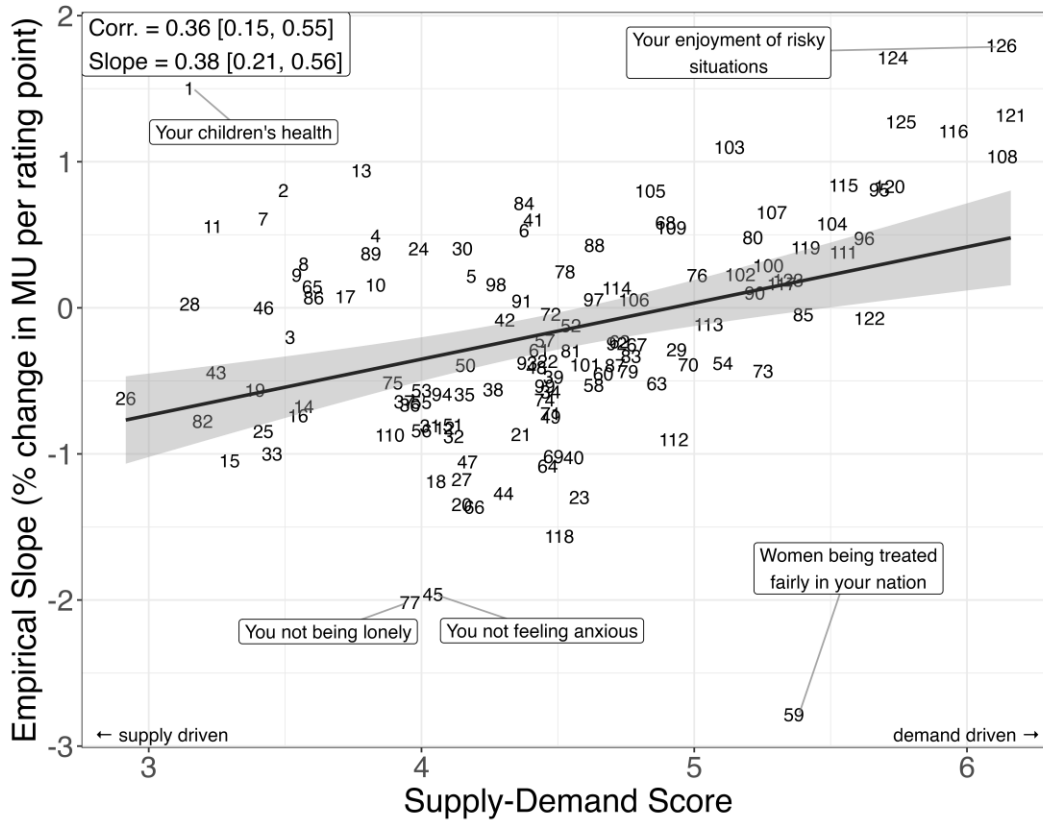
Alternative measures of importance. While our stated-preference questions are designed to quantitatively measure marginal utility, they may be cognitively challenging for respondents to answer. We investigated three simpler questions (in a separate survey). (1) “How important is this aspect to you?” While simple to state, it has the disadvantage of being ambiguous: it could be interpreted, for example, as eliciting total or average, rather than marginal, utility from the aspect. We find that responses correlate 0.85 with our (unadjusted) marginal utility estimates (**Extended Data Figure 2, panel i**).

Our two other questions are designed to more clearly elicit marginal utilities: (2) “Think about your own life as it is. Given how much you have of this aspect, how eager would you be to get more? In particular, compared to other aspects of life, how valuable would it be to have a little bit more of this aspect?” and, even more explicit about tradeoffs, (3) “Imagine you could gain more of this aspect by giving up other things. How much would you be willing to give up?” Responses to these questions correlate 0.87 and 0.94, respectively, with our unadjusted estimated marginal utilities (**Extended Data Figure 2, panel ii and panel iii**). Although we cannot reject the null hypothesis that these three high correlations are equal ($t(125) = 1.75, P = 0.08$ for the third versus first correlation), their progressive increase is consistent with how we interpret them. Taken together, we view the evidence from these simpler questions as providing additional support for the evidence from our marginal utility estimates. We also find informative differences across these alternative measures (e.g., aspects such as ***You not being discriminated against for any reason*** have systematically lower rankings based on explicit tradeoffs than they do based on simpler questions); see **Supplementary Table 1** and **Supplementary Note 5.1**.

Understanding heterogeneity across individuals. For every aspect we study, the standard deviation of marginal utilities across individuals is substantial relative to the mean (**Table 2** column 4). For some aspects, respondents with higher levels have higher marginal utilities, whereas for other aspects, the reverse holds (**Supplementary Figure 1**). We estimate 126 aspect-specific regressions of respondents' log marginal utilities on their levels. We find that 14% of aspects have a significant (at $P < 0.01$) positive slope and 46% have a significant (at $P < 0.01$) negative slope. On average, the estimated slopes correspond to a 0.5 percent increase and 0.7 percent decrease, respectively, in marginal utility per one-point change in level.

According to a supply-and-demand framework (detailed in **Extended Data Figure 3**), aspects with positive slopes should reflect predominantly demand (i.e., taste) differences across individuals, whereas aspects with negative slopes should reflect predominantly supply (i.e., constraint) differences. To test this prediction, we asked respondents to rate each aspect (on a 9-point Likert scale) on whether people differ predominantly in supply or demand: "...think about **people you know who don't care much about getting more of [the aspect]**. They might not care much about getting more for either of two reasons: (1) They already have plenty of it. (2) They don't value it much. For each aspect below, of people who don't care much about getting more, what fraction have each reason?" (bolded as shown to respondents). High scores correspond to an aspect being demand-driven; low scores to being supply-driven. **Figure 3** compares each aspect's supply-demand score (on a 9-point scale) with its slope (from the 126 regressions) of the relationship between log marginal utility and level. Consistent with the supply-and-demand framework, the correlation between the two is positive: 0.36 ($t(124) = 4.33$, $P = 3.0 \times 10^{-5}$). Looking at some examples at the extremes, the aspect ***Your enjoyment of risky situations*** (unadjusted rank #126 in **Table 2**) has the steepest positive slope and among the highest supply-demand scores; intuitively, people differ in how much they enjoy risky situations, while the supply of risky situations in life is not a constraint. At the other extreme, the aspects ***You not being lonely*** (#77) and ***You not feeling anxious*** (#45), with the second and third most *negative* slopes, are also in the bottom quarter of supply-demand scores; intuitively, nobody wants to be lonely or feel anxious, but people differ in the constraints they face when attempting to avoid loneliness and anxiety. We highlight, however, that the correlation is far below 1.00, and, accordingly, many aspects are far from the regression line (some are labeled in **Figure 3**).

Figure 3. Supply- and Demand-Driven Differences in Marginal Utility (MU) Estimates



Notes: Horizontal axis: supply-demand 9-point rating of aspects, i.e., stating why some people “don’t care much about getting more” (1 = almost all already have plenty, 3 = most already have plenty, 5 = both reasons equally common, 7 = most don’t value it much, 9 = almost all don’t value it much). Based on 364 Aspect-Characteristics Survey respondents rating subsets of 126 aspects ($N = 11,582$), with aspect-level characteristic ratings averaged across respondent (see **Methods**). Vertical axis: empirical log marginal utility-level slopes from regressing individual log MU estimates m_{ij} on levels for aspect j ; estimated slope is multiplied by 100 to facilitate interpretation as percentage change in marginal utility for a one-point change in level. OLS regression line shown with 95% confidence interval. Points are marked by unadjusted log MU rank (column 3 in **Table 2**).

Demographic heterogeneity. As a simple way to investigate demographic differences, we re-estimate the model separately within binary sample splits (e.g., men versus women) and test aspect-level differences in unadjusted average log marginal utility (i.e., without social-desirability correction) using each group’s posterior mean and posterior SD, assuming independence across subgroup estimates (**Supplementary Figure 2**). We find remarkable consistency in relative marginal utilities across splits by sex, income, age, race, education, number of children, employment status, relationship status, political stance, religiosity, and numeracy. For example, the correlation of mean-male with mean-female marginal utilities is 0.979 (not statistically distinguishable from 1.00; $t(124) = -1.15$, $P = 0.25$). Although perhaps

surprising, these findings are consistent with the limited demographic heterogeneity in preferences reported by Benjamin et al. (2014b).

Despite overall concordance across demographic groups, there are several aspects of well-being that exhibit meaningful group differences in marginal utility (labeled points in **Supplementary Figure 2**). The largest are: Democrats place higher marginal utility on *Women being treated fairly* (difference = 1.95); Republicans, on *Your culture and traditions being honored* (diff = 1.33); religious respondents, on *You being part of a grand story* (diff = 1.07) and *You showing people the right way to live by your words and actions* (diff = 0.86); employed respondents, on *You having a life outside of work* (diff = 0.96) and *You feeling like you have work-life balance* (diff = 0.77); lower-income respondents, on *You not worrying about money* (diff = 0.86) and *You feeling that you have enough money for the things that are most important to you* (diff = 0.63); and older respondents, on *Your physical health* (diff = 0.63). The supply-and-demand lens helps interpret some of these differences. For Democrats and lower-income respondents, these higher marginal utilities coincide with lower reported levels for the mentioned aspects, consistent with lower perceived supply in these groups. In contrast, religious respondents report *higher* levels and have higher marginal utility, implying higher demand for these aspects.

Discussion

We have provided systematic evidence on the relative marginal utilities of a wide range of aspects of well-being. We draw five main conclusions. First, existing theories of what people want may have underemphasized some aspects that we find are most highly valued, related to family, health, and money, such as *Your children's health* (the top aspect by a large margin) and *Your financial security*. Second and relatedly, like other recent work (Bernheim et al., 2024; Benjamin et al., 2023; Benjamin et al., 2014b), we find that happiness and life satisfaction, which are often assumed to capture all of what people care about, have far lower relative marginal utilities than many other aspects of well-being. This finding also extends to other aspects that have received increasing attention, including affective SWB measures (e.g., enjoying life, avoiding depression, and avoiding sadness) and eudaimonic ones (e.g., having a sense of purpose, control, and meaning). Third, our results confirm that social desirability may bias stated preferences; while the general findings above hold both before and after adjustment for such

bias, more specific findings change, including the relative importance of money versus health and family, and that of mental versus physical health. Fourth, while there is substantial heterogeneity in preferences across individuals, little of it can be identified with the sociodemographic variables we measured. Finally, heterogeneity can be driven primarily by supply or demand, depending on the aspect. Moreover, the supply-demand framework seems to explain several of the few cases where we do find heterogeneity in preferences across demographic groups.

Our empirical results can be compared to the answers proposed in religious, philosophical, and other scholarly writings to the questions: What do people want? What should people want? **Supplementary Table 2** lists selected scholarly theories and highlights aspects most closely matching each theory and their ranks, both before and after adjustment. While constructing the table involves judgment and is merely illustrative, we highlight a few intriguing observations. The importance of family-related aspects, involving children, spouses, or loved ones, appears aligned with Adler's (1933) emphasis on contribution to one's kin and close relations, and relative income's high rank after adjustment appears aligned with Duesenberry's (1949) emphasis on relative income and social status concerns. However, overall, the most striking pattern is that, by and large, the theories focus on aspects that we estimate to have relatively low marginal utilities. For example, the middling marginal utilities of happiness and life satisfaction are in tension with views of happiness as the ultimate goal (Bentham, 1789; Mill, 1863; Layard, 2005)—at least as measured via self-reports. The middling marginal utilities of the aspects that religious thinkers say people *should* want—cessation of grasping desire for Buddhism, accumulating wisdom for Avicenna, loving neighbors for Martin Luther—may help explain why people struggle to live according to these prescriptions.

In addition to providing empirical evidence relevant to theoretical accounts of human goals, our results have direct practical relevance. One strand of the burgeoning “Beyond GDP” movement explores measuring well-being through surveys (Daly & Cobb, 1989; Dolan & Metcalfe, 2012; Kubiszewski et al., 2013); our findings can help inform what should be measured by such surveys, by identifying the most highly valued aspects, such as family well-being, health (including mental health), and financial security. Prioritizing policies that increase these aspects may generate the largest improvements in well-being, and our analysis by

demographic groups, where we find limited preference differences, suggests such policies could have broad support.

A limitation is that we rely on stated preferences. While we use novel approaches to probe our results, including comparing our tradeoff-based marginal utilities to importance ratings and adjusting for social-desirability bias, stated preferences could differ from “true” preferences for various reasons, including simply that people may not be able to fully imagine changes in certain aspects of well-being. By finding real-world situations in which people face tradeoffs for a small set of aspects of well-being (as in Benjamin et al. 2014a), future work with real choices could provide an independent test of our methodology. Another limitation is that our results are from one particular online sample (although, again, we find limited preference heterogeneity by the observable demographics we measured). Future research should explore how preferences may differ across cultural contexts and evolve over time. Finally, developing methodologies to translate these marginal utility estimates into practical welfare indices could provide valuable tools for policy evaluation.

Methods

Survey Design and Data Collection. We conducted four surveys, fielded to the same pool of respondents. The first survey, which we call Baseline, included the demographic questions which we use for analysis here, among other questions (see **Supplementary Methods section 1.2.ii.** for details). To be invited to participate in any of the other three surveys, respondents had to first complete the Baseline survey and pass the quality control filter; the quality control procedure is described in **Supplementary Methods section 1.3.**

The first of the follow-up surveys, which we call the Multi-Block Survey, is our primary data source. The survey structure allowed us to ask each respondent many questions; for full details, see **Supplementary Methods section 1.2.** The self-reported well-being (SWB) and tradeoff questions appeared in design using random triples of aspects: after each set of three SWB questions, the respondent completes six tradeoffs with those three aspects. (This provides a lot of information about the variance of each individual's response errors.) Every respondent in our main sample answered SWB questions, as in **Figure 1a**, for all 126 aspects of well-being whose marginal utilities are reported in Table 2. These 126 aspects of well-being were selected from an initial list of over 2000 candidate aspects, as described in the main text. Every respondent also faced 648 stated-preference questions, as in **Figure 1b**, in which the starting point for each aspect's change is the level that the respondent had reported earlier in the SWB question. To prevent default responding, the submit button was disabled until the slider was moved. The magnitudes of the changes were randomly chosen from the integers 1 to 8, with equal probability (in **Figure 1b**, for example, these magnitudes are +7 and +4, respectively), and separately randomized to be either both positive or both negative changes with equal probability. Before answering tradeoff questions, respondents were given extensive instruction and practice questions.

We call the second follow-up survey the Aspect-Characteristics Survey. Respondents were asked to rate aspects (using a 9-point Likert scale, see **Extended Data Figure 1**) on 28 characteristics including importance, breadth, susceptibility to social desirability biases, and reasons why people might value the aspect differently. See **Supplementary Methods 1.2.v.** for design details. We refer to the final survey as the Life and Psych Survey. It included socio-demographic and behavioral questions and standard psychology measures, including the Marlowe-Crowne Short Form scale for social desirability bias (Reynolds, 1982).

Sample. We collected data on Amazon’s Mechanical Turk (MTurk) platform, with strict quality control filters and prior to the popularity of LLM agents. As identified by MTurk, respondents were aged 18 or older, with at least 100 previously approved HITs and a HIT approval rating of $\geq 95\%$, and located in the United States. The Baseline survey was fielded between June and December, 2022; the Multi-Block Survey was fielded between September 2022 and January 2023; the Life and Psych Survey was fielded between June 2022 and December 2022; the Aspect-Characteristics Surveys used in this paper comes from sixteen tranches fielded between September 2024 and January 2025.

Our main analysis sample consists of 896 respondents who (i) passed the Baseline quality control filter and (ii) completed at least the first 9 blocks of the Multi-Block Survey, which included SWB and stated-preference questions about all 126 aspects. (The Multi-Block Survey had 30 blocks in total.) The Life and Psych Survey sample for this paper consists of the 811 respondents among these 896 who also (iii) completed the Life and Psych follow-up survey upon invitation. We split the Aspect-Characteristics Survey into 16 tranches of 301 questions each (43 aspects rated on 7 characteristics) to keep the survey length reasonable; our sample consists of 1088 total respondents who (i) passed the Baseline quality control filter and (ii) completed a tranche of the survey upon invitation. Only some of them (83%) also completed the Multi-Block Survey. Each aspect-characteristic cell was rated by 84 (SD: 10) respondents, on average.

Although similar to the U.S. population on most demographics, our respondents are more likely to have completed college, be younger than 50, and be unemployed; and less likely to have annual household income above \$120,000 and to be Black or Hispanic/Latino (**Extended Data Table 1**). As a self-selected sample, they are likely to systematically differ from the general population also on additional, unobservable characteristics. However, our finding of remarkably similar stated preferences across demographic groups within our sample helps alleviate the concern that a more representative sample would have yielded dramatically different results.

Econometric Model and Inference. Our econometric model assumes that, when answering stated-preference questions, respondents maximize utility except for a response error, modeled as an additive random error to log marginal utility. Denote individual i ’s level of aspect j by w_{ij} , i ’s vector of levels by $\mathbf{w}_i$, individual i ’s utility function by $u_i(\mathbf{w}_i)$, and marginal utility

(MU) by $M_{ij} \equiv \frac{du_i}{dw_{ij}}$. The change in utility that results from a small change in the level of aspect j for individual i , Δw_{ij} , can be approximated as linear:

$$\Delta u_i = u_i(w_{ij} + \Delta w_{ij}, \mathbf{w}_{i,-j}) - u_i(w_{ij}, \mathbf{w}_{i,-j}) \approx M_{ij} \Delta w_{ij},$$

where $\mathbf{w}_{i,-j}$ is the subvector that excludes j .

Tradeoff questions ask respondents to choose between two options: a (randomized) increase in aspect j , Δw_{ij} and a (randomized) increase in a different aspect j' , $\Delta w_{ij'}$. Preference maximization implies that a respondent chooses the option changing aspect j if and only if $M_{ij} \Delta w_{ij} > M_{ij'} \Delta w_{ij'}$. Taking logs (and denoting $\ln(M_{ij})$ by m_{ij}), rearranging, and allowing for response error $e_{ijj'q}$ that can be different for each specific tradeoff question q :

$$m_{ij} - m_{ij'} + \ln\left(\frac{\Delta w_{ij}}{\Delta w_{ij'}}\right) + e_{ijj'q} > 0.$$

To allow for and estimate preference heterogeneity, we assume a hierarchical model. Let μ_j and σ_j be hyperparameters representing the mean and variance of log marginal utility in the population for aspect j . Individual log marginal utilities are modeled as random draws:

$$m_{ij} \sim N(\mu_j, \sigma_j).$$

Since the scale of marginal utilities is not identified, we impose the normalization that the average across log marginal utility means, $\bar{\mu}$, equals 0. Therefore, all log marginal utility estimates can be interpreted as log marginal rates of substitution relative to the average aspect.

We model response error as

$$e_{ijj'q} \sim N(0, \sigma_{e_i}),$$

whose variance is individual-specific with a log-normal distribution in the population,

$$\sigma_{e_i} \sim \text{lognormal}(\mu_e, \sigma_e),$$

where μ_e and σ_e are hyperparameters. Note that by allowing the variance of response error to vary by individual, our estimates of μ_j and σ_j will weight more heavily responses from respondents who make more consistent choices.

The conditional likelihood of observing any given binary choice is

$$\mathbb{P}[r_{ijj'} = 1 | \Delta w_{ij}, \Delta w_{ij'}, m_{ij}, m_{ij'}, \sigma_{e_i}] = \Phi\left(\frac{m_{ij} - m_{ij'} + \ln\left(\frac{\Delta w_{ij}}{\Delta w_{ij'}}\right)}{\sigma_{e_i}}\right),$$

where $r_{ijj'}$ is an indicator for choosing the option associated with increasing aspect j and $\Phi[\cdot]$ is the standard normal cumulative distribution function.

We estimate the model using Hamiltonian Markov Chain Monte Carlo (HMC). Since this estimation method imposes a Bayesian framework, we follow standard practice (Gelman, 2006) and specify weakly informative priors for our hyperparameters

$$\begin{aligned}\mu_j &\sim N(0, 10) \\ \sigma_j &\sim \text{cauchy}(0, 2) \\ \mu_e &\sim N(0, 10) \\ \sigma_e &\sim \text{cauchy}(0, 2).\end{aligned}$$

We run eight chains of 500 iterations each (500 warm-up).

All reported log marginal utility estimates are posterior means, and uncertainty is quantified using 95% credible intervals. For assessing differences in marginal utilities across demographic subgroups, we estimate the model separately in each subgroup. For each aspect j , we compute $\Delta_j = \hat{\mu}_{j,1} - \hat{\mu}_{j,0}$ and $SE(\Delta_j) = \sqrt{sd_{j,1}^2 + sd_{j,0}^2}$, where $sd_{j,g}^2$ is the posterior SD of the subgroup mean and subgroup estimates are treated as independent. We classify an aspect as different across groups when $\Phi(-|\Delta_j/SE(\Delta_j)|) < .001$ (confidence level 0.999). In Supplementary Figure 2, labels additionally require significance relative to the total least squares (TLS) line, using the aspect's orthogonal distance from the TLS line and its propagated uncertainty at the same confidence level.

For aspects that only apply to a subset of the population (e.g., child-related aspects), we model them as distinct aspects for those to whom they do not apply (e.g., non-parents for child-related aspects). These individuals' parameters are estimated separately and excluded from the main results. This allows the primary estimates to reflect valuations by respondents for whom the aspect is relevant, while preserving model structure.

Aspect-Characteristics Survey Analysis. On each screen of the survey, respondents saw a two-part prompt describing an aspect characteristic (e.g. **Characteristic:** *it's complex*) and corresponding rating question (e.g., **Question:** *How complex is this aspect?*), followed by a list of aspects to rate on 9-point integer response scales (e.g., *1 Extremely simple, 9 Extremely complex*) (see **Extended Data Figure 1**).

For analysis, we excluded respondents for any of three reasons: submitting a duplicate survey, rating every aspect at the same level for a given characteristic, or failing an attention check. For the attention check, each survey had two aspect rating screens (randomly selected out of the 7 total) where one of the text slots for an aspect of well-being said, “To ensure you’re paying attention, please choose [a number randomly chosen from 1 to 9] for this aspect.” With this quality control exclusion restriction, we retained 86.1% to 93.8% of completes per tranche.

