NBER WORKING PAPER SERIES

# WE ARE ALL BEHAVIORAL, MORE OR LESS:
# A TAXONOMY OF CONSUMER DECISION MAKING

Victor Stango
Jonathan Zinman

Working Paper 28138
http://www.nber.org/papers/w28138

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2020

Thanks to Hannah Trachtman and Sucitro Dwijayana Sidharta for outstanding research assistance, and to the Sloan/Sage Working Group on Behavioral Economics and Consumer Finance, the Roybal Center (grant # 3P30AG024962), and the National University of Singapore for funding and patience. We thank Shachar Kariv and Dan Silverman for helping us implement their (with Choi and Muller) interface for measuring choice consistency, Charlie Sprenger for help with choosing the certainty premium elicitation tasks and with adapting the convex time budget tasks, Georg Weizsacker for help in adapting one of the questions we use to measure narrow bracketing, Julian Jamison for advice on measuring ambiguity aversion, Josh Schwartzstein for many conversations, and Doug Staiger and Erik Snowberg for advice re: measurement error. Special thanks to Joanne Yoong for collaboration on the Round 1 survey design and implementation. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research. This paper supersedes three archived working papers that use only a single cross-section of data or focus on behavioral summary statistics; please see https://sites.dartmouth.edu/jzinman/working-papers/ for details.

We are all Behavioral, More or Less: A Taxonomy of Consumer Decision Making
Victor Stango and Jonathan Zinman
NBER Working Paper No. 28138
November 2020
JEL No. C36,C81,D90,E70

## ABSTRACT

We examine how 17 behavioral biases relate to each other, to other decision inputs, and to decision outputs. Most consumers exhibit multiple biases in our nationally representative panel data. There is substantial heterogeneity across consumers, even within similar demographic/skill groups. Biases are positively correlated within person, especially after adjusting for measurement error, and less correlated with other inputs—risk aversion, patience, cognitive skills, and personality traits—with some expected exceptions. Accounting for this correlation structure, we reduce our 29 decision inputs to eight common factors. Seven common factors load on at least two biases, six on clusters of theoretically related biases, and two or three are distinctly behavioral. All but one common factor is distinct from cognitive skills. Factor scores strongly conditionally correlate with decisions and outcomes in various domains. We discuss several potential implications of this taxonomy for various approaches to modeling influences of behavioral biases on decision making.

Victor Stango
Graduate School of Management
University of California, Davis
One Shields Avenue
Davis, CA 95616
vstango@ucdavis.edu

Jonathan Zinman
Department of Economics
Dartmouth College
314 Rockefeller Hall
Hanover, NH 03755
and NBER
jzinman@dartmouth.edu

*"Everything should be as simple as it can be, but no simpler."*

(Attributed to Albert Einstein)

Despite the growing impact of behavioral economics on social science research and applications, little is known about how the many potential behavioral biases fit into a taxonomy of consumer decision making. How common is it for people to exhibit multiple behavioral biases, and how heterogenous is the consumer-level portfolio of biases across consumers? How are biases correlated within-consumer, and how distinct are biases from other inputs to decision making? After accounting for such relationships among decision inputs, how do biases and other inputs correlate with decision outputs in various domains?

We address such questions with evidence, from nationally representative panel data, on the nature and economic importance of multiple behavioral biases. For each of 800+ consumers, at two points in time three years apart, we measure a large set of decision inputs and outputs. Specifically, we measure 17 oft-cited biases using standard lab-style elicitations, conceptually distinguishing a bias as something that can be measured against a clear classical benchmark (e.g., time-consistent discounting, dominance avoidance, mean-zero expectation error). We also measure classical decision inputs such as patience and risk aversion, cognitive skills, and personality traits; demographics; and decision outputs including actual financial decisions and various measures of financial condition and subjective well-being. The panel structure of our data helps us account for measurement error when estimating relationships among and between biases, other decision inputs, demographics, and decision outputs.

Our first finding is that biases are more rule than exception. The median consumer exhibits 10 of 17 potential biases. No one exhibits all 17, but almost everyone exhibits multiple biases; e.g., the 5$^{th}$ percentile is 6.

Our second finding is that cross-consumer heterogeneity in biases is substantial. The standard deviation of the number of biases exhibited is about 20% of its mean, and several results below suggest that this variance is economically meaningful and not substantially inflated by measurement error.

Our third finding is that cross-consumer heterogeneity in biases is poorly explained by even a "kitchen sink" of other consumer characteristics, including classical decision inputs, demographics, and measures of survey effort. Most strikingly, we find more bias variance within classical sub-groups widely thought to proxy for behavioral biases than across them. E.g., we find more bias variation with the highest-education group than across the highest- and lowest-education groups.

Our fourth finding is that our 17 biases are positively correlated with each other within-consumer, especially after accounting for measurement error following Gillen et al. (2019).[1] Across all biases, the average pairwise correlation is 0.13, and 18% have p-values < 0.001. Within six theoretically-related groups of biases (present-biased discounting, inconsistent and/or dominated choices, risk biases, overconfidence, math biases, and limited attention/memory), the average pairwise correlation is 0.25 and 29% have $p < 0.001$.[2]

Our fifth finding is that there are also some important correlations between biases and classical inputs. Classical inputs and demographics may not explain much of the variance in biases (per finding #3), but some of them are correlated with biases in patterns that align with prior work.[3] Most notably, the average pairwise correlation between cognitive skills and biases is -0.25. Cognitive skills are strongly negatively correlated with most biases, but positively correlated with loss aversion and ambiguity aversion.[4] Other classical inputs are relatively weakly correlated with biases, except for a few expected links between patience and present bias,[5] risk aversion and aversion to uncertainty and losses, and risk aversion and math biases that can lead to undervaluation of returns to risk-taking.

---

[1] The measurement error instrumental variables strategy we use here is feasible because biases are temporally stable within-consumer (Stango and Zinman 2020).

[2] See also Dean and Ortoleva (2019) and Chapman et al. (2019a) on relationships among biases re: risk attitudes, and Chapman et al. (2019a) on relationships among overconfidence varieties.

[3] See, e.g., Benjamin et al. (2013), Chapman et al. (2019a), Dean and Ortoleva (2019), and Li et al. (2013). Burks et al. (2009) and Dohmen et al. (2010) are seminal papers in the related literature on relationships among classical characteristics.

[4] Chapman, Snowberg, Wang and Camerer (2019) also find a positive correlation between loss aversion and cognitive ability. The positive correlation between cognitive skills and ambiguity aversion is more surprising in light of prior empirical evidence (see Prokosheva (2016) for a review), although to our knowledge that evidence comes from student populations.

[5] We use "present bias" instead of "present focus" (Bernheim and Taubinsky 2018) for consistency with our labeling of other behavioral inputs to decision making as "biases" that deviate from a normative classical benchmark.

Our sixth finding is that our 29 measures of decision inputs can be reduced to 8 dimensions. We do this using factor analysis (FA), which takes the previous five findings into account. FA facilitates exploration of how factor inputs (our 29 measures) are related through an underlying, latent structure (of "common factors"). If behavioral biases were all connected somehow-- along the lines of the single common factor found for various cognitive skills and thought to reflect "*g*", generalized intelligence (Jensen 1998)-- then FA would "load" all biases onto a single common factor. And if, in turn, biases were simply reflections of *g*, all biases and cognitive skills would load on the same common factor. In practice, cognitive skills load onto a single common factor as expected, but the taxonomy of behavioral biases is multi-dimensional and mostly distinct from cognitive skills.

Our seventh finding is that behavioral biases feature prominently in the common factor structure. Seven of the eight common factors load strongly on at least two biases, two or three of the common factors appear to be distinctly behavioral, and six load on clusters of related biases.

Our eighth finding is that no single decision input, among our 29, has outsized importance in driving the factor structure. No single input measure loads strongly on more than one to three common factors, depending on one's preferred definition of "strong".

Our ninth finding is that the decision input taxonomy suggested by our factor structure is coherent, at least for the top four common factors by eigenvalue, which together explain 45% of the variance in our 29 decision input measures. The first common factor captures cognitive skills and related biases in math and overconfidence. The second captures our three inconsistent and/or dominated choice biases, plus limited memory (suggesting a link from limited memory to limited learning to persistently poor choice quality). The third captures our two present-biased discounting measures and impatience, suggesting a distinct intertemporal choice factor. The fourth captures two of our three risk biases (preference for certainty and loss aversion, but not ambiguity aversion), and the Gambler's Fallacy (which presumably affects perceived returns to risk-taking). It also loads somewhat on our presumed-classical measures of risk aversion, suggesting that this factor is clearly related to risk attitudes.

Our tenth finding is that risk-related inputs appear in several common factors, and in patterns that are not easily explained by or reflected in current models. This adds to other recent evidence suggesting that models of consumer decision making under uncertainty may require enrichment.[6]

Our eleventh finding is that the common factors, which as discussed above are meant to summarize decision inputs, suggest that behavioral biases are strongly conditionally correlated with a range of decision outputs, in expected directions for the most part. Some clusters of biases and related decision inputs are more strongly correlated with outputs than other clusters.

Turning to key takeaways, altogether our results suggest that behavioral biases are commonplace and not limited to certain demographic groups, heterogeneous in the cross-section of consumers, distinct from classical decision inputs in some important ways but linked in others, and economically important. On bias prevalence and variance, we add consumer-level evidence on multiple biases to the body of work estimating population parameters for one or two biases at a time.[7] On relationships between biases and other consumer traits, we add evidence on fit and multiple biases to a "Who is behavioral?" literature that has focused on correlations and piecemeal biases (see footnote 3 for references). On relationships among various biases, we add the additional litmus test of whether related clusters of biases correlate with decision outputs to two other papers examining relationships among a relatively large number of biases and other decision inputs.[8]

Another key takeaway is that myriad potential biases and other decision inputs can be tractably summarized with relatively few parameters.[9] A related point is that the approach we use here—

[6] See, e.g., Chapman et al. (2019a; 2019b); Chapman, Snowberg, Wang, and Camerer (2019); Dean and Ortoleva (2019); O'Donoghue and Somerville (2018).

[7] For reviews see e.g., the Handbook of Behavioral Economics (Bernheim, DellaVigna, and Laibson 2018; 2019).

[8] Dean and Ortoleva (2019) examine correlations among 11 decision inputs, including several behavioral biases, in a student sample. Chapman et al. (2019a) examine relationships among 21 decision inputs, including several behavioral biases, in a nationally representative sample. There are several differences across the papers in decision input coverage; e.g., we focus more on decision quality, limited attention/memory, and personality, they more on social preferences (which we do not consider at all) and richer measurement of attitudes re: uncertainty.

[9] "Relatively few" parameters may still imply "more than customary" for many applications, unless additional evidence linking decision inputs is discovered. The current standard practice in behavioral economics is to consider biases one or two at a time, assuming separability from other potential biases and many other potential decision inputs. (For critiques of this practice and/or of model proliferation in behavioral economics see, e.g., Fudenberg (2006), Levine (2012), Koszegi (2014), Bernheim (2016), Ghisellini and Chang (2018).) Our results— including our correlation matrix—provide a reference for

describing the prevalence and taxonomy of model inputs, including their links to real-world outputs—has been productively applied in many other fields. In physics, research since the 1930s has integrated its "Particle Zoo", and three of the four known fundamental forces in the universe, into the "Standard Model". In social science, a century of painstaking research has reduced the seemingly countless intelligences of Galton and Spearman's era to low-dimensional constructs like fluid and crystalized intelligence, and myriad personality attributes to low-dimensional constructs like the Big Five. Even chemistry, with its 112 known elements in the periodic table, has relied on discovering relationships *among* the elements to develop tractable models.[10] We expect that the relationships uncovered here will help behavioral economics along a similarly productive track.

## 1. Research design

In this section we describe our sample, survey design and elicitation methods for measuring behavioral biases, other decision inputs and demographics. We postpone providing details on measurement of decision outputs— financial decisions, financial condition, and subjective well-being—until we first use those measures, in Section 4-C.

### *A. The American Life Panel*

We administered our surveys through the RAND American Life Panel (ALP). The ALP is an online survey panel that was established, in collaboration between RAND and the University of Michigan, to study methodological issues of Internet interviewing. Since its inception in 2003, the ALP has expanded to approximately 6,000 members aged 18 and older. The ALP takes great pains to obtain a nationally representative sample, combining standard sampling techniques with offers of hardware and a broadband connection to potential participants who lack adequate Internet access. ALP sampling weights match the distribution of age, sex, ethnicity, and income to the Current Population Survey.[11]

assessing such separability assumptions. For other approaches and discussions re: relationships among behavioral biases (and other decision inputs) and the modeling challenges they present, see, e.g., Benjamin et al. (2016), Chetty (2015), Ericson (2017), Heidhues et al. (2018), Mullainathan et al. (2012).

[10] E.g., Scerri (2011) states: "Fortunately, the periodic table allows chemists to function by mastering the properties of a handful of typical elements…."

[11] We report unweighted results and find similar results using the ALP's population weights.

*B. Research design and sample*

Three principles guided our research design. First, measure the richest set of individual characteristics possible, to minimize potential confounds from omitted variables and to allow exploration of relationships between behavioral biases and classical covariates such as demographics, cognitive skills, personality traits and standard measures of patience and risk aversion. Second, use standard elicitations and survey questions wherever possible, although in many cases we shorten lab-style elicitations to free up time and budget for measuring a broader set of characteristics. Third, take repeated measurements at different points in time, to describe the temporal stability of behavioral biases (Stango and Zinman 2020) and to help account for measurement error in biases and other decision inputs (Gillen, Snowberg, and Yariv 2019).

To those ends, we designed a "round" of about an hour's worth of elicitations and survey questions. We then split this round into two modules designed to be administered roughly two weeks apart, to minimize survey fatigue and per the ALP's advice re: module length and invitation sequencing. After extensive piloting and refinements, we had the ALP field our two Round 1 modules starting in November 2014. We targeted 1,500 working-age respondents, sending 2,103 initial invitations, and ultimately received 1,515 responses to Round 1 Module 1 (ALP #315), with 1,427 subsequently completing Round 1 Module 2 (ALP #352). 95% of respondents completing both modules did so by the end of February 2015.

We undertook Round 2 in October 2017, by inviting the universe of 1,308 panelists who completed both Round 1 modules and remained empaneled to take ALP #474, which is a replica of Round 1 Module 1 (but was not advertised as such). We received 967 responses and then invited those panelists to take the second module. ALP #472 is a replica of Round 1 Module 2 with some additional questions added at the end (but was not advertised as such). We received 845 responses to this second module, creating our sample of 845 panelists who responded to both modules in both rounds.

In refining our elicitations for each potential bias source and other consumer characteristics, we were mindful of the reality that research budgets force tradeoffs between the depth and breadth of measurements, incentives, and sample size. Per standard ALP practice, we paid panelists $10 per completed module. Beyond that, all but one of our elicitations are unincentivized on the margin (limited prospective memory being the exception; see Table 1 for details). Scrutiny of usual motivations for paying marginal incentives casts doubt on their value, given our research objectives, relative to spending research funds on measuring a broader set of consumer characteristics, on a broader sample, at multiple points in time. Researchers often hypothesize that subjects find stylized tasks unpleasant and hence need marginal incentives to engage with the tasks, but the ALP measures panelist engagement and finds evidence to the contrary.[12] Researchers often hypothesize that unincentivized elicitations change inferences, but that hypothesis is not robustly supported empirically (e.g., Von Gaudecker, Van Soest, and Wengström 2011; Gneezy, Imas, and List 2015; Branas-Garza et al. 2020) and there is a long tradition of using unincentivized lab-style elicitations in surveys (e.g., Barsky et al. 1997; Falk et al. 2018; Bauer, Chytilová, and Miguel 2020). Researchers often assume that marginal incentive mechanisms are the best way to mimic real-world stakes, but this is not generally true for behavioral consumers (Azrieli, Chambers, and Healy 2018), and tasks with hypothetical rewards like ours can offer some conceptual advantages (e.g., Montiel Olea and Strzalecki 2014). In any case, our repeated elicitations and measurement error models should suffice to address concerns about noise.

### C. *Measures of decision inputs: Biases, patience and risk aversion, cognitive skills, personality*

We measure 17 potential sources of behavioral biases and 12 “classical” decision inputs: three standard measures of patience and risk aversion that we label “presumed-classical” because they do not allow for straightforward identification of biased tendencies, four standard measures of cognitive skills, and standard measures of the “Big Five” personality traits. Tables 1 and Table 2 summarize these 29 constructs and their measures. The Data Appendix provides additional details on the data, elicitation methods, and their theoretical and methodological antecedents.

As Table 1 makes evident, a key conceptual distinction between behavioral biases and other decision inputs is that biases are defined relative to a clear, normative classical benchmark (e.g., time-consistent discounting, accurate perceptions of one’s own skills). In contrast, it would be more difficult to define someone as behavioral or not based on, e.g., their level of patience or cognitive skills: where is the cutoff?

[12] For example, each ALP survey ends with “Could you tell us how interesting or uninteresting you found the questions in this interview?” and roughly 90% of our sample replies that our modules are “Very interesting” or “Interesting”, with only 3% replying “Uninteresting” or “Very uninteresting,” and 7% “Neither interesting nor uninteresting”.

We measure biases that have featured prominently in behavioral economics, are amenable to compact elicitation in online surveys, and are easily subsetted into one of six "related" groups based on priors. Two relate to *present-bias*, over consumption and money discounting (Read and van Leeuwen 1998; Andreoni and Sprenger 2012). Three relate to *inconsistent and/or dominated choices:* inconsistency with GARP, the General Axiom of Revealed Preference (Choi et al. 2014) measured two ways, and narrow bracketing (Rabin and Weizsäcker 2009).[13] Another three relate to *risk biases* (re: preferences or attitudes toward uncertainty): loss aversion or small-stakes risk aversion (Fehr and Goette 2007), preference for certainty (Callen et al. 2014) and ambiguity aversion (Dimmock et al. 2016). Another three measure varieties of *overconfidence* (Moore and Healy 2008): in level performance, in the precision of one's expectations, and in performance relative to peers. Four sources of *math biases* include two statistical fallacies—gambler's and non-belief in the law of large numbers (Dohmen et al. 2009; D. Benjamin, Moore, and Rabin 2017; D. Benjamin, Rabin, and Raymond 2016)—and exponential growth biases over borrowing costs and investment returns (Stango and Zinman 2009; Levy and Tasoff 2016). Finally, we elicit two measures of *limited attention and limited memory*, in ways that are motivated by behavioral models of inattention (K. Ericson 2011; Bronchetti et al. 2020).[14]

Many potential bias sources are bi-directional: one can be present- or future-biased, one can underestimate or overestimate future values, and so on. Here, including both directions for a given source adds little, since our focus is on correlations among different biases, and between biases and other decision inputs, and bi-directional biases are mechanically correlated with their flipsides. E.g., if one direction appears in a common factor, its other direction does too, with essentially the opposite weight. We therefore focus on just one direction for each potential bias source, choosing the one that is more prominent in theory or more common empirically.[15]

Our bias data readily yields measures of the level of each bias, on the extensive margin: biased or not. Measuring the extensive margin is feasible in a wide range of settings because it does not

[13] These and some of our other biases relate to "deviations from rational thinking" studied by other social scientists. Those literatures, like economics, are only beginning to grapple with relationships among decision inputs (e.g., Stanovich 2016).
[14] Following a common delineation in behavioral economics, we do not measure social preferences. See Dean and Ortoleva (2019) and Chapman et al. (2019a) for evidence on relationships between behavioral biases and social preferences.
[15] We examine bi-directional biases in more detail in Stango and Zinman (2020).

require an elaborate elicitation or granular bias measures. It is desirable because bias indicators are unit-free in cross-bias comparisons and map well into models with a mix of biased and unbiased agents.

Our bias data also readily yields at least somewhat granular measures of intensity for 15 of the 17 biases (Table 1 Column 4). We focus on the respondent's cross-sectional percentile rank,[16] measured as the percentile into which the respondent falls for that bias and survey round.[17] Rank has several conceptual and practical advantages as a metric for studying individual differences (cross-consumer heterogeneity), as we discuss in Section 3-A. Our median bias has 11 distinct ranks.

Turning to classical decision inputs, we measure risk aversion with the adaptive lifetime income gamble task developed by Barsky et al. (1997) and the financial risk-taking scale from Dohmen et al. (2010; 2011), and patience using the average savings rate across the 24 choices in our version of the Convex Time Budget task (Andreoni and Sprenger 2012). We measure cognitive skills using standard tests for general/fluid intelligence (McArdle, Fisher, and Kadlec 2007), numeracy (Banks and Oldfield 2007), financial literacy (crystalized intelligence for financial decision making) per Lusardi and Mitchell (2014), and executive function/working memory (MacLeod 1991). In our second round of surveying, we add elicitations of the Big Five personality traits to the end of our second module (Rammstedt and John 2007).[18]

*D. Demographics, response times*

We also consider standard demographic variables in some of our analysis (see Appendix Table 1 for the complete list). The ALP elicits these characteristics-- gender, age, race/ethnicity, education, etc.-- when a panelist first registers, and then refreshes them quarterly.