We construct an aspect-level measure for each characteristic by averaging ratings across respondents for each aspect-characteristic pair. To compare tradeoff-derived marginal utilities with aspect-characteristic ratings, we treat both the log marginal utility means and the mean characteristic ratings at the aspect level as data. Unless noted as a Spearman rank correlation, all correlations are Pearson correlations, with bootstrapped 95% confidence intervals. For each of the three importance-related characteristics: (1) “importance to you,” (2) “marginal value given current level,” and (3) “willingness to give things up to get more,” we report correlations with log marginal utility. To test whether the strength of correlation differs across the three characteristics, we stack aspect-level data across characteristics and regress standardized log marginal utility on standardized characteristic ratings interacted with indicators for each characteristic, clustering standard errors at the aspect level. Because both variables are standardized, the coefficient on each characteristic-specific term is equal to the correlation between log marginal utility and that characteristic. The stacked regression structure enables joint inference and comparison of correlation across the different characteristics.

To compute empirical log marginal utility-level slopes for each aspect, we treat the mean of posterior draws for individual log marginal utilities, m_{ij} , as data. We regress these posterior means on reported levels. For aspects that only apply to a subset of the population (e.g., child-related aspects), we restrict the data used for calculating these slopes to respondents for whom the aspect applies (e.g. respondents with children). To correct for Bayesian shrinkage, we multiply slopes by $\frac{\sigma_j^2}{\text{Var}(m_{ij})}$ where $\text{Var}(m_{ij})$ is the variance across mean posterior estimates of individuals’ log MU and σ_j^2 is the model-estimated variance of individuals’ log marginal utility for aspect j . All tests were two-sided.

References

- Adler, A.** (1933). *Der Sinn des Lebens [The meaning of life]*. Wien und Leipzig: Verlag Dr. Rolf Passer.
- Benjamin, D.J., Cooper, K., Heffetz, O. and Kimball, M.,** (2023). From happiness data to economic conclusions. *Annual Review of Economics*, 16.
- Benjamin, D. J., Heffetz, O., Kimball, M. S., & Rees-Jones, A.** (2012). What do you think would make you happier? What do you think you would choose?. *American Economic Review*, 102(5), 2083-2110.
- Benjamin, D. J., Heffetz, O., Kimball, M. S., & Rees-Jones, A.** (2014a). Can marginal rates of substitution be inferred from happiness data? Evidence from residency choices. *American Economic Review*, 104(11), 3498-3528.
- Benjamin, D. J., Heffetz, O., Kimball, M. S., & Szembrot, N.** (2014b). Beyond happiness and satisfaction: Toward well-being indices based on stated preference. *American economic review*, 104(9), 2698-2735.
- Bentham, J.** (1789). *An introduction to the principles of morals and legislation*. London: Printed for T. Payne.
- Bernheim, B. D.** (2016). The Good, the Bad, and the Ugly: A Unified Approach to Behavioral Welfare Economics. *Journal of Benefit-Cost Analysis*, 7(1), 12–68.
- Bernheim, B. D., Kim, K., & Taubinsky, D.** (2024). *Welfare and the Act of Choosing* (No. w32200). National Bureau of Economic Research.
- Daly, H. & Cobb, J.** (1989). *For the Common Good*. Beacon Press, Boston.
- Diener, Ed, & Martin E. P. Seligman.** 2004. “Beyond Money: Toward an Economy of Well-Being.” *Psychological Science in the Public Interest* 5 (1): 1–31.
- Dolan, P., & Metcalfe, R.** (2012). Measuring Subjective Wellbeing: Recommendations on Measures for use by National Governments. *Journal of Social Policy*, 41(2), 409–427. doi:10.1017/S0047279411000833.
- Duesenberry, J. S.** (1949). *Income, Saving, and the Theory of Consumer Behavior*. Cambridge, MA: Harvard University Press.
- Edwards, A. L.** (1957). *The Social Desirability Variable in Personality Assessment and Research*. New York, NY: Dryden Press.

- Gelman, A.** (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3), 515–533. doi:10.1214/06-BA117A.
- Kubiszewski, I., Costanza, R., Franco, C., Lawn, P., Talberth, J., Jackson, T., Aylmer, C.** (2013). Beyond GDP: Measuring and achieving global genuine progress. *Ecological Economics*, 93: 57–68.
- Layard, R.** (2005). *Happiness: Lessons from a New Science*. London: Allen Lane.
- Mill, J. S.** (1863). *Utilitarianism*. London: Parker, Son, and Bourn.
- Reynolds, W. M.** (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, 38(1), 119–125.

Appendix

<i>Extended Data: Tables and Figures.....</i>	2
(Extended Data) Figure 1 – Annotated Screenshot of Aspect-Characteristics Survey	2
(Extended Data) Figure 2 – Comparing Log MU to Direct Importance Elicitation	3
(Extended Data) Figure 3 – Supply- and Demand-Driven Differences in MU Estimates ..	4
(Extended Data) Figure 4 – Social Desirability vs. Difference in Log MU (Marlowe Crowne split)	6
(Extended Data) Table 1 – Demographics.....	7
(Extended Data) Table 2 – MUs in units of standard deviations of aspect levels.....	8
<i>Supplementary Information.....</i>	11
1. Supplementary Methods	11
1.1 Overview of Survey Flow	11
1.2 Survey Design.....	11
1.2.i. Baseline and Multi-Block Surveys	11
1.2.ii. Levels and Tradeoff Questions in <i>Triple</i> Design	13
1.2.iii. Aspects of Well-Being in the Multi-Block Survey	15
1.2.iv. Life and Psych Survey Design	20
1.2.v. Aspect-Characteristics Survey Design	23
1.3. Quality Control Procedure	27
1.3.i. Baseline Quality Control.....	27
1.3.ii. Prescreening Survey Details	30
1.3.iii. Results of Quality Control Procedure	31
1.4 Survey Fielding	34
1.4.i. Baseline Survey Fielding.....	34
1.4.ii. Multi-Block Survey Fielding.....	35
1.4.iii. Life and Psych Survey Fielding.....	35
1.4.iv. Aspect-Characteristics Survey Fielding.....	35
1.5 References for Supplementary Methods	36
2. Supplementary Tables	37
Supplementary Table 1 – Largest Discrepancies Between Revealed and Stated Importance	37
Supplementary Table 2 – Selected Scholarly Theories of What People Do Want or Should Want.....	38
3. Supplementary Discussion	40
3.1 Subset/Superset Pairs	40
3.2 MU Spreads for High- and Low-Numeracy Respondents	41

4.	Supplementary Equations	42
5.	Supplementary Note.....	43
	5.1 Importance Ratings vs. Trade-Based MU Estimates	43
6.	Supplementary Figures	44
	Supplementary Figure 1 – Level percentiles and log MU for each aspect.....	44
	Supplementary Figure 2 – Log MU Averages within Demographic Groups.....	48

Extended Data: Tables and Figures

(Extended Data) Figure 1 – Annotated Screenshot of Aspect-Characteristics

Survey

Flag Characteristic → **Characteristic:** it's complex

Question: How complex is this aspect? ← **Flag Question Text**

Your ability to climb a flight of stairs ← **Aspect**

1 2 3 4 5 6 7 8 9
 Extremely Simple Simple In Between Complex Extremely Complex

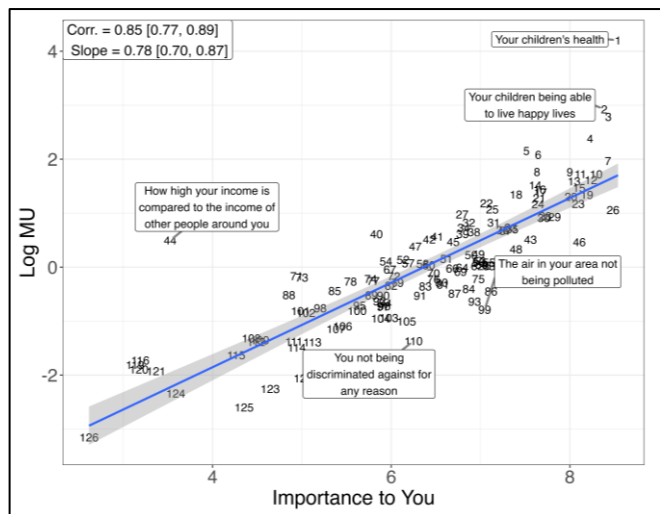
Response Options

Your nation doing the right thing in domestic policy

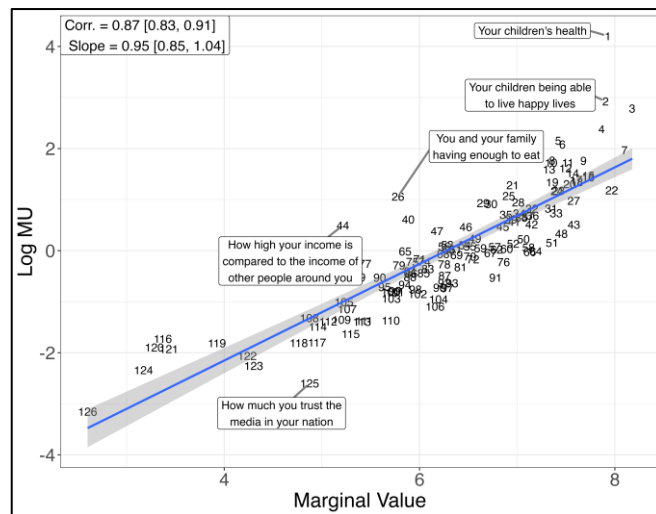
1 2 3 4 5 6 7 8 9
 Extremely Simple Simple In Between Complex Extremely Complex

(Extended Data) Figure 2 – Comparing Log MU to Direct Importance Elicitation

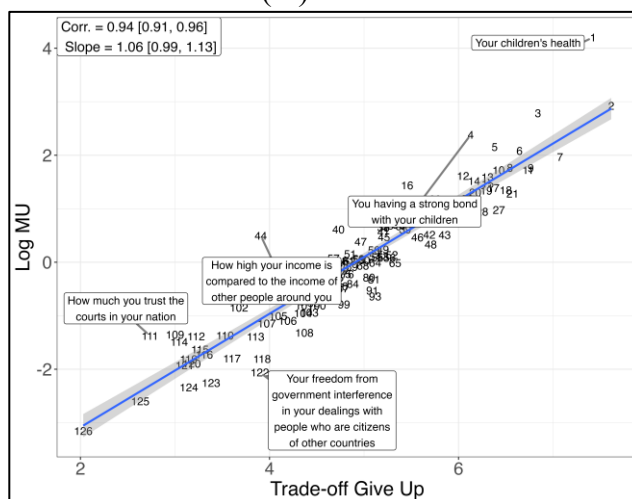
(i)



(ii)



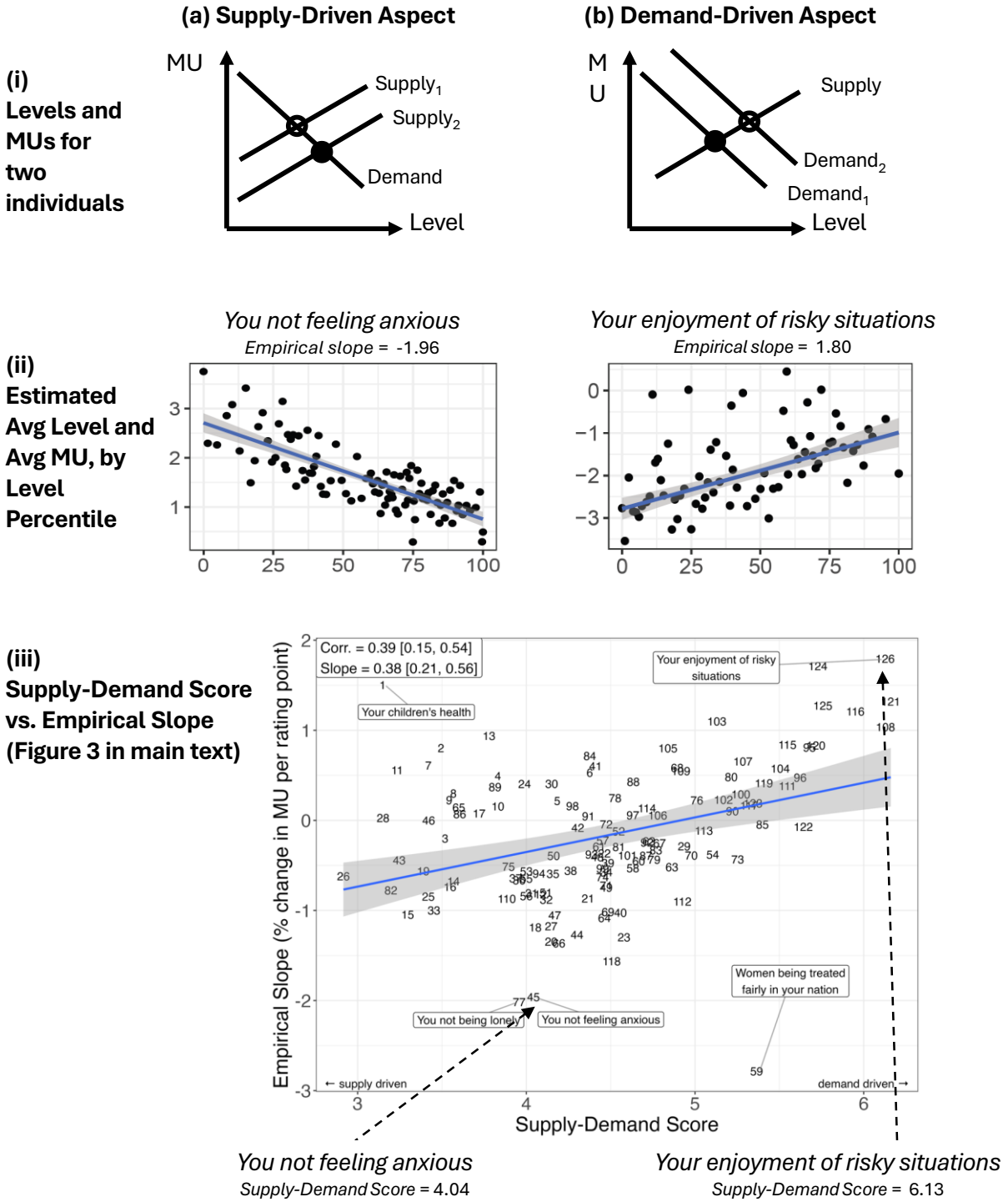
(iii)



Notes: Each panel compares unadjusted log marginal utility estimates (μ_j) with average ratings from an aspect-characteristic question. Panel (i): “How important is this aspect to you?” (281 respondents rating 126 aspects, $N = 13,056$). Panel (ii): “Think about your own life as it is. Given how much you have of this aspect, how eager would you be to get more? In particular, compared to other aspects of life, how valuable would it be to have a little bit more of this aspect?” (364 respondents rating 126 aspects, $N = 16,018$). Panel (iii): “Imagine you could gain more of this aspect by giving up other things. How much would you be willing to give up?” (281 respondents rating 126 aspects, $N = 13,056$). For response option labels, see **Table SI.1.2. Characteristic and Question Details by Aspect-Characteristics Survey Tranche**. Aspect-level characteristic ratings averaged across respondents as described in Methods. Labeled aspects are the five furthest from the OLS regression line. Points are marked by uncorrected log MU rank (column 3 in **Table 2**).

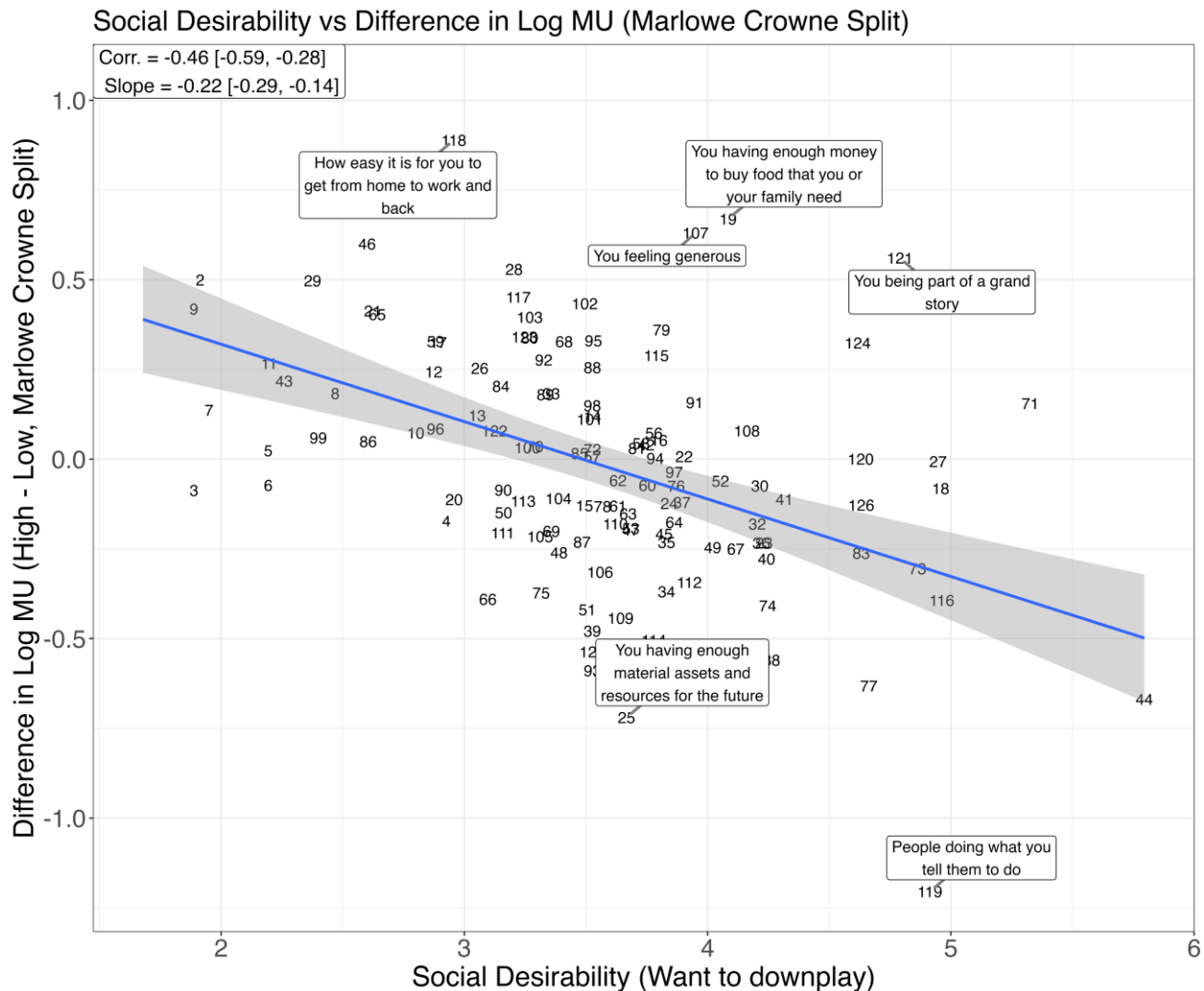
(Extended Data) Figure 3 – Supply- and Demand-Driven Differences in MU Estimates

To formalize the discussion of **Figure 3** in the main text, **Supplementary Figure 5** illustrates differences between supply-driven and demand-driven aspects. Row (i) shows a theoretical data-generating process for the level and marginal utility (MU) for two different individuals. In panel (a), the individuals have different supply curves for an aspect; the level and MU are negatively correlated. In panel (b), the individuals have different demand curves; the level and MU are positively correlated. In row (ii), we turn to our Multi-Block Survey data; we chose ***You not feeling anxious*** as the representative supply-driven aspect (panel a) and ***Your enjoyment of risky situations*** as the representative demand-driven aspect (panel b). For each aspect, we binned respondents by percentile of reported aspect level. The x-axis (y-axis) is the average level (average MU) by bin. The estimated empirical slopes for these resulting figures are -1.96 and 1.80 for the supply-driven and demand-driven aspects, respectively. Row (iii) reproduces **Figure 3**, highlighting the two aspects shown above. From the Aspect-Characteristics survey, the supply-driven and demand-driven aspects have average supply-demand scores of 4.04 and 6.13 , respectively.



Notes: Panel (ii): For each percentile of the self-reported aspect level (x-axis), the mean marginal utility across respondents; based on 896 respondents who completed the first nine blocks of the Multi-Block Survey. Regression line and slope are estimated on the individual (not binned) data. Panel (iii): See **Figure 3**.

(Extended Data) Figure 4 – Social Desirability vs. Difference in Log MU
(Marlowe Crowne split)



Notes: Each point compares the difference in unadjusted log marginal utility estimates between respondents in the high and low social desirability groups (as measured by the Marlowe Crowne scale) with average aspect-characteristic ratings from the question, “Would someone downplay how much they want this aspect because openly wanting it might make them look bad?” The difference in log marginal utility is computed as the high-group estimate minus the low-group estimate, using separate hierarchical-model estimates for the two groups (N = 354 high; N = 106 low). Aspect-level characteristic ratings are averaged across respondents as described in Methods. Numbers mark uncorrected log MU rank. Labeled aspects are those above the 95th percentile of absolute residuals from the OLS regression line.