We use demographics selectively when describing consumer-level heterogeneity in behavioral biases in Section 2, and more comprehensively as control variables when estimating links between

[16] There are various ways of measuring rank, and they are very highly correlated with each other in our data.
[17] If the panelist is unbiased, we set the percentile is a "0" for that bias/round. For biases with less-granular measures, the percentiles simply take on values corresponding to the cumulative frequencies of each value.
[18] We initially decided against eliciting personality measures, given our resource constraints and the lack of prior evidence of correlations between them and behavioral biases (see, e.g., Becker et al.'s (2012) review article).

our estimated decision input factor structure and outcomes in Section 4. We do not include demographics in most of our analyses of relationships among decision inputs. Practically, omitting demographics conserves space in our already dense tables describing correlations among decision inputs (see especially Table 3 and Table 4). Conceptually, demographics tend to be more readily observable than our other decision inputs; accounting for measurement error is more of a concern for the latter. Also, some key demographics like education and income are plausibly endogenous—more output than input—in many contexts.[19]

The ALP also tracks and records survey response time, screen-to-screen, and we use this to construct flexibly parameterized measures of survey effort for inclusion in "kitchen sink" specifications using a relatively comprehensive set of consumer covariates to try to explain bias variance (Section 2-B) or as control variables when estimating links between decision outputs and inputs (Section 4-A).

## 2. Prevalence and heterogeneity of multiple behavioral biases

### *A. Summarizing biases at the consumer level*

Is exhibiting multiple behavioral biases common or anomalous? How biased are consumers, are how much heterogeneity in biases is there across consumers? We address these questions using a simple, consumer-level summary statistic that counts the number of biases exhibited on the extensive margin: the "B-index".

Table 2 Panel A shows that the exhibiting many biases is the rule, not the exception. The mean B-index is 10, whether we use all round 1 data (i.e., panelists who completed both of our round 1 modules; N=1427), round 1 data only for panelists who went on to complete round 2 (N=845), or round 2 data (N=845). The median, not shown in the table, also equals 10 in each case. Everyone exhibits at least one bias (Column 4), although no one exhibits the maximum possible 17 (Column 5). Panel B suggests that item non-response is rare and does not materially complicate interpretation of B-index variation.[20]

---

[19] For some applications, like life-cycle modeling, it may be informative to include demographics in factor structure estimation in future work.

[20] On average, only about 1 out of maximum possible 17 biases is missing due to non-response (Column 1), with a standard deviation of about 1.5. Every panelist completes at least five of the bias elicitations (Column 5).

Column 2 suggests that cross-consumer heterogeneity is the number of biases exhibited is substantial. The B-index standard deviation is about 20% of its mean. One might wonder if these variance estimates are substantially upward-biased by measurement error, but Figures 1a and 1b provide some reassurance: comparing the figures shows that dispersion in the Full B-index for our full sample is only modestly greater than for the sub-sample with identical B-indexes across our two rounds. Below we develop several additional pieces of evidence that cross-consumer heterogeneity in multiple behavioral biases is important, and largely distinct from heterogeneity in other decision inputs.

The key findings from this sub-section are that almost everyone exhibits multiple biases, with substantial heterogeneity across people in how many.

*B. Biases are not simply functions of other consumer characteristics*

Are biases meaningfully distinct from other consumer characteristics?

One metric of distinctness starts with correlations; before undertaking that analysis below, we start by examining how well other characteristics *fit* our bias measures. This exercise adds to the "Who is (more) behavioral" literature (e.g., D. Benjamin, Brown, and Shapiro 2013), by adding evidence on fit to the prior focus on correlations, and by adding evidence based on consumer-level metrics of behavioral tendencies to a literature that has considered behavioral biases piecemeal.

Figure 2a plots raw, consumer-level variation in the "B-proportion": the share of possible biases a consumer displays. Using the proportion instead of the level B-index accounts for missingness without overfitting. Figure 2b plots consumer-level residuals from regressing the B-proportion on the "kitchen sink" of other decision inputs (Table 1), demographics and other covariates in our data (Appendix Table 1). These residuals are rescaled to the mean of the raw B-proportion in Figure 2a for comparability. Comparing the figures illustrates how little variation in the B-index is explained by our complete set of other covariates; although partialing out variation explained by other covariates does produce a more normal B-index distribution, it does little to reduce dispersion. E.g., the raw vs. residualized interquartile ranges are [0.53, 0.71] vs. [0.56, 0.69].

Figures 3a-3d provide some simple univariate comparisons further highlighting that the B-index does not simply proxy for other covariates found to correlate with behavioral biases.[21] These show distributions of the B-proportion, broken out for paired groups at the opposite ends of the income, risk aversion, education, and cognitive skills distributions. These do show the expected level differences on average; e.g., the B-index distribution is shifted rightward for those in the lowest cognitive skills quartile relative to the highest. But also noteworthy is that the B-index varies substantially <u>within</u> each of the sub-groups we examine. Indeed, within-group variation in the B-index dwarfs cross-group variation, even between groups that are very different by construction.

Appendix Table 2 Column 1 formalizes the inference that other consumer characteristics explain relatively little of the variance in biases, by showing R-squareds and partial R-squareds from OLS regressions of the B-proportion on demographics, cognitive skills, non-cognitive skills, traditional measures of preferences, and measures of survey effort.[22] Appendix Table 3 shows similarly poor fit for each of the 17 bias ranks described in Table 1.

The key finding from this sub-section is that biases are largely distinct consumer traits, in the related senses that biases are not limited to particular demographic or skill groups, and that cross-consumer variance in bias is poorly explained by classical decision inputs and other characteristics.

## 3. Relationships among behavioral biases, and between biases and other characteristics

### *A. Estimating rank correlations that account for measurement error*

We start by estimating pairwise rank correlations among and between the biases, cognitive skills, personality traits, and traditional measures of risk and time preferences described above. These correlations are interesting descriptively in their own right, and they also serve as inputs to our factor analyses starting in Section 3-D.

---

[21] See, e.g., Benjamin et al. (2013), Burks et al. (2009), Cesarini et al. (2012), Chapman et al. (2019a), Dean and Ortoleva (2019), Frederick (2005), Li et al. (2013). See also Dohmen et al. (2018) on the relationship between measures of presumed-classical preferences/attitudes and cognitive skills.

[22] Appendix Table Column 2 repeats the exercise for a sub-index counting just the math/statistical biases, since one might predict that cognitive skills and demographics (including education and income) would explain more of the variation in these biases. That prediction is accurate for demographics, but the increase is only 4pp.

Our decision making inputs of interest are latent characteristics, and so accounting for measurement error from elicitations of them is essential for obtaining unbiased estimates of correlations and common factors (Fuller 2009). The panel structure of our data allows us to use a measurement error instrumental variables strategy that exploits the temporal stability of biases and other characteristics (Stango and Zinman 2020). Specifically, we use univariate Obviously Related Instrumental Variables (ORIV) regressions to estimate pairwise correlations, following Gillen (2019). These instrument for the consumer's Round 2 value of the characteristic with the consumer's Round 1 value of the characteristic and vice versa, for each characteristic in the pair, inflating the two observations we have per consumer to four replicates and clustering standard errors at the consumer (i.e., the panelist) level.[23]

For the rest of our analyses, we scale each variable as a percentile rank unless noted otherwise,[24] because rank offers both conceptual and practical advantages relative to alternative parametrizations.[25] Unlike binary measures (e.g., biased vs. unbiased, or high vs. low intelligence), rank captures both extensive and intensive margins, and its relative smoothness makes it more amenable to successful measurement error IV than binary variables.[26] Unlike structural parameter estimates, rank is conceptually defined and practically measurable for each characteristic in our data, with comparable units across characteristics.

*B. Correlations among biases*

Table 3 shows our ORIV estimates of all pairwise correlations between our 29 decision input measures, and Table 4 summarizes key patterns.

---

[23] Thanks to Erik Snowberg for providing the Stata code "ORIVcorrelation.do".

[24] Some of our variables are continuous, permitting percentiles to take on the full range of values from 1 to 100. For discrete-response variables, the percentiles take on fewer values but still measure where a panelist stands in the distribution relative to others. [24] For example, loss aversion takes on four values: unbiased, and then three ordered responses (whether the individual respondent rejects a compound but not a single lottery, rejects a single but not a compound lottery, or rejects both) coded as 1/2/3. Any respondent accepting both lotteries receives a 0 (meets the classical benchmark), and 37% of individuals share that response. Anyone with the smallest deviation from the benchmark therefore is in the $37^{th}$ percentile, and 13% of responses fall into that category. Summing, anyone in the next category is in the $50^{th}(=37^{th}+13^{th})$, and so on

[25] Indeed, psychometricians tend to prefer rank when studying individual differences (e.g., Schildberg-Hörisch 2018).

[26] Binary variables are prone to non-classical misclassification error (e.g., Black, Berger, and Scott 2000).

Starting with correlations among all biases, the first row of Table 4 Panel A shows that biases tend to be correlated with each other: 50% and 22% of the bias pairwise estimates are significantly different from zero using p-value cutoffs of 0.10 or 0.001. Moreover, these correlations tend to be positive: being (more) biased on one dimension makes you more likely to be (more) biased on another dimension, as 37% (18%) are positive with $p$ <=0.100 (<=0.001). The negative correlations tend to involve risk biases, with loss aversion accounting for largest magnitudes and the only negative correlations with $p$ <=0.001. The average correlation across all bias pairs is 0.13 (Panel B).

Focusing on correlations within the six groups of related biases defined in Section 1-C, Table 4 Panel A shows that these are relatively strong statistically, with 53% and 29% of these 17 pairwise correlations having $p$ <=0.100 and <=0.001. The second column of Table 4 Panel B summarize within-group correlation magnitudes, with the average being 0.25 (vs. 0.11 between biases in other groups). By way of comparison to other characteristics that have been theoretically and empirically grouped in prior work, the average correlation among our two classical risk aversion measures is 0.29, and among our four cognitive skills measures is 0.58. Among the bias groups, the two discounting biases are particularly strongly correlated with each other, and only the preferences for uncertainty group does not exhibit robustly positive and larger within-group correlations. This foreshadows our factor structure findings in Sections 3-D and 4-C that related biases tend to cluster together, except for preference for certainty, loss aversion, and ambiguity aversion. Those biases appear in different common factors that also have different correlations with outcomes.

The key findings in this sub-section are that biases tend to be positively correlated with each other, with risk biases (principally loss aversion) a noteworthy exception. Correlations are substantially stronger within subsets of related biases grouped by theory.

*C. Correlations between biases and other characteristics*

Turning to correlations between biases and other characteristics, the results in Tables 3 and 4 mostly align with prior work. Patience is negatively correlated with the discounting biases (e.g., Dean and Ortoleva 2019), and with limited attention (Gabaix 2019), but is mostly uncorrelated with other biases. Risk aversion is positively correlated with loss aversion (e.g., Chapman et al. 2019a; Mrkva et al. 2020) and ambiguity aversion (e.g., Dean and Ortoleva 2019), and with statistical biases that distort perceived returns to risk-taking (e.g., D. Benjamin, Rabin, and Raymond 2016; Rabin and Vayanos 2010; Levy and Tasoff 2016), but mostly uncorrelated with other biases. Cognitive skills are strongly negatively correlated with most biases (see footnote 3 for references), but positively correlated with loss aversion (Chapman, et al. 2019) and ambiguity aversion. Personality traits are weakly correlated with biases overall (Becker et al. 2012): summarizing the bottom-left portion of Table 3, the average point estimate between biases and each of the Big Five ranges from 0.00 to 0.05 (and |0.04| to |0.09|), and only two of these 85 pairwise point estimates between biases and the personality measures have $p$ <0.001 (Table 4 Panel A).

The key findings in this sub-section are that classical inputs are only weakly correlated with biases overall, with some important exceptions that align with prior work.

*D. Common factors and their loadings*

To identify structure underlying our set of biases and other decision inputs, we use our correlation matrix to extract a set of common factors. Selecting the number of common factors to retain is as much art as science, but we settle on eight. Each of these eight have eigenvalues strictly greater than one, while the other common factors explain at most barely more variation in the data than would a single variable.[27] Furthermore, the difference between the eigenvalues is greater between common factors 8 and 9 than for any number beyond 6, and so our choice fits with a Scree approach that looks for breaks in eigenvalue levels. Using Scree could instead justify retaining one or five factors. But results below support erring on the side of expansiveness, with the eight-factor model loadings showing theoretical appeal (Table 5) and some of the lower-eigenvalue factors having strong links to decision outputs (Table 6).[28]

---

[27] Appendix Table 4 shows eigenvalues for 20 common factors that collectively explain 96 percent of variation in our 29 measures of decision inputs.
[28] We have estimated models where we retain five or seven factors instead of eight, and those deliver results similar to the eight-factor model.

Table 5 shows the rotated factor loadings from our 8-factor model.[29] As is customary, we order factors from the highest eigenvalue and most variance explained ("Factor 1") to lowest and least ("Factor 8").

Factor 1 loads most strongly on the cognitive skills variables (negatively), and on a majority of biases from each of two groupings: overconfidence and math biases. The link between low cognitive skills and math biases is quite intuitive, since math skills are a component of cognitive ability, and greater math skills might de-bias the statistical and price perception biases in our "math bias" grouping.[30] The link between low cognitive skills and overconfidence has been documented in several studies (largely outside of economics, to our knowledge).[31] We worried that the link between overconfidence and math biases could be mechanical, but robustness checks suggest otherwise.[32]

Factor 2 loads most strongly on the three decision quality biases, and limited memory (limited attention also loads positively, although relatively weakly). To us this suggests a link between limited memory and limited learning that allows low decision quality to persist. Notably, this common factor is quite distinct from-- does not load on-- other decision inputs that one might conjecture are linked to inconsistent, dominated, or otherwise inattentive choices: cognitive skills, math biases, patience, and conscientiousness.

Factor 3 loads most strongly on the two present biases and impatience. This common factor seems to cover intertemporal choice and nothing else.

Factor 4 is clearly risk-related, but in only a partly intuitive way. It loads most strongly on two of the three risk biases, and on one of the math biases that affects risk perceptions (Cold Hand bias). It also loads, albeit less strongly, on the two presumed-classical risk aversion measures.

---

[29] Loadings scale the importance of each variable in the common factor, between -1 and 1. We use a promax rotation for this table; promax is an oblique rotation, which allows for correlation between factors. We have used other rotations, including orthogonal rotations such as varimax, with indiscernible effects on the results.
[30] This conjecture is consistent with, e.g., the negative correlations between exponential growth bias and cognitive skills found in Almenberg and Gerdes (2012) and Goda et al. (2019).
[31] See, e.g., Ehrlinger et al. (2008).
[32] Specifically, our concern was that the two overconfidence measures with strong loadings here are based on panelists' self-assessments of two of their cognitive skill measures. However, we find similar loadings even when we exclude those two components of cognitive skills from the factor analysis, suggesting a genuine link between cognitive skills and overconfidence.

Perhaps most strikingly, the loads on the two risk-related biases are in opposite directions (positive on preference for certainty, negative on loss aversion), and the load on ambiguity aversion is approximately zero. This adds to recent evidence, from various approaches, motivating a rethinking of how to best specify models of decision making under uncertainty (see footnote 6 for references).

Factor 5 loads most strongly on four biases and two classical factors: present-biased consumption, limited memory and attention, lack of overconfidence in precision, lack of conscientiousness, and risk tolerance at large stakes. The relationship between the attention biases, lack of conscientiousness, and present-biased consumption is at least somewhat consistent with some theories of myopia and present-focus (K. M. Ericson and Laibson 2019; Gabaix 2019), although the lack of an empirical link with patience here is complicating.

Factor 6 loads most strongly on two personality traits (extraversion +, neuroticism -) and financial risk tolerance. There are also loadings above 0.20 on two other risk-related inputs: lack of loss aversion, and risk tolerance at large stakes. It may be that Factor 6 is another risk-related factor, along with Factor 3 and Factor 7, which loads most strongly on ambiguity aversion and financial risk aversion. As we shall see, these different common factors have different relationships to decision outputs as well. Factor 8 is the most difficult to parse. Relative to the other common factors, it does not load particularly strongly on any decision input: its largest load magnitude is |0.54|, whereas each of the other common factors has at least one input loading >=|0.56|, and six of the other common factors has at least one loading >= |0.76|. Factor 8 loads most strongly on an ungainly mix of four biases and one personality trait (Openness).

We glean four key findings from this pattern of common factor loadings.

First, behavioral biases feature prominently in the factor structure. Seven of the eight common factors (all but Factor 6) load strongly on at least two biases. Two or three of the common factors appear to be distinctly behavioral (Factors 2 and 4, possibly also Factor 3 if one views impatience as resulting from present-bias). Six of the common factors load on clusters of related biases (all but Factors 6 and 7). Overall, we estimate 17x8=136 loadings of biases on common factors; 9% of them exceed 0.6 and 21% exceed 0.35. The comparable results for presumed-classical patience and risk aversion are 4% and 10%, for cognitive skills 13% and 13%, and for personality traits 2% and 10%.

Second, no single decision input across our 29 looms especially large. None of them loads more than 0.60 on more than one common factor, and only one—present-biased consumption—loads more than 0.35 on more than two common factors. Cognitive skills load together very strongly on the powerful first factor, but so do several biases.

Third, the decision input taxonomy suggested by our factor structure is coherent, at least for the top four common factors by eigenvalue, which together explain 45% of the variance in our 29 decision input measures. The cognitive skills variables load together and strongly on a single common factor, as one might expect based on a century of research on generalized intelligence (Jensen 1998). That common factor also loads strongly on biases previously linked to cognitive skills: statistical fallacies and exponential growth bias, and overconfidence. (Perhaps smarts affect economic decision making primarily by debiasing perceptions of returns to risk-taking and saving.) The second common factor captures something distinct about decision quality, loading strongly only on our three inconsistent and/or dominated choice biases, plus limited memory. (Perhaps limited memory is linked to limited learning that allows poor choice quality to persist.) The third common factor captures something distinct about intertemporal choice, loading strongly only on our two present-biased discounting measures and impatience. (Perhaps this factor isolates the time preference component of discounting, with expectations of returns, etc. appearing in several other factors.) The fourth common factor relates to something distinct about risk attitudes, loading strongly only on two of our three risk biases and the Gambler's Fallacy (which presumably affects perceived returns to risk-taking). It also loads somewhat on our presumed-classical measures of risk aversion.

Fourth, risk-related inputs appear in several common factors, and in patterns that are not easily explained by or reflected in current models. This adds to other recent evidence suggesting that models of consumer decision making under uncertainty may require enrichment.

## 4. Correlations between decision input common factors and outcomes

We now examine correlations between our common factors and various outcomes: objective financial condition, subjective financial well-being, overall life satisfaction, happiness and health status. These results are interesting descriptively, and provide additional litmus tests for which factors researchers should focus on, or can safely ignore, when modeling relationships between decision inputs and outputs for a given outcome domain.

## A. *Specification and variable construction*

We investigate links between common factors and outcomes by estimating models of the form:

$$(1)\ Outcome_{it} = f\left(CF_{it}^{j}, X_{it}\right) + \varepsilon_{it}$$

where *Outcome* is one of the measures detailed in Section 4-C for each panelist *i* and survey round *t*. We correlate each outcome with factor scores $CF_{it}^{j}$ , where *j* indexes our eight common factors, as well as a set of demographic and other controls $X_{it}$ (see Appendix Table 1 for details).[33]

For each factor *j*, the score *CF* is just the sum of the value of each of the 29 decision input measures indexed by *k*, $c_{ikt}$ (recall that each is scaled as a percentile rank), weighted by its estimated loading $\omega_{jk}$, for each panelist and survey date:

$$CF_{it}^{j} = \sum\nolimits_{jk} \omega_{jk} c_{ikt}$$

## B. *Econometric issues*

There are three potential obstacles to obtaining unbiased estimates of *f(.)* in equation (1).