(Extended Data) Table 1 – Demographics

Category	Population	Baseline Sample (<i>N</i> =5466)	Passed QC, Invited to Surveys (<i>N</i> =3358)	Main Analysis Sample: Completed ≥ 9 Blocks of Multi- Block Survey (<i>N</i> =896)	Completed an Aspect- Characteristics Survey (<i>N</i> =1088)	Completed Life & Psych Survey (<i>N</i> =811)	Census
Gender	Male	49.6	43.5	45.6	44.5	45.4	49.0
	Female	49.6	55.4	53.8	55.1	54.1	51.0
	Nonbinary	0.8	1.1	0.6	0.4	0.5	
Marital status	Married	65.7	50.4	52.0	47.8	53.1	51.1
	Never married	24.6	35.4	34.2	37.6	33.3	33.5
	Other	9.6	14.3	13.8	14.6	13.6	15.4
Highest education level completed	High school grad	7.9	10.4	10.8	10.7	10.7	28.3
	Some college	16.6	25.7	24.1	26.2	23.4	27.1
	Bachelor's degree	52.6	46.2	48.0	45.7	48.5	22.2
	Graduate degree	22.6	17.2	16.9	18.9	17.3	12.8
Age	18–29	18.7	14.9	12.2	7.8	12.1	20.6
	30–39	35.3	33.6	32.8	30.7	32.3	17.5
	40–49	21.6	23.4	25.3	23.0	25.5	16.0
	50–64	19.2	20.7	24.3	27.8	24.9	24.9
	65 and older	5.2	7.4	5.4	10.7	5.2	21.1
Household Income	Less than \$30,000	14.2	17.3	15.6	18.7	16.0	22.1
	\$30,000–\$49,999	21.7	21.1	23.7	20.7	23.6	15.7
	\$50,000–\$69,999	22.2	20.1	19.5	17.8	19.2	13.4
	\$70,000–\$89,999	18.4	15.4	15.0	15.7	14.8	10.8
	\$90,000–\$119,999	13.5	12.8	12.6	11.5	12.3	12.0
	\$120,000 and above	10.0	13.3	13.6	15.6	14.1	26.0
Region	Midwest	22.8	22.9	22.9	21.9	22.6	20.8
	Northeast	16.6	18.1	19.9	19.5	19.4	17.4
	South	38.0	38.5	37.6	36.3	38.2	38.1
	West	22.6	20.5	19.6	22.3	19.9	23.7
Race	White, and other	79.1	78.6	80.1	79.7	79.5	63.3
	Black	6.8	8.1	6.9	8.3	7.0	12.1
	Hispanic/Latino	9.8	7.3	6.8	5.5	7.0	18.7
	Asian	4.3	6.0	6.1	6.5	6.4	5.9
Household size	1	15.6	20.7	21.0	24.4	19.6	28.5
	2	27.9	30.3	28.5	34.0	29.1	35.0
	3	19.0	19.5	20.3	16.1	20.3	15.0
	4 and above	37.6	29.5	30.2	25.6	30.9	21.5
Employment status	Employed	86.6	79.7	80.0	77.2	79.6	57.6
	Unemployed	6.4	10.0	9.8	10.9	10.0	3.8
	Not in labor force	7.0	10.3	10.2	11.9	10.4	38.5

Notes: All numbers are percentages. “Passed QC” refers to respondents who passed quality control. Sources include authors’ survey and 2020 US Census.

(Extended Data) Table 2 – MUs in units of standard deviations of aspect levels

(1) Aspect	(2) M_j (LB - UB)	(3) Rank (M_j)	(4) M_j per SD (LB - UB)	(5) Rank (M_j per SD)
Your children's health	66.9 (42.4 - 115.5)	1	1731.5 (1098.0 - 2990.5)	1
Your children being able to live happy lives	18.6 (13.7 - 26.3)	2	461.7 (340.3 - 650.5)	2
Your spouse/partner's health	16.2 (11.5 - 23.8)	3	400.7 (284.1 - 587.8)	3
You having a strong bond with your children	10.8 (7.8 - 15.8)	4	291.6 (208.8 - 424.4)	4
The health of your parents and siblings	8.6 (6.8 - 11.0)	5	185.8 (146.0 - 236.1)	5
The people who have been most important to you in your life still being alive	8.0 (6.4 - 10.1)	6	176.7 (140.9 - 222.9)	6
The overall well-being of your family	7.2 (5.9 - 8.6)	7	119.1 (98.9 - 142.8)	9
Your family living a wonderful life	5.8 (5.0 - 6.8)	8	119.3 (102.4 - 139.9)	8
The happiness of you and your family	5.8 (5.0 - 6.8)	9	108.0 (93.2 - 125.3)	13
Your ability to protect your loved ones	5.6 (4.6 - 6.8)	10	107.9 (89.0 - 130.2)	14
The overall well-being of you and your family	5.6 (4.8 - 6.7)	11	100.9 (85.7 - 119.8)	16
Your health	5.0 (4.3 - 5.9)	12	99.7 (85.1 - 116.9)	17
The happiness of your family	4.9 (4.4 - 5.5)	13	88.4 (79.3 - 99.0)	19
You being able to support your family financially	4.6 (4.1 - 5.1)	14	124.1 (111.5 - 139.1)	7
Your financial security	4.3 (3.7 - 5.2)	15	117.9 (100.2 - 141.5)	10
You having enough money to pay for healthcare and/or medicines that you or your family need	4.2 (3.5 - 5.1)	16	109.8 (91.8 - 133.5)	12
Your ability to take care of your family	4.0 (3.5 - 4.6)	17	90.5 (78.4 - 104.4)	18
You not worrying about money	3.8 (3.3 - 4.5)	18	112.5 (97.7 - 130.7)	11
You having enough money to buy food that you or your family need	3.8 (3.2 - 4.6)	19	81.4 (68.2 - 98.5)	22
Your physical health	3.7 (3.3 - 4.1)	20	72.0 (64.9 - 79.9)	26
Your children being kind and caring about other people	3.6 (2.9 - 4.6)	21	86.3 (68.2 - 110.8)	20
You having enough financial resources for your retirement years	3.3 (2.8 - 3.8)	22	104.0 (89.8 - 121.4)	15
Your mental health	3.2 (3.0 - 3.5)	23	78.1 (71.7 - 85.0)	23
Your relationship with your spouse, partner, or closest friend being stronger than ever	3.2 (2.7 - 3.9)	24	82.8 (70.1 - 99.4)	21
You having enough material assets and resources for the future	2.9 (2.6 - 3.3)	25	77.1 (67.6 - 88.2)	24
You and your family having enough to eat	2.9 (2.3 - 3.6)	26	49.9 (40.1 - 62.1)	36
You having enough money to do everything you want to do	2.7 (2.3 - 3.1)	27	74.5 (64.8 - 86.1)	25
You not having to worry about where your family's next meal will come from	2.6 (2.2 - 3.1)	28	57.8 (48.5 - 69.4)	31
Your children growing up to follow the golden rule— treating others as they would like to be treated	2.5 (2.0 - 3.2)	29	59.1 (46.9 - 75.0)	30
The quality of your spouse/partner as a mate	2.5 (2.1 - 3.0)	30	70.6 (59.0 - 84.9)	27
You not having depression	2.3 (2.0 - 2.6)	31	64.8 (56.6 - 73.9)	28
You feeling that you have enough money for the things that are most important to you	2.3 (2.0 - 2.6)	32	61.7 (54.0 - 69.8)	29
Your freedom from physical pain	2.1 (1.8 - 2.4)	33	47.1 (41.1 - 54.0)	37
The absence of worry in your life	2.1 (1.9 - 2.3)	34	57.0 (50.8 - 64.3)	32
How happy you feel	2.0 (1.9 - 2.1)	35	45.6 (42.9 - 48.4)	38
How satisfied you are with your life	2.0 (1.9 - 2.1)	36	45.4 (43.2 - 47.8)	39
How much you enjoy your life	2.0 (1.9 - 2.1)	37	42.6 (40.4 - 45.0)	41
You not feeling downhearted and depressed	1.9 (1.7 - 2.1)	38	50.8 (45.7 - 56.4)	35
The absence of stress in your life	1.9 (1.6 - 2.1)	39	51.1 (45.3 - 57.9)	34
Your rating of your life on a ladder where the lowest rung is "worst possible life for you" and the highest rung is "best possible life for you"	1.8 (1.7 - 2.0)	40	41.2 (38.4 - 44.2)	43
The quality of your romantic relationships, marriage, love life or sex life	1.8 (1.5 - 2.1)	41	54.5 (45.2 - 65.7)	33
You enjoying every day	1.7 (1.5 - 1.8)	42	37.6 (34.3 - 41.5)	46
You getting the health care you need	1.7 (1.4 - 1.9)	43	37.8 (32.3 - 44.1)	44
How high your income is compared to the income of other people around you	1.6 (1.4 - 2.0)	44	42.1 (35.2 - 50.4)	42
You not feeling anxious	1.6 (1.5 - 1.7)	45	43.5 (39.8 - 47.5)	40
Your physical safety and security	1.6 (1.4 - 1.9)	46	25.2 (21.8 - 29.4)	56
The absence of sadness in your life	1.5 (1.4 - 1.6)	47	37.8 (35.2 - 40.5)	45
You feeling calm and peaceful	1.4 (1.3 - 1.6)	48	31.3 (28.2 - 34.8)	48
Your sense of control over your life	1.3 (1.2 - 1.3)	49	28.3 (26.6 - 30.0)	50

You feeling well-rested	1.2 (1.0 - 1.3)	51	27.3 (23.9 - 31.2)	52
You liking what you do every day	1.1 (1.0 - 1.3)	52	25.3 (22.9 - 27.9)	55
You having people you can turn to in time of need	1.1 (1.0 - 1.3)	53	27.0 (23.2 - 31.1)	53
You reaching most of your goals over the past 12 months	1.1 (1.0 - 1.3)	54	27.5 (24.2 - 31.5)	51
You having many options and possibilities in your life and the freedom to choose among them	1.1 (1.0 - 1.2)	55	25.4 (22.3 - 28.7)	54
You not having to worry about being unemployed	1.1 (0.9 - 1.3)	56	32.4 (27.5 - 37.7)	47
You feeling active and productive every day	1.1 (1.0 - 1.2)	57	24.4 (22.0 - 27.1)	60
You having a lot of energy	1.1 (0.9 - 1.2)	58	25.2 (22.2 - 28.4)	57
Women being treated fairly in your nation	1.1 (0.9 - 1.3)	59	23.1 (18.8 - 28.6)	63
Your life being more than just getting things done	1.0 (0.9 - 1.2)	60	24.7 (22.0 - 27.9)	59
The extent to which you feel the things you do in your life are worthwhile	1.0 (1.0 - 1.1)	61	22.6 (21.4 - 23.8)	64
Your knowledge and skills	1.0 (0.9 - 1.2)	62	16.4 (14.4 - 18.9)	78
Your sense of purpose	1.0 (1.0 - 1.1)	63	24.3 (22.7 - 26.0)	61
You not feeling tired all the time	1.0 (0.9 - 1.1)	64	25.0 (22.0 - 28.4)	58
Your living environment not being spoiled by crime and violence	1.0 (0.8 - 1.2)	65	20.3 (17.2 - 24.0)	69
You getting enough sleep	1.0 (0.8 - 1.1)	66	23.6 (20.2 - 27.6)	62
How much of the time the things you do seem meaningful	1.0 (0.9 - 1.1)	67	21.9 (19.6 - 24.4)	66
You being a good person	0.9 (0.9 - 1.0)	68	14.8 (13.4 - 16.2)	81
You feeling that you have enough time for the things that are most important to you	0.9 (0.9 - 1.0)	69	18.4 (17.3 - 19.7)	73
You being able to turn your good intentions into actions	0.9 (0.8 - 1.0)	70	17.9 (15.7 - 20.2)	75
How satisfied you are with your standard of living compared to the people you spend time with	0.8 (0.8 - 1.0)	71	18.9 (16.7 - 21.3)	70
How often you smile or laugh	0.8 (0.7 - 1.0)	72	18.4 (16.0 - 21.1)	72
You being a winner in life	0.8 (0.7 - 1.0)	73	21.6 (18.9 - 24.9)	67
Your spouse/partner taking on their fair share of responsibility for the household	0.8 (0.7 - 1.0)	74	22.2 (18.3 - 26.6)	65
Your home being comfortable	0.8 (0.7 - 0.9)	75	15.5 (13.6 - 17.7)	80
The wisdom you have been able to accumulate	0.8 (0.7 - 0.9)	76	13.6 (12.0 - 15.6)	86
You not being lonely	0.8 (0.7 - 0.9)	77	20.4 (18.5 - 22.4)	68
Your friends and family giving you positive energy every day	0.8 (0.7 - 0.9)	78	17.2 (15.2 - 19.6)	77
You taking care of others instead of just yourself	0.8 (0.6 - 0.9)	79	16.2 (13.9 - 18.6)	79
The absence of anger in your life	0.8 (0.7 - 0.9)	80	18.7 (16.3 - 21.5)	71
You having people in your life who see the good in you	0.7 (0.6 - 0.8)	81	13.9 (12.1 - 15.7)	83
Your ability to climb a flight of stairs	0.7 (0.6 - 0.9)	82	13.7 (11.3 - 16.5)	85
You not being bothered by having little interest or pleasure in doing things	0.7 (0.6 - 0.8)	83	17.9 (15.8 - 20.2)	74
You being a good friend and doing your duty by your friends	0.7 (0.6 - 0.8)	84	13.0 (11.2 - 14.9)	89
You learning or doing something interesting every day	0.6 (0.6 - 0.7)	85	13.7 (12.0 - 15.7)	84
You not having things stolen from you	0.6 (0.5 - 0.8)	86	11.3 (9.6 - 13.3)	94
You feeling like you have work-life balance	0.6 (0.5 - 0.7)	87	13.6 (12.0 - 15.4)	87
You always making time for regular trips or vacations with friends and family	0.6 (0.5 - 0.7)	88	17.5 (14.9 - 20.5)	76
How much fun you have when you get together with friends	0.6 (0.5 - 0.7)	89	11.2 (9.8 - 12.8)	95
You finding meaning in the world	0.6 (0.5 - 0.7)	90	13.4 (11.8 - 15.3)	88
You being satisfied with your job	0.6 (0.5 - 0.7)	91	14.7 (12.7 - 16.8)	82
You getting to use your strengths to do what you do best every day	0.5 (0.5 - 0.6)	92	11.5 (10.1 - 13.0)	92
You having a life outside of work	0.5 (0.5 - 0.6)	93	11.9 (10.4 - 13.7)	91
How much you like your home (house or apartment)	0.5 (0.4 - 0.6)	94	11.2 (9.7 - 12.9)	96
You showing people the right way to live by your words and actions	0.5 (0.4 - 0.6)	95	9.6 (8.3 - 11.1)	101
The ability of ordinary citizens to influence your national government	0.5 (0.4 - 0.6)	96	12.2 (9.7 - 14.9)	90
The overall quality of your experience at work	0.5 (0.4 - 0.6)	97	10.5 (9.0 - 12.1)	99
You being able to spend a lot of time with your friends	0.5 (0.4 - 0.5)	98	11.4 (9.7 - 13.3)	93
The air in your area not being polluted	0.5 (0.4 - 0.5)	99	9.8 (8.0 - 11.8)	100
You having the frequent opportunity to see things of beauty	0.4 (0.4 - 0.5)	100	8.2 (7.2 - 9.4)	103

You living in a city or area that is perfect for you	0.4 (0.4 - 0.5)	101	10.9 (9.3 - 12.6)	98
You having someone in your life who always encourages you to be healthy	0.4 (0.4 - 0.5)	102	10.9 (9.3 - 12.8)	97
The pleasure you get from helping people	0.4 (0.3 - 0.5)	103	7.2 (6.2 - 8.4)	106
How much you can trust most people in your nation	0.4 (0.3 - 0.5)	104	8.9 (7.4 - 10.5)	102
You doing your duty	0.4 (0.3 - 0.4)	105	7.0 (5.9 - 8.1)	107
At work, you getting to use your strengths to do what you do best every day	0.3 (0.3 - 0.4)	106	7.7 (6.6 - 8.9)	104
You feeling generous	0.3 (0.3 - 0.4)	107	6.1 (5.2 - 7.2)	111
You having an impact on the world	0.3 (0.2 - 0.3)	108	7.2 (5.9 - 8.7)	105
How much you trust the police in your nation	0.3 (0.2 - 0.3)	109	6.9 (5.4 - 8.6)	109
You not being discriminated against for any reason	0.3 (0.2 - 0.3)	110	5.6 (4.5 - 6.8)	112
How much you trust the courts in your nation	0.3 (0.2 - 0.3)	111	6.3 (5.1 - 7.7)	110
You having the courage to talk to people you don't already know	0.3 (0.2 - 0.3)	112	6.9 (5.7 - 8.3)	108
You being satisfied with your city	0.2 (0.2 - 0.3)	113	5.5 (4.7 - 6.5)	114
How easy it is for you to express who you are through your lifestyle, culture, and beliefs	0.2 (0.2 - 0.3)	114	4.2 (3.6 - 5.0)	116
You having a leader in your life who makes you enthusiastic about the future	0.2 (0.2 - 0.2)	115	5.6 (4.5 - 6.7)	113
You having others remember your accomplishments long after your death	0.2 (0.1 - 0.2)	116	5.1 (4.1 - 6.2)	115
Your understanding of the society you live in	0.2 (0.1 - 0.2)	117	3.0 (2.5 - 3.6)	121
How easy it is for you to get from home to work and back	0.2 (0.1 - 0.2)	118	3.3 (2.7 - 4.0)	120
People doing what you tell them to do	0.2 (0.1 - 0.2)	119	3.6 (2.8 - 4.6)	119
Your stature in your community	0.2 (0.1 - 0.2)	120	3.8 (3.1 - 4.6)	118
You being part of a grand story	0.1 (0.1 - 0.2)	121	4.1 (3.3 - 5.0)	117
Your freedom from government interference in your dealings with people who are citizens of other countries	0.1 (0.1 - 0.2)	122	2.7 (2.2 - 3.4)	122
Your cultures and traditions being honored	0.1 (0.1 - 0.1)	123	2.2 (1.7 - 2.9)	123
You having good taste and being up on the latest things	0.1 (0.1 - 0.1)	124	2.1 (1.7 - 2.6)	124
How much you trust the media in your nation	0.1 (0.1 - 0.1)	125	2.0 (1.5 - 2.7)	125
Your enjoyment of risky situations	0.0 (0.0 - 0.1)	126	1.2 (0.8 - 1.6)	126

Notes: Bayesian hierarchical model estimates using Hamiltonian Monte Carlo, based on 574,282 tradeoff responses from 896 respondents who completed the first nine blocks of the survey and passed quality control. Standard subjective well-being questions are highlighted in purple. Orange rows indicate aspects applicable only to individuals with children, and green rows indicate aspects applicable only to those with a romantic partner, respectively. Sample sizes are 423 having at least one child and a partner, 213 have neither. 79 have children but no partner, and 181 have partners but no children. Columns: (1) Aspect name; (2) Raw marginal utility per point, $M_j = e^{\mu_j}$ with 95% credible intervals in parentheses; (3) Rank based on M_j ; (3) Marginal utility per point multiplied by the standard deviation of raw aspect ratings for aspect j, indicating the marginal utility gain from a one standard deviation increase in aspect ratings, with 95% credible intervals; (4) Rank based on the SD-adjusted marginal utility. Because the average marginal utility across aspects equals zero, raw M_j in column (2) can be interpreted as the marginal utility of a one-point increase relative to the geometric mean across all aspects. The SD-adjusted estimates in column (4) account for differences in rating variability across aspects, showing the marginal utility gain from a standardized one standard deviation improvement in each aspect.

Supplementary Information

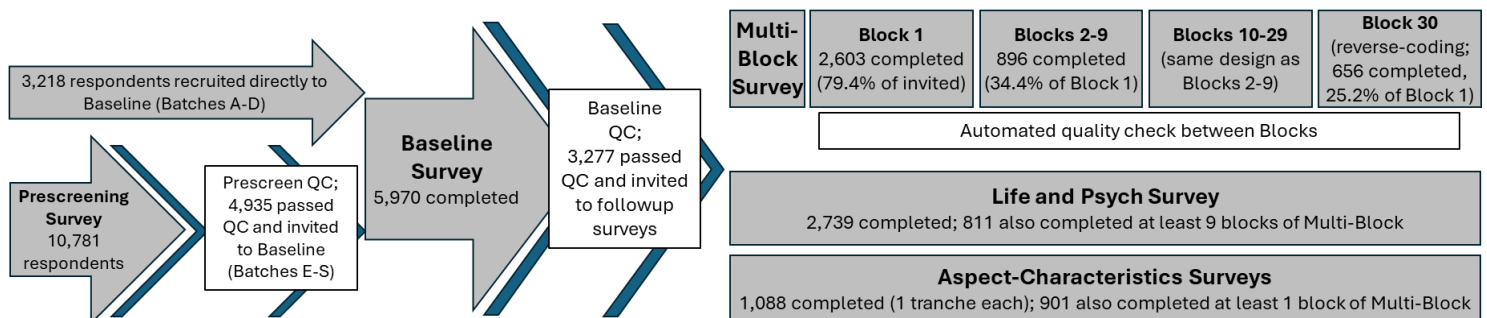
1. Supplementary Methods

This section details the design, pre-registered quality control procedure, and fielding of the four surveys whose data are used in the main text of the paper – Baseline, Multi-Block Survey, Life and Psych Survey, and Aspect-Characteristics Surveys – as well as our preliminary Prescreening Survey.

1.1 Overview of Survey Flow

To be invited to the Multi-Block, Life and Psych, or Aspect-Characteristics Surveys, respondents had to first complete our Baseline survey and pass a quality control (QC) procedure which is described in **Supplementary Information 1.3**. About half of the Baseline respondents also passed a preliminary QC procedure, based on a Prescreening survey. Figure SI.1.1 shows the progression of respondents through all surveys. As described in the main text of the paper, the sample used for marginal utility estimates is the 896 respondents who completed at least 9 blocks of the Multi-Block survey.

Figure SI 1.1. Overview of survey flow and recruitment



1.2 Survey Design

1.2.i. Baseline and Multi-Block Surveys

The two key types of questions in our Baseline and Multi-Block Surveys are self-reported well-being (SWB) questions about aspect levels, as shown in **Figure 1a**, and tradeoff questions, as shown in **Figure 1b**. Both surveys also include calibration questions, which are designed to have the same objective answer across respondents; the calibration questions are not analyzed in this paper so we omit further description of them.

For Baseline, the survey flow is: (i) consent form; (ii) basic demographic questions (age, gender, household income, ZIP code); (iii) instructions; (iv) 36 SWB questions and 72 tradeoff

questions (in a *triple* design, described further below) and (v) calibration questions ((iv) and (v) in randomized order); (vi) additional demographic, behavioral, and psychological questions; and (vii) exit questions about how the respondent approached the survey. From the Baseline survey, only the demographic questions are analyzed in this paper. Screenshots of the survey can be found in the Web Appendix.

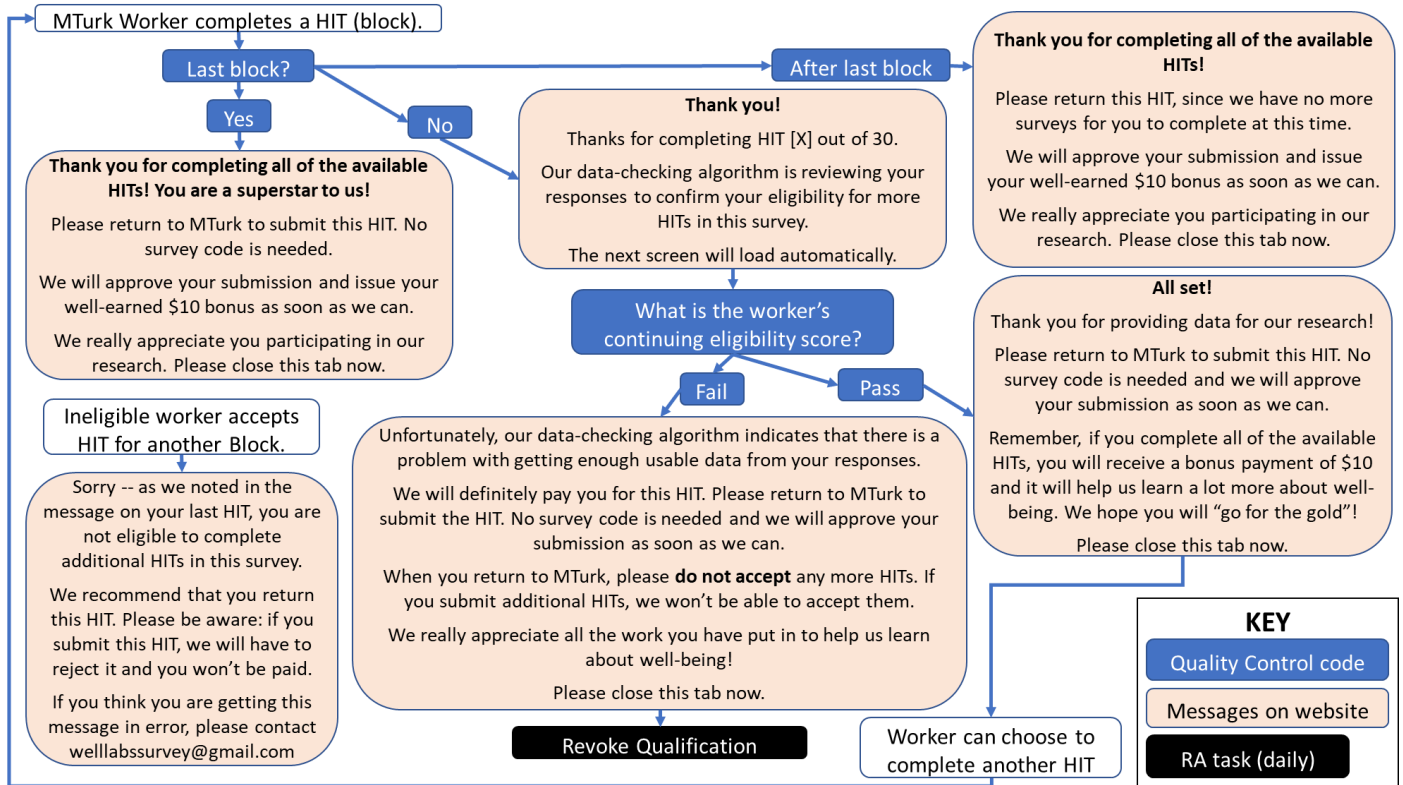
The Multi-Block Survey consists of 30 sequential “blocks,” each of which corresponds to a single Human Intelligence Task (HIT) on MTurk. (In previous work we described this as the “Bottomless” survey because of its long length if respondents complete all available blocks.) For Blocks 2-29, the survey flow is: (i) consent form; (ii) abbreviated instructions; (iii) one or more SWB questions on “alternative scales” (with response options other than the 0-100 scale from **Figure 1a**) or (at the start of Blocks 8 and 9) questions about remembered or expected SWB at ages 50 and 75¹; (iv) 36 SWB and 72 tradeoff questions (again, in a *triple* design and as depicted in **Figure 1a** and **Figure 1b**); and (v) calibration questions. Block 1 is the same as this, without the alternative scale frame SWB question (iii). Block 30 is distinct, with no tradeoff questions: it consists of the five Satisfaction With Life Scale (SWLS) questions from Diener et al. (1985), followed by 48 SWB questions and 36 calibration questions; half of the questions of each type are “reverse-coded” to study scale use tendencies for SWB and calibration question ratings. In this paper, we only analyze the SWB and tradeoff questions from Blocks 1-9. To maintain attention, at the end of some blocks we asked questions about survey interpretation and effort, such as, “When you made the decisions, were you focusing most on which aspects of life were changing, or on how much the aspects were changing?”

Respondents had to answer every SWB question and tradeoff question to advance through the surveys. However, there was a “loophole” which allowed respondents to advance much faster: if one of the aspects in a triple is rated below 8 and another one of the aspects is rated above 92, we skip the tradeoff questions for that triple. To ensure respondents were not doing this, we used another quality control check between each block of the Multi-Block survey: to maintain eligibility for additional Blocks, the respondent had to provide data for at least 54 out of the 72 tradeoffs in each block. The whole flow chart of respondent messaging between blocks

¹ The SWB questions come from the European Social Survey (ESS), General Social Survey (GSS), Health and Retirement Study (HRS), the global life satisfaction question in Kapteyn, Smith, and van Soest (2009) (KSV), World Values Survey (WVS), Gallup World Poll, and UK Office of National Statistics (ONS).

of the Multi-Block Survey is shown in Figure SI.1.2.

Figure SI.1.2. Quality Check and Messages Between Blocks of Multi-Block Survey



1.2.ii. Levels and Tradeoff Questions in *Triple Design*

A *triple* consists of three SWB questions followed by six pairwise tradeoffs corresponding to those SWB questions; if we call the aspects corresponding to the SWB questions A, B, C then the tradeoffs are A vs B, B vs C, A vs C, B vs A, C vs B, and C vs A, in randomized order. The starting point for each aspect's change in a tradeoff is the level that the respondent had reported earlier in the SWB question. In the Baseline survey and each block of the Multi-Block Survey, there are 11 unique triples and 12 triples in total; the 11th triple is a repeat of the 2nd triple as a check for data quality. The order of the SWB questions (i.e., their assignment to triples within a block) is randomized at the respondent level.

We now describe the tradeoff questions in detail, since they are the focus of this paper. Each tradeoff question asks, "Which option do you think you would choose?" Below the question are two panels. The top of the left panel has a large button labeled "Option 1;" the top of the right panel has a large button labeled "Option 2." Below each option button is an aspect of well-being followed by the word "increases" or "decreases" and a verbal representation of the

change (e.g., “+4 from 58 to 62” or “−5 from 68 to 63”). The starting point for the change is the aspect level which the respondent has previously reported in a rating question; we explain how the directions and magnitudes of the changes are determined further below. The representation of the change (“+4” or “−5”) appears in a contrasting circle for further emphasis. The final element of each panel is an image of the slider bar from the rating questions. The slider bar is colored up to the level of the starting point for the change (mimicking what the respondent saw at the end of the rating question for this aspect). A darker shade of the same color fills the area between the final and initial levels of the change, and a circle appears around this area. We also emphasize the shaded area of the change by putting a black arrow that points in the direction of the change within the slider bar on that area.

To progress through the survey, a respondent must click on the Option 1 button or the Option 2 button. Below both panels, respondents also have the option of checking a box next to the statement, “This decision does not make sense.” If the respondent checks the box, a question appears: “Why does this decision not make sense?” (see **Figure SI.1.3.** for screenshot). Based on pilot testing, where we had an open text box for replies to this question, the following options are given: The aspect on the left does not apply to my life; The aspect on the right does not apply to my life; Both aspects do not apply to my life; These two aspects of life are too similar; These two aspects of life go hand in hand (you can’t change one without the other changing as well); I am indifferent between these two options. Other: [text box].² The respondent can progress to the next screen either by choosing one of the two options or clicking the “submit” button in the “This decision does not make sense” dropdown.

The magnitudes of the changes in levels depicted in the tradeoffs range from 1 to 8. The changes are chosen as follows:

1. For a given triple, the changes in the tradeoff questions all go in the same direction³, and that direction is randomized with 50% probability.⁴

² Below the question, “Why does this decision not make sense?” we include a note: “Note: If you dislike both options, please choose the option you dislike the least.” In pilot testing, we also found that a significant fraction of respondents who checked the box wanted to express that they did not prefer either option more than the status quo.

³ At the respondent level of each pairwise tradeoff, there is a 2% probability that the aspects move in opposite directions. We use this “opposite direction trap” as a data quality check.

⁴ There are some exceptions. If a respondent has rated one of the aspects in the tradeoff below 8 or above 92, it would be possible for the realized value of the tradeoff to put the rating below 0 or above 100, respectively. To avoid this outcome, if one of the three aspects in the triple was rated above 92, we “force” the direction of the changes in the tradeoffs to be negative (i.e., we do not randomize the direction of changes for the triple); if one of the three aspects was rated below 8, we “force” the direction of tradeoffs to be positive. If one of the aspects was

2. For each aspect in a tradeoff, the magnitude of the change is randomly chosen from integers 1 to 8, with equal probability.⁵

Figure SI.1.3. Screenshot of “This decision does not make sense” options

☒ This decision does not make sense

Why does this decision not make sense?

Note: If you dislike both options, please choose the option you dislike the least.

- ☐ The aspect on the left does not apply to my life.
- ☐ The aspect on the right does not apply to my life.
- ☐ Both aspects do not apply to my life.
- ☐ These two aspects of life are too similar.
- ☐ These two aspects of life go hand in hand (you can't change one without the other changing as well).
- ☐ I am indifferent between these two options.
- ☐ Other

1.2.iii. Aspects of Well-Being in the Multi-Block Survey

The Multi-Block Survey includes SWB questions for 33 aspects of well-being in each block; this number was determined primarily by survey length constraints. **Table SI.1.1.** lists the 33 SWB questions in Block 1 of the Multi-Block Survey, in alphabetical order. (This list is also identical to the SWB questions in the Baseline Survey.) In designing this list, our primary goal was to try to cover dimensions of personal well-being as comprehensively as possible, informed by well-being measurement initiatives from the UK, New Zealand, Gallup, the OECD, and our own previous findings. To increase comprehensiveness, we included four aspects to capture things for which people would trade off SWB measures, as found by Benjamin et al. (2012). Our

rated below 8 *and* another one of the aspects was rated above 92, we skip the tradeoff questions for that triple. We do not tell respondents about this feature of the survey. In our Multi-Block survey data for 2534 respondents covering 336,486 triples, only 0.7% of triples were skipped.

⁵ For the practice questions (in the Baseline survey instructions), the magnitudes of the tradeoffs are always 4 and 5; we randomize which aspect has the bigger change. The first example tradeoff shows decreases of -4 and -5 . The second and third example tradeoffs show increases of $+4$ and $+5$. Exceptions can arise if a respondent rates an aspect below 8 or above 92 in the practice questions; if both aspects in a practice tradeoff question were rated below 8 (above 92), the relevant tradeoff question is always shown as an increase (decrease). If a respondent has rated one aspect below 8 and the other above 92, the practice tradeoff screen says, “There is no practice decision about the two aspects you just rated since your ratings were extraordinarily high and extraordinarily low.”

other major consideration was choosing aspects that had high estimated relative marginal utilities in Benjamin et al. (2014) and/or our pilot data.⁶

The 16 SWB questions with stars are also repeated in every block of the Multi-Block Survey; we needed an “interlocking” design such as this one in order to compare the marginal utilities (MUs) of aspects across blocks. (MUs are only defined up to a multiplicative constant, so one needs a normalization to define MU=1; our approach is to normalize so that the average MU for all aspects =1, which works as long as there are no “islands” of aspects which are not compared to others.) We chose these 16 to repeat because we are studying these aspects in more depth in other work; we can estimate their MUs somewhat more precisely with additional tradeoff data for them.

Table SI.1.1 SWB Questions in Block 1 of Multi-Block Survey

<i>Thinking about the past year, how would you rate...</i>
How happy you feel *
How much you can trust most people in your nation
How much you enjoy your life *
How satisfied you are with your life *

⁶ Here we provide more detail on the design of the Block 1/Baseline aspect list. First, we describe the selection of the 16 aspects repeated in each block. We began with 10 concepts from the SWB literature: happiness, anxiety, worthwhileness, and life satisfaction from the UK ONS; two "hedonic" aspects (sadness, enjoyment) and an alternative global evaluative aspect (Ladder) inspired by Gallup; and three aspects based on the New Zealand General Social Survey: physical health, mental health, and loneliness. We added three Benjamin et al. (2012) - inspired aspects to capture the things for which people would trade off SWB measures: sense of purpose, control over one's life, and family happiness (to capture these concepts, we chose an aspect from each category based on high MU's in Benjamin et al. (2014) and pilot data). We included ***You being able to support your family financially*** because it was highest in pilot data and we hypothesized a priori that it will move opposite SWB for some choices. We included ***You feeling that you have enough time for the things that are most important to you*** because we hypothesized that time scarcity will move opposite to financially-oriented measures. We also wanted an aspect to reflect the OECD Better Life Initiative (BLI) domain of "Work-Life Balance." For the last slot, we chose ***You being a good person*** because a priori it seemed to have little overlap with the others.

We now describe design considerations for the 17 other aspects in Block 1/Baseline. We included 13 aspects to cover dimensions from other wellbeing research: three from Gallup (worry, stress, anger), seven for the remaining OECD BLI domains, and three for remaining domains from the New Zealand General Social Survey. Among these are two “local public goods” aspects: ***The air in your area not being polluted*** and ***Your living environment not being spoiled by crime and violence***. We included the personal and policy aspects with the highest estimated MUs from Benjamin et al. (2014), ***The overall wellbeing of you and your family*** and ***You having many options and possibilities in your life and the freedom to choose among them***. We included a potential “necessity” aspect for exploratory analysis: ***You and your family having enough to eat***. Finally, we sought an aspect to capture the concept of relative status which was found to be important in Benjamin et al. (2012). Benjamin et al. (2014) included two status-related aspects, ***Your social status*** and ***Your power over other people***, but their estimated MUs were negative so we did not consider them. We looked for alternatives in a list of 116 status-related aspects in our pilot data. Based on our criteria that the aspect had to be (1) general (i.e., not specifically about financial status), and (2) comparative, we chose the aspect with the 9th highest MU from that list: ***You being a winner in life***.

The ability of ordinary citizens to influence your national government

The absence of anger in your life

The absence of sadness in your life *

The absence of stress in your life

The absence of worry in your life

The air in your area not being polluted

The extent to which you feel the things you do in your life are worthwhile *

The happiness of your family *

The overall well-being of you and your family *

You and your family having enough to eat

You being a good person

You being a winner in life

You being able to support your family financially *

You feeling that you have enough time for the things that are most important to you *

You having many options and possibilities in your life and the freedom to choose among them

You having people you can turn to in time of need

You not being lonely *

You not feeling anxious *

You not having to worry about being unemployed

Your cultures and traditions being honored

Your home being comfortable

Your knowledge and skills

Your living environment not being spoiled by crime and violence

Your mental health *

Your physical health *

Your physical safety and security

Your rating of your life on a ladder where the lowest rung is “worst possible life for you” and the highest rung is “best possible life for you” *

Your sense of control over your life *

Your sense of purpose *

In Blocks 2-29 of the Multi-Block Survey, 17 additional aspects appear in each block, in the “triple” format. The aspects were selected for various reasons.⁷ The Web Appendix includes a list of all aspects by Block. For this paper, we only include aspects from Blocks 1-9: 169 aspects in total. Our normalization for $MU = 1$ is based on that list. (As noted above, these aspects’ marginal utilities cannot be separately estimated since they are overlapping with each other in the data.) However, we do not report the MUs for some aspects, because they were not studied to determine “what do people want.” The excluded aspects are: 2 aspects which had been included as placebos (because they have low MUs); 4 aspects chosen for local public goods to be validated with objective data; 22 aspects to study variant wording of aspects; 13 aspects which correspond to calibration questions; 2 aspects hypothesized to have extreme values on characteristics ratings. The 43 aspects in Blocks 1-9 which are included in the estimation but not reported in the main paper text can be found in **Table SI.1.2.** in alphabetical order.

⁷ In Block 2, we algorithmically developed an aspect list which with maximal “comprehensiveness.” Based on pilot data for 2187 aspects of well-being, we chose the 17 additional aspects for Block 2 with a stepwise procedure; we maximized an R^2 criterion with marginal utility weights to see which of the other aspects added the most explanatory power to the 33 aspects already in the Block 1 list. We also did the procedure with marginal utility-squared weights and without marginal utility weights; the aspects identified with the alternate procedures which did not already appear in the Multi-Block Survey are included in later Blocks (6-8).

In the other Blocks, we included other aspects of well-being from the pilot survey’s long list of aspects, for various reasons. The following list of rationales corresponds to aspects in Blocks 3 through 12. For the first 12 blocks, every respondent sees the same 33 aspects in that block. (1) Two aspects are included to study dimensions of well-being with low importance; they were identified by low marginal utilities in pilot data. (2) Fifteen aspects are included because we hypothesize that they are closely related to choices; we plan to study these in another project studying the multidimensionality of well-being. (3) Four aspects are included because they reflect local public goods; their ratings can be compared with external data. (4) Three aspects are included because they correspond to questions we ask about expected future aspect levels; we plan to study these for studying theories about the shape of well-being over the life course. (5) We include 35 aspects because they correspond, as nearly as possible, to questions asked by major SWB surveys (Gallup and statistical agencies). (6) We include 22 aspects as variant wording of Block 1 aspects, to test the sensitivity of small wording changes. (7) We include 13 aspects because they correspond to the aspects used for personal calibration questions. This is another source of data for exploring general vs. specific scale use. (8) We include 53 aspects which reflect extremes in different categories – such as social desirability, ease of purchasing with money, or “objectivity” – based on another preliminary pilot survey. Across the remaining 17 regular Blocks, the order of the 289 aspects is randomized. These aspects were also chosen from the long pilot list for various reasons. We include 7 aspects related to political efficacy and life satisfaction which correspond as closely as possible to the vignette literature. We include 66 aspects which are considered local public goods, in order to further probe potential correlations with external “objective” measures. We include 47 aspects which are considered national or global public goods; from among all such aspects in the long list, we looked for the best possible ones to use as “quasi-calibration questions”: hypothesized to have minimal interpersonal differences in definition and perception (for example, trying to avoid aspects connected to political ideology). We include 82 aspects with simple and compound forms – such as *Your feeling of independence*, *Your feeling of self-sufficiency*, and *Your feeling of independence and self-sufficiency* – to test hypotheses about wording and conceptual overlap between aspects. We also include 14 aspects which we wrote since the pilot to test conceptual overlap related to trust. Finally, we include 130 new aspects – developed since the pilot – which are hypothesized to have high marginal utilities either directly or indirectly (through having a high marginal product), and 10 new aspects which are hypothesized to be highly predictive of other aspect levels.

Table SI.1.2 Aspects of well-being included in estimation, not marginal utility results

Freedom of the press in your nation
How confident you are about your personal appearance
How easy your life is
How little violence there is in the world
How often the weather is nice outside where you live
How often you feel good mentally
How often you wake up feeling good physically
The condition of the natural environment
The craftsmanship you experience in the things that you own
The health of the ecosystems in the world
The leaders of your state government not being corrupt
The quality and quantity of green spaces in your area
The water in your area not being polluted
You being able to rise to the challenges you face
You being able to sleep well at night
You being able to spend enough time with the ones you love
You being content with your life
You feeling in control of your own destiny
You feeling reasonably happy, all things considered
You feeling satisfied about your life as a whole
You feeling stylish
You feeling that your work has value
You feeling the things you do in your life are worthwhile
You having enough to eat
You having someone you can always turn to
You knowing you are a good person
You not being trapped in physical pain
You not being worried about money
You not feeling lonely
You not feeling too much anger
You not feeling too much sadness
You not having to worry about being personally unemployed

Your understanding your purpose in life
Your ability to breathe in and out easily
Your ability to hear
Your ability to remember things
Your ability to see
Your ability to walk several blocks
Your access to information
Your family having enough to eat
Your knowledge, skills, and access to information
Your overall well-being
Your sense of connection with other people

1.2.iv. Life and Psych Survey Design

After the consent form, respondents to the “Life and Psych” Survey answer a series of questions about life circumstances and opinions, as well as standard psychological measures. For brevity we just show here the relevant two sets of questions. Full screenshots can be found in the Web Appendix. **Figure SI.1.4** shows a screenshot of the math questions we use to calculate numeracy, from screen 5 of the survey.

Figure SI.1.4: Screenshot of Math Questions for Numeracy Measure

Survey Completion
0% 100%

←

Instructions: If you don't know the answer to one of these questions, please make your best guess.

What is $2 - 8$?

What is $\sqrt{169}$?

What is $\frac{(5 \times 5) - (7 \times 2)}{2}$?

2 is the first prime number. What is the fifth prime number?

What is the next number in this series? [100, 93, 86, 79]

What is the next number in this series? [1, 2, 4, 8, 16, 32, 64]

What is the next number in this series? [1, 3, 6, 10, 15]

← →

Powered by Qualtrics

Figure SI.1.5. shows a screenshot of the social desirability scale we use, from Reynolds (1982). It's screen 34 of the survey, following two other psychological batteries.

Figure SI.1.5: Screenshot of Marlowe-Crowne Social Desirability Scale – Short Form

Survey Completion

0% 100%

←

Listed below are a number of statements concerning personal attitudes and traits. Read each item and decide whether the statement is true or false as it pertains to you personally.