The first is measurement error, because the factor scores use not only the loadings (which are derived from the measurement error IV estimates of the correlation matrix in Table 3) but also the values $c_{it}$ of each characteristic *k*-- which are themselves measured with error. To address this we exploit the fact that we have multiple elicitations of each characteristic, one in each survey round, and therefore multiple estimates of the factor score for each panelist.[34] These factor scores are temporally stable within-panelist across rounds and thus one round's factor score can instrument for the other (Fuller 2009).[35] The standard ME-IV approach to such estimation would two "single IV" models for estimation, where for each consumer we have:

---

[33] In principle, one could estimate the factor structure and the links between factors and outcomes simultaneously, but in practice we faced convergence problems in any structural model we specified.
[34] Exceptions are our five personality traits, which we elicit only once.
[35] Across the 8 common factors, the mean cross-round within-panelist factor score correlation is 0.44, with a minimum of 0.18 and a maximum of 0.77. Only one correlation falls below 0.30.

| Round | Outcome | Covariates | Factor score | Factor score IV |
|---|---|---|---|---|
| 1 | Round 1 | Round 1 | Round 1 | Round 2 |
| 2 | Round 2 | Round 2 | Round 2 | Round 1 |

We go beyond the single IV approach by implementing the "both-ways" approach of Obviously Related Instrumental Variables (Gillen, Snowberg, and Yariv 2019). ORIV stacks the data, using both the first elicitation to instrument for the second *and* the second elicitation to instrument for the first. Accordingly we inflate our two observations per person to four "replicates" (per Gillen et al.):

| Replicate | Outcome | Covariates | Factor score | Factor score IV |
|---|---|---|---|---|
| 1 | Round 1 | Round 1 | Round 1 | Round 2 |
| 2 | Round 1 | Round 1 | Round 2 | Round 1 |
| 3 | Round 2 | Round 2 | Round 2 | Round 1 |
| 4 | Round 2 | Round 2 | Round 1 | Round 2 |

As with the single IV approach, ORIV will produce an unbiased estimate of the partial correlation between a common factor and outcome and *Y* if the measurement errors in the common factors are uncorrelated across rounds. [36] We cluster standard errors by panelist.

The second potential obstacle to obtaining unbiased estimates of *f(.)* is collinearity, as each common factor is a linear transformation of the same underlying 29 variables. To mitigate this, in our main specification, we exclude variables with a loading below |0.35| from each factor score (see Table 5 for loadings). The results are similar if we include variables with a weight>=|0.20|, or all variables regardless of weight, in the factor scores, although these alternative specifications tend to produce less precise estimates.

[36] This model also imposes a coefficient restriction. We first estimate the model separately for replicates 1 and 2 ("round 1 ORIV") and compare those estimates to those obtained using replicates 3 and 4 ("round 2 ORIV"). We do not reject the restriction that the empirical relationships are identical for round 1 ORIV and round 2 ORIV.

A third potential obstacle is over-controlling. Some potential control variables in *X*, like income and education, might themselves be functions of the common factors in a model like (1). If that is the case, then including those endogenous variables in *X* risks mis-estimating the links between common factors and the outcomes of interest. Different researchers will take different stances on the optimal specification—e.g., on whether variables like income and education should be included in *X*, in the factor structure (see Section 1-D), or not at all—depending on priors and data constraints. As such, we present results for three different specifications of *X*. Our "parsimonious" specification, presented in Table 6, includes only plausibly exogenous demographics: age, gender, state of residence, immigrant status, and race and ethnicity. Appendix Table 5 presents results for our "no controls" and "kitchen sink" specifications, with the latter including income, education, marital status, household size, employment status and survey time variables along with exogenous demographics.

*C. Results*

Table 6 shows the coefficients between each of the eight factor scores, from our main specification, and each of six outcomes: one ORIV regression per outcome. We normalize every factor score, so the coefficients are the marginal effect of a one-sd change in the factor score (in the cross-section of individuals) on the outcome.

We provide details on outcome measurement in Data Appendix Section 3 and summarize these measures here. We scale all outcomes on the [0,1] interval, with higher values indicating better outcomes. The first outcome (Column 1) is an index of five positively correlated measures of objective financial condition: net worth, retirement assets, stockholding, recent saving, and lack of recent severe hardship. The second outcome (Column 2) is an index of four positively correlated measures of subjective financial condition: financial satisfaction, savings adequacy for retirement and other purposes, and lack of financial stress. The third and fourth outcomes are standard measures of life satisfaction. The fifth and sixth are indices of standard measures of happiness and health status. The last three outcomes are drawn from other modules administered during our study period.[37]

[37] In deciding which measures to merge in from other modules, we define "study period" as post-our Round 1 (we could not find any relevant measure post-our Round 2 at the time we conducted our analyses), and select questions that have: a) been used in other studies; b) measure highly rated "aspects" of subjective

One key pattern in Table 6 is that common factors tend to be strongly conditionally correlated with outcomes. Statistically, 20 of the 48 estimates have $p<0.01$. Economically, the point estimates tend to imply 5 to 10 percent changes in the outcome variable per one standard deviation change in the common factor.[38]

A second key pattern is that the strength and direction of correlations with outcomes varies across common factors. Factors 1, 5, and 6 are strongly correlated with five or more of the six outcomes. Factors 2 and 3 have no p-values<0.01 (although Factor 3 looks correlated with both financial outcomes). Overall, there is a non-monotonic relationship between the strength of a common factor's correlations with the outcomes we measure and the amount of variance in decision inputs it explains (highest for Factor 1, descending to Factor 8).

A third key pattern is that, for a given common factor, the direction of any correlation is weakly consistent across outcomes. For each factor, the six correlations are either all statistically distinguishable from zero with the same sign (Factor 6), or a mix of statistically significant correlations with the same sign and correlations that are indistinguishable from zero. We find little evidence that any cluster of decision inputs pushes different outcomes in different directions.

A fourth key pattern is that behavioral biases tend to negatively correlate with outcomes (more biased <-> worse outcomes). The strongest candidates for exceptions seem to be loss aversion (in Factor 4) and ambiguity aversion (in Factor 7).

A fifth key pattern is that classical factors tend to have their expected relationship to outcomes (e.g., higher cognitive skills <-> better outcomes in Factor 1; more patience <-> better financial outcomes in Factor 3). The strongest candidate for an exception again comes from Factor 7, which suggests that higher large-stakes risk aversion is associated with better outcomes.

---

well-being in the marginal utility sense per Benjamin, Heffetz, Kimball, and Szembrot (2014); c) are answered at least once by at least 2/3 of our sample.

[38] Appendix Table 5 shows that results are similar with the no controls specification (19 correlations with $p<0.01$, similar magnitudes) and not quite as strong with the kitchen sink (12 correlations with $p<0.01$, and somewhat smaller magnitudes for the most part).

A sixth key pattern is that the overall strength of correlations does not vary wildly across outcomes. E.g., the number of correlations with $p<0.01$ ranges from 3 for several outcomes to 5 for subjective financial condition.[39]

Our key takeaways from this sub-section are that behavioral biases are strongly conditionally correlated with a range of outcomes, in expected directions for the most part, with some biases and clusters of related decision inputs more strongly correlated with outcomes than other clusters.

## 5. Conclusion

We develop eleven sets of findings on the taxonomy of consumer decision making: on how 17 behavioral biases relate to each other, to other decision inputs—patience and risk aversion, cognitive skills, and personality traits-- and to decision outputs. Our key empirical key takeaways are that behavioral biases are commonplace and not limited to certain demographic groups, heterogeneous in the cross-section of consumers, distinct from classical decision inputs in some important ways but linked in others, and economically important.

One key next step empirically is to conduct similar analyses with measures of yet more biases. A related next step is to do so on larger samples, while continuing to develop methods that account for measurement error in elicitation data (including model selection approaches). And for some applications, it may make sense to include demographics in the set of decision inputs.

Our results suggest many potential implications and next steps for theory development, and we recap a few here. We provide empirical references and a methodological guide for assessing the separability assumptions undergirding behavioral economics' standard practice of modeling one or two biases at a time. We add to the growing body of empirical evidence motivating enrichment and/or rethinking of how economists model decision making under uncertainty. And we inform various efforts to capture multiple behavioral influences with one or few parameters (e.g., Mullainathan, Schwartzstein, and Congdon 2012; Gabaix 2014); our results suggest that these approaches are on the right track but could improve their predictive power by accommodating cross-sectional heterogeneity and/or additional parameters.

[39] Appendix Table 5 shows a similar pattern within each of our two alternative specifications.

We expect that continued interplay between innovations in elicitation design, measurement error modeling, and theory will be crucial for advancing the science of consumer decision making.

## REFERENCES


Almenberg, Johan, and Christer Gerdes. 2012. "Exponential Growth Bias and Financial Literacy." *Applied Economics Letters* 19 (17): 1693–96.

Andreoni, James, and Charles Sprenger. 2012. "Estimating Time Preferences from Convex Budgets." *The American Economic Review* 102 (7): 3333–56.

Azrieli, Yaron, Christopher P. Chambers, and Paul J. Healy. 2018. "Incentives in Experiments: A Theoretical Analysis." *Journal of Political Economy* 126 (4): 1472–1503.

Banks, J., and Z. Oldfield. 2007. "Understanding Pensions: Cognitive Function, Numerical Ability, and Retirement Saving." *Fiscal Studies* 28 (2): 143–70.

Barsky, Robert B, F. Thomas Juster, Miles S Kimball, and Matthew D Shapiro. 1997. "Preference Parameters and Behavioral Heterogeneity; An Experimental Approach in the Health and Retirement Study." *Quarterly Journal of Economics* 112 (2): 537–79.

Bauer, Michal, Julie Chytilová, and Edward Miguel. 2020. "Using Survey Questions to Measure Preferences: Lessons from an Experimental Validation in Kenya."

Becker, Anke, Thomas Deckers, Thomas Dohmen, Armin Falk, and Fabian Kosse. 2012. "The Relationship Between Economic Preferences and Psychological Personality Measures." *Annual Review of Economics* 4 (1): 453–78.

Benjamin, Daniel, Sebastian Brown, and Jesse Shapiro. 2013. "Who Is 'Behavioral'? Cognitive Ability and Anomalous Preferences." *Journal of the European Economic Association* 11 (6): 1231–55.

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.

Benjamin, Daniel, Don Moore, and Matthew Rabin. 2017. "Biased Beliefs about Random Samples: Evidence from Two Integrated Experiments."

Benjamin, Daniel, Matthew Rabin, and Collin Raymond. 2016. "A Model of Nonbelief in the Law of Large Numbers." *Journal of the European Economic Association* 14 (2): 515–44.

Bernheim, B. Douglas. 2016. "The Good, the Bad, and the Ugly: A Unified Approach to Behavioral Welfare Economics1." *Journal of Benefit-Cost Analysis* 7 (1): 12–68.

Bernheim, Douglas, Stefano DellaVigna, and David Laibson. 2018. *Handbook of Behavioral Economics - Foundations and Applications 1, Volume 1*. Elsevier.

———. 2019. *Handbook of Behavioral Economics - Foundations and Applications 2, Volume 2*. Elsevier.

Bernheim, Douglas, and Dmitry Taubinsky. 2018. "Behavioral Public Economics." In *Handbook of Behavioral Economics: Applications and Foundations 1*, edited by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, 1:381–516. North-Holland.

Black, Dan, Mark C. Berger, and Frank Scott. 2000. "Bounding Parameter Estimates With Non-Classical Measurement Error." *Journal of The American Statistical Association* 95 (451): 739–48.

Branas-Garza, Pablo, Lorenzo Estepa, Diego Jorrat, Victor Orozco, and Ericka Rascon Ramirez. 2020. "To Pay or Not to Pay: Measuring Risk Preferences in Lab and  Eld."

Bronchetti, Erin, Judd Kessler, Ellen Magenheim, Dmitry Taubinsky, and Eric Zwick. 2020. "Is Attention Produced Rationally?"

Burks, S. V., J. P. Carpenter, L. Goette, and A. Rustichini. 2009. “Cognitive Skills Affect Economic Preferences, Strategic Behavior, and Job Attachment.” *Proceedings of the National Academy of Sciences* 106 (19): 7745–50.

Callen, Michael, Mohammad Isaqzadeh, James D Long, and Charles Sprenger. 2014. “Violence and Risk Preference: Experimental Evidence from Afghanistan.” *The American Economic Review* 104 (1): 123–48.

Cesarini, David, Magnus Johannesson, Patrik K. E. Magnusson, and Björn Wallace. 2012. “The Behavioral Genetics of Behavioral Anomalies.” *Management Science* 58 (1): 21–34.

Chapman, Jonathan, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer. 2019a. “Econographics.”

———. 2019b. “Willingness to Pay and Willingness to Accept Are Probably Less Correlated than You Think.”

Chapman, Jonathan, Erik Snowberg, Stephanie Wang, and Colin Camerer. 2019. “Loss Attitudes in the U.S. Population: Evidence from Dynamically Optimized Sequential Experimentation (DOSE).”

Chetty, Raj. 2015. “Behavioral Economics and Public Policy: A Pragmatic Perspective.” *American Economic Review* 105 (5): 1–33.

Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman. 2014. “Who Is (More) Rational?” *American Economic Review* 104 (6): 1518–50.

Dean, Mark, and Pietro Ortoleva. 2019. “The Empirical Relationship between Non-Standard Economic Behaviors.” *Proceedings of the National Academy of Science* 116 (33): 16262–67.

Dimmock, Stephen, Roy Kouwenberg, Olivia S. Mitchell, and Kim Peijnenburg. 2016. “Ambiguity Aversion and Household Portfolio Choice Puzzles: Empirical Evidence.” *Journal of Financial Economics* 119 (3): 559–77.

Dohmen, Thomas, Armin Falk, David Huffman, Felix Marklein, and Uwe Sunde. 2009. “Biased Probability Judgment: Evidence of Incidence and Relationship to Economic Outcomes from a Representative Sample.” *Journal of Economic Behavior & Organization* 72 (3): 903–15.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2010. “Are Risk Aversion and Impatience Related to Cognitive Ability?” *American Economic Review* 100 (3): 1238–60.

———. 2018. “On the Relationship between Cognitive Ability and Risk Preference.” *Journal of Economic Perspectives* 32 (2): 115–34.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. “Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences.” *Journal of the European Economic Association* 9 (3): 522–50.

Ehrlinger, Joyce, Kerri Johnson, Matthew Banner, David Dunning, and Justin Kruger. 2008. “Why the Unskilled Are Unaware: Further Explorations of (Absent) Self-Insight among the Incompetent.” *Organizational Behavior and Human Decision Processes* 105 (1): 98–121.

Ericson, Keith. 2011. “Forgetting We Forget: Overconfidence and Memory.” *Journal of the European Economic Association* 9 (1): 43–60.

Ericson, Keith Marzilli. 2017. “On the Interaction of Memory and Procrastination: Implications for Reminders, Deadlines, and Empirical Estimation.” *Journal of the European Economic Association* 15 (3): 692–719.

Ericson, Keith Marzilli, and David Laibson. 2019. “Intertemporal Choice.” In *Handbook of Behavioral Economics: Applications and Foundations 1*, edited by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, 2:1–67. Handbook of Behavioral Economics - Foundations and Applications 2. North-Holland.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. “Global Evidence on Economic Preferences.” *The Quarterly Journal of Economics* 133 (4): 1645–92.

Fehr, Ernst, and Lorenz Goette. 2007. “Do Workers Work More If Wages Are High? Evidence from a Randomized Field Experiment.” *American Economic Review* 97 (1): 298–317.

Frederick, Shane. 2005. “Cognitive Reflection and Decision Making.” *Journal of Economic Perspectives* 19 (4): 25–42.

Fudenberg, Drew. 2006. “Advancing Beyond Advances in Behavioral Economics.” *Journal of Economic Literature* 44 (3): 694–711.

Fuller, Wayne A. 2009. *Measurement Error Models*. Vol. 305. John Wiley & Sons.

Gabaix, Xavier. 2014. “A Sparsity-Based Model of Bounded Rationality.” *The Quarterly Journal of Economics* 129 (4): 1661–1710.

———. 2019. “Behavioral Inattention.” In *Handbook of Behavioral Economics- Foundations and Applications 2, Volume 2*, edited by Douglas Bernheim, Stefano DellaVigna, and David Laibson. Elsevier.

Ghisellini, Fabrizio, and Beryl Y. Chang. 2018. *Behavioral Economics: Moving Forward*. Palgrave Macmillan.

Gillen, Ben, Erik Snowberg, and Leeat Yariv. 2019. “Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study.” *Journal of Political Economy* 127 (4): 1826–63.

Gneezy, Uri, Alex Imas, and John List. 2015. “Estimating Individual Ambiguity Aversion: A Simple Approach.”

Goda, Gopi Shah, Matthew Levy, Colleen Flaherty Manchester, Aaron Sojourner, and Joshua Tasoff. 2019. “Predicting Retirement Savings Using Survey Measures of Exponential-Growth Bias and Present Bias.” *Economic Inquiry* 57 (3): 1636–58.

Heidhues, Paul, Botond Kőszegi, and Philipp Strack. 2018. “Unrealistic Expectations and Misguided Learning.” *Econometrica* 86 (4): 1159–1214.

Jensen, Arthur Robert. 1998. *The g Factor: The Science of Mental Ability*. Praeger.

Kőszegi, Botond. 2014. “Behavioral Contract Theory.” *Journal of Economic Literature* 52 (4): 1075–1118.

Levine, David K. 2012. *Is Behavioral Economics Doomed?: The Ordinary versus the Extraordinary*. Open Book Publishers.

Levy, Matthew, and Joshua Tasoff. 2016. “Exponential-Growth Bias and Lifecycle Consumption.” *Journal of the European Economic Association* 14 (3): 545–83.

Li, Ye, Martine Baldassi, Eric J. Johnson, and Elke U. Weber. 2013. “Complementary Cognitive Capabilities, Economic Decision Making, and Aging.” *Psychology and Aging* 28 (3): 595–613.

Lusardi, Annamaria, and Olivia S. Mitchell. 2014. “The Economic Importance of Financial Literacy: Theory and Evidence.” *Journal of Economic Literature* 52 (1): 5–44.

MacLeod, Colin M. 1991. “Half a Century of Research on the Stroop Effect: An Integrative Review.” *Psychological Bulletin* 109 (2): 163.

McArdle, John J., Gwenith G. Fisher, and Kelly M. Kadlec. 2007. "Latent Variable Analyses of Age Trends of Cognition in the Health and Retirement Study, 1992-2004." *Psychology and Aging* 22 (3): 525–45.
Montiel Olea, J. L., and T. Strzalecki. 2014. "Axiomatization and Measurement of Quasi-Hyperbolic Discounting." *The Quarterly Journal of Economics* 129 (3): 1449–99.
Moore, Don A., and Paul J. Healy. 2008. "The Trouble with Overconfidence." *Psychological Review* 115 (2): 502–17.
Mrkva, Kellen, Eric J. Johnson, Simon Gächter, and Andreas Herrmann. 2020. "Moderating Loss Aversion: Loss Aversion Has Moderators, But Reports of Its Death Are Greatly Exaggerated." *Journal of Consumer Psychology* 30 (3): 407–28.
Mullainathan, Sendhil, Joshua Schwartzstein, and William J. Congdon. 2012. "A Reduced-Form Approach to Behavioral Public Finance." *Annual Review of Economics* 4 (1): 511–40.
O'Donoghue, Ted, and Jason Somerville. 2018. "Modeling Risk Aversion in Economics." *Journal of Economic Perspectives* 32 (2): 91–114.
Prokosheva, Sasha. 2016. "Comparing Decisions under Compound Risk and Ambiguity: The Importance of Cognitive Skills." *Journal of Behavioral and Experimental Economics*, Cognitive Abilities and Economic Behavior, 64 (October): 94–105.
Rabin, Matthew, and Dimitri Vayanos. 2010. "The Gambler's and Hot-Hand Fallacies: Theory and Applications." *Review of Economic Studies* 77 (2): 730–78.
Rabin, Matthew, and Georg Weizsäcker. 2009. "Narrow Bracketing and Dominated Choices." *American Economic Review* 99 (4): 1508–43.
Rammstedt, Beatrice, and Oliver P. John. 2007. "Measuring Personality in One Minute or Less: A 10-Item Short Version of the Big Five Inventory in English and German." *Journal of Research in Personality* 41 (1): 203–12.
Read, Daniel, and Barbara van Leeuwen. 1998. "Predicting Hunger: The Effects of Appetite and Delay on Choice." *Organizational Behavior and Human Decision Processes* 76 (2): 189–205.
Scerri, Eric R. 2011. "The Evolution of the Periodic System." *Scientific American*, January 21.
Schildberg-Hörisch, Hannah. 2018. "Are Risk Preferences Stable?" *Journal of Economic Perspectives* 32 (2): 135–54.
Stango, Victor, and Jonathan Zinman. 2009. "Exponential Growth Bias and Household Finance." *The Journal of Finance* 64 (6): 2807–49.
———. 2020. "Behavioral Biases Are Temporally Stable."
Stanovich, Keith E. 2016. "The Comprehensive Assessment of Rational Thinking." *Educational Psychologist* 51 (1): 23–34.
Von Gaudecker, Hans-Martin, Arthur Van Soest, and Erik Wengström. 2011. "Heterogeneity in Risky Choice Behavior in a Broad Population." *The American Economic Review* 101 (2): 664–94.

**Figure 1. B-index variance does seem to reflect true heterogeneity**

Figure 1a. B-index=Count of biases exhibited, for panelists completing both rounds

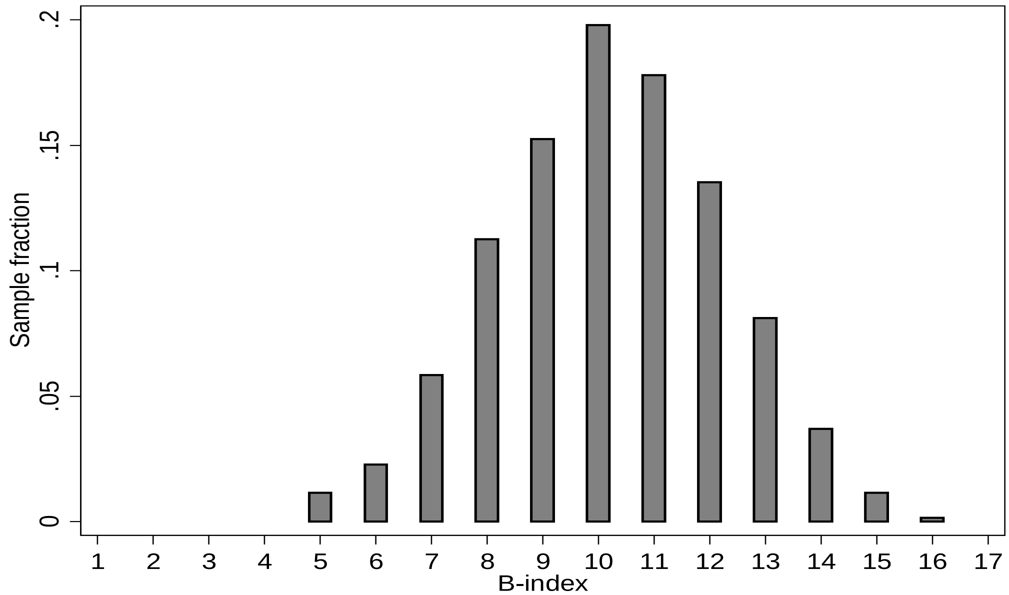
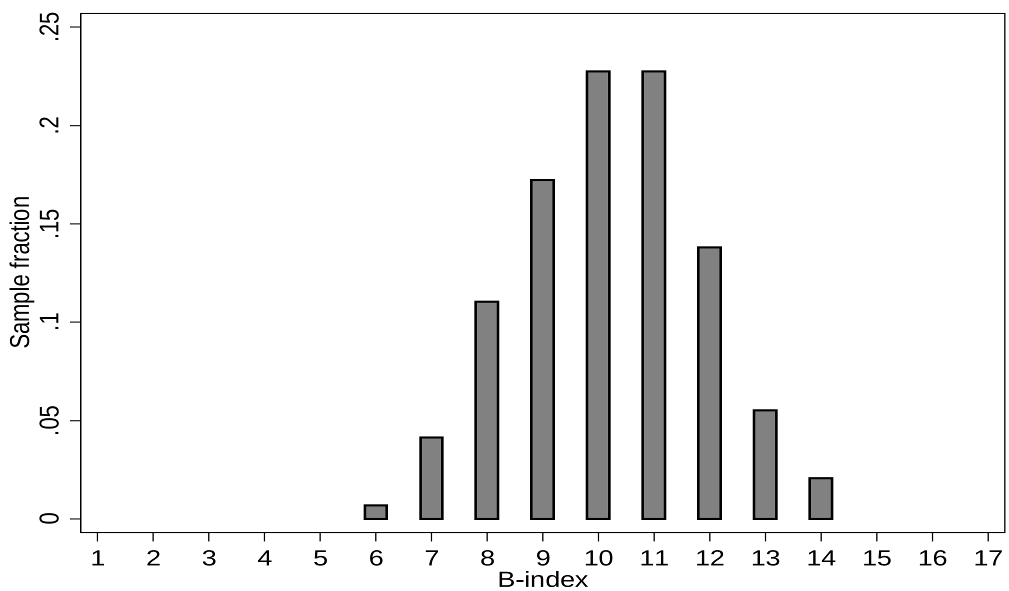


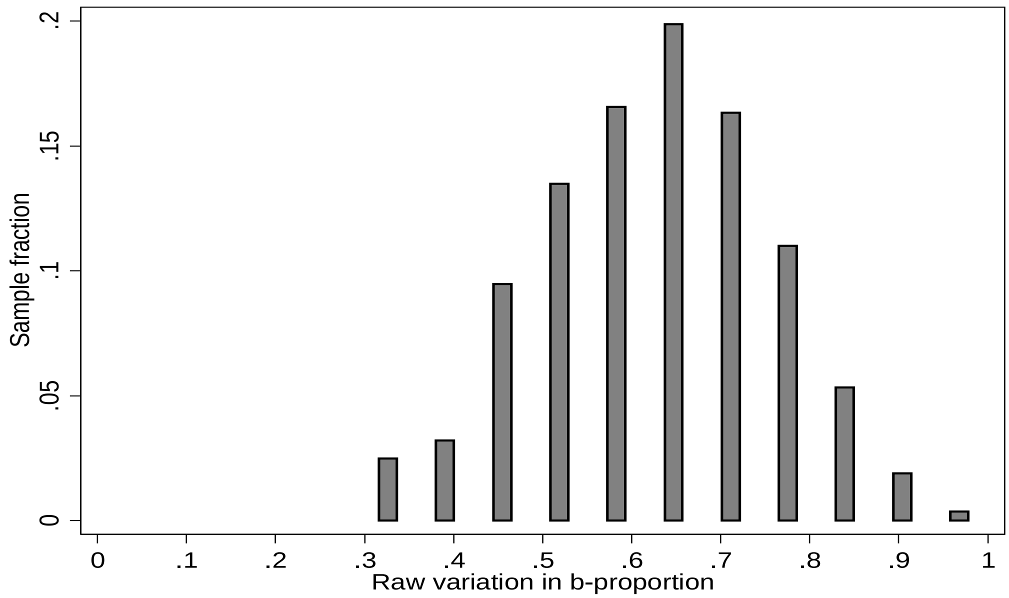

Figure 1b. Only panelists with equal counts across rounds




Round 1 B-index. We omit panelists with missing data on 2 or more of our 17 potential sources of behavioral biases, to mitigate spurious variance from variance in missingness. This leaves sample sizes of 702 individuals in the top panel and 145 in the bottom.

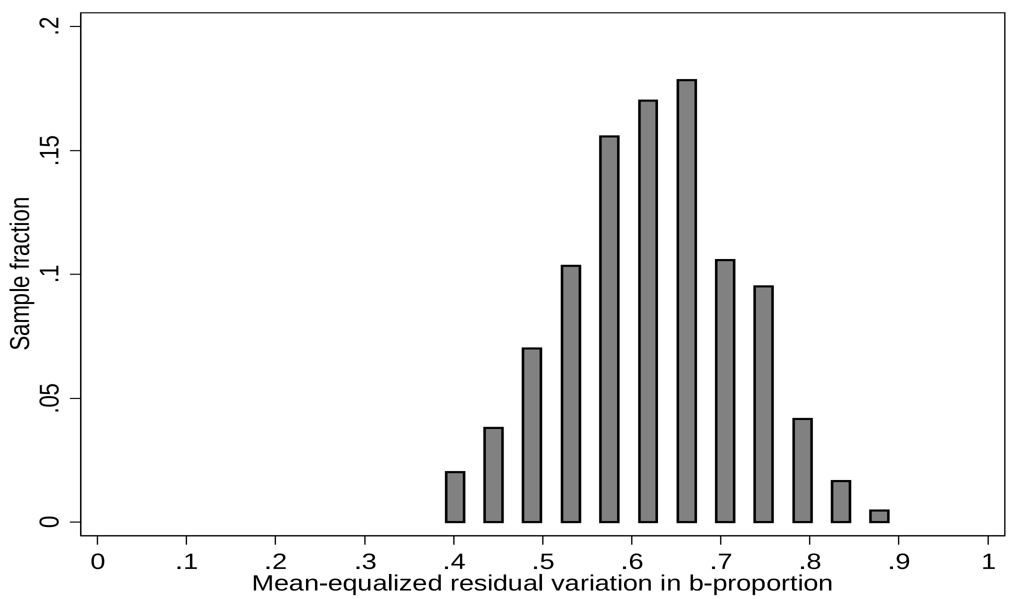
**Figure 2. The B-index is not well-explained by other covariates**

Figure 2a. B-index unconditional variation




On X-axis, B-proportion is the ratio of the B-index to the count of non-missing bias measures. We use B-proportion instead of the level B-index for comparability to Figure 2b. Interquartile range here is [0.53, 0.71] and $5^{th}/95^{th}$ percentiles are [0.41, 0.82].

Figure 2b. B-index residual variation




X-axis shows distribution of residuals from regression of B-index proportion on full set of covariates (Appendix Table 2 Col 1 reports R-squared). We use B-index proportion instead of B-index to avoid overfitting. Mean of residuals is set equal to the mean of the B-proportion from Figure 2a, for comparability. Interquartile range here is [0.56, 0.69] and $5^{th}/95^{th}$ percentiles are [0.46, 0.78].

**Figure 3. B-index variation within- and across- key sub-groups**

Figure 3a. B-index variation by top vs. bottom income quartiles

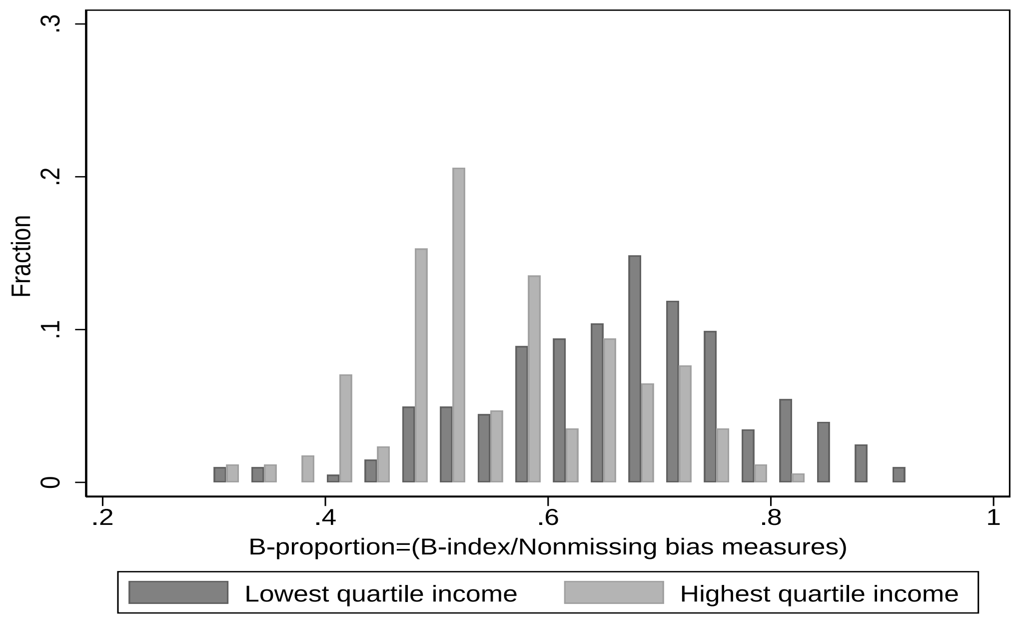
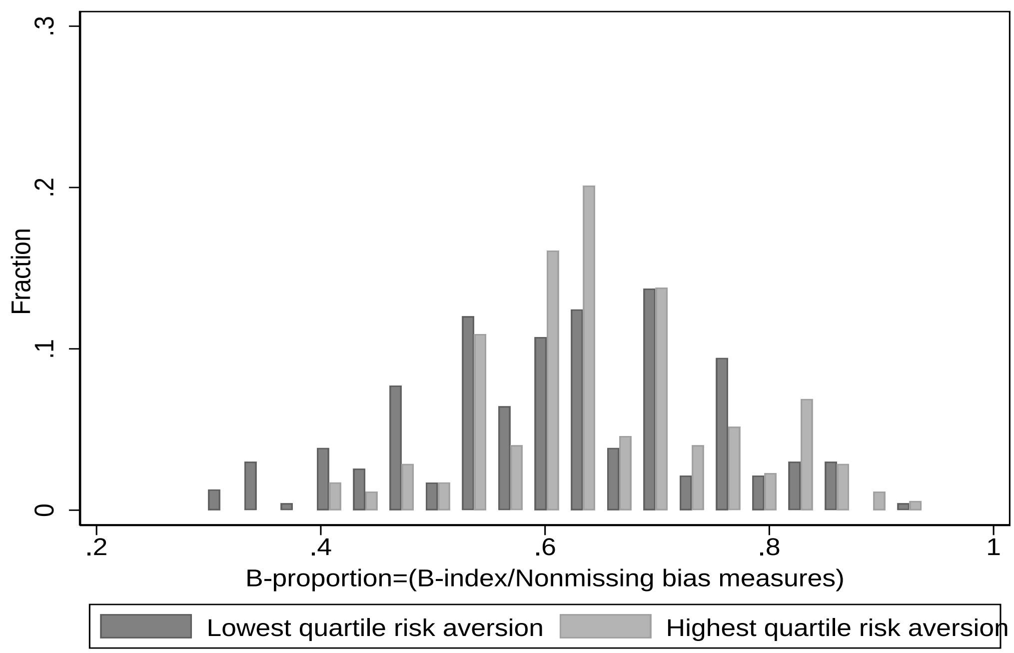



Figure 3b. B-index variation by top vs. bottom risk aversion quartiles




**(Notes at bottom of next page, following Figure 3d.)**

Figure 3c. B-index variation by top vs. bottom education bins

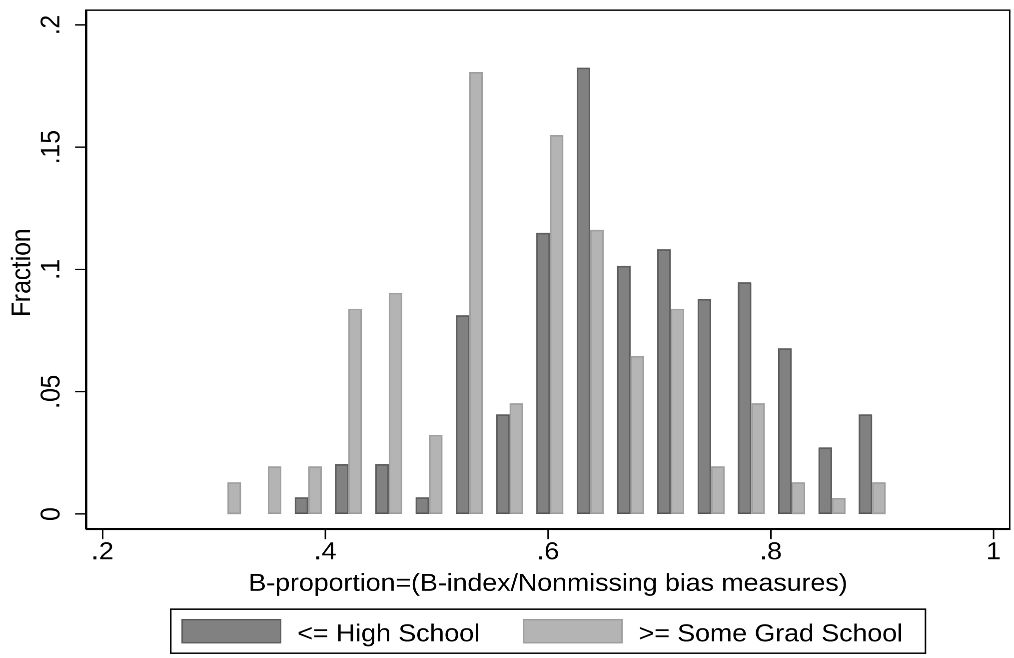
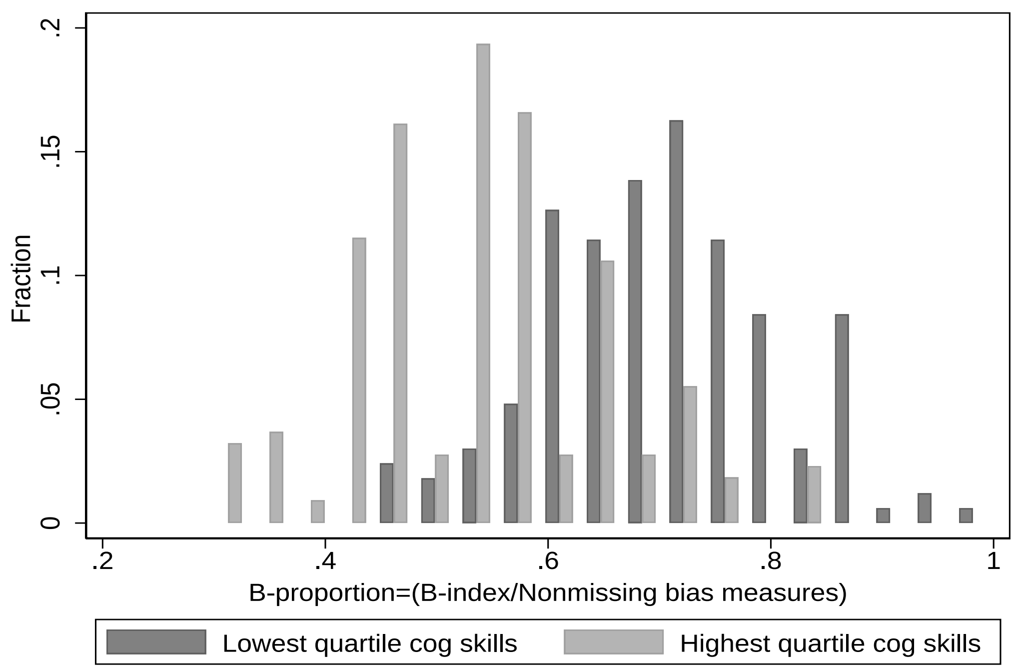



Figure 3d. B-index variation by top vs. bottom cognitive skill quartiles




Round 1 data only. On X-axis, we use B-proportion instead of B-index to allow for item non-response to vary across sub-groups. Cognitive skills measured here with the 1st principal component of our four test scores. Risk aversion measured here with the 1st principal component of the Dohmen et al and Barsky et al measures. See Table 1 and Data Appendix Section 2 for details on component test score and risk aversion measures.

**Table 1. Measuring decision inputs**

| | Definition | Classical benchmark | Related bias group | Unique non-miss ranks | Non-missing obs. Round 1 | Round 2 |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Behavioral biases | | | | | | |
| Present-bias: Money discounting | Discounts more when sooner date is today vs. 5 weeks from today, across 24 choices | Same discount rate at both sooner dates | Present-bias | 28 | 803 | 819 |
| Present-bias: Consumption discounting | Chooses less healthy snack today and more healthy snack for 5 weeks from now | Chooses same snack on both dates | Present-bias | 2 | 835 | 829 |
| Violations of General Axiom of Revealed Preference | Inconsistent across 11 choices under uncertainty subject to budget constraint | Choices are consistent with GARP | Inconsistent/dominated choices | 54 | 774 | 799 |
| Violations of GARP and dominance avoidance | See above, also counts dominated choices as inconsistent | Choices are consistent with GARP and dominance avoidance | Inconsistent/dominated choices | 97 | 774 | 799 |
| Narrow bracketing | Makes dominated choice(s) given implications of an earlier decision | Avoid dominated choices | Inconsistent/dominated choices | 5 | 827 | 837 |
| Preference for certainty | Certainty premium > 0, from 20 MPL choices between certain payoffs and lotteries | Certainty premium = 0 | Risk biases | 20 | 620 | 620 |
| Loss aversion | Chooses certain $0 over pos expected value gamble(s) with potential for small loss | Chooses gambles with small positive expected value | Risk biases | 4 | 843 | 845 |
| Ambiguity aversion | Prefers lower but known expected payoff to unknown payoff | Chooses higher expected value gamble | Risk biases | 17 | 784 | 751 |
| Overconfidence in level performance | Self-assessment > actual score on numeracy quiz | Accurate self-assessment | Overconfidence | 4 | 829 | 813 |
| Overconfidence in precision | Indicates 100% certainty about quiz performance and/or future income change | Acknowledges imperfect certainty | Overconfidence | 3 | 793 | 775 |
| Overconfidence in relative performance | Greater diff between self-assessed and actual intelligence test rank relative to others | Accurate self-assessment | Overconfidence | 78 | 844 | 818 |
| Underestimates convergence: Non-belief in the law of large numbers | Overestimates variance in sample of 1000 coin flips | Accurately estimates variance | Math biases | 21 | 833 | 819 |
| Cold Hand Gambler's Fallacy | After 10 straight "heads," thinks prob. "tails" >50% | Thinks probability is 50% | Math biases | 10 | 842 | 817 |
| Underestimates APR: Exponential growth bias, loan-side | Underestimates loan APR given other terms | Accurately estimates APR | Math biases | 50 | 778 | 783 |
| Underestimates future value: Exponential growth bias, asset-side | Underestimates future value given other terms | Accurately estimates future value | Math biases | 11 | 761 | 735 |
| Limited attention | Regrets paying too little attention to finances, taking opportunity cost into account | No regret re: attention | Limited attention/memory | 5 | 832 | 829 |
| Limited memory | Says will complete short survey for $10 tomorrow but does not complete | Says will complete and does | Limited attention/memory | 2 | 825 | 803 |
| Presumed-classical patience and risk aversion | | | | | | |
| Patience | Average savings rate from money discounting questions | | | 75 | 803 | 819 |
| Risk aversion: financial | -1*(Self-assessed willingness to take financial risks on 100-point scale) | | | 52 | 842 | 823 |
| Risk aversion: large stakes | # of times chooses less risky salary over higher expected value salary | | | 6 | 840 | 840 |
| Cognitive skills | | | | | | |
| Fluid intelligence | # correct answers on number series test | | | 13 | 845 | 819 |
| Numeracy | # correct answers on two questions re: division and percent | | | 3 | 832 | 813 |
| Financial literacy | # correct answers on three questions about interest rates, inflation, diversification | | | 4 | 843 | 823 |
| Executive function | # correct answers on two-minute Stroop test | | | 76 | 816 | 797 |
| Personality traits | | | | | | |
| Extraversion | More answers indicating being energetic, talkative, assertive | | | 9 | n/a | 813 |
| Agreeableness | More answers indicating being kind, affectionate, sympathetic | | | 9 | n/a | 812 |
| Conscientiousness | More answers indicating being organized and thorough | | | 9 | n/a | 813 |
| Neuroticism | More answers indicating emotional instability and negative emotions | | | 9 | n/a | 812 |
| Openness | More answers indicating enjoyment from learning new things and new experiences | | | 9 | n/a | 812 |

See Section 1 and Data Appendix for details on elicitation methods and their antecedents.