	True	False
1. It is sometimes hard for me to go on with my work if I am not encouraged.	☐	☐
2. I sometimes feel resentful when I don't get my way.	☐	☐
3. On a few occasions, I have given up doing something because I thought too little of my ability.	☐	☐
4. There have been times when I felt like rebelling against people in authority even though I knew they were right.	☐	☐
5. No matter who I'm talking to, I'm always a good listener.	☐	☐
6. There have been occasions when I took advantage of someone.	☐	☐
7. I'm always willing to admit it when I make a mistake.	☐	☐
8. I sometimes try to get even rather than forgive and forget.	☐	☐
9. I am always courteous, even to people who are disagreeable.	☐	☐
10. I have never been irked when people expressed ideas very different from my own.	☐	☐
11. There have been times when I was quite jealous of the good fortune of others.	☐	☐
12. I am sometimes irritated by people who ask favors of me.	☐	☐
13. I have never deliberately said something that hurt someone's feelings.	☐	☐

←
→

Powered by Qualtrics [↗](#)

1.2.v. Aspect-Characteristics Survey Design

The Aspect-Characteristics Survey is a series of many survey tranches, each one completed by a different group of respondents, all having the same structure. The goal is to elicit descriptions or characterizations of aspects of well-being in a scientifically systematic way, by averaging the elicitations across many respondents. The structure of the data from the survey can be thought of as a matrix, where the rows are aspects of well-being and the characteristics are columns; our survey design populates that matrix with enough observations in each cell to calculate the cell mean, which is the aspect's estimated level for that characteristic.

First, we describe the structure of a survey tranche. (We describe the aspects and characteristics studied in the sixteen tranches further below.) After the consent form and one screen of instructions, the survey consists of 14 screens of aspect-characteristic rating questions. (We don't ask demographic questions because all the respondents recruited to the survey already took the Baseline survey.) As described in the main paper text, in a given tranche, each respondent rates 43 aspects for 7 characteristics. We split the 43 aspects across two screens to make the survey less repetitive. The characteristic and corresponding question to be answered are at the top of the screen, and remain there as the respondent scrolls down to answer the questions. Each screen has 22-24 questions in total. In addition to 21 or 22 questions of interest (totaling 43 across the two screens), there are 2 questions for "calibrator aspects." We determined calibrator aspects for each characteristic through a combination of a priori judgment and pilot testing to have high and low values, respectively, of the characteristic in question. We include the calibrators to encourage respondents to use the entire range of the response scale. Finally, each screen may include an attention check question, as described in **Methods**. The screens on which the attention checks are located are randomized at the tranche level. In sum, there are three randomizations at the respondent level: (i) order of aspects within each characteristic; (ii) placement of calibrator aspects within each characteristic; and (iii) the order of the 14 screens.

Across all survey tranches, we study 172 aspects: the 126 studied in the main text and the rest for other purposes. Those other purposes are: aspects corresponding to calibration questions, aspects corresponding to SWB questions and CQs in another survey we fielded on the Understanding America Study, and aspects to study alternative wording and placebo tests.

Table SI.1.2 lists all 28 characteristics in the 16 tranches of data collected, including the characteristic, the question shown, the labels on the response options, the high and low calibrators, and a short question name for reference. The five characteristics analyzed in this paper (social desirability, supply versus demand, and three versions of questions about importance) are highlighted in yellow. Recall, there are four tranches per characteristic because we only had 43 aspects per tranche and we needed to cover all 172 aspects.

Table SI.1.2. Characteristic and Question Details by Aspect-Characteristics Survey Tranche

Tranches	Characteristic	Question	Response Option Labels for 1, 3, 5, 7, 9	Low Calibrator	High Calibrator	Question Name (Short)
2.1, 2.2, 2.3, 2.4	how much it covers	Question: Of everything people want for their lives, how much does this aspect cover?	covers almost nothing; covers little; covers some; covers much; covers everything	the pleasure you get from gourmet hot beverages	how rewarding your life is	coverage
2.1, 2.2, 2.3, 2.4	it's broad and general	Question: Is this aspect narrow and specific or broad and general? (For example, on the scale below, "you being able to use your legs" might be a 1 and "the overall well-being of you and your family" might be a 9.)	extremely narrow and specific; narrow; in between; broad; extremely broad and general	the pleasure you get from gourmet hot beverages	the well-being of the people in your nation	generality
2.1, 2.2, 2.3, 2.4	it's objective	Question: How objective is this aspect? (Could it be measured and proved in court?)	not at all objective; not very objective; somewhat objective; mostly objective; completely objective	you knowing a lot of interesting people	your ability to bathe yourself	objective
2.1, 2.2, 2.3, 2.4	higher or lower standards for themselves than for others	What this question is about: Sometimes people have higher standards for themselves or what happens to them and lower standards for others. And sometimes people have higher standards for others or what happens to others and lower standards for themselves. Question: In particular, suppose people were rating this aspect in their own lives as well as for others. To what extent do you think they would use higher or lower standards for themselves or things that happened to them than for others (even if they had the same situation for this aspect)?	much lower standards for themselves; somewhat lower standards for themselves; same standards for themselves and others; somewhat higher standards for themselves; much higher standards for themselves	you feeling that you are not alone in this world	you being a loyal person	response_consistency_2
2.1, 2.2, 2.3, 2.4	people overstate it in order to look good or feel good	Question: Would someone overstate their level of this aspect in order to look good or feel good about themselves ?	not overstate at all; overstate a little; overstate somewhat; overstate quite a bit; overstate a lot	how bicycle-friendly your area is	you being able to impress other people	social_des_level_2
2.1, 2.2, 2.3, 2.4	it's open to interpretation	Question: To what extent is this aspect open to interpretation ?	not at all open to interpretation; a little open to interpretation; somewhat open to interpretation; quite open to interpretation; completely open to interpretation	your ability to bathe yourself	you knowing a lot of interesting people	subjective
2.1, 2.2, 2.3, 2.4	people define it differently	Question: How different would the meaning of this aspect be to different people? (For example, if they came from a different background, culture, gender, or life experience?)	same meaning regardless of background; slightly different meanings; somewhat different meanings; quite different meanings; extremely different meanings for different backgrounds	your ability to climb several flights of stairs	your life mattering to the people with power	subjective_demographics

3.1, 3.2, 3.3, 3.4	having at least a little is critical	How critical is it to you to have at least a little of this aspect in your life rather than none?	not at all critical: it wouldn't be so bad to have zero of this; somewhat critical: having zero is less bad than for most aspects; moderately critical, like most aspects of life; very critical: having zero is worse than for most aspects; extremely critical: losing the last bit would be devastating	your life making a good novel or movie	you having hope in your life	Inada
3.1, 3.2, 3.3, 3.4	it's desirable for its own sake	Is this aspect desirable to you mainly because it helps you get something else you want, mainly for its own sake, or some of both? For example, <i>You having a dishwasher</i> might be a 1 and <i>You having fun</i> might be a 9.	want it only because it gets you something else you want; want it mostly because it gets you something else you want; want it for both reasons about equally; want it mostly for its own sake; want it only for its own sake	people like you mattering to the people with power	the relaxation you get from hot showers and baths	intermediate_vs_final
3.1, 3.2, 3.3, 3.4	having more would be especially valuable	Think about your own life as it is. Given how much you have of this aspect, how eager would you be to get more? In particular, compared to other aspects of life, how valuable would it be to have a little bit more of this aspect?	not at all valuable to have more; a little valuable to have more; somewhat valuable to have more; quite valuable to have more; extremely valuable to have more	the pleasure you get from drinking fruit juice	the quality of your sleep	marginal_value
3.1, 3.2, 3.3, 3.4	a bad mood would make you rate it less positively	To what extent would you rate this aspect of your life in a less positive way if you were in a bad mood when you started the survey?	mood doesn't affect rating at all; mood affects rating a little; mood affects rating somewhat; mood affects rating moderately; mood affects rating a lot	the well-being of the vegetation in the world	you not feeling frustrated	mood_effect
3.1, 3.2, 3.3, 3.4	it would make many other aspects better	How many other aspects of life would be better because of this aspect improving?	very few other aspects; a small number of other aspects; a medium number of other aspects; many other aspects; a huge number of other aspects	the amount of wonderful new kinds of toys being created each year	you being confident	other_aspects_improved
3.1, 3.2, 3.3, 3.4	people downplay wanting it because that might make them look bad	Would someone downplay how much they want this aspect because openly wanting it might make them look bad?	not downplay wanting at all; downplay wanting a little; downplay wanting somewhat; downplay wanting quite a bit; downplay wanting a lot	you being kindhearted and good to others	you being the center of attention	social_desire_to_want
3.1, 3.2, 3.3, 3.4	why some people don't care much about getting more	For each aspect, think about people you know who don't care much about getting more of it . They might not care much about getting more for either of two reasons: 1. They already have plenty of it. 2. They don't value it much. For each aspect below, of people who don't care much about getting more, what fraction have each reason?	almost all of them already have plenty; most of them already have plenty; both reasons equally common; most of them don't value it much; almost all of them don't value it much	you having access to water	how quickly space exploration is advancing	supply_vs_demand

4.1, 4.2, 4.3, 4.4	it would take a long time to describe what's going on with this	How long would it take you to describe to a friend in detail what is going on in this aspect of your life?	less than a minute; a few minutes; about 15 minutes; about half an hour; an hour or more	your ability to taste	your knowledge of your family's history	details_described
4.1, 4.2, 4.3, 4.4	it's important to you	How important is this aspect to you?	not important at all; somewhat important; moderately important; very important; extremely important	the pleasure you get from drinking fruit juice	you having people in your life who care about you	importance_you
4.1, 4.2, 4.3, 4.4	how much of other things you would give up to get more of this	Imagine you could gain more of this aspect by giving up other things. How much would you be willing to give up?	give up almost nothing; give up a little; give up a moderate amount; give up quite a bit; give up a lot	your enjoyment from owning rare things	you having time to pursue the activities you love	trade-off_give_up
4.1, 4.2, 4.3, 4.4	people rate it differently because they have different standards	If someone had the exact same amount of this aspect as you, do you think the other person might rate their amount of this aspect quite differently than you on a scale of 0-100 because their standards are different?	definitely rate the same; probably rate the same; could go either way; probably rate differently; definitely rate differently	your ability to eat solid food	the wisdom of those who are in authority	use_different_standards
4.1, 4.2, 4.3, 4.4	people would use numbers differently for this	If someone had the exact same amount of this aspect as you, do you think the other person might rate their amount of this aspect quite differently than you on a scale of 0-100 because they use the numbers differently?	definitely rate the same; probably rate the same; could go either way; probably rate differently; definitely rate differently	your ability to eat solid food	the wisdom of those who are in authority	use_numbers_differently
4.1, 4.2, 4.3, 4.4	values cause big differences between people	People differ in how much they have of each aspect of life. Sometimes it's because of differences in the options available to them, because of differences in resources or abilities. Sometimes it's because of differences in values (how much they prioritize this aspect relative to other aspects of life). Suppose everyone had the same options available to them. How much would people still differ in how much they have of this aspect because of differences in their values?	almost no differences due to values; small differences due to values; moderate differences due to values; large differences due to values; huge differences due to values	your success at accomplishing your goals	your spirituality	var_demand_shifter_q
4.1, 4.2, 4.3, 4.4	resources and abilities cause big differences between people	People differ in how much they have of each aspect of life. Sometimes it's because of differences in the options available to them, because of differences in resources or abilities. Sometimes it's because of differences in values (how much they prioritize this aspect relative to other aspects of life). Suppose everyone had the same values. How much would people still differ in how much they have of this aspect because of differences in resources or abilities (and therefore differences in what options are available to them)?	almost no differences due to resources and abilities; small differences due to resources and abilities; moderate differences due to resources and abilities; large differences due to resources and abilities; huge differences due to resources and abilities	your determination	your success at accomplishing your goals	var_supply_shifter_q

5.1, 5.2, 5.3, 5.4	losing it would be bad	How crucial is what you have of this aspect to your life? In particular, compared to other aspects of life, how bad would it be to lose almost all of this aspect?	not too bad to lose; somewhat bad to lose; quite bad to lose; extremely bad to lose; catastrophic to lose	the beauty of the products that you use	how glad you are that you are alive	average_value
5.1, 5.2, 5.3, 5.4	you could imagine the experience from a description	Given a description of someone else's situation that is different from your own, how vividly could you imagine what it's like to experience their level of this aspect?	couldn't imagine at all from a description; could imagine somewhat well from a description; could imagine fairly well from a description; could imagine quite well from a description; could imagine very well from a description	you being able to be virtuous	the relaxation you get from hot showers and baths	imaginable
5.1, 5.2, 5.3, 5.4	how important it is to most people	How important is each aspect to most people?	not important at all; somewhat important; moderately important; very important; extremely important	the pleasure you get from drinking fruit juice	you being loved by your friends and family	importance_most people
5.1, 5.2, 5.3, 5.4	you can tell which of two people an individual is closer to, in their level of this	If you knew three people well, how easy would it be to rank them on this aspect and say where the person in the middle is, relative to the other two —for example, closer to the bottom person or closer to the top person?	not easy at all; somewhat easy; moderately easy; very easy; extremely easy	your ability to keep up with technology	you fulfilling your destiny	interval_scale
5.1, 5.2, 5.3, 5.4	losing some would be especially painful	Think about your own life as it is, and how much you have of this aspect. Compared to other aspects of life, how painful would it be to lose some of this aspect?	not at all painful to lose some; a little painful to lose some; somewhat painful to lose some; quite painful to lose some; extremely painful to lose some	the beauty of the products that you use	how glad you are that you are alive	marginal_loss
5.1, 5.2, 5.3, 5.4	how hard it is to decide how much of other things you would give up to get more of this	Remember the trade-off question, where you had to decide how much you would give up of one aspect to get more of another. If you consider how important this aspect is in your life, how hard is it to decide how much of other things you would give up to get more of this aspect?	can't decide how much to give up at all; very hard to decide; somewhat hard to decide; not too hard to decide; not at all hard to decide	people not treating you unfairly	you feeling that almost all of your time is productive	trade-off_give_up_decide
5.1, 5.2, 5.3, 5.4	there's a yardstick for this	To what extent is there a yardstick for this? That is, how easy is it to say how much there is of this in a way that somebody else could understand?	no possible yardstick for this; only a lousy yardstick for this; only a mediocre yardstick for this; an okay yardstick for this; a good yardstick for this	you being able to transcend the physical and material	you being close to the weight you would like to be	yardstick

1.3. Quality Control Procedure

1.3.i. Baseline Quality Control

This subsection describes the “quality control” (QC) criteria for inviting respondents to the three surveys after Baseline. (We also excluded individuals from invitations if they opted out of being contacted for follow-up surveys.) The goal of the criteria, overall, is to exclude respondents whom we suspect of being outside the U.S., being a bot (non-human respondent), and/or putting forth extremely low effort. This procedure is especially important given the concerns that many researchers have raised about the quality of MTurk samples (e.g., Kennedy et al., 2020; Peer et al., 2022).

Using the Center for Open Science’s Open Science Framework, we pre-registered a project that is closely related to this one, six times, as we prepared to collect the survey data and conduct the analyses for the other paper. (We describe these plans as “pre-registrations” since they were created before most or all of the data were collected.) These pre-registrations can be

found on our project page using this link (which is view-only, with anonymized contributor names): https://osf.io/qk5ta/?view_only=c2c05eb8d51d4c088c363db7a26d2f15.

Baseline QC (Exclusion criterion for invitation to follow-up surveys)

Our Baseline QC criterion was established in Registration 4 (6/10/2022), before the recruitment of respondents who were invited to follow-up surveys. The exclusion criterion is based on three factors: IP address, textbox responses, and survey completion time. We weight the factors using points, as described below. If a respondent has three or more total points, they are excluded from invitations to follow-up surveys.

- I. 3 points: Baseline survey was completed in less than 10 minutes. This based on (a) the fact that the fastest we can physically complete the survey is approximately 10 minutes and (b) 10 minutes is approximately the bottom percentile (fastest survey takers) of Baseline respondents in the pilot data who would not have otherwise been excluded based on the other criteria.
- II. 2 points: IP Address originated from a virtual private server or from a non-US server. This is checked using an IP-checking service, DB-IP. We use an R package, `rgeolocate`, to access this service.
- III. 1 point: IP Address appeared more than once in the dataset.
- IV. 2 or 3 points: Poor responses to two textbox questions. The questions were:
 - i. When you rated: “how satisfied you are with your life”, what was your thought process?
 - ii. Think of a ladder where the lowest rung is ‘worst possible life for you’ and the highest rung is ‘best possible life for you’. Describe the life you imagine on the middle rung of the ladder.

For both questions, the prompt in the textbox is: “Insert one or more sentences. Please use complete sentences.”

Free-response textbox answers were scored by a team of three or four undergraduate research assistants, hereafter raters, with worse answers getting more points. (We started with four raters, but one found they did not have enough time for the project and quit.) For the QC criterion, we use the largest point value at least half of the raters give a participant. For example, if there were three raters and the rating values were (0, 0, 3), the points used would be a 0. If

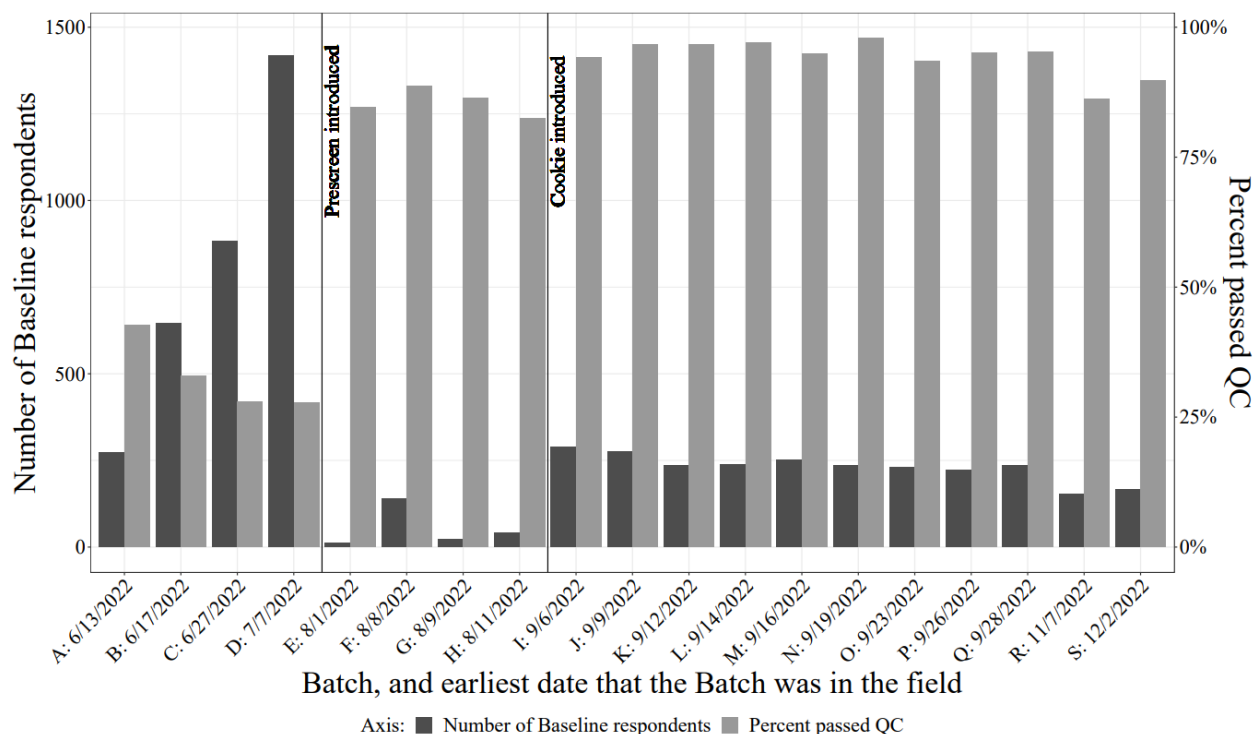
there were four raters and the rating values were (0, 0, 2, 3), the points used for exclusion would be a 2. This rule is robust to the number of raters changing.

To develop the textbox rating procedure, we began by rating 597 text boxes ourselves from the pilot sample. We gave the raters written instructions for rating textbox responses, including examples of graded responses. They rated the pilot sample text boxes based on these instructions. Then, we provided feedback and modified the instructions based on the raters' points assignments, to bring their ratings closer to our ratings. Finally, the raters completed the textbox scoring for the non-pilot survey sample. To make sure the raters evaluated only the textbox responses, not other quality metrics or scale use tendencies, the raters did not work on any analysis tasks nor have access to any other data when they rated the textboxes.

Quality Control Results and Addition of Prescreening Survey

We fielded the Baseline Survey in batches, with increasing size, to ensure our recruitment procedure was working. Through the first four batches, we found that the percentage of respondents who passed QC declined. This was especially troubling because the percentage passing QC in Batch A was already lower than it had been in a pilot. To avoid paying so many subjects for surveys with unusable data, we introduced a "Prescreening" survey, described further below. After testing the Prescreening survey and QC procedure with a few small batches (E-H), we introduced our final tool for quality control: using Qualtrics' "prevent multiple submissions" feature on the Prescreening survey. With this feature enabled on a Qualtrics survey, Qualtrics places a cookie on the respondent's browser when they submit the survey, and Qualtrics does not permit the respondent to take that survey again. **Figure SI.1.6.** shows the total number of Baseline respondents and the percentage of those respondents who passed the Baseline QC, by fielding Batch. It also notes the points at which the Prescreening survey and cookie were introduced. The effect of the Prescreening survey on our percentage of respondents passing Baseline QC is a clear increase, from Batch D to E. The Baseline QC pass rate also increased when the cookie was introduced, from Batch H to I. The results of the QC procedure on our survey data are discussed further in section A.3.iii.

Figure SI.1.6. Quality Control Results for Baseline Survey, by Batch



1.3.ii. Prescreening Survey Details

The Prescreening Survey and related QC procedure are described in Registration 6 (7/25/2022). The Prescreening Survey has 11 questions in total. We conducted the survey using Qualtrics. The outline is: (a) Consent form; (b) Abbreviated version of the Baseline instructions; (c) One “triple” (three SWB questions followed by six tradeoff questions) featuring aspects of well-being chosen randomly from the 20 Baseline aspects that are not “public” or “double negative” (to make the survey instructions as simple as possible); (d) Two free-response (textbox) questions about what the respondent was thinking about while answering survey questions.

Prescreening QC (Exclusion criteria for invitation to Baseline)

The Prescreening QC criterion is based on IP address and textbox responses in the Prescreening survey. If a respondent has three or more total points, they are not invited to the Baseline survey. (Note that our payment of subjects was not tied to these criteria. Respondents were compensated \$0.75 for completing the Prescreening survey, regardless of whether or not they passed our Prescreening QC.)

- I. 2 points: IP Address originated from a virtual private server or from a non-US server. This is checked using an IP-checking service, DB-IP. We use an R package, *rgeolocate*, to access this service.
- II. 1 point: IP Address appeared more than once in the dataset.
- III. 2 or 3 points: Poor responses to the two free-response questions in the Prescreening survey. These questions are:
 - i. In the rating questions, 0 was the “Lowest level possible” and 100 was the “Highest level possible.” How would you describe how you interpreted the word ‘possible’?
 - ii. When you made the decisions, what were you thinking about? As a reminder, here is an example decision. [Screen includes a cropped screenshot of a tradeoff question.]

For both questions, the prompt in the textbox is: “Insert one or more sentences here. Please use complete sentences.”

The Prescreening textboxes were scored by the same team of three or four undergraduate research assistants who scored the Baseline QC textboxes.

1.3.iii. Results of Quality Control Procedure

Table SI.1.3. reports respondent-level quality control variables by fielding group and QC results. Overall, the results reassure us that our quality control procedure led to a sample of respondents who are real humans and putting forth at least a minimum level of effort. Each variable (row) in the table is reported so that a higher number corresponds to a higher level of response quality.

The first three columns are three respondent groups from before the Prescreening survey was implemented (Batches A-D): all who completed Baseline, those who failed Baseline QC, and those who passed Baseline QC. Looking down the rows for these columns, we see that the respondents who passed QC (column 3) have consistently higher quality than those who failed QC (column 2). This relationship must hold by definition for the variables that are directly related to the QC criterion. However, the other response-quality variables show a similar pattern.

Note that the state comparison quality check—whether the respondent’s state derived from their IP address matches the state we derived from their ZIP code—has a relatively low percentage of respondents passing the check in all columns. However, we did not expect these

percentages to be as high as those for other quality checks. There would be some mismatches even if respondents behaved perfectly, for three reasons: IP address is an approximate geographic variable; ZIP codes do not map uniquely to states; and respondents could be traveling away from home while taking the survey. Still, we see the same patterns of the quality control procedure improving that variable as it improves the others.

The final row is based on our tradeoff data. Here, we report the percentage of respondents who never fail an “opposite-direction trap,” i.e., they never choose a decrease in one aspect level over an increase in another aspect level. (As noted in footnote 3, the Baseline survey is designed so that tradeoff questions show opposite changes in aspect levels with 2% probability at the respondent level.) Before the Prescreening survey (“Prescreen”) was implemented, the percentage of Baseline respondents who passed all the opposite-direction traps they faced was roughly twice as high in the passed-QC sample (23.5%, Column 3) than in the failed-QC sample (11.9%, Column 2). After the Prescreen was introduced, that percentage rose to 38.7%, for all Baseline respondents (Column 6). Because opposite-direction traps are rare, note that the sample sizes corresponding to the final row for Columns 7 and 8 are very small (N=1 and N=4, respectively). Therefore, the percentage of respondents who never failed an opposite-direction trap is basically unchanged between Column 6 (all who took Baseline after Prescreen) and Column 9 (the subset of those who took Prescreen after the cookie was introduced and passed Baseline QC).

We now turn back to general comparisons across columns. Comparing Column 1 to Column 6, we see the intended effect of introducing the Prescreening survey: the quality level for respondents completing Baseline is better on every metric. The improvement in the Baseline textbox grade is especially striking, with only 25.5% receiving the highest grade before the Prescreening QC criterion was introduced and 90.5% receiving the highest grade after it was introduced.

Column 7 shows the response-quality variables for the group of respondents who passed Prescreening QC but went on to fail Baseline QC. The quality measures in Column 7 are worse than those for all the other groups who passed the Prescreening survey (Columns 6, 8, and 9), with the exception of two variables with no difference (Unique IP address and Prescreening completion time greater than or equal to 3 minutes). This suggests that the Prescreening QC criterion had some “type 2 error” in the sense of inaccurately allowing some low-quality

respondents to proceed to Baseline. This also shows that there was value added by the second (Baseline) round of the QC procedure.

Table SI.1.3. QC Variables by Respondent Group

	1	2	3	4	5	6	7	8	9
Respondents who...	All who took Baseline before Prescreen	Failed Baseline QC before Prescreen	Passed Baseline QC before Prescreen	All Prescreen	Failed Prescreen QC	All who took Baseline after Prescreen	Failed Baseline QC after Prescreen	Passed Baseline QC after Prescreen, before cookie	Passed Baseline QC after Prescreen, after cookie
Sample size	3,218	2,250	968	10,781	5,846	2,752	162	221	2,369
Variables related to QC criterion									
Unique IP Address	86.7%	82.5%	96.5%	98.1%	96.9%	99.6%	100.0%	100.0%	99.6%
Not using a virtual private server	90.7%	87.6%	98.0%	95.7%	93.3%	98.5%	95.1%	97.3%	98.9%
Best possible textbox grade, Prescreening survey				41.2%	0.0%	90.6%	65.4%	85.1%	92.8%
Best possible textbox grade, Baseline	25.5%	0.4%	83.9%			90.5%	1.2%	91.0%	96.5%
Baseline completion time >= 10 minutes	96.9%	95.5%	100.0%			99.9%	98.8%	100.0%	100.0%
Other response quality variables									
Prescreening completion time >= 3 minutes				83.0%	74.9%	93.4%	93.2%	93.2%	93.4%
State derived from IP Address matches state derived from zip code	20.2%	6.1%	56.4%			78.6%	50.4%	65.5%	81.8%
CQ responses have correct ranking of high and low stimuli: region size trio	73.7%	64.1%	95.9%			97.6%	88.3%	96.8%	98.4%
CQ responses have correct ranking of high and low stimuli: darkness of circle trio	74.2%	64.6%	96.4%			97.2%	87.0%	97.3%	97.8%
CQ responses have correct ranking of high and low stimuli: curvy line trio	68.4%	58.4%	91.7%			94.5%	84.6%	93.7%	95.2%
CQ responses have correct ranking of high and low stimuli: 'your access to information'	56.1%	47.6%	75.6%			84.8%	70.4%	80.5%	86.2%
CQ responses have correct ranking of high and low stimuli: 'your ability to remember'	56.3%	47.2%	77.5%			81.3%	68.5%	79.2%	82.4%
CQ responses have correct ranking of high and low stimuli: 'your living area not being spoiled by crime and violence'	53.8%	46.1%	71.7%			78.8%	69.8%	73.3%	79.9%
Test-retest correlation of repeated SWB ratings within Baseline survey (3 per person)	0.73	0.57	0.83			0.84	0.74	0.86	0.84
Never failed an "opposite direction trap" in tradeoff questions (choosing a decrease in one direction of SWB instead of an increase in another dimension of SWB)	15.3%	11.9%	23.5%			38.7%	0.0%	50.0%	38.6%

Note: For calibration questions (CQs), see Baseline screens in Web Appendix.

We can also evaluate the success of our quality control procedure by comparing three of our quality metrics to similar metrics from Douglas et al. (2023), who compare respondent quality between five online survey platforms: MTurk, Prolific, CloudResearch, SONA, and Qualtrics. They construct a variable that indicates a high-quality respondent based on 11 measures: five attention checks, unique worker ID, unique IP address, unique geolocation,

meaningful or blank open response, sufficient time taken, and self-reported high data quality. Here we will just discuss their MTurk and Prolific results.

Douglas et al. find that 97.2% of respondents have unique IP addresses in their MTurk sample and 98.8% have unique IP addresses on Prolific. As seen in Table 4, our percentage of Baseline respondents with a unique IP address increases from 86.7% without any QC procedure (Column 1) to 96.5% with the Baseline QC procedure (Column 3) and 99.6% with the Prescreening survey and cookie (Column 9). For Douglas et al.'s open response question, a graduate student RA and an undergrad RA both graded the answer to "Do you have any additional comments for us?" A worker's textbox is classified as meaningful if (a) both graders rate the answer as meaningful, or (b) the worker leaves the textbox blank. Otherwise the textbox is classified as meaningless. On MTurk, 82.2% of their respondents have a blank or meaningful response. On Prolific, it is 99.2%. We believe that a high-quality response is more difficult for our textbox questions, since we ask about question interpretation and we do not accept blank responses (a blank response is assigned the worst score of 3 points). Nevertheless, looking at the percentage of respondents with the best possible textbox grade on Baseline, we see an increase from 25.5% before any QC procedure (Column 1) to 83.9% with the Baseline QC (Column 3) and 96.5% after the Prescreening survey and cookie (Column 9).

Finally, Douglas et al. asked participants the same survey question twice, 21 questions apart in the survey, to assess the test-retest correlation of participant responses. They find a correlation of 0.54 for MTurk and 0.87 for Prolific. Recall that in our Baseline survey, the 11th triple is a repeat of the 2nd triple as a check for data quality; this means that each respondent rates three SWB questions twice within the same survey. We find that the test-retest correlation is 0.73 before any QC procedure (Column 1), 0.83 after the Baseline QC (Column 3), and 0.84 after the Prescreening QC and cookie (Column 9). Given these three comparisons (though they are admittedly rough), we conclude that our survey sample's quality is closer to Douglas et al.'s Prolific sample quality than their MTurk quality.

1.4 Survey Fielding

1.4.i. Baseline Survey Fielding

Using Amazon's Mechanical Turk (MTurk) platform, we collected data between June 13 and December 7, 2022. As **Figure SI.1.6** shows, we proceeded in 19 batches, which we labeled Batch A through Batch S. For all batches, we restricted eligibility to MTurk workers located in

the United States (as identified by MTurk), with a HIT approval rating of at least 95%, and at least 100 previously approved HITs. MTurk restricts all participation to workers who are at least 18 years of age. Respondents were recruited to Baseline with a HIT titled “Academic survey about what is important in life,” either directly (batches A-D) or after passing an initial prescreening survey (batches E-S). Respondents were compensated \$4.50 for completing Baseline. The median completion time for those who passed our stringent quality control checks was 36 minutes.

1.4.ii. Multi-Block Survey Fielding

Using MTurk, we collected data on the Multi-Block Survey (Multi-Block) between September 20, 2022, and January 3, 2023. Respondents who passed the Baseline QC procedure and indicated they were interested in follow-up surveys were recruited by email to a HIT titled, “INVITE ONLY: Follow-up survey on what is important in life.” The median length of time between completing Baseline and the first block of the Multi-Block survey was about 7 weeks (46 days), with a range from 6 days to 164 days. Respondents were compensated \$1.50 per Multi-Block block completed, with a bonus of \$15 if all 30 blocks were completed. The median completion time for Block 1 was 18 minutes; Blocks 2-29, 13 minutes; and Block 30, 13 minutes. Of the 3,277 respondents who passed Baseline quality control and were invited to follow-up surveys, $N = 2,603$ (79.4 percent) completed the first block of Multi-Block and $N = 656$ (25.2 percent of those who completed Block 1) completed the survey through all 30 blocks, taking an average of 7.2 hours to complete them. The sample used for this paper consists of the $N = 896$ respondents who completed Blocks 1-9 (34.4 percent of those who completed Block 1); they took an average of 1.5 hours to complete the 9 blocks.

1.4.iii. Life and Psych Survey Fielding

Using MTurk, we collected data on the Life Circumstances and Psychological Factors (“Life and Psych”) Survey between June 2022, and December 2022. The median length of time between completing Baseline and the Life and Psych survey was 3 weeks, with a range from 0 days to 104 days. Respondents were compensated \$5.75 and the median completion time was 41 minutes.

1.4.iv. Aspect-Characteristics Survey Fielding

Using MTurk again, we collected data on the Aspect-Characteristics Surveys between September 2024, and January 2025 – about 24 months after respondents had completed Baseline. As described in the main paper text, the aspects and characteristics were split across survey

tranches to keep survey length manageable, and each respondent was invited to one tranche at most. Compensation was \$5 (first four tranches) or \$5.50 (later 12 tranches) and the median completion time was 31 minutes.

1.5 References for Supplementary Methods

Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Alex Rees-Jones. 2012. “What Do You Think Would Make You Happier? What Do You Think You Would Choose?” *American Economic Review* 102 (5): 2083–2110.

Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Nichole Szembrot. 2014. “Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference.” *American Economic Review* 104 (9): 2698–2735.

Diener, Ed. Robert A. Emmons, Randy J. Larsen, and Sharon Griffin. 1985. “The Satisfaction With Life Scale.” *Journal of Personality Assessment* 49 (1): 71-75.

Douglas, Benjamin D., Patrick J. Ewell, and Markus Brauer. 2023. “Data Quality in Online Human-Subjects Research: Comparisons Between MTurk, Prolific, CloudResearch, Qualtrics, and SONA.” *PLoS ONE* 18 (3): e0279720.

Kennedy, Ryan, Scott Clifford, Tyler Burleigh, Philip D. Waggoner, Ryan Jewell, and Nicholas J. G. Winter. 2020. “The Shape of and Solutions to the MTurk Quality Crisis.” *Political Science Research and Methods* 8 (4): 614-629.

Peer, Eyal, David Rothschild, Andrew Gordon, Zak Evernden, and Ekaterina Damer. 2022. “Data Quality of Platforms and Panels for Online Behavioral Research.” *Behavior Research Methods* 54 (4): 1643-1662.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, 38(1), 119–125. [https://doi.org/10.1002/1097-4679\(198201\)38:1<119::AID-JCLP2270380118>3.0.CO;2-I](https://doi.org/10.1002/1097-4679(198201)38:1<119::AID-JCLP2270380118>3.0.CO;2-I)

2. Supplementary Tables

Supplementary Table 1 – Largest Discrepancies Between Revealed and Stated Importance

Aspect	Log MU	Residual
<i>Panel A: Revealed importance higher than predicted</i>		
How high your income is compared to the income of other people around you	0.50	1.51
Your children's health	4.20	1.49
The health of your parents and siblings	2.16	1.30
Your rating of your life on a ladder from "worst possible" to "best possible"	0.61	1.25
The people who have been most important to you in your life still being alive	2.08	1.00
Your family living a wonderful life	1.77	0.92
You not feeling anxious	0.47	0.77
You learning or doing something interesting every day	−0.44	0.75
Your spouse/partner's health	2.79	0.74
You reaching most of your goals over the past 12 months	0.11	0.70
<i>Panel B: Revealed importance lower than predicted</i>		
You not having things stolen from you	−0.45	−0.76
Your physical safety and security	0.47	−0.81
Your spouse/partner taking on their fair share of responsibility for the household	−0.22	−0.85
You feeling like you have work-life balance	−0.49	−0.90
How much you trust the media in your nation	−2.60	−0.92
The quality of your spouse/partner as a mate	0.92	−0.99
You having a life outside of work	−0.64	−1.03
How easy it is for you to get from home to work and back	−1.81	−1.06
Your cultures and traditions being honored	−2.25	−1.12
You not being discriminated against for any reason	−1.37	−1.59

Notes: This table shows the ten aspects with the largest positive (Panel A) and negative (Panel B) residuals from a regression of mean log marginal utility estimates on aspect-level ratings of importance to oneself, importance to most people, social desirability, and the log fraction of respondents to whom each aspect applies (N = 126 aspects). Positive residuals indicate aspects that are more important in tradeoff choices than direct ratings predict; negative residuals indicate aspects that are less important. Log MU is the estimated log marginal utility from tradeoff choices.

Supplementary Table 2 – Selected Scholarly Theories of What People Do Want or Should Want

Thinker	Book/Paper	Theory	Ranks of aspects most closely matching the theory
Panel A: Should Want			
Buddhism	[No single canonical book accepted by all Buddhists]	Wanting things and having desires leads to suffering; ceasing to have these desires will end the suffering.	(39 → 44) The absence of stress in your life, (48 → 53) You feeling calm and peaceful
Avicenna (Ibn Sīnā)	<i>Kitāb al-shifā’ (The Cure)</i> (11 th century CE)	“Real happiness is the perfection of the rational soul through knowledge.”	(62 → 63) Your knowledge and skills, (76 → 65) The wisdom you have been able to accumulate, (85 → 83) You learning or doing something interesting every day
Martin Luther	<i>The Freedom of a Christian</i> (1520)	People should share God’s love not only with God but also with their neighbors, doing so sincerely and not for personal gain.	(79 → 82) You taking care of others instead of just yourself
Panel B: Do Want			
Thomas Hobbes	<i>Leviathan</i> (1651)	In natural conditions people compete and want to protect themselves from others. They therefore forgo liberty for a sovereign that ensures safety.	(46 → 70) Your physical safety and security, (65 → 88) Your living environment not being spoiled by crime and violence, (86 → 105) You not having things stolen from you
Alfred Adler	<i>Social interest: A challenge to mankind</i> (1933)	The drive to overcome feelings of inferiority, if unsuccessful, can lead to an inferiority complex. Contributing to the welfare of others—including one’s family—is an important way to overcome such feelings.	(17 → 29) Your ability to take care of your family, (79 → 82) You taking care of others instead of just yourself, (120 → 112) Your stature in your community
Viktor Frankel	<i>Man’s search for meaning</i> (1946)	People search for meaning in life.	(61 → 62) The extent to which you feel the things you do in your life are worthwhile, (63 → 60) Your sense of purpose, (67 → 50) How much of the time the things you do seem meaningful, (90 → 95) You finding meaning in the world
James Duesenberry	<i>Income, saving and the theory of consumer behavior</i> (1949)	People aim to enhance their social status. This goal often manifests in the desire for a higher standard of living relative to others.	(44 → 3) How high your income is compared to the income of other people around you, (71 → 27) How satisfied you are with your standard of living compared to the people you spend time with, (120 → 112) Your stature in your community

Abraham Maslow	<i>Motivation and personality</i> (1954)	Hierarchy of needs: physiological, safety, love, esteem, self-actualization. Once a level is satisfied, people tend to focus on the next.	(12 → 20) Your health, (24 → 28) Your relationship with your spouse, partner, or closest friend being stronger than ever, (26 → 37) You and your family having enough to eat, (46 → 70) Your physical safety and security, (73 → 40) You being a winner in life, (92 → 94) You getting to use your strengths to do what you do best every day
----------------	--	---	---

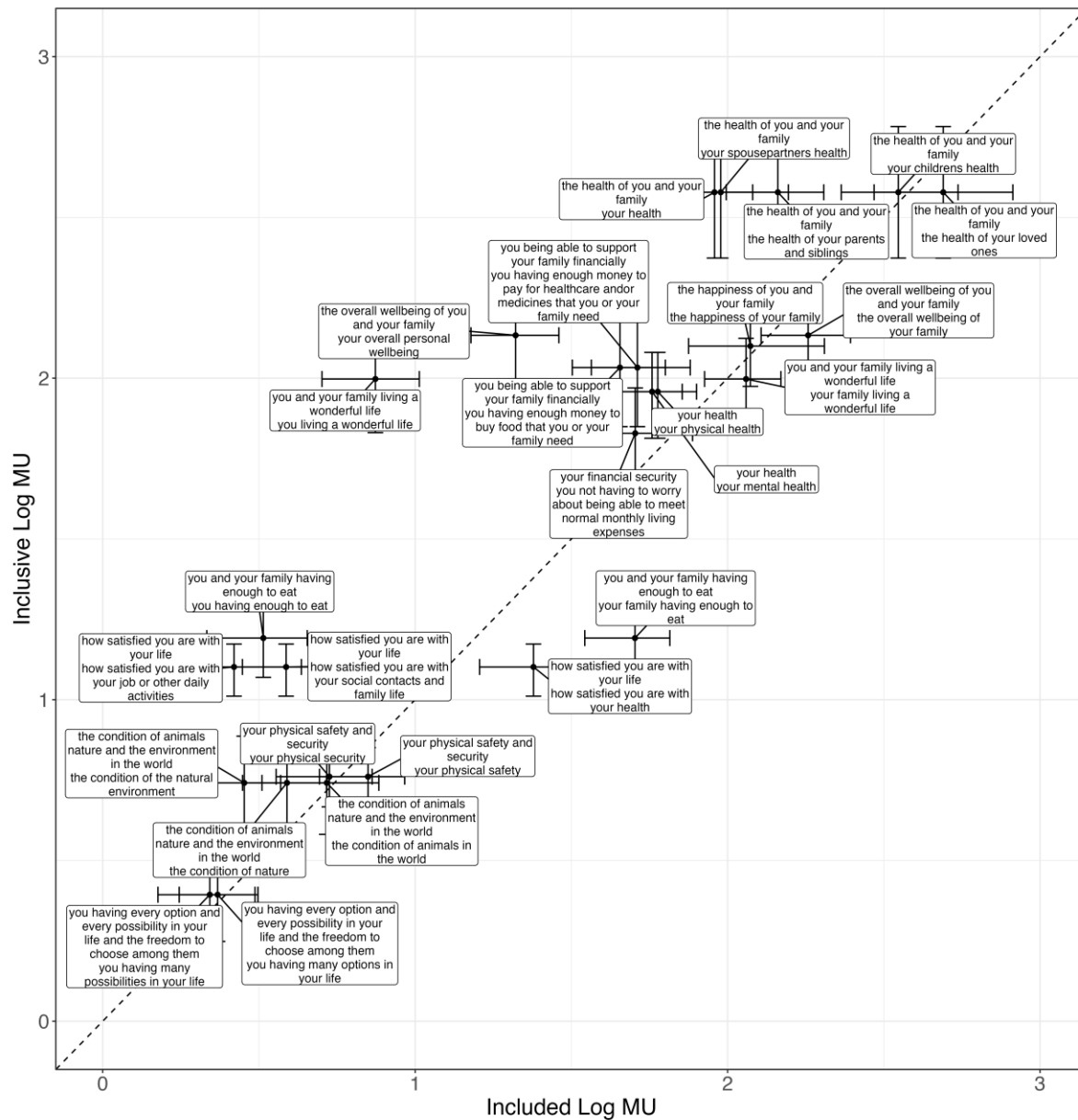
Panel C: Both Should and Do Want

Utilitarians (Jeremy Bentham, John Stuart Mill) and modern happiness economists	<i>An Introduction to the Principles of Morals and Legislation</i> (1789) <i>Utilitarianism</i> (1863) <i>Happiness lessons from a new science</i> (2005)	Happiness—maximizing pleasure and minimizing pain—is the sole end of human action. The goal of legislation should be greatest happiness of the greatest number.	(33 → 43) Your freedom from physical pain, (35 → 35) How happy you feel, (37 → 34) How much you enjoy your life
---	---	---	---

Notes: The table includes only a selection of thinkers who are influential and whose theories make claims about what people do want or should want. We excluded thinkers whose views do not map well onto our aspect list (e.g., Plato, Aristotle, Confucius, Augustine, al-Fārābī, Maimonides, Aquinas) and those whose theories primarily focus on groups rather than individuals (e.g., Marx and Engels). The last column includes aspects that closely fit the theory, with their (*Unadjusted rank*) → [**Adjusted rank**]. For the full list of our aspects and their ranks see **Table 2**, column 3.