**Table 2. Descriptive statistics for multiple behavioral biases**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Mean | SD | 5th, 95th percentiles | Share>0 | Max (possible) | Mean Proportion |
| **Panel A: Consumer-level bias count: the B-index** | | | | | | |
| Round 1 (N=1427) | 10.04 | 2.16 | 6,13 | 1.00 | 16 (17) | 0.63 |
| Round 1, in Round 2 (N=845) | 10.06 | 2.02 | 7,13 | 1.00 | 16 (17) | 0.62 |
| Round 2 (N=845) | 9.92 | 2.22 | 6,13 | 1.00 | 16 (17) | 0.62 |
| **Panel B. Count of missing inputs to B-index** | | | | | | |
| Round 1 | 1.02 | 1.71 | 0,4 | 0.49 | 12 (17) | 0.06 |
| Round 1, in Round 2 | 0.75 | 1.19 | 0,3 | 0.43 | 9 (17) | 0.04 |
| Round 2 | 0.89 | 1.76 | 0,4 | 0.41 | 11 (17) | 0.05 |

Our data consist of two survey rounds, of two modules each, conducted 3 years apart. We include only those panelists who took both modules in Round 1 (N=1427) or all four modules across both rounds (N=845). "Round 1, in Round 2" refers to Round 1 data for panelists who also completed survey Round 2. B-index components and definitions are summarized in Table 1 and Section 1-C; please see Data Appendix Section 1 for details. Column 6 Panel A is the "B-proportion" we use in other figures and tables for analyzing fit: (B-index)/(Number of nonmissing inputs to B-index). Column 6 Panel B = (Count missing inputs)/17 .

**Table 3. Correlations among behavioral biases, and between biases and other decision inputs**

| | PB money | PB cons'n | Viols. GARP | Viols. GARP+ | Narr. brack. | Pref. cert. | Loss aver. | Ambig. aver. | OC level | OC prec. | OC rel. | NBLLN | Cold Hand | EGB APR | EGB asset | Lim. att. | Lim. mem. | Patience | Risk av. fin | Risk av. large | Fluid int. | Num. | Fin. lit. | Exec. fcn. | Extrav. | Agree. | Consc. | Neur. | Open. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Present-bias | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Present-bias: Money discounting | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Present-bias: Consumption discounting | 0.47 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Inconsistent and/or dominated choices | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Violations of General Axiom of Revealed Preference | 0.30 | 0.31 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Violations of GARP and dominance avoidance | 0.14 | 0.32 | 1.00 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | | |
| Narrow bracketing | 0.29 | 0.02 | 0.31 | 0.24 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | |
| Risk biases | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Preference for certainty | 0.47 | 0.07 | 0.51 | 0.56 | 0.13 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | |
| Loss aversion | 0.09 | -0.27 | -0.38 | -0.28 | -0.40 | -0.06 | 1.00 | | | | | | | | | | | | | | | | | | | | | | |
| Ambiguity aversion | -0.02 | 0.25 | -0.01 | -0.14 | -0.10 | -0.21 | 0.06 | 1.00 | | | | | | | | | | | | | | | | | | | | | |
| Overconfidence | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Overconfidence in level performance | -0.02 | 0.07 | 0.34 | 0.38 | 0.22 | 0.37 | -0.18 | -0.25 | 1.00 | | | | | | | | | | | | | | | | | | | | |
| Overconfidence in precision | -0.15 | -0.06 | 0.02 | 0.04 | -0.02 | 0.00 | 0.00 | -0.02 | 0.19 | 1.00 | | | | | | | | | | | | | | | | | | | |
| Overconfidence in relative performance | 0.34 | 0.13 | 0.41 | 0.47 | 0.22 | 0.22 | -0.33 | -0.16 | 0.46 | 0.06 | 1.00 | | | | | | | | | | | | | | | | | | |
| Math biases | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Underestimates convergence: Non-belief in the law of large numbers | 0.36 | 0.14 | 0.47 | 0.56 | 0.33 | 0.26 | -0.12 | -0.13 | 0.54 | 0.05 | 0.41 | 1.00 | | | | | | | | | | | | | | | | | |
| Cold Hand Gambler's Fallacy | 0.30 | 0.04 | 0.41 | 0.48 | 0.22 | 0.46 | -0.26 | 0.01 | 0.29 | 0.03 | 0.45 | 0.50 | 1.00 | | | | | | | | | | | | | | | | |
| Underestimates APR: Exponential growth bias, loan-side | 0.12 | 0.05 | -0.20 | -0.08 | -0.16 | -0.13 | 0.10 | -0.22 | -0.09 | 0.02 | -0.01 | 0.09 | -0.01 | 1.00 | | | | | | | | | | | | | | | |
| Underestimates future value: Exponential growth bias, asset-side | 0.27 | 0.15 | 0.28 | 0.38 | 0.33 | 0.21 | -0.22 | -0.20 | 1.06 | -0.09 | 0.38 | 0.42 | 0.35 | -0.01 | 1.00 | | | | | | | | | | | | | | |
| Limited attention/memory | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Limited attention | -0.09 | 0.23 | 0.18 | 0.11 | 0.07 | 0.15 | -0.01 | -0.16 | 0.11 | -0.14 | 0.02 | 0.11 | 0.06 | -0.19 | 0.14 | 1.00 | | | | | | | | | | | | | |
| Limited memory | -0.01 | 0.41 | 0.51 | 0.50 | -0.01 | 0.23 | -0.08 | 0.04 | 0.03 | -0.12 | 0.17 | 0.30 | 0.09 | -0.11 | 0.07 | 0.22 | 1.00 | | | | | | | | | | | | |
| Presumed-classical patience and risk aversion | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Patience | -0.57 | -0.19 | -0.05 | -0.12 | 0.03 | -0.07 | -0.09 | -0.10 | 0.02 | 0.02 | 0.00 | 0.05 | 0.04 | -0.01 | -0.01 | -0.16 | 0.17 | 1.00 | | | | | | | | | | | |
| Risk aversion: financial | 0.11 | 0.02 | -0.12 | -0.05 | -0.05 | -0.10 | 0.35 | 0.22 | -0.07 | -0.08 | -0.15 | 0.03 | 0.12 | 0.04 | 0.09 | 0.09 | -0.14 | -0.06 | 1.00 | | | | | | | | | | |
| Risk aversion: large stakes | 0.09 | 0.01 | 0.11 | 0.28 | 0.13 | 0.21 | 0.09 | 0.04 | 0.21 | 0.14 | 0.24 | 0.29 | 0.31 | -0.02 | 0.22 | -0.10 | 0.02 | 0.08 | 0.29 | 1.00 | | | | | | | | | |
| Cognitive skills | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Fluid intelligence | -0.39 | -0.13 | -0.46 | -0.51 | -0.31 | -0.36 | 0.28 | 0.17 | -0.50 | 0.05 | -1.05 | -0.54 | -0.56 | 0.05 | -0.54 | -0.08 | -0.18 | -0.05 | 0.01 | -0.33 | 1.00 | | | | | | | | |
| Numeracy | -0.33 | -0.20 | -0.38 | -0.40 | -0.24 | -0.35 | 0.22 | 0.35 | -0.98 | 0.14 | -0.61 | -0.50 | -0.40 | 0.08 | -0.57 | -0.15 | -0.35 | -0.18 | 0.01 | -0.35 | 0.75 | 1.00 | | | | | | | |
| Financial literacy | -0.34 | -0.15 | -0.41 | -0.42 | -0.33 | -0.32 | 0.26 | 0.09 | -0.48 | 0.00 | -0.47 | -0.58 | -0.41 | 0.02 | -0.63 | -0.16 | -0.32 | -0.09 | -0.12 | -0.35 | 0.62 | 0.77 | 1.00 | | | | | | |
| Executive function | -0.18 | -0.02 | -0.31 | -0.46 | -0.17 | -0.19 | 0.24 | 0.13 | -0.40 | -0.05 | -0.44 | -0.34 | -0.32 | -0.11 | -0.31 | 0.04 | -0.07 | -0.06 | 0.04 | -0.32 | 0.51 | 0.52 | 0.29 | 1.00 | | | | | |
| Personality traits | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Extraversion | -0.05 | 0.12 | 0.01 | 0.11 | 0.09 | 0.00 | -0.03 | -0.04 | 0.10 | 0.02 | 0.14 | 0.17 | 0.03 | -0.04 | 0.06 | 0.07 | 0.12 | 0.02 | -0.11 | -0.03 | -0.07 | -0.08 | -0.04 | -0.06 | 1.00 | | | | |
| Agreeableness | 0.01 | -0.08 | 0.06 | 0.13 | -0.01 | 0.04 | -0.06 | 0.07 | -0.02 | 0.12 | 0.07 | 0.07 | 0.11 | 0.07 | -0.01 | -0.02 | 0.03 | -0.05 | -0.01 | 0.12 | -0.09 | -0.01 | -0.05 | -0.09 | 0.08 | 1.00 | | | |
| Conscientiousness | 0.17 | -0.21 | 0.02 | 0.11 | -0.04 | 0.04 | 0.04 | 0.10 | 0.04 | 0.21 | 0.13 | 0.02 | 0.11 | 0.00 | 0.01 | -0.12 | -0.10 | -0.08 | 0.03 | 0.16 | -0.06 | 0.12 | 0.07 | -0.02 | 0.20 | 0.17 | 1.00 | | |
| Neuroticism | 0.05 | 0.15 | 0.01 | -0.01 | -0.02 | 0.02 | 0.12 | -0.05 | -0.07 | -0.13 | -0.08 | 0.05 | 0.01 | 0.04 | 0.04 | 0.14 | 0.05 | 0.00 | 0.23 | -0.04 | -0.02 | -0.01 | -0.09 | 0.06 | -0.17 | -0.20 | -0.14 | 1.00 | |
| Openness | -0.15 | -0.02 | 0.01 | 0.02 | 0.03 | 0.11 | 0.04 | 0.04 | 0.01 | 0.04 | -0.03 | -0.01 | -0.08 | -0.02 | 0.01 | -0.01 | -0.06 | 0.04 | -0.01 | -0.05 | 0.08 | -0.02 | 0.04 | 0.10 | 0.08 | 0.00 | 0.12 | 0.01 | 1.00 |

Pairwise rank correlation point estimates, each estimated using the ORIV method described in Section 3-A. Standard errors (not shown but summarized in Table 4) are clustered by panelist. Estimated correlations can be outside [-1, 1] because of our regression-based estimator, but none of the estimates are significantly different from |1| at any conventional p-value threshold.

**Table 4. Correlations among and between biases and other decision inputs: Summary of pairwise estimates**

**Panel A. Signs and p-values**

| | Count of pairwise correlations estimated | Share of pairwise correlation estimates | | | |
|---|---|---|---|---|---|
| | | >0 | | <0 | |
| | | p<=0.10 | p<=0.001 | p<=0.10 | p<=0.001 |
| Biases: All | 136 | 0.37 | 0.18 | 0.13 | 0.04 |
| Biases: Related only | 17 | 0.47 | 0.29 | 0.06 | 0.00 |
| Biases and patience | 17 | 0.00 | 0.00 | 0.06 | 0.00 |
| Biases and risk aversion | 34 | 0.35 | 0.18 | 0.03 | 0.00 |
| Biases and cognitive skills | 68 | 0.12 | 0.07 | 0.65 | 0.49 |
| Biases and personality | 85 | 0.19 | 0.02 | 0.06 | 0.00 |

See Table 3 for each pairwise point estimate. P-values are from tests of the null hypothesis that an estimated pairwise correlation is zero.

**Panel B. Magnitudes**

| | Average correlation with | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Biases | | | | Other decision inputs | | | |
| | Related category | All | Related | Other | Patience | Risk Av. | Cog. Skills | Personality* |
| Present-bias: Money discounting | Present-bias | 0.16 | 0.47 | 0.14 | -0.57 | 0.10 | -0.31 | 0.08 |
| Present-bias: Consumption discounting | Present-bias | 0.14 | 0.47 | 0.12 | -0.19 | 0.01 | -0.13 | 0.12 |
| Violations of General Axiom of Revealed Preference | Inconsistent/dominated choices | 0.27 | 0.66 | 0.22 | -0.05 | 0.00 | -0.39 | 0.02 |
| Violations of GARP and dominance avoidance | Inconsistent/dominated choices | 0.28 | 0.62 | 0.23 | -0.12 | 0.12 | -0.45 | 0.07 |
| Narrow bracketing | Inconsistent/dominated choices | 0.15 | 0.27 | 0.07 | 0.03 | 0.04 | -0.26 | 0.04 |
| Preference for certainty | Risk biases | 0.18 | -0.13 | 0.23 | -0.07 | 0.06 | -0.30 | 0.04 |
| Loss aversion | Risk biases | -0.14 | 0.10 | -0.17 | -0.09 | 0.22 | 0.25 | 0.06 |
| Ambiguity aversion | Risk biases | -0.09 | -0.07 | -0.10 | -0.10 | 0.13 | 0.19 | 0.06 |
| Overconfidence in level performance | Overconfidence | 0.22 | 0.32 | 0.21 | 0.02 | 0.07 | -0.59 | 0.05 |
| Overconfidence in precision | Overconfidence | -0.01 | 0.12 | 0.07 | 0.02 | 0.03 | 0.03 | 0.11 |
| Overconfidence in relative performance | Overconfidence | 0.19 | 0.26 | 0.16 | 0.00 | 0.05 | -0.64 | 0.09 |
| Underestimates convergence: Non-belief in the law of large numbers | Math biases | 0.25 | 0.25 | 0.24 | 0.05 | 0.16 | -0.49 | 0.06 |
| Cold Hand Gambler's Fallacy | Math biases | 0.21 | 0.28 | 0.20 | 0.04 | 0.21 | -0.42 | 0.07 |
| Underestimates APR: Exponential growth bias, loan-side | Math biases | -0.06 | 0.00 | -0.07 | -0.01 | 0.01 | 0.01 | 0.03 |
| Underestimates future value: Exponential growth bias, asset-side | Math biases | 0.21 | 0.21 | 0.21 | -0.01 | 0.15 | -0.51 | 0.03 |
| Limited attention | Attention/memory | 0.06 | 0.22 | 0.04 | -0.16 | -0.01 | -0.09 | 0.07 |
| Limited memory | Attention/memory | 0.14 | 0.22 | 0.14 | 0.17 | -0.06 | -0.23 | 0.07 |
| Cross-bias average | | 0.13 | 0.25 | 0.11 | -0.06 | 0.08 | -0.25 | 0.06 |

We report the average absolute value of correlations with personality, across the five traits, because different traits might be expected to correlate with biases in different directions. We summarize average correlations between biases and each of the five personality traits in Section 3-C. Cross-bias average is unweighted mean of the correlations in that column in Panel B.

**Table 5. Rotated 8-factor models and loadings of decision inputs on common factors**

| | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | Factor 6 | Factor 7 | Factor 8 |
|---|---|---|---|---|---|---|---|---|
| Behavioral biases | | | | | | | | |
| Present-bias: Money discounting | 0.17 | 0.07 | 0.87 | 0.02 | 0.03 | -0.09 | 0.05 | -0.21 |
| Present-bias: Consumption discounting | 0.12 | 0.15 | 0.37 | -0.13 | 0.58 | 0.12 | 0.38 | -0.01 |
| Violations of General Axiom of Revealed Preference | -0.01 | 0.94 | 0.06 | -0.12 | 0.10 | -0.09 | 0.01 | 0.06 |
| Violations of GARP and dominance avoidance | 0.06 | 0.91 | 0.03 | 0.02 | 0.03 | -0.01 | -0.05 | 0.06 |
| Narrow bracketing | -0.07 | 0.70 | 0.21 | 0.10 | -0.05 | -0.09 | -0.32 | 0.14 |
| Preference for certainty | -0.24 | -0.11 | 0.07 | 0.85 | -0.07 | -0.05 | -0.03 | 0.07 |
| Loss aversion | 0.27 | 0.08 | 0.15 | -0.64 | -0.19 | -0.27 | 0.05 | 0.19 |
| Ambiguity aversion | -0.19 | -0.06 | 0.02 | -0.05 | 0.03 | 0.05 | 0.84 | 0.05 |
| Overconfidence in level performance | 0.73 | -0.04 | -0.05 | -0.01 | -0.11 | 0.18 | -0.25 | 0.44 |
| Overconfidence in precision | -0.14 | 0.17 | -0.12 | -0.03 | -0.52 | 0.13 | -0.02 | 0.03 |
| Overconfidence in relative performance | 0.70 | 0.05 | 0.11 | -0.12 | -0.07 | 0.23 | -0.01 | -0.17 |
| Underestimates convergence: Non-belief in the law of large numbers | 0.50 | 0.29 | 0.08 | 0.10 | 0.00 | -0.01 | -0.04 | 0.07 |
| Cold Hand Gambler's Fallacy | 0.38 | 0.27 | 0.03 | 0.67 | -0.03 | 0.00 | -0.01 | -0.10 |
| Underestimates APR: Exponential growth bias, loan-side | 0.12 | -0.30 | 0.24 | 0.18 | 0.02 | 0.01 | -0.32 | -0.54 |
| Underestimates future value: Exponential growth bias, asset-side | 0.79 | -0.16 | 0.12 | -0.03 | 0.00 | 0.06 | -0.12 | 0.40 |
| Limited attention | 0.09 | 0.16 | 0.00 | 0.15 | 0.49 | 0.03 | -0.07 | 0.47 |
| Limited memory | 0.03 | 0.59 | -0.23 | 0.05 | 0.56 | 0.12 | 0.15 | 0.02 |
| Presumed-classical patience and risk aversion | | | | | | | | |
| Patience | 0.26 | -0.04 | -0.82 | -0.04 | -0.01 | -0.09 | 0.02 | -0.05 |
| Risk aversion: financial | 0.18 | -0.24 | 0.08 | 0.34 | -0.03 | -0.42 | 0.45 | 0.15 |
| Risk aversion: large stakes | 0.33 | 0.12 | -0.11 | 0.19 | -0.41 | -0.24 | 0.29 | -0.01 |
| Cognitive skills | | | | | | | | |
| Fluid intelligence | -0.87 | -0.07 | -0.05 | 0.05 | -0.01 | -0.03 | -0.04 | 0.10 |
| Numeracy | -0.96 | 0.09 | 0.14 | 0.01 | -0.18 | -0.05 | 0.10 | -0.08 |
| Financial literacy | -0.86 | -0.02 | 0.03 | 0.08 | -0.14 | 0.11 | -0.11 | -0.06 |
| Executive function | -0.54 | -0.17 | 0.14 | 0.05 | 0.14 | 0.01 | 0.11 | 0.18 |
| Personality traits | | | | | | | | |
| Extraversion | 0.24 | -0.13 | 0.01 | 0.09 | 0.12 | 0.76 | 0.09 | 0.30 |
| Agreeableness | 0.04 | 0.07 | 0.00 | 0.15 | -0.20 | 0.29 | 0.10 | -0.17 |
| Conscientiousness | -0.07 | 0.03 | 0.21 | 0.02 | -0.53 | 0.33 | 0.22 | 0.09 |
| Neuroticism | -0.02 | 0.07 | 0.03 | 0.03 | 0.28 | -0.59 | 0.00 | 0.06 |
| Openness | -0.03 | 0.01 | -0.10 | 0.03 | -0.10 | 0.23 | 0.09 | 0.54 |
| Eigenvalue | 6.81 | 2.19 | 2.12 | 1.88 | 1.64 | 1.42 | 1.34 | 1.23 |
| Variation explained | 0.23 | 0.08 | 0.07 | 0.07 | 0.06 | 0.05 | 0.05 | 0.04 |
| Cumulative variation explained | 0.23 | 0.31 | 0.38 | 0.45 | 0.50 | 0.55 | 0.60 | 0.64 |

Notes: Promax rotation. See Appendix Table 4 for eigenvalues for, and variation explained by, additional common factors.