3. Supplementary Discussion

3.1 Subset/Superset Pairs



Notes: Bayesian hierarchical model estimates using Hamiltonian Monte Carlo, based on 1,998,177 tradeoff responses for 533 aspects from 2524 respondents who completed any block from the Multi-Block Survey and passed quality control. Plot shows 27 pairs of aspects where one aspect is “Inclusive” (e.g., *The health of you and your family*) and the other is “Included” (e.g., *Your health*), in that it is a conceptual subset of the inclusive aspect. Estimates for the log marginal utility of the inclusive and included aspects are shown on the vertical and horizontal axis, with 95% credible intervals. Dotted line is the identity (45 degree) line.

Overall, we find that marginal utility estimates for each pair follow the expected pattern: the superset aspect (“inclusive”) receives higher marginal utility. The two pairs that do not follow this trend are (*Your family having enough to eat, You and your family having enough to eat*) and (*How satisfied you are with your health, How satisfied you are with your life*). We speculate that for these two pairs, the superset framing may reduce the salience of dimensions that are particularly important. Family welfare and health are among the highest-valued aspects in our data, and expanding the scope to additional elements may dilute attention from these core aspects. Additionally, respondents may interpret the superset as requiring them to average across components, so adding less-valued elements (e.g., personal welfare) to higher-valued ones (e.g., family welfare) lowers the overall marginal utility of improvement, rather than raising it.

3.2 MU Spreads for High- and Low-Numeracy Respondents

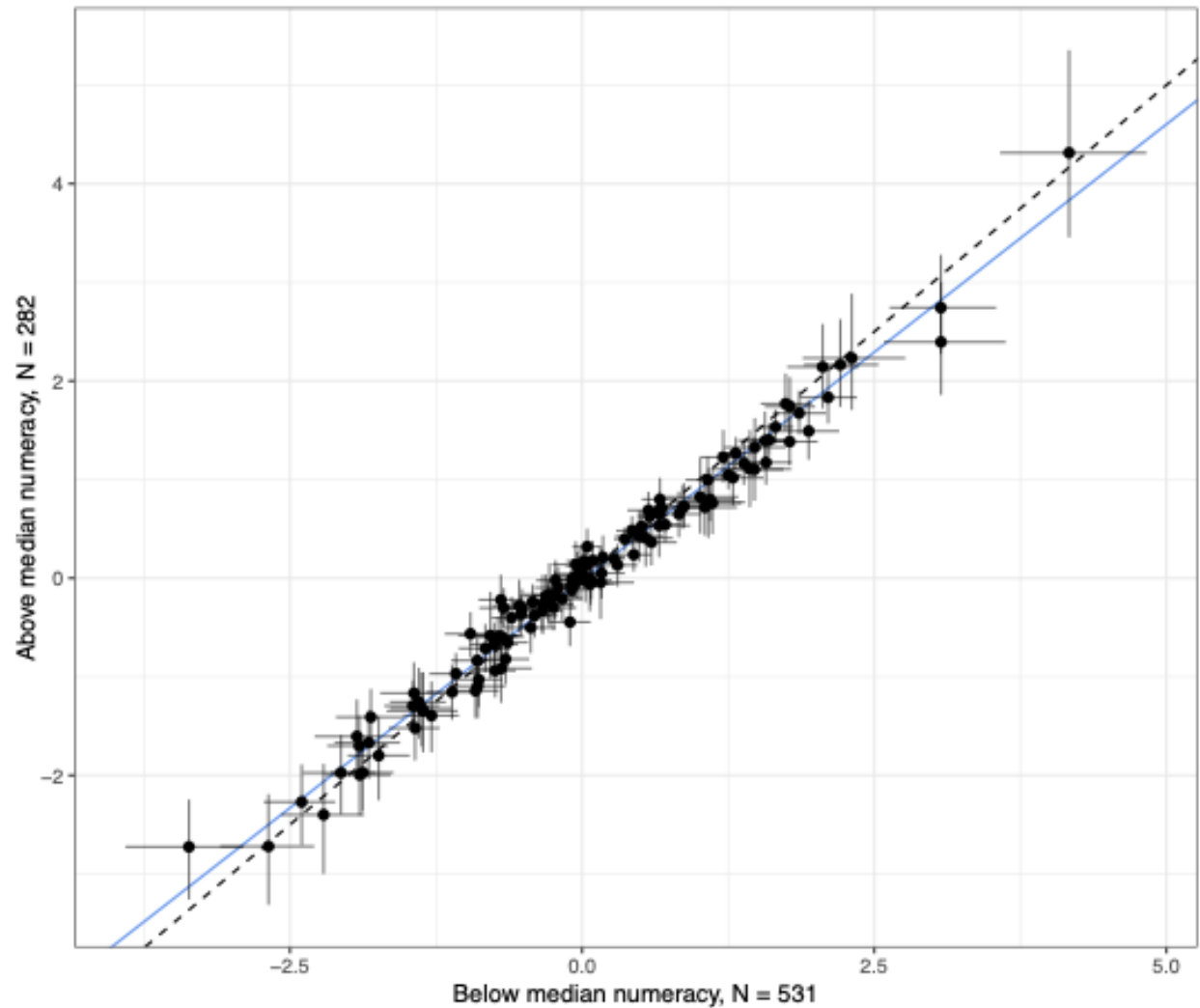
Under our model, we assume that respondents weigh marginal utilities and level changes in the same manner when evaluating tradeoffs. However, if certain individuals systematically place more or less weight on the numerical magnitudes of changes, we might observe differences in the dispersion of estimated marginal utilities that reflect response style rather than true preferences. For instance, a respondent who focuses primarily on numerical magnitudes when choosing between options would generate relatively bunched estimates of marginal utility across aspects, while a respondent who largely ignores the numbers would generate more dispersed estimates.

To test whether such heterogeneity affects our results, we estimated marginal utilities separately for respondents with high versus low numeracy. Numeracy was measured in the Life and Psych Survey using seven math questions (e.g., arithmetic operations, identifying prime numbers, and completing number sequences). Because the distribution of scores was concentrated at the top, we split respondents into those who answered all seven questions correctly ($N = 282$) versus those who made at least one error ($N = 531$). If higher-numeracy respondents place greater weight on numerical changes, we would expect their marginal utility estimates to be less dispersed. Instead, we find that the variance of log marginal utility estimates is similar across the two groups, and regressing high-numeracy estimates on low-numeracy estimates yields a slope indistinguishable from one. This suggests that our main results are not driven by differential attention to numerical information across respondents.

Comparing Log MU estimates – sample subsets

Dashed black: identity; blue: TLS; error bars are credible intervals

Label if both tests pass at 99.9% (45-degree diff + TLS-distance); color reflects 45-degree direction



Notes: Bayesian hierarchical model estimates of log marginal utility for the list of 126 aspects using Hamiltonian Monte Carlo. Vertical axis shows estimates from the above numeracy sample (181,687 tradeoffs from 282 respondents), and horizontal axis shows estimates from the below numeracy sample (339,847 tradeoffs from 531 respondents). Crosses show 95% credible intervals.

4. Supplementary Equations

None

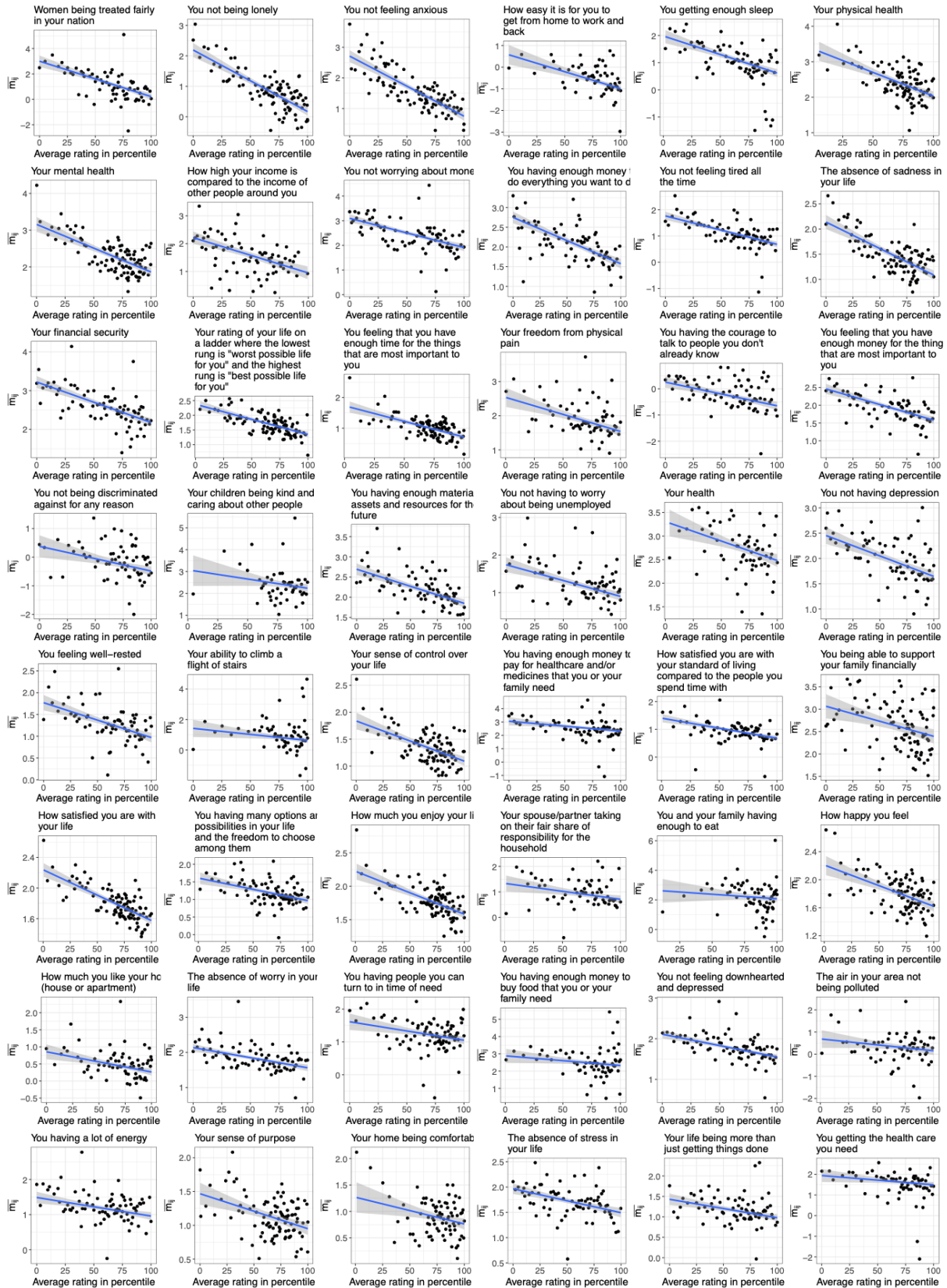
5. Supplementary Note

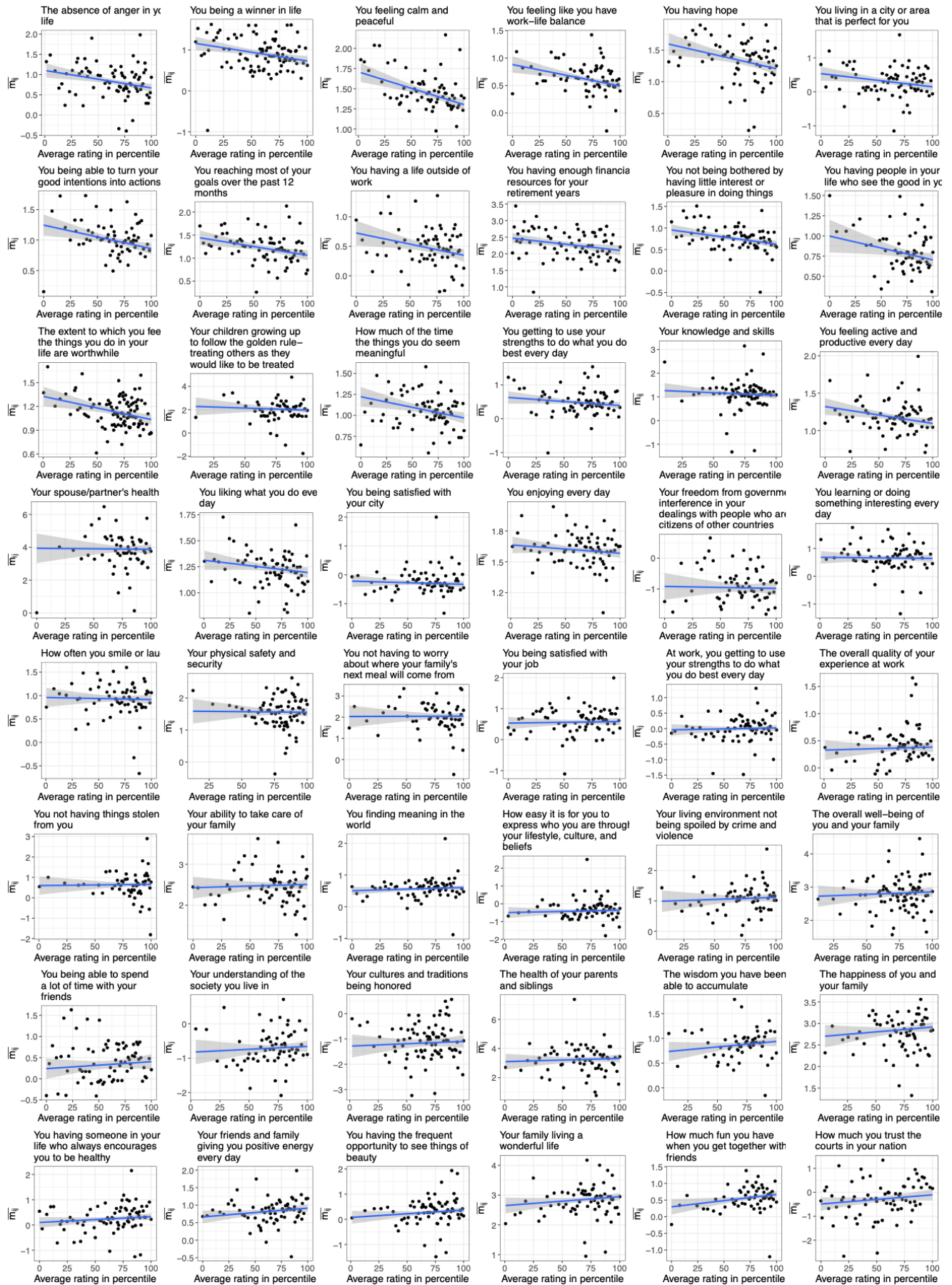
5.1 Importance Ratings vs. Trade-Based MU Estimates

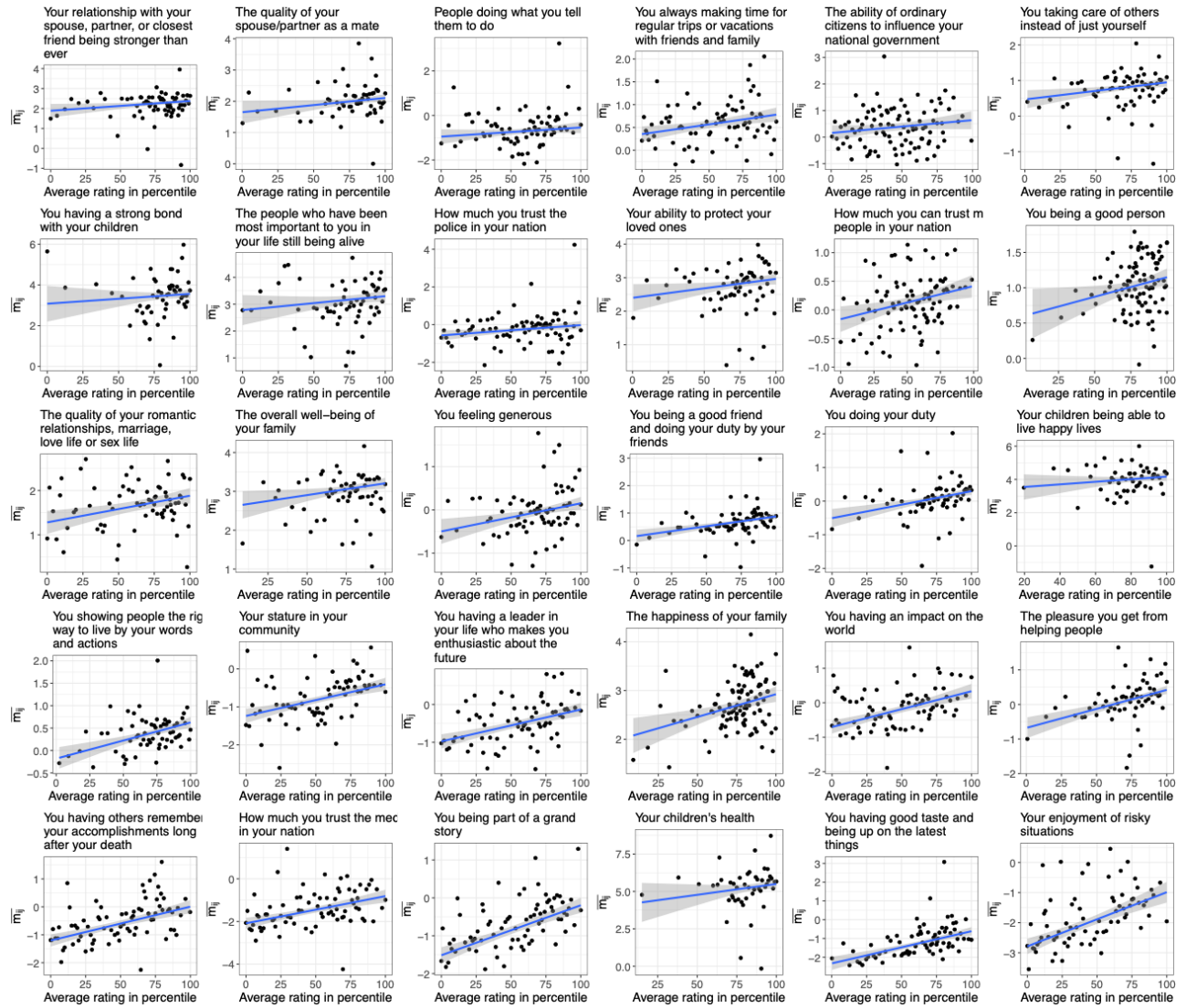
On average, people's direct reports of aspect importance, from the Aspect-Characteristics survey, are a very good predictor of what we find from people's choices in tradeoff questions. But there are some systematic differences. We ran a regression of the log marginal utility as measured from the tradeoff choices on the three direct questions just above about people's sense of the importance of different aspects, controlling for social desirability and the logarithm of the fraction of people who have a particular relationship—since it is unclear, for example, whether people answer how important children's health is for those who have kids or for people in general, including those who have no children. (This procedure yields the same residuals whether it is adjusted or unadjusted log marginal utility that is predicted.) These regressions predict 81% of the variance for unadjusted and 76% of the variance for adjusted log marginal utility. Yet the residuals for the biggest outliers, shown in **Supplementary Table 1**, are also large. If we take the tradeoff choices as the gold standard, the ten biggest positive residuals show aspects being 2.0 to 4.5 times as important to people than predicted. The ten biggest negative residuals similarly show aspects being 2.1 to 4.8 times less important to people than predicted. These ten biggest negative outliers are all aspects of well-being that are frequently talked about in the public sphere (*e.g.*, ***You not being discriminated against for any reason***) or “around the water cooler” (*e.g.*, ***How easy it is for you to get from work and back***). Thus, they may be particularly salient to people, potentially causing them to overestimate how important they would be to people faced with a tough choice. The ten biggest positive surprises do not appear to similarly have a single such explanation. Our respondents underestimated how much people—people like them—care when faced with a tough choice about relative status (*e.g.* relative income or standing on the ladder of life), the health of family members, and not being beset by negative emotions.