**Table 6. ORIV correlations between decision outputs and decision input common factors: Main specification**

| *Outcome variable* | *Objective fin. condition* | *Subjective fin. condition* | *Life satisfaction I* | *Life satisfaction II* | *Happiness* | *Health status* |
|---|---|---|---|---|---|---|
| Common factor: loading variables included | (1) | (2) | (3) | (4) | (5) | (6) |
| Fac 1: Cog skills (-), EGB FV, OC level & rel, NBLLN, GF | -0.110 | -0.035 | -0.037 | -0.038 | -0.020 | -0.046 |
| | (0.017) | (0.013) | (0.013) | (0.014) | (0.015) | (0.014) |
| Fac 2: GARP violations, narrow bracketing, limited memory | -0.029 | -0.004 | 0.011 | 0.008 | -0.003 | 0.013 |
| | (0.016) | (0.012) | (0.011) | (0.011) | (0.013) | (0.012) |
| Fac 3: Present-biases, patience (-) | -0.027 | -0.028 | -0.014 | 0.003 | -0.011 | -0.004 |
| | (0.014) | (0.011) | (0.010) | (0.010) | (0.012) | (0.011) |
| Fac 4: Preference for certainty, loss aversion (-), GF | -0.010 | -0.016 | -0.027 | -0.003 | -0.039 | -0.014 |
| | (0.017) | (0.012) | (0.012) | (0.012) | (0.014) | (0.013) |
| Fac 5: Lim mem & attent, OC precision, PB cons'n, consc (-), classical risk av (-) | -0.001 | -0.038 | -0.034 | -0.028 | -0.035 | -0.037 |
| | (0.013) | (0.011) | (0.009) | (0.010) | (0.011) | (0.011) |
| Fac 6: Extraversion, neuroticism (-), classical fin risk aversion (-) | 0.066 | 0.047 | 0.044 | 0.032 | 0.056 | 0.038 |
| | (0.012) | (0.009) | (0.009) | (0.009) | (0.011) | (0.009) |
| Fac 7: Ambiguity averse, risk averse, PB consumption | 0.025 | 0.023 | 0.031 | 0.005 | 0.032 | 0.030 |
| | (0.015) | (0.011) | (0.011) | (0.011) | (0.013) | (0.011) |
| Fac 8: EGB APR (-), EGB FV, limited attention, OC level, openness | -0.051 | -0.059 | 0.001 | 0.000 | 0.018 | -0.005 |
| | (0.016) | (0.012) | (0.012) | (0.012) | (0.014) | (0.013) |
| Loading cutoff for inclusion in factor scores | \|0.35\| | \|0.35\| | \|0.35\| | \|0.35\| | \|0.35\| | \|0.35\| |
| Controls | parsimonious | parsimonious | parsimonious | parsimonious | parsimonious | parsimonious |
| Mean of LHS | 0.53 | 0.50 | 0.68 | 0.64 | 0.70 | 0.61 |
| Observations | 1684 | 1684 | 1681 | 1612 | 1567 | 1673 |

Each column presents results from a single ORIV regression of the outcome variable described in the column heading on the factor scores and other variables described in the rows, with standard errors clustered on panelist. N consists of data for 845 panelists, surveyed twice in 2014 and 2017. N varies across outcomes due to missing outcome data; see Data Appendix Section 4 for details on outcome measurement. Higher values indicate better outcomes. All factor scores are normalized such that a coefficient shows the marginal effect of a one-SD change in its factor score on an outcome.

**Appendix Table 1. Additional Variables**

| | |
|---|---|
| Outcomes used in Section 4 | |
| Objective financial condition index | Mean across indicators of postive net worth, positive retirement assets, holding equities, having a positive savings rate over the prior 12 months, and not having any of four severe financial hardships during the prior 12 months. Please see Data Appendix for question scripts and sources. |
| Subjective financial condition index | Mean across standardized measures of financial satisfaction, retirement savings adequacy, non-retirement savings adequacy, and lack of financial stress. Please see Data Appendix for question scripts and sources. |
| Life satisfaction I | Standard "… how satisfied are you with your life as a whole these days?" asked in many surveys worldwide. |
| Life Satisfaction II | Within-panelist average of non-missing responses across six ALP modules subsequent to our round 1 modules. Please see Data Appendix for details on module coverage. |
| Happiness index | Within-panelist average of non-missing responses, to two standard questions on happiness in general and in the last 30 days, across five ALP modules subsequent to our Round 1 modules. Please see Data Appendix for details on module coverage. |
| Health status | Within-panelist average of non-missing responses, to standard "Would you say your health is excellent, very good, good, fair, or poor?" question,  across eight ALP modules subsequent to our Round 1 modules. Please see Data Appendix for details on module coverage. |
| Demographics and other additional covariates used to fit biases in Section 2 and as controls in Section 4 | |
| Gender | Indicator, "1" for female. |
| Age | Four categories: 18-34, 35-45, 46-54, 55+ |
| Education | Four categories: HS or less, some college/associates, BA, graduate |
| Income | The ALP's 17 categories (collapsed into deciles in some specifications) |
| Race/ethnicity | Three categories: White, Black, or Other; separate indicator for Hispanic |
| Marital status | Three categories: married/co-habitating; separated/divorced/widowed; never married |
| Household size | Five categories for count of other members: 0, 1, 2, 3, 4+ |
| Employment status | Five categories: working, self-employed, not working, disabled, missing |
| Immigrated to USA | Indicator, "1" for immigrant |
| State of residence | Fixed effects |
| Time spent on questions | Measured for each behavioral elicitation (and other variables), included as decile indicators relative to other respondents |
| Item non-response | Indicators for variables with non-trivial rates of non-response (although all are <5%): Income, employment status, etc. |

**Appendix Table 2. Distinctness: B-indexes are not well-explained by other covariates**

| | (1) | (2) |
|---|---|---|
| *LHS: B-index proportion* | *Full* | *Math* |
| R-squared: All variables below | 0.33 | 0.33 |
| Partial R-squared: Demographics, not including state of residence | 0.19 | 0.23 |
| Partial R-squared: State of residence | 0.05 | 0.07 |
| Partial R-squared: Presumed-classical patience and risk aversion | 0.04 | 0.02 |
| Partial R-squared: Cognitive skills | 0.24 | 0.24 |
| Partial R-squared: Personality traits | 0.02 | 0.01 |
| Partial R-squared: Time spent on behavioral q's, deciles | 0.00 | 0.01 |
| | | |
| mean(LHS) | 0.62 | 0.69 |

LHS variable is a B-proportion: a B-index scaled by the count of its potential behavioral biases with nonmissing data. Each cell presents results from a single OLS regression, using the two observations per panelist from our two rounds of surveying (except for personality traits, where we only have Round 2 data), of the LHS variable described in the column heading on the RHS variables described in the row labels. See Appendix Table 1 for details on demographic variables. We limit the sample to the 845 panelists who completed both of our rounds, and so each regression here has approximately 1,690 observations except for those estimating the personality traits partial r-squareds.

**Appendix Table 3. Distinctness: Single behavioral bias ranks are not well-explained by other consumer characteristics**

| | Cross-round mean of R-squared | | | | | |
|---|---|---|---|---|---|---|
| | Unadj. R-sq | Adjusted R-sq | | | | |
| | All variables | | Subsets of variables | | | |
| **Panel A. Within Bias** | | | Demos | Risk & Time | Cog skills | Personality |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Present-bias: Money discounting | 0.13 | 0.01 | 0.00 | 0.00 | 0.02 | 0.00 |
| Present-bias: Consumption discounting | 0.13 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 |
| Violations of General Axiom of Revealed Preference | 0.15 | 0.03 | 0.03 | -0.01 | 0.03 | 0.01 |
| Violations of GARP and dominance avoidance | 0.21 | 0.10 | 0.06 | 0.00 | 0.08 | 0.02 |
| Narrow bracketing | 0.16 | 0.04 | 0.02 | 0.01 | 0.03 | 0.01 |
| Preference for certainty | 0.16 | 0.00 | 0.01 | 0.00 | 0.03 | -0.02 |
| Loss aversion | 0.18 | 0.07 | 0.03 | 0.05 | 0.02 | 0.00 |
| Ambiguity aversion | 0.15 | 0.03 | 0.02 | 0.00 | 0.01 | 0.00 |
| Overconfidence in level performance | 0.22 | 0.12 | 0.06 | 0.03 | 0.09 | 0.00 |
| Overconfidence in precision | 0.19 | 0.07 | 0.04 | 0.02 | 0.01 | 0.02 |
| Overconfidence in relative performance | 0.30 | 0.22 | 0.11 | 0.07 | 0.16 | 0.03 |
| Underestimates convergence: Non-belief in the law of large numbers | 0.23 | 0.12 | 0.08 | 0.02 | 0.10 | 0.02 |
| Cold Hand Gambler's Fallacy | 0.23 | 0.12 | 0.08 | 0.03 | 0.10 | 0.00 |
| Underestimates APR: Exponential growth bias, loan-side | 0.12 | -0.01 | 0.00 | -0.01 | 0.00 | 0.00 |
| Underestimates future value: Exponential growth bias, asset-side | 0.31 | 0.20 | 0.10 | 0.03 | 0.18 | 0.02 |
| Limited attention | 0.18 | 0.07 | 0.03 | 0.01 | 0.01 | 0.02 |
| Limited memory | 0.14 | 0.02 | 0.01 | 0.00 | 0.01 | 0.01 |
| **Panel B. Across biases** | | | | | | |
| Mean across biases | 0.19 | 0.07 | 0.04 | 0.01 | 0.05 | 0.01 |
| Mean across non-binary bias measures | 0.20 | 0.08 | 0.05 | 0.02 | 0.06 | 0.01 |

Each cell shows the unweighted mean R-squared or adjusted R-squared across two OLS regressions, one per round, of the bias rank described in the row label on flexibly parameterized measures (one bin per response value or decile) of: plausibly exogenous demographics (education, age, gender, immigration, and race and ethnicity), patience and risk aversion (the risk&time column refers to presumed-classical measures of risk aversion and patience), cognitive skills, and personality traits. Money discounting bias regressions drop patience, level overconfidence regressions drop numeracy, and the relative overconfidence regression drops number series (fluid intelligence) so that we are not overfitting by using RHS variables created from same elicitation as the bias measure on the LHS. We only have binary measures of biased discounting of snacks, and of limited memory, and we exclude those measures from the second cross-bias average (reported in the last row). 845 observations per round, but sample sizes are lower than 845 here due to item non-response for biases. Cross-bias statistics in Panel B are unweighted.

**Appendix Table 4. Decision input common factor eigenvalues and explanatory power**

| Common Factor | Eigenvalue | Difference | Proportion | Cumulative |
|---|---|---|---|---|
| Factor 1 | 6.81 | 4.61 | 0.23 | 0.23 |
| Factor 2 | 2.19 | 0.07 | 0.08 | 0.31 |
| Factor 3 | 2.12 | 0.23 | 0.07 | 0.38 |
| Factor 4 | 1.88 | 0.25 | 0.07 | 0.45 |
| Factor 5 | 1.64 | 0.21 | 0.06 | 0.50 |
| Factor 6 | 1.42 | 0.08 | 0.05 | 0.55 |
| Factor 7 | 1.34 | 0.11 | 0.05 | 0.60 |
| Factor 8 | 1.23 | 0.13 | 0.04 | 0.64 |
| Factor 9 | 1.10 | 0.06 | 0.04 | 0.68 |
| Factor 10 | 1.04 | 0.03 | 0.04 | 0.72 |
| Factor 11 | 1.00 | 0.11 | 0.03 | 0.75 |
| Factor 12 | 0.90 | 0.01 | 0.03 | 0.78 |
| Factor 13 | 0.88 | 0.05 | 0.03 | 0.81 |
| Factor 14 | 0.84 | 0.06 | 0.03 | 0.84 |
| Factor 15 | 0.78 | 0.13 | 0.03 | 0.87 |
| Factor 16 | 0.64 | 0.02 | 0.02 | 0.89 |
| Factor 17 | 0.62 | 0.11 | 0.02 | 0.91 |
| Factor 18 | 0.51 | 0.05 | 0.02 | 0.93 |
| Factor 19 | 0.46 | 0.02 | 0.02 | 0.94 |
| Factor 20 | 0.44 | 0.08 | 0.02 | 0.96 |

Common factors estimated using the full set of 29 decision inputs described in Table 1. "Difference" shows the gap between the listed eigenvalue and the next highest. "Proportion" shows the share of variance in characteristics explained by that factor. "Cumulative" shows the share of variance in characteristics explained by all factors together, from the highest on the list to the listed factor.

**Appendix Table 5. ORIV correlations between decision outputs and decision input common factors: Kitchen sink and no controls specifications**

| *Outcome variable* | *Objective financial condition* | | *Subjective financial condition* | | *Life satisfaction I* | | *Life satisfaction II* | | *Happiness* | | *Health status* | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Common factor: loading variables included | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Fac 1: Cog skills (-), EGB FV, OC level & rel, NBLLN, GF | -0.031 | -0.120 | -0.011 | -0.037 | -0.013 | -0.034 | -0.006 | -0.040 | 0.008 | -0.017 | -0.005 | -0.051 |
| | (0.016) | (0.017) | (0.013) | (0.013) | (0.013) | (0.012) | (0.014) | (0.013) | (0.016) | (0.015) | (0.013) | (0.013) |
| Fac 2: GARP violations, narrow bracketing, limited memory | -0.017 | -0.026 | 0.001 | -0.002 | 0.014 | 0.017 | 0.013 | 0.013 | -0.004 | 0.008 | 0.015 | 0.020 |
| | (0.012) | (0.016) | (0.011) | (0.012) | (0.011) | (0.011) | (0.011) | (0.011) | (0.012) | (0.013) | (0.011) | (0.012) |
| Fac 3: Present-biases, patience (-) | -0.026 | -0.016 | -0.026 | -0.026 | -0.014 | -0.011 | 0.001 | 0.005 | -0.008 | -0.004 | 0.003 | -0.005 |
| | (0.011) | (0.014) | (0.010) | (0.011) | (0.010) | (0.010) | (0.010) | (0.010) | (0.011) | (0.012) | (0.010) | (0.011) |
| Fac 4: Preference for certainty, loss aversion (-), GF | 0.008 | -0.018 | -0.007 | -0.023 | -0.019 | -0.028 | 0.006 | -0.005 | -0.027 | -0.035 | 0.000 | -0.018 |
| | (0.014) | (0.018) | (0.012) | (0.012) | (0.012) | (0.012) | (0.012) | (0.013) | (0.014) | (0.014) | (0.012) | (0.013) |
| Fac 5: Lim mem & attent, OC precision, PB cons'n, consc (-), classical risk av (-) | -0.017 | -0.020 | -0.045 | -0.045 | -0.041 | -0.037 | -0.035 | -0.031 | -0.040 | -0.037 | -0.044 | -0.032 |
| | (0.011) | (0.013) | (0.010) | (0.010) | (0.008) | (0.009) | (0.009) | (0.009) | (0.011) | (0.011) | (0.010) | (0.010) |
| Fac 6: Extraversion, neuroticism (-), classical fin risk aversion (-) | 0.027 | 0.066 | 0.031 | 0.048 | 0.031 | 0.037 | 0.018 | 0.025 | 0.042 | 0.046 | 0.021 | 0.030 |
| | (0.010) | (0.012) | (0.009) | (0.009) | (0.008) | (0.008) | (0.009) | (0.009) | (0.011) | (0.010) | (0.009) | (0.009) |
| Fac 7: Ambiguity averse, risk averse, PB consumption | 0.014 | 0.024 | 0.021 | 0.022 | 0.028 | 0.028 | 0.003 | 0.006 | 0.024 | 0.030 | 0.022 | 0.030 |
| | (0.012) | (0.016) | (0.010) | (0.012) | (0.010) | (0.011) | (0.010) | (0.011) | (0.012) | (0.013) | (0.010) | (0.011) |
| Fac 8: EGB APR (-), EGB FV, limited attention, OC level, openness | -0.041 | -0.057 | -0.047 | -0.059 | 0.001 | 0.001 | 0.001 | 0.000 | 0.016 | 0.018 | -0.003 | -0.005 |
| | (0.013) | (0.016) | (0.012) | (0.012) | (0.011) | (0.012) | (0.012) | (0.012) | (0.013) | (0.014) | (0.012) | (0.013) |
| Loading cutoff for inclusion in factor scores | \|0.35\| | \|0.35\| | \|0.35\| | \|0.35\| | \|0.35\| | \|0.35\| | \|0.35\| | \|0.35\| | \|0.35\| | \|0.35\| | \|0.35\| | \|0.35\| |
| Controls | kitchen sink | none | kitchen sink | none | kitchen sink | none | kitchen sink | none | kitchen sink | none | kitchen sink | none |
| Mean of LHS | 0.53 | | 0.50 | | 0.68 | | 0.64 | | 0.70 | | 0.61 | |
| Observations | 1679 | 1684 | 1679 | 1684 | 1677 | 1682 | 1609 | 1614 | 1565 | 1570 | 1669 | 1674 |

Each column presents results from a single ORIV regression of the outcome variable described in the column heading on the factor scores (and other variables) described in the rows, with standard errors clustered on panelist. N consists of data for 845 panelists, surveyed twice in 2014 and 2017. N varies across outcomes due to missing outcome data; see Data Appendix Section 4 for details on outcome measurement. Higher values indicate better outcomes. N falls slightly when income, education, employment status, household size, marital status and survey response time are included in the "kitchen sink" specification, due to missing values for some of those variables. All factor scores are normalized such that a coefficient shows the marginal effect of a one-SD change in its factor score on an outcome.

## Data Appendix

### 1. Measuring Behavioral Biases

This section details, for each of the 17 potential sources of behavioral bias we measure:

i) The motive for eliciting that potential source of bias (B-factor) and the mechanism through which that factor might affect financial condition;

ii) our elicitation method and its key antecedents;

iii) data quality indicators, including item non-response;

iv) sample size (as it compares to that for other B-factors);

v) definitions and prevalence estimates of behavioral *indicators*, with background on the distinctions between expected direction (standard) vs. less-expected (non-standard) direction biases where applicable;

vi) descriptions of the *magnitude* and *heterogeneity* of behavioral deviations, including descriptions of the distribution and—where the data permit—estimates of key parameters used in behavioral models;

Since our empirical work here is purely descriptive, we focus on our Round 1 data (ALP modules 315 and 352) to get the largest possible sample of panelists. We provide comparisons to prior work wherever possible.

*A. Present- or future-biased discounting (money)*

Time-inconsistent discounting has been linked, both theoretically and empirically, to low levels of saving and high levels of borrowing (e.g., Laibson 1997; Meier and Sprenger 2010; Toubia et al. 2013).

We measure discounting biases with respect to money using the Convex Time Budgets (CTB) method created by Andreoni and Sprenger (2012). In our version, fielded in ALP module 315 (the first of our two surveys), subjects make 24 decisions, allocating 100 hypothetical tokens each between (weakly) smaller-sooner and larger-later amounts. See Data Appendix Figure 1 for an example. The 24 decisions are spread across 4 different screens with 6 decisions each. Each screen varies start date (today or 5 weeks from today) x delay length (5 weeks or 9 weeks); each decision within a screen offers a different yield on saving. Among the 1,515 individuals who take our first module in Round 1, 1,502 subjects make at least one CTB choice, and the 1,422 who complete at least the first and last decisions on each of the 4 screens comprise our CTB sample.

The CTB already has been implemented successfully in field contexts in the U.S. (Barcellos and Carvalho 2014; Carvalho, Meier, and Wang 2016) and elsewhere (Giné et al. 2018). In exploring data quality and prevalence below we focus on comparisons to Andreoni and Sprenger (2012), and Barcellos and Carvalho (2014).[1] AS draw their sample from university students. BC's sample is drawn from the ALP, like ours (module 212 in their case), but they use a different adaptation of the CTB.

Indicators of response quality are encouraging for the most part. Interior allocations are more common in our sample than in AS, and comparable to BC. More of our subjects exhibit some variance in their allocations than AS or BC. Our subjects are internally consistent overall—e.g., exhibiting strong correlations in choices across different screens and delay dates—but 41% do exhibit some upward-sloping demand among 20 pairs of decisions, a figure that is within the range commonly found in discount rate elicitations but high compared to the 8% in AS.[2]

We calculate biased discounting, for each individual, by subtracting the consumption rate when the sooner payment date is five weeks from today from the consumption rate when the sooner payment date is today, for each of the two delay lengths. We then average the two differences to get a continuous measure of biased discounting. In keeping with AS, BC and several other recent papers (including Carvalho, Meier, and Wang (2016) and Goda et al. (2019)), we find little if any present-bias on average, with a median discount bias of zero, and a 1pp mean tilt toward future bias.[3]

[1] Carvalho, Meier, and Wang use the American Life Panel like we and Barcello and Carvalho, but on a lower-income sample (ALP module 126).

[2] High rates of non-monotonic demand are not uncommon in discount rate elicitation: Andreoni and Sprenger (2012) report rates ranging from 10 to 50 percent in their literature review. In Barcellos and Carvalho 26% of subjects exhibit some upward-sloping demand, among only 4 pairs of decisions. In our sample non-monotonic demand is strongly correlated within-subject across the four screens, and decreases slightly by the final screen, suggesting that responses are picking up something systematic.

[3] See also Imai et al's (2020) meta-analysis of average estimates (imposing homogeneity in a given sample) of the quasi-hyperbolic discounting model's present-bias parameter. They find "many studies did *not* find strong evidence to reject the null of PB = 1…" (see, e.g., their Figure 1). Bradford et al. (2017) do find present-bias on average in their Qualtrics sample, classifying >50% as present-biased and 26% as future-biased.

Indicators of behavioral deviations here are bi-directional: we label someone as present-biased (future-biased) if the average difference is >0 (<0). We deem present-bias the "standard" direction, since future-bias is relatively poorly understood.[4] Counting any deviation from time-consistent discounting as biased, 26% of our sample is present-biased and 36% is future-biased. These prevalence estimates fall substantially if we set a higher threshold for classifying someone as behavioral; e.g., if we count only deviations > |20|pp, then only 3% of the sample is present-biased and 5% future-biased.

Our prevalence estimates are similar to those from other studies of broad populations that allow for the possibility of future- or present-bias. E.g., BC's CTB elicitation in the ALP shows 29% with any present-bias, and 37% with any future-bias. Carvalho et al (2019) find 28% with any present-bias and 31% with any future-bias in a sample of account aggregation software users in Iceland.[5]

*B. Present- or future-biased discounting (food)*

In light of evidence that discounting can differ within-subject across domains (e.g., Augenblick, Niederle, and Sprenger 2015), we also obtain a coarse measure of discounting biases for consumption per se, by asking two questions that follow Read and van Leeuwen (1998) : *"Now imagine that you are given the choice of receiving one of two snacks for free, [right now/five weeks from now]. One snack is more delicious but less healthy, while the other is healthier but less delicious. Which would you rather have [right now/five weeks from now]: a delicious snack that is not good for your health, or a snack that is less delicious but good for your health?* We fielded these questions in our second Round1 module.

Of the 1427 persons taking our second survey, 1423 answer one of the two snack questions, and 1404 respond to both. 61% choose the healthy snack for today, while 68% choose it for five weeks in the future, with 15% exhibiting present bias (consume treat today, plan to eat healthy in the future) and 7% future bias (consume healthy today, plan to eat treat in the future).[6] Barcellos

[4] Although see Koszegi and Szeidl (2013) for a theory of future-biased discounting.
[5] Goda et al. use a different elicitation method—a "time-staircase" multiple price list (Falk et al. 2018)—and classify 55% of their nationally representative sample (from the ALP and another online panel) as present-biased. In the AS sample 14% exhibit any present-bias and 12% any future-bias.
[6] If we limit the sample to those who did not receive the informational/debiasing treatment about self-control in ALP module 212 (Barcellos and Carvalho), we find 15% with present bias and 8% with future bias (N=748).

and Carvalho's ALP subjects answered similar questions in their baseline survey, albeit with only a one-week instead of a five-week delay, with 6% exhibiting present-bias and 9% future-bias. Read and van Leeuwen (1998) offer actual snacks to a convenience sample of employees in Amsterdam but do not calculate individual-level measures of bias. They do find substantial present-bias on average. We do not know of any prior work estimating correlations between measures of consumption discounting biases and field outcomes.

*C. Inconsistency with General Axiom of Revealed Preference (and dominance avoidance)*

Our third and fourth behavioral factors follow Choi et al. (2014), which measures choice inconsistency with standard economic rationality. Choice inconsistency could indicate a tendency to make poor (costly) decisions in field contexts; indeed, Choi et al. (2014) find that more choice inconsistency is conditionally correlated with less wealth in a representative sample of Dutch households.

We use the same task and user interface as in Choi et al. (2014) but abbreviate it from 25 decisions to 11.[7] Each decision confronts respondents with a linear budget constraint under risk: subjects choose a point on the line, and then the computer randomly chooses whether to pay the point value of the x-axis or the y-axis. 1,270 of the 1,427 individuals taking our second Round 1 module make all 11 decisions, and comprise our sample for measuring choice inconsistency.[8] See Data Appendix Figure 2 for an example.

Following Choi et al., we average across these 11 decisions, within-consumer, to benchmark choices against two different standards of rationality. One benchmark is a complete and transitive preference ordering adhering to the General Axiom of Revealed Preference (GARP), as captured by the Afriat (1972) Critical Cost Efficiency Index. 1-CCEI can be interpreted as the subject's degree of choice inconsistency: the percentage points of potential earnings "wasted" per the GARP standard. But as Choi et al. discuss, consistency with GARP is not necessarily the most appealing measure of decision quality because it allows for violations of monotonicity with

[7] We were quite constrained on survey time and hence conducted a pilot in which we tested the feasibility of capturing roughly equivalent information with fewer rounds. 58 pilot-testers completed 25 rounds, and we estimated the correlation between measures of choice inconsistency calculated using the full 25 rounds, and just the first 11 rounds. These correlations are 0.62 and 0.88 for the two key measures.

[8] 1424 individuals view at least one of the instruction screens, 1,311 are recorded as completing at least one round of the task, and 1,270 are recorded as completing each of the 11 rounds.

respect to first-order stochastic dominance (FOSD).[9] Hence, again following Choi et al., our second measure captures inconsistency with both GARP and FOSD.[10] Note that these measures of inconsistency are unidirectional: there is no such thing as being *overly* consistent.

Our distribution of individual-level CCEI estimates is nearly identical to Choi et al.'s— if we use only the first 11 rounds of choices from Choi et al. to maximize comparability to our setup. Our median (1-CCEI) is 0.002, suggesting nearly complete consistency with GARP. The mean is 0.05. The median (1-combined-CCEI), capturing FOSD violations as well, is 0.10, with a mean of 0.16. Choice inconsistency is substantially higher when using the full 25 rounds in both our pilot data and Choi et al. (e.g., mean CCEI of 0.12 in both samples), and we have verified that this is a mechanical effect (more rounds means more opportunities to exhibit inconsistency) rather than deterioration in consistency as rounds increase, by finding that CCEIs measured over small blocks of consecutive rounds remain constant as the average round number of those blocks increases.

Our prevalence estimates are also nearly identical to those from the Choi et al (2014) data. In our data, 53% of subjects exhibit any inconsistency with GARP, and 96% exhibit any inconsistency with GARP or FOSD. If we set a 20pp threshold for classifying someone as inconsistent, only 7% are inconsistent with GARP, and 31% are inconsistent with GARP or FOSD. Looking more directly at heterogeneity, we see standard deviations of 0.08 and 0.18, and $10^{th}$-$90^{th}$ percentile ranges of 0.16 and 0.41.

*D. Risk attitude re: certainty (certainty premium)*

Behavioral researchers have long noted a seemingly disproportionate preference for certainty (PFC) among some consumers and posited various theories to explain it: Cumulative Prospect Theory (Daniel Kahneman and Tversky 1979; Tversky and Kahneman 1992), Disappointment Aversion (Bell 1985; Loomes and Sugden 1986; Gul 1991), and u-v preferences (Neilson 1992;

---

[9] E.g., someone who always allocates all tokens to account X is consistent with GARP if they are maximizing the utility function U(X, Y)=X. Someone with a more normatively appealing utility function—that generates utility over tokens or consumption per se—would be better off with the decision rule of always allocating all tokens to the cheaper account.

[10] The second measure calculates 1-CCEI across the subject's 11 actual decisions and "the mirror image of these data obtained by reversing the prices and the associated allocation for each observation" (Choi et al. p. 1528), for 22 data points per respondent in total.

Schmidt 1998; Diecidue, Schmidt, and Wakker 2004). PFC may help to explain seemingly extreme risk averse behavior, which could in turn lead to lower wealth in the cross-section.

We use Callen et al.'s (2014) two-task method[11] for measuring a subject's *certainty premium* (CP).[12] Similar to Holt and Laury tasks, in one of the Callen et al. tasks subjects make 10 choices between two lotteries, one a (p, 1-p) gamble over X and Y > X , (p; X, Y), the other a (q, 1-q) gamble over Y and 0, (q; Y, 0). Both Callen et al. and we fix Y and X at 450 and 150 (hypothetical dollars in our case, hypothetical Afghanis in theirs), fix p at 0.5, and have q range from 0.1 to 1.0 in increments of 0.1. In the other task, p = 1, so the subject chooses between a lottery and a certain option. Our two tasks are identical to Callen et al.'s except for the currency units. But our settings, implementation, and use of the elicited data are different. Callen et al. administer the tasks in-person, using trained surveyors, at polling centers and homes in Afghanistan. They use the data to examine the effects of violence on risk preferences.

1,463 of 1,505 (97%) of our subjects who started the tasks completed all 20 choices (compared to 977/1127 = 87% in Callen et al.). As is typical with Holt-Laury tasks, we exclude some subjects whose choices indicate miscomprehension of or inattention to the task. 11% of our subjects multiple-switch on our two-lottery task (compared to 10% in Callen et al.), and 9% of our subjects multiple-switch on the lottery vs. certain option tasks (compared to 13% in Callen et al.). 14% of our subjects switch too soon for monotonic utility in the two-lottery—in rows [2, 4] in the two-lottery task—compared to 13% in Callen et al. All told, 19% of our subjects exhibit a puzzling switch (17% in Callen et al.), leaving us with 1,188 usable observations. Of these subjects, 1,049 switch on both tasks, as is required to estimate CP. Of these 1,049, only 30% switch at the same point on both tasks, in contrast to 63% in Callen et al.

We estimate CP for each respondent i by imputing the likelihoods q* at which i expresses indifference as the midpoint of the q interval at which i switches, and then using the two likelihoods to estimate the indirect utility components of the CP formula. As Callen et al. detail, the CP "is defined in probability units of the high outcome, Y, such that one can refer to certainty of X being worth a specific percent chance of Y relative to its uncertain value." We estimate a

[11] Callen et al. describes its task as "a field-ready, two-question modification of the uncertainty equivalent presented in Andreoni and Sprenger (2016)."

[12] The Callen et al. tasks also elicit non-parametric measures of classical risk aversion: a higher switch point indicates greater risk aversion. We discuss these measures in Section 1-D of the paper.

mean CP of 0.16 in our sample (SD=0.24, median =0.15), compared to 0.37 (SD=0.15) in Callen et al. Their findings suggest that much of the difference could be explained by greater exposure to violence in their sample.

As Callen et al. detail, the sign of CP also carries broader information about preferences. CP = 0 indicates an expected utility maximizer. CP>0 indicates a preference for certainty (PFC), as in models of disappointment aversion or u-v preferences. We classify 77% of our sample as PFC type based on an any-deviation threshold. This falls to 73%, 60%, or 42% if we count only larger deviations >0 (5pp, 10pp, or 20pp) as behavioral. In Callen et al. 99.63% of the sample exhibits PFC. CP<0 indicates a cumulative prospect theory (CPT) type, and we classify 23%, 20%, 13% or 7% as CPT under the different deviation thresholds. We denote PFC as the standard bias, simply because CP>0 is far more common than CP<0 in both our data and Callen et al.'s.

*E. Loss aversion/small-stakes risk aversion*

Loss aversion refers to placing higher weight on losses than gains, in utility terms. It is one of the most influential concepts in the behavioral social sciences, with seminal papers—e.g., Tversky and Kahneman (1992) and Benartzi and Thaler (1995)—producing thousands of citations. Loss aversion has been implicated in various portfolio choices (Barberis 2013) and consumption dynamics (Kőszegi and Rabin 2009) that can lead to lower wealth.

We measure loss aversion using the two choices developed by Fehr and Goette (2007) in their study of the labor supply of bike messengers (see Abeler et al. (2011) for a similar elicitation method). Choice 1 is between a lottery with a 50% chance of winning $80 and a 50% chance of losing $50, and zero dollars. Choice two is between playing the lottery in Choice 1 six times, and zero dollars. As Fehr and Goette (FG) show, if subjects have reference-dependent preferences, then subjects who reject lottery 1 have a higher level of loss aversion than subjects who accept lottery 1, and subjects who reject both lotteries have a higher level of loss aversion than subjects who reject only lottery 1. In addition, if subjects' loss aversion is consistent across the two lotteries, then any individual who rejects lottery 2 should also reject lottery 1 because a rejection of lottery 2 implies a higher level of loss aversion than a rejection of only lottery 1. Other researchers have noted that, even in the absence of loss aversion, choosing Option B is compatible with small-stakes risk aversion.[13] We acknowledge this but use "loss aversion" instead of "loss aversion and/or small-stakes risk aversion" as shorthand. Small-stakes risk aversion is also often classified as behavioral because it is incompatible with expected utility theory (Rabin 2000).

Response rates suggest a high level of comfort with these questions; only two of our 1,515 subjects skip, and only two more who answer the first question do not answer the second. 37% of our 1,511 respondents reject both lotteries, consistent with relatively extreme loss aversion, compared to 45% of FG's 42 subjects. Another 36% of our subjects accept both lotteries, consistent with classical behavior, compared to 33% in FG. The remaining 27% of our subjects (and 21% of FG's) exhibit moderate loss aversion, playing one lottery but not the other, with our main difference from FG being that 14% of our subjects (vs. only 2% of theirs) exhibit the puzzling behavior of playing lottery 1 but not lottery 2. Although one wonders whether these 14% misunderstood the questions, we find only a bit of evidence in support of that interpretation: those playing the single but not compound lottery have slightly lower cognitive skills than other loss averters, conditional on our rich set of covariates, but actually have higher cognitive skills than the most-classical group. And playing the single but not the compound lottery is uncorrelated with our measure of ambiguity aversion, pushing against the interpretation that the compound lottery is sufficiently complicated as to appear effectively ambiguous (Dean and Ortoleva 2019).

All told 64% of our subjects indicate some loss aversion, defined as rejecting one or both small-stakes lotteries, as do 67% in FG. In Abeler et al.'s (2011) student sample, 87% reject one or more of the four small-stakes lotteries with positive expected value. The Abeler et al. questions were also fielded in an ALP module from early 2013 used by Hwang (2016); 70% of that sample exhibits some loss aversion. In von Gaudecker et al.'s nationally representative Dutch sample, 86% exhibit some loss aversion, as inferred from structural estimation based on data from multiple price lists. We also order sets of deviations to indicate greater degrees of loss aversion, based on whether the individual respondent rejects the compound but not the single lottery, rejects the single but not the compound lottery, or rejects both.

[13] A related point is that there is no known "model-free" method of eliciting loss aversion (Dean and Ortoleva 2019).

## F. *Narrow bracketing and dominated choice*

Narrow bracketing refers to the tendency to make decisions in (relative) isolation, without full consideration of other choices and constraints. Rabin and Weizsacker (2009) show that narrow bracketing can lead to dominated choices—and hence expensive and wealth-reducing ones—given non-CARA preferences.

We measure narrow bracketing and dominated choice (NBDC) using two of the tasks in Rabin and Weizacker (2009). Each task instructs the subject to make two decisions. Each decision presents the subject with a choice between a certain payoff and a gamble. Each decision pair appears on the same screen, with an instruction to consider the two decisions jointly. RW administer their tasks with students and, like us, in a nationally representative online panel (Knowledge Networks in their case). Like us, payoffs are hypothetical for their online panel.

Our first task follows RW's Example 2, with Decision 1 between winning $100 vs. a 50-50 chance of losing $300 or winning $700, and Decision 2 between losing $400 vs. a 50-50 chance of losing $900 or winning $100.[14] As RW show, someone who is loss averse and risk-seeking in losses will, in isolation (narrow bracketing) tend to choose A over B, and D over C. But the combination AD is dominated with an expected loss of $50 relative to BC. Hence a broad-bracketer will never choose AD. 29% of our subjects choose AD, compared to 53% in the most similar presentation in RW.

Our second task reproduces RW's Example 4, with Decision 1 between winning $850 vs. a 50-50 chance of winning $100 or winning $1,600, and Decision 2 between losing $650 vs. a 50-50 chance of losing $1,550 or winning $100. As in task one, a decision maker who rejects the risk in the first decision but accepts it in the second decision (A and D) violates dominance, here with an expected loss of $75 relative to BC. 23% of our subjects choose AD, compared to 36% in the most similar presentation in RW. As RW discuss, a new feature of task two is that AD sacrifices expected value in the second decision, not in the first. This implies that for all broad-bracketing risk averters AC is optimal: it generates the highest available expected value at no

[14] Given the puzzling result that RW's Example 2 was relatively impervious to a broad-bracketing treatment, we changed our version slightly to avoid zero-amount payoffs. Thanks to Georg Weizsacker for this suggestion.

variance. 50% of our subjects choose AC, compared to only 33% in the most similar presentation in RW. I.e., 50% of our subjects do NOT broad-bracket in this task, compared to 67% in RW.

Reassuringly, responses across our two tasks are correlated; this is especially reassuring given that the two tasks appear non-consecutively in the survey, hopefully dampening any tendency for a mechanical correlation. E.g., the unconditional correlation between choosing AD across the two tasks is 0.34.

1,486 subjects complete both tasks (out of the 1,515 who respond to at least one of our questions in module 315). Putting the two tasks together to create summary indicators of narrow bracketing, we find 59% of our subjects exhibiting some narrow bracketing in the sense of not broad-bracketing on both tasks, while 13% narrow-bracket on both tasks. These are uni-directional indicators: we either classify someone as narrow-bracketing, or not. RW do not create summary indicators across tasks, but, as noted above, their subjects exhibit substantially more narrow bracketing at the task level than our subjects do.

*G. Ambiguity aversion*

Ambiguity aversion refers to a preference for known uncertainty over unknown uncertainty—preferring, for example, a less-than-50/50 gamble to one with unknown probabilities. It has been widely theorized that ambiguity aversion can explain various sub-optimal portfolio choices, and Dimmock et al. (2016) find that it is indeed conditionally correlated with lower stockholdings and worse diversification in their ALP sample (see also Dimmock, Kouwenberg, and Wakker (2016)).

We elicit a coarse measure of ambiguity aversion using just one or two questions about a game that pays $500 if you select a green ball. The first question offers the choice between a Bag One with 45 green and 55 yellow balls vs. a Bag Two of unknown composition. 1,397 subjects respond to this question (out of 1,427 who answer at least one of our questions on ALP module 352). 73% choose the 45-55 bag, and we label them ambiguity averse. The survey then asks these subjects how many green balls would need to be in Bag One to induce them to switch.[15] We subtract this amount from 50, dropping the 99 subjects whose response to the second

[15] Because not everyone answers the second question, we measure time spent responding to the ambiguity aversion elicitation using only the first question.

question is >45 (and the 10 subjects who do not respond), to obtain a continuous measure of ambiguity aversion that ranges from 0 (not averse in the first question) to 50 (most averse=== the three subjects who respond “zero” to the second question). The continuous measure (N=1,288) has a mean of 14 (median=10), and a SD of 13. If we impose a large-deviation threshold of 10 (20% of the max) for labeling someone as ambiguity averse, 50% of our sample exceeds this threshold and another 16% are at the threshold. Our elicitation does not distinguish between ambiguity-neutral and ambiguity-seeking choices (for more comprehensive but still tractable methods see, e.g., Dimmock, Kouwenberg et al. (2016), Dimmock, Kouwenberg, and Wakker (2016), Gneezy et al. (2015)), and so our measure of deviation from ambiguity-neutrality is one-sided.

Despite the coarseness of our elicitation, comparisons to other work suggest that it produces reliable data. Our ambiguity aversion indicator correlates with one constructed from Dimmock et al.’s elicitation in the ALP (0.14, p-value 0.0001, N=789), despite the elicitations taking place roughly 3 years apart. Prevalence at our 10pp large-deviation cutoff nearly matches that from Dimmock, Kouwenberg et al.’s (2016) ALP sample and Butler et al.’s (2014) Unicredit Clients’ Survey sample from Italy, and our prevalence of any ambiguity aversion, 0.73 is similar to Dimmock, Kouwenberg, and Wakker’s (2016) 0.68 from the Dutch version of the ALP .

*H. Overconfidence: Three varieties*

Overconfidence has been implicated in excessive trading (Daniel and Hirshleifer 2015), over-borrowing on credit cards (Ausubel 1991), paying a premium for private equity (Moskowitz and Vissing-Jorgensen 2002; although see Kartashova 2014), and poor contract choice (Grubb 2015), any of which can reduce wealth and financial security.

We elicit three distinct measures of overconfidence, following e.g., Moore and Healy (2008).