6. Supplementary Figures

Supplementary Figure 1 – Level percentiles and log MU for each aspect

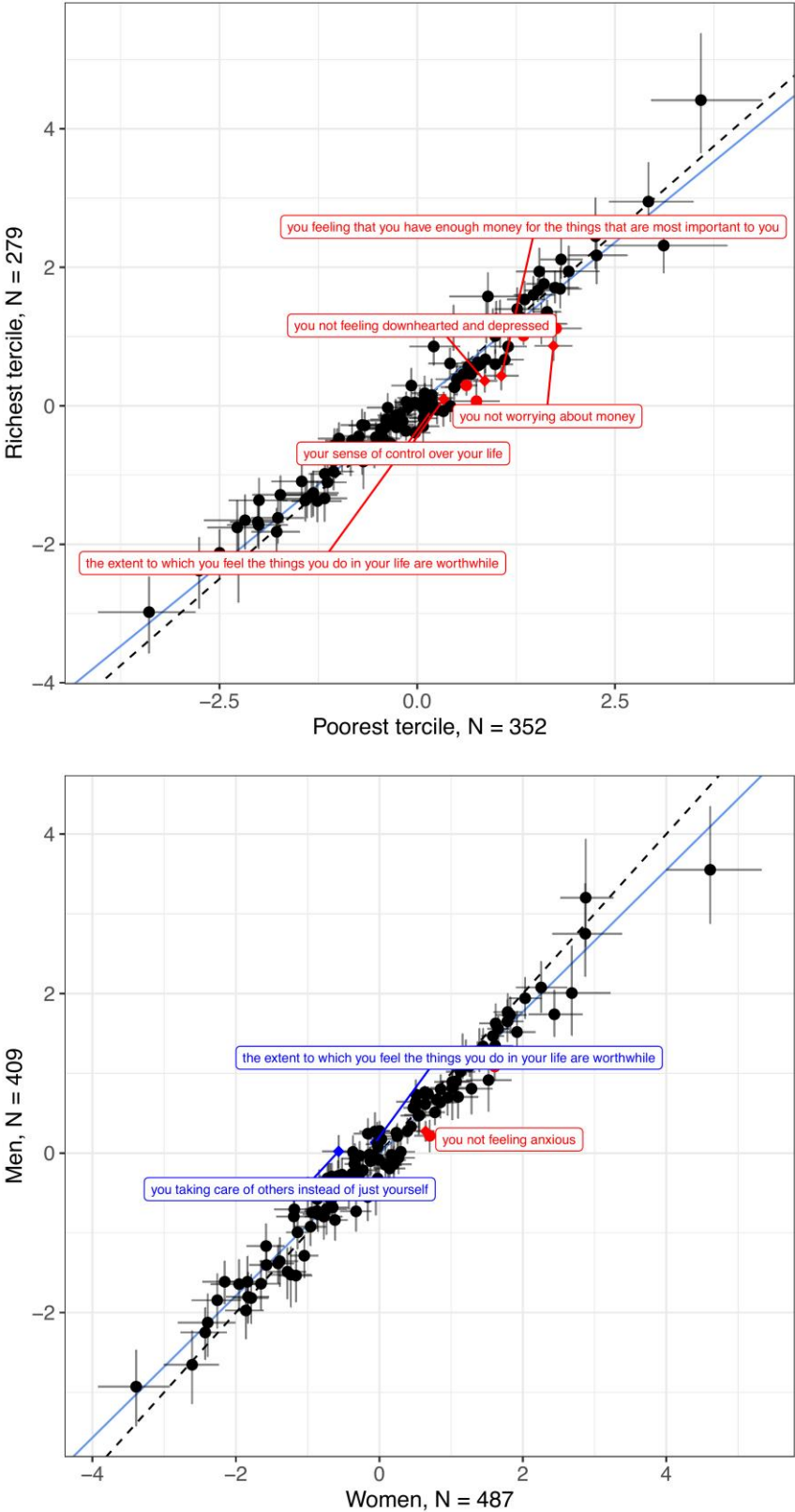


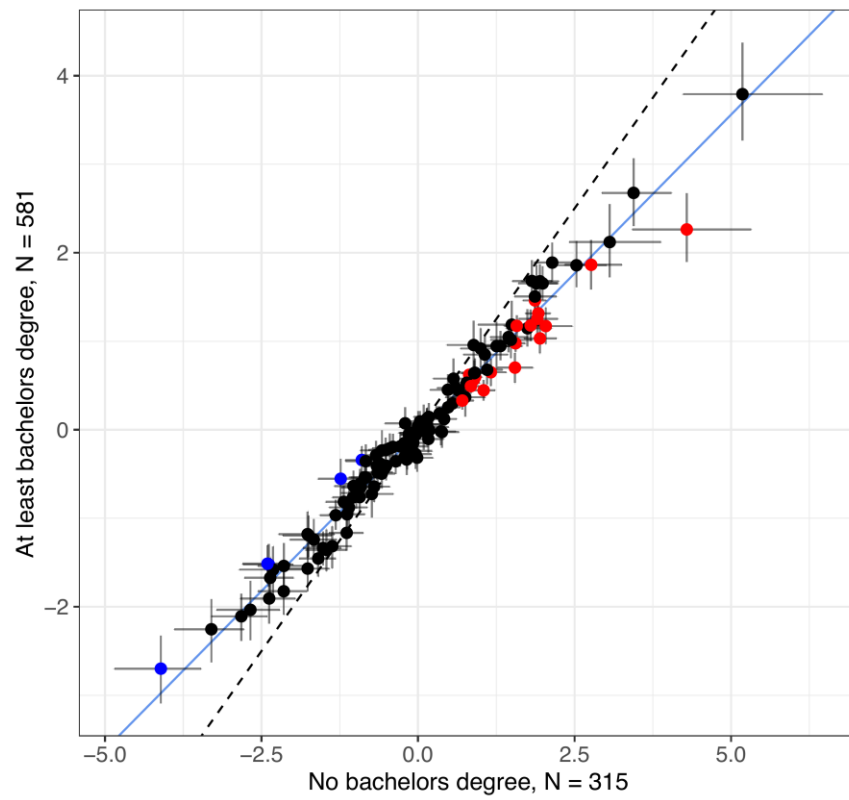
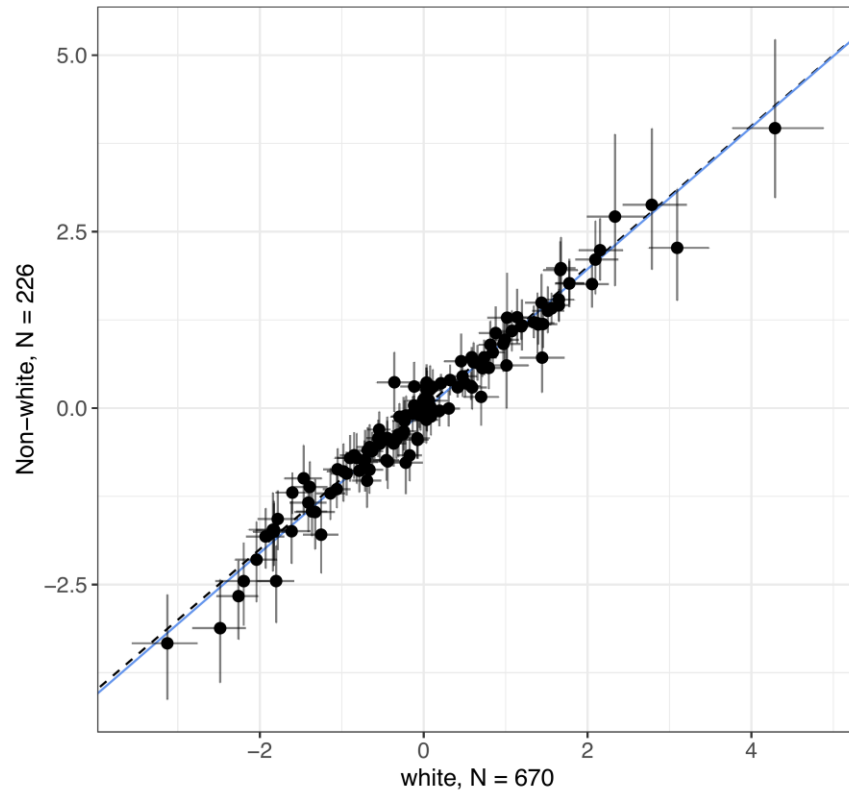


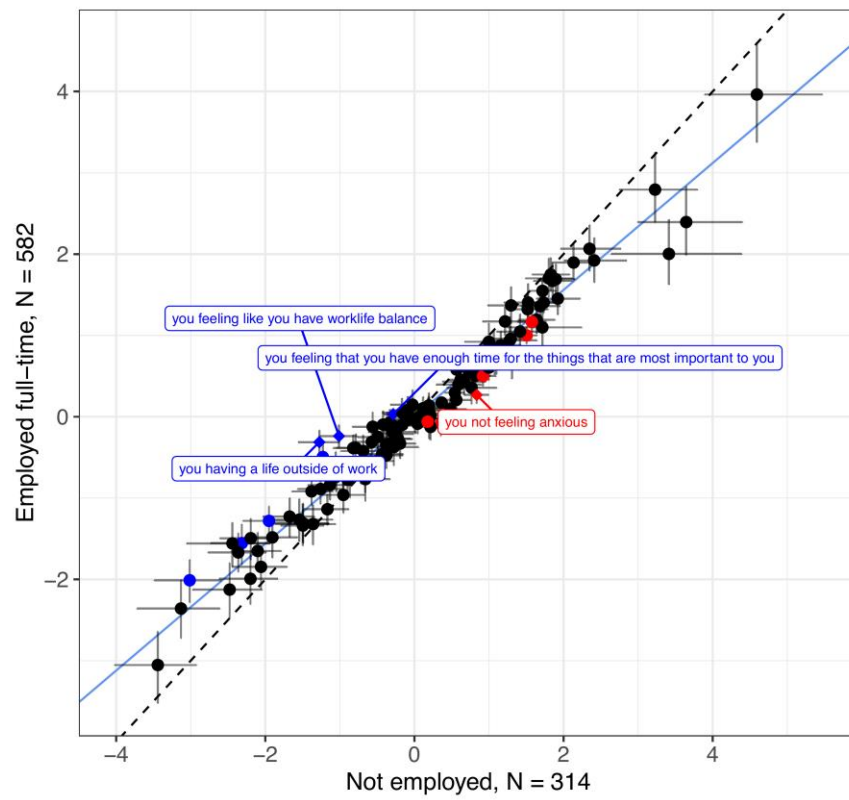
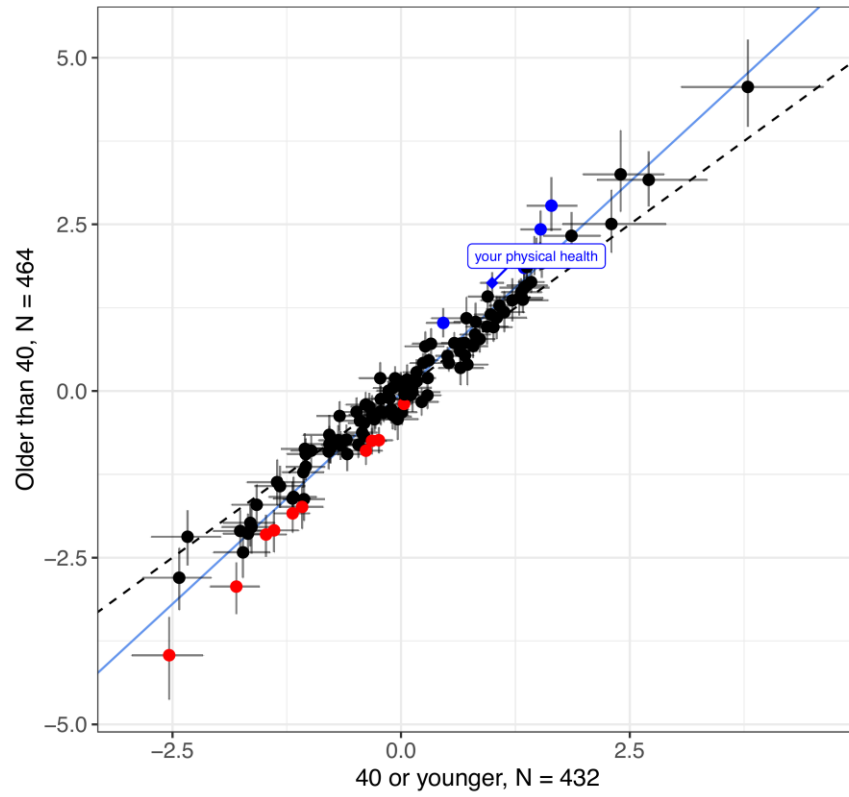


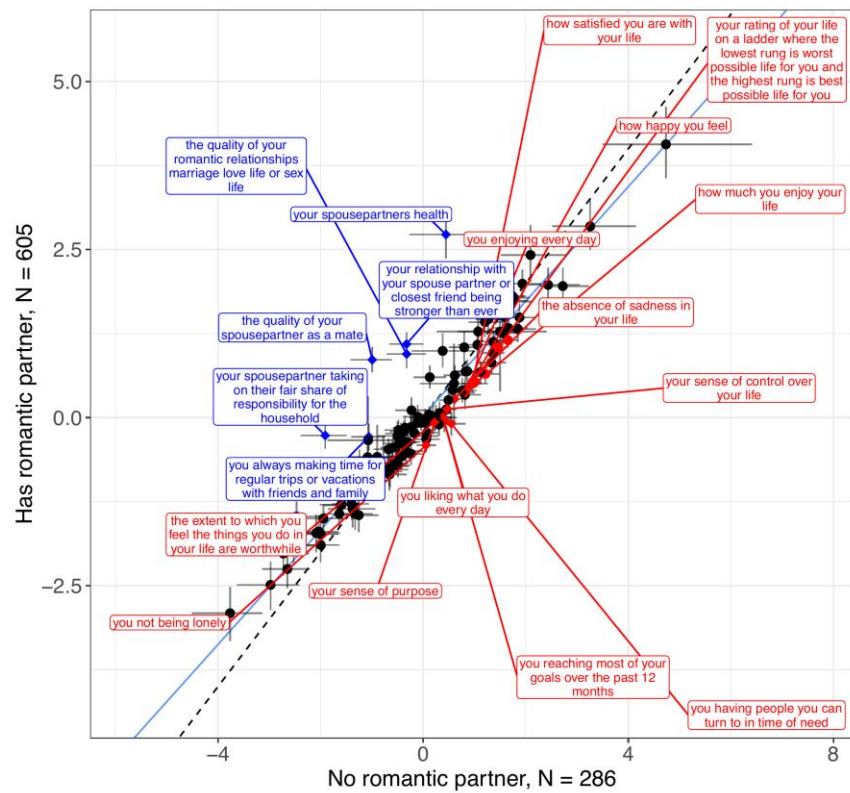
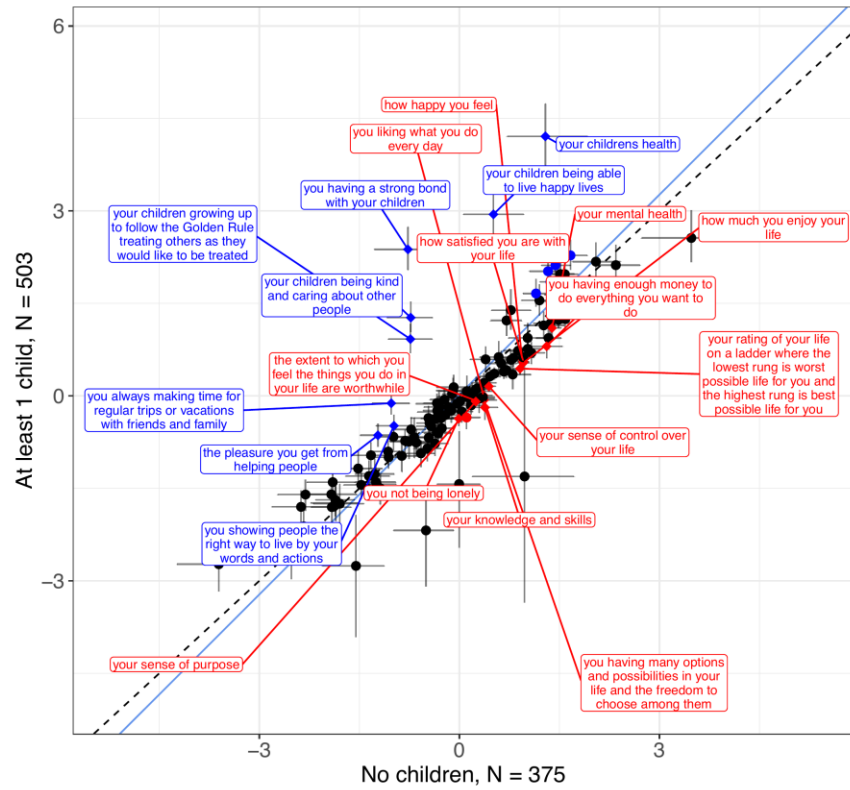
Notes: For each of the 126 aspects, respondents are divided into percentiles based on their self-reported level of the aspect. Each panel displays the mean aspect level against the mean log marginal utility, averaging within the percentile group. Panels are sorted by most negative to most positive slope. Based on 896 respondents who completed the first nine blocks of the Multi-Block Survey and passed quality control. The regression line in the figures and the average slopes reported in the main text are estimated on the individual (not binned) data.

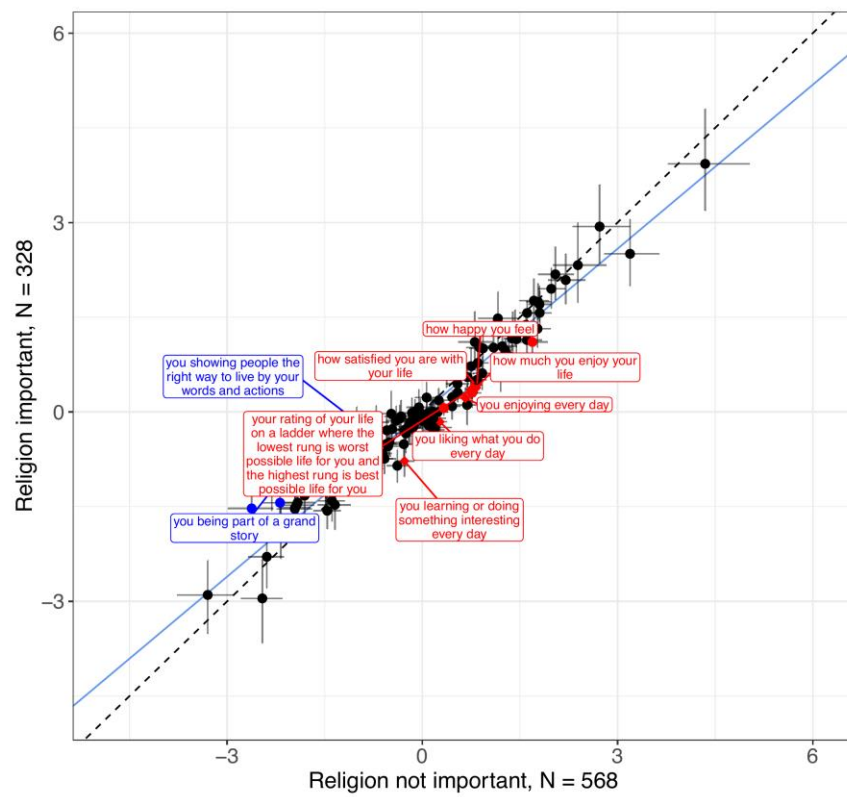
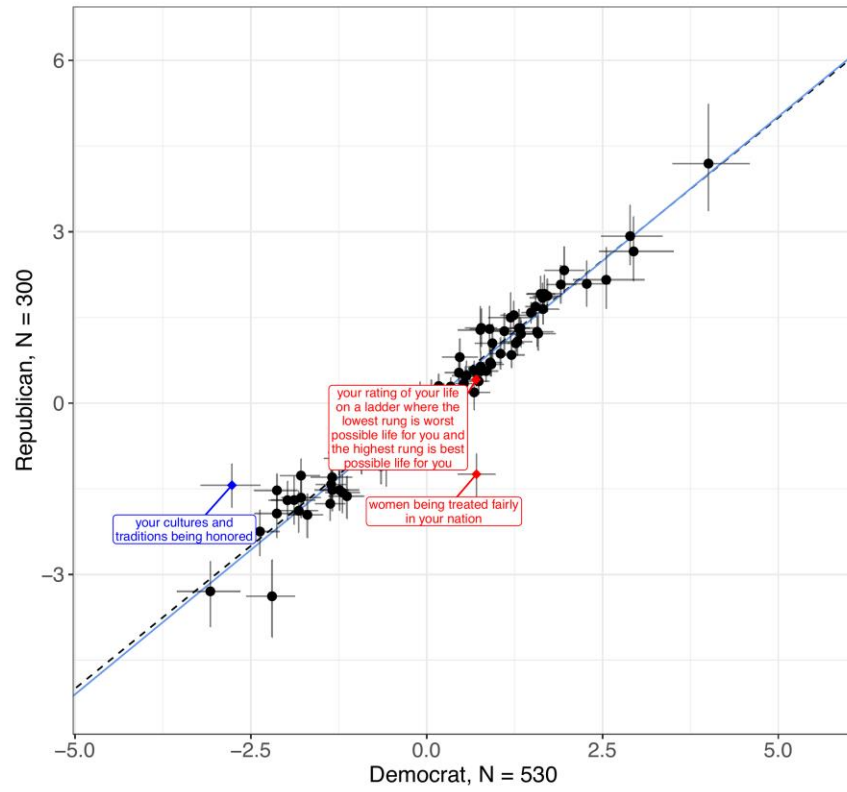
Supplementary Figure 2 – Log MU Averages within Demographic Groups











Notes: Comparison of group-averaged log marginal utilities across subgroup splits, shown in the following order: poorest vs richest income tercile; female vs male; white vs non-white; less than bachelor's degree vs bachelor's degree or more; age 40 or younger vs older than 40; employed full-time vs not employed; no children vs at least one child; no romantic partner vs has romantic partner; Democrat vs Republican; religion less important vs more important. Each point is an aspect, with x- and y-coordinates equal to subgroup-specific posterior means of average log marginal utility from separate hierarchical-model runs (same 896 respondents as in Table 2); error bars show 95% credible intervals. The dashed 45-degree line indicates equality across groups, and the blue line is the total least squares (TLS) fit, which captures the common rescaling/spread relationship between subgroup estimates. Aspects are labeled when they pass a 0.999 confidence-level significance rule based on (i) a difference-in-means statistic that uses each group's posterior mean and posterior SD, assuming independence across subgroup estimates, and (ii) a TLS-distance test: the aspect's signed orthogonal distance from the TLS line must exceed its propagated uncertainty band at the same confidence level (treating the TLS line as fixed). Aspects are colored when passing the difference-in-means test, with blue and red points indicating significantly above and below the 45-degree line, respectively. The main text highlights only the largest absolute subgroup gaps and reports their magnitudes. Estimates shown here are unadjusted for social desirability.