The first measures it in level/absolute terms, by following the three Banks and Oldfield numeracy questions, in our second Round 1 module, with the question: *“How many of the last 3 questions (the ones on the disease, the lottery and the savings account) do you think you got correct?”* We then subtract the respondent’s assessment from her actual score. 39% of 1,366 subjects are overconfident (“overestimation” per Moore and Healy) by this measure (with 32% overestimating by one question), while only 11% are underconfident (with 10% underestimating by one question). Larrick et al. (2007), Moore and Healy, and other studies use this method for measuring overestimation, but we are not aware of any that report individual-level prevalence estimates (they instead focus on task-level data, sample-level summary statistics, and/or correlates of cross-sectional heterogeneity in estimation patterns).

The second measures overconfidence in precision, as indicated by responding "100%" on two sets of questions about the likelihoods (of different possible Banks and Oldfield quiz scores or of future income increases). This is a coarse adaptation of the usual approaches of eliciting several confidence intervals or subjective probability distributions (Moore and Healy). In our data 34% of 1,345 responding to both sets respond 100% on >=1 set, and 10% on both.

The third measures confidence in placement (relative performance), using a self-ranking elicited before taking our number series test: *"We would like to know what you think about your intelligence as it would be measured by a standard test. How do you think your performance would rank, relative to all of the other ALP members who have taken the test?"* We find a better-than-average effect in the sample as a whole (70% report a percentile>median) that disappears when we ask the same question immediately post-test, still not having revealed any scores (50% report a percentile>median). We also construct an individual-level measure of confidence in placement by subtracting the subject's actual ranking from his pre-test self-ranking (N=1,395). This measure is useful for capturing individual-level heterogeneity ordinally, but not for measuring prevalence because the actual ranking is based on a 15-question test and hence its percentiles are much coarser than the self-ranking.

*I. Non-belief in the Law of Large Numbers*

Under-weighting the importance of the Law of Large Numbers (LLN) can affect how individuals treat risk (as in the stock market), or how much data they demand before making decisions. In this sense non-belief in LLN (a.k.a. NBLLN) can act as an "enabling bias" for other biases like loss aversion (Benjamin, Rabin, and Raymond 2016).

Following Benjamin, Moore, and Rabin (see also D Kahneman and Tversky 1972; Benjamin, Rabin, and Raymond 2016), we measure non-belief in law of large numbers (NBLLN) using responses to the following question:

*… say the computer flips the coin 1000 times, and counts the total number of heads. Please tell us what you think are the chances, in percentage terms, that the total number of heads will lie within the following ranges. Your answers should sum to 100.*

The ranges provided are [0, 480], [481, 519], and [520, 1000], and so the correct answers are 11, 78, 11.

1,375 subjects respond (out of the 1,427 who answer at least one of our questions in Module 352),[16] with mean (SD) responses of 27 (18), 42 (24), and 31 (20). We measure NBLLN using the distance between the subject's answer for the [481, 519] range and 78. Only one subject gets it exactly right. 87% underestimate; coupled with prior work, this result leads us to designate underestimation as the "standard" directional bias. The modal underestimator responds with 50 (18% of the sample). The other most-frequent responses are 25 (10%), 30 (9%), 33 (8%), and 40 (7%). Few underestimators—only 4% of the sample—are within 10pp of 78, and their mean distance is 43, with an SD of 17. 9% of the sample underestimates by 20pp or less. 13% overestimate relative to 78, with 5% of the sample quite close to correct at 80, and another 5% at 100. Benjamin, Moore, and Rabin (2017) do not calculate individual-level measures of underestimation or overestimation in their convenience sample, but do report that the sample means are 35%, 36%, and 29% for the three bins. The comparable figures in our data are 27%, 42%, and 31%.

### *J. Gambler's Fallacies*

The Gambler's Fallacies involve falsely attributing statistical dependence to statistically independent events, in either expecting one outcome to be less likely because it has happened recently (recent reds on roulette make black more likely in the future) or the reverse, a "hot hand" view that recent events are likely to be repeated. Gambler's fallacies can lead to overvaluation of financial expertise (or attending to misguided financial advice), and related portfolio choices like the active-fund puzzle, that can erode wealth (Rabin and Vayanos 2010).

We take a slice of Benjamin, Moore, and Rabin's (2017) elicitation for the fallacies:

[16] Only 26 subjects provide responses that do not sum to 100 after a prompt, and each response for an individual range is [0, 100], so we do not exclude any subjects from the analysis here.

*"Imagine that we had a computer "flip" a fair coin… 10 times. The first 9 are all heads. What are the chances, in percentage terms, that the 10th flip will be a head?"*

1,392 subjects respond, out of the 1,427 respondents to module 352. The cold-hand fallacy implies a response < 50%, while the hot-hand fallacy implies a response > 50%. Our mean response is 45% (SD=25), which is consistent with the cold-hand but substantially above the 32% in Benjamin, Moore, and Rabin. Another indication that we find less evidence of the cold-hand fallacy is that, while they infer that "at the individual level, the gambler's fallacy [cold-hand] appears to be the predominant pattern of belief" (2013, p. 16), we find only 26% answering < "50." 14% of our sample responds with >"50" (over half of these responses are at "90" or "100"). So 60% of our sample answers correctly. Nearly everyone who responds with something other than "50" errs by a substantial amount—e.g., only 2 % of the sample is [30, 50) or (50, 70]. Sixteen percent of our sample answers "10,"[17] which Benjamin, Moore, and Rabin speculates is an indicator of miscomprehension; we find that while subjects with this indicator do have significantly lower cognitive skills than the unbiased group, they actually have higher cognitive skills than the rest of subjects exhibiting a gambler's fallacy.

Dohmen et al. (2009) measure the fallacies using a similar elicitation that confronts a representative sample of 1,012 Germans, taking an in-person household survey, with:

*Imagine you are tossing a fair coin. After eight tosses you observe the following result: tails-tails-tails-heads-tails-heads-heads-heads. What is the probability, in percent, that the next toss is "tails"?*

986 of Dohmen et al.'s respondents provide some answer to this question, 95 of whom say "Don't know." Among the remaining 891, 23% exhibit cold-hand (compared to 26% in our sample), and 10% exhibit hot-hand (compared to 14% in our sample). Conditional on exhibiting cold-hand, on average subjects err by 29pp (40 pp in our sample). Conditional on exhibiting hot-hand, the mean subject error is 27pp (39pp in our sample).

[17] 34% of the sample in Benjamin, Moore, and Raymond respond "10%" on one or more of their ten questions.

*K. Exponential growth bias: Two varieties*

Exponential growth bias (EGB) produces a tendency to underestimate the effects of compounding on costs of debt and benefits of saving. It has been linked to a broad set of financial outcomes (Levy and Tasoff 2016; Stango and Zinman 2009).

We measure EGB, following previous papers, by asking respondents to solve questions regarding an asset's future value or a loan's implied annual percentage rate. Our first measure of EGB follows in the spirit of Stango and Zinman (2009; 2011) by first eliciting the monthly payment the respondent would expect to pay on a $10,000, 48 month car loan. The survey then asks "… What percent rate of interest does that imply in annual percentage rate ("APR") terms?" 1,445 panelists answer both questions, out of the 1,515 respondents to Module 315. Most responses appear sensible given market rates; e.g., there are mass points at 5%, 10%, 3%, 6% and 4%.

We calculate an individual-level measure of "debt-side EGB" by comparing the difference between the APR *implied* by the monthly payment supplied by that individual, and the *perceived* APR as supplied directly by the same individual. We start by binning individuals into under-estimators (the standard bias), over-estimators, unbiased, and unknown (15% of the sample).[18] The median level difference between the correct and stated value is 500bp, with a mean of 1,042bp and SD of 1,879bp. Among those with known bias, we count as biased 51% and 34% as negatively biased (overestimating APR) under error tolerance of zero. This is less EGB than Stango and Zinman (2009; 2011) see from questions in the 1983 Survey of Consumer Finances, where 98% of the sample underestimates, and the mean bias is 1,800bp or 3,800bp depending on the benchmark. The time frames of the questions differ, which may account for the difference (and is why we do not estimate an EGB structural model parameter to compare with our prior work or that of Levy and Tasoff).

Our second measure of EGB comes from a question popularized by Banks and Oldfield (2007) as part of a series designed to measure basic numeracy: "Let's say you have $200 in a savings account. The account earns 10 percent interest per year. You don't withdraw any money

[18] Non-response is relatively small, as only 4% of the sample does not respond to both questions. 7% state payment amounts that imply a negative APR, even after being prompted to reconsider their answer. We also classify the 4% of respondents with implied APRs >=100% as having unknown bias.

for two years. How much would you have in the account at the end of two years?" 1,389 subjects answer this question (out of the 1,427 respondents to Module 352), and we infer an individual-level measure of "asset-side EGB" by comparing the difference between the correct future value ($242), and the future value supplied by the same individual.[19] We again bin individuals into underestimators (the standard bias), overestimators, unbiased, and unknown (14% of the sample).[20] Among those with known bias (N=1,222), the median bias is $0, with a mean of $2 and SD of $14.[21] 44% of our sample provides the correct FV. 47% of our sample underestimates by some amount, with most underestimators (29% of the sample) providing the linearized (uncompounded) answer of $240. Nearly all other underestimates provide an answer that fails to account for even simple interest; the most common reply in this range is "$220." Only 9% of our sample overestimates the FV, with small mass points at 244, 250, 400, and 440.

Other papers have used the Banks and Oldfield question, always—to our knowledge—measuring accuracy as opposed to directional bias and then using a 1/0 measure of correctness as an input to a financial literacy or numeracy score (e.g., James Banks, O'Dea, and Oldfield 2010; Gustman, Steinmeier, and Tabatabai 2012). Our tabs from the 2014 Health and Retirement Study suggest, using only the youngest HRS respondents and our oldest respondents to maximize comparability (ages 50-60 in both samples), that there is substantially more underestimation in the HRS (74%, vs. 48% in our sample). 14% overestimate in the HRS among those aged 50-60, vs. 9% in our sample.

Goda et al. (2019) and Levy and Tasoff (2016) measure asset-side EGB using more difficult questions in their representative samples. They find that 9% and 11% overestimate FVs, while 69% and 85% underestimate. We do not construct an EGB parameter to compare to theirs, because our questions lack their richness and yield heavy mass points at unbiased and linear-biased responses.

[19] Responses to this question are correlated with responses to two other questions, drawn from Levy and Tasoff (2016), that can also be used to measure asset-side EGB, but our sample sizes are smaller for those two other questions and hence we do not use them here.

[20] We label as unknown the 8% of the sample answering with future value < present value, the 3% of the sample answering with a future value > 2x the correct future value, and the 3% of the sample who skip this question.

[21] For calculating the mean and SD we truncate bias at -42 for the 4% sample answering with future values 284<FV<485, to create symmetric extrema in the bias distribution since our definition caps bias at 42.

*L. Limited attention and limited memory*

Prior empirical work has found that limited attention affects a range of financial decisions (e.g., Barber and Odean 2008; DellaVigna and Pollet 2009; Karlan et al. 2016; Stango and Zinman 2014). Behavioral inattention is a very active line of theory inquiry as well (e.g., Bordalo, Gennaioli, and Shleifer forthcoming; Kőszegi and Szeidl 2013; Schwartzstein 2014).

In the absence of widely used methods for measuring limited attention and/or memory, we create our own, using five simple questions and tasks.

The first three ask, “Do you believe that your household's [horizon] finances… would improve if your household paid more attention to them?” for three different horizons: “day-to-day (dealing with routine expenses, checking credit card accounts, bill payments, etc.)” “medium-run (dealing with periodic expenses like car repair, kids’ activities, vacations, etc.)” and “long-run (dealing with kids' college, retirement planning, allocation of savings/investments, etc.)” Response options are the same for each of these three questions: “Yes, and I/we often regret not paying greater attention” (26%, 23%, and 35%), “Yes, but paying more attention would require too much time/effort” (8%, 11%, and 12%), “No, my household finances are set up so that they don't require much attention” (15%, 16%, and 13%), and “No, my household is already very attentive to these matters” (52%, 51%, and 41%). We designed the question wording and response options to distinguish behavioral limited inattention (“Yes… I/we often…”)—which also includes a measure of awareness thereof in “regret”—from full attention (“… already very attentive”), rational inattention, and/or a sophisticated response to behavioral inattention (“Yes, but… too much time/effort”; “… set up so that they don’t require much attention”).

Responses are strongly but not perfectly correlated (ranging 0.56 to 0.69 among pairwise expressions of regret). A fourth measure of limited attention is also strongly correlated with the others, based on the question: “Do you believe that you could improve the prices/terms your household typically receives on financial products/services by shopping more?”[22] 18% respond “Yes, and I/we often regret not shopping more,” and the likelihood of this response is correlated 0.25 with each of the regret measures above. 1,483 subjects answer all four questions, out of the

[22] This question is motivated by evidence that shopping behavior strongly predicts borrowing costs (Stango and Zinman 2016).

1,515 respondents to Module 315. Summing the four indicators of attentional regret, we find that 49% of subjects have one or more (earning a classification of behavioral inattention), 29% have two or more, 19% three or more, and only 6% have all four.

We also seek to measure limited prospective memory, following previous work suggesting that limited memory entails real costs like forgetting to redeem rebates (e.g., Ericson 2011). We offer an incentivized task to subjects taking module 352: "The ALP will offer you the opportunity to earn an extra $10 for one minute of your time. This special survey has just a few simple questions but will only be open for 24 hours, starting 24 hours from now. During this specified time window, you can access the special survey from your ALP account. So we can get a sense of what our response rate might be, please tell us now whether you expect to do this special survey." 97% say they intend to complete the short survey, leaving us with a sample of 1,358. Only 14% actually complete the short survey.

Our indicator of behavioral limited memory— (not completing the follow-up task conditional on intending to complete)—is a bit coarse. We suspect that some noise is introduced because our elicitation makes it costless to express an intention to complete (in future research we plan to explore charging a small "sign up" fee), thereby including in the indicator's sample frame some subjects who rationally do not complete the task. Relatedly, although we set the payoff for task completion to be sufficiently high to dominate any attention/memory/time costs in *marginal* terms for most subjects (the effective hourly wage is in the hundreds of dollars), it may well be the case that the *fixed* cost exceeds $10 for some respondents.

## 2. Measuring Classical Decision Inputs

*A. Patience and Risk Aversion*

We measure patience using the average savings rate across the 24 choices in our version of the Convex Time Budget task described in Data Appendix Section 1-A.

One risk aversion measure comes from Barsky et al. (1997), a leading example of the "lottery-choice" class of risk elicitations (e.g., Mata et al. 2018). This task starts with: "…. Suppose that you are the only income earner in the family. Your doctor recommends that you move because of allergies, and you have to choose between two possible jobs. The first would guarantee your current total family income for life. The second is possibly better paying, but the income is also less certain. There is a 50% chance the second job would double your current total family income for life and a 50% chance that it would cut it by a third. Which job would you take—the first job or the second job?" Those taking the risky job are then faced with a 50% probability that it cuts it by one-half (and, if they still choose the risky job, by 75%). Those taking the safe job are then faced with lower expected downsides to the risky job (50% chance of 20% decrease, and then, if they still choose the safe job, a 50% chance of a 10% decrease). We create separate bins for each possible combination of choices and use either a linear scale (with more higher values indicating more risk aversion) or the separate bins, depending on the specification.

Our second risk aversion measure comes from Dohmen et al. (2010; 2011), a leading example of the "stated" or "self-report" class of risk aversion elicitations (e.g., Mata et al. 2018). The question asks: "How do you see yourself: Are you generally a person who is fully prepared to take financial risks", and we transform the 100-point response scale so that higher values indicate greater risk aversion.

*B. Cognitive Skills*

We measure fluid intelligence using a 15-question, non-adaptive number series (McArdle, Fisher, and Kadlec 2007). Number series scores correlate strongly with those from other fluid intelligence tests like IQ and Raven's.

We measure numeracy using: “If 5 people split lottery winnings of two million dollars ($2,000,000) into 5 equal shares, how much will each of them get?” and “If the chance of getting a disease is 10 percent, how many people out of 1,000 would be expected to get the disease?” (Banks and Oldfield 2007). Response options are open-ended. These questions have been used in economics as numeracy and/or financial literacy measures since their deployment in the 2002 English Longitudinal Study of Ageing, with subsequent deployment in the Health and Retirement Study and other national surveys.

We measure financial literacy using Lusardi and Mitchell’s (2014) “Big Three”: “Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow?”; “Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, how much would you be able to buy with the money in this account?”; and “Please tell me whether this statement is true or false: "Buying a single company's stock usually provides a safer return than a stock mutual fund." Response options are categorical.

We measure executive function using a two-minute Stroop task (MacLeod 1991). Our version displays the name of a color on the screen (red, blue, green, or yellow) and asks the subject to click on the button corresponding to the color the word is printed in (red, blue, green, or yellow; not necessarily corresponding to the color name). Answering correctly tends to require using conscious effort to override the tendency (automatic response) to select the name rather than the color. The Stroop task is sufficiently classic that the generic failure to overcome automated behavior (in the game “Simon Says,” when an American crosses the street in England, etc.) is sometimes referred to as a “Stroop Mistake” (Camerer 2007). Before starting the task, the computer shows demonstrations of two choices (movie-style)—one with a correct response, and one with an incorrect response—and then gives the subject the opportunity to practice two choices on her own. After practice ends, the task lasts for two minutes.

*C. Personality traits*

We use the validated 10-item version of the Big Five inventory for extraversion, agreeableness, conscientiousness, neuroticism and openness (Rammstedt and John 2007).

## 3. Outcomes: Measuring Decision Outputs

We scale all outcomes on the [0,1] interval, with higher values indicating better outcomes.

### *A. Objective financial condition index*

This index is the unweighted mean of five indicators coded from responses to standard questions:

1. Positive net worth, based on two questions drawn from the National Longitudinal Surveys:
    a. "Please think about all of your household assets (including but not limited to investments, other accounts, any house/property you own, cars, etc.) and all of your household debts (including but not limited to mortgages, car loans, student loans, what you currently owe on credit cards, etc.) Are your household assets worth more than your household debts?"
    b. "You stated that your household's [debts/assets] are worth more than your household's [assets/debts]. By how much?"
2. Any retirement assets, based on two questions asking whether someone has one or more IRA accounts and one or more workplace plans, followed in each case by questions on amounts in such accounts. Questions like these are asked in the Survey of Consumer Finances, the Health and Retirement Study, and many other surveys.
3. Any stockholding, based on three questions on stock mutual funds in IRAs, stock mutual funds in 401ks/other retirement accounts, and direct holdings. Questions like these are asked in the Survey of Consumer Finances, the Health and Retirement Study, and many other surveys.
4. Any saving in the last 12 months, based on the Survey of Consumer Finances question: "Over the past 12 months, how did your household's spending compare to your household's income? If the total amount of debt you owe decreased, then count yourself as spending less than income. If the total amount of debt you owe increased, then count yourself as spending more than income." Response options are: "Spent more than income", "Spent same as income", and "Spent less than income".
5. No severe hardship in the last 12 months, based on questions from the National Survey of American Families re: late/missed payment for rent, mortgage, heat, or

electric; moved in with other people because could not afford housing/utilities; postponed medical care due to financial difficulty; adults in household cut back on food due to lack of money. Response options for each of the four are Yes or No.

These five index components are strongly positively correlated with each other: the pairwise correlation range is 0.35 to 0.56.

*B. Subjective financial condition index*

This index is the unweighted mean of responses to four questions: about retirement savings adequacy, non-retirement savings adequacy, overall financial satisfaction, and financial stress.

1. Financial satisfaction, which follows standard life and economic satisfaction question wording: "How satisfied are you with your household's overall economic situation?"; responses on a 100-point scale (input using slider or text box).
2. Retirement savings adequacy: "Using any number from one to five, where one equals not nearly enough, and five equals much more than enough, do you feel that your household is saving and investing enough for retirement? Please consider the income you and any other members of your household expect to receive from Social Security, 401(k) accounts, other job retirement accounts and job pensions, and any additional assets you or other members of your household have or expect to have." This question is a variant on a standard one asked in many surveys, but in our version the 5 response options are framed to encourage people to recognize tradeoffs between saving and consumption: any response that includes "saving more" also includes "and borrowing/spending less", and vice versa. In mapping the 5 responses into the variables used here, we code: saved-enough, more-than-enough, and much-more-than-enough as 1 (the latter two responses are rare: only 3% of the sample); saved < enough as 0.5; saved << enough as 0.
3. Non-retirement savings adequacy. We placed this question in a different module than retirement savings adequacy, with different wording, to mitigate mechanical correlations. It reads: "Now, apart from retirement savings, please think about how your household typically uses the money you have: how much is spent and how much is saved or invested. Now choose which statement best describes your household". This question is a variant on a standard one asked re: retirement savings in many surveys, but in our version

the 5 response options are framed to encourage people to recognize tradeoffs between saving and consumption: any response that includes "saving more" also includes "and borrowing/spending less", and vice versa. In mapping the 5 responses into the variables used here, we code: saved-enough, more-than-enough, and much-more-than-enough as 1 (the latter two responses are rare: only 4% of the sample); saved < enough as 0.5; saved << enough as 0.

4. Financial stress, question taken from The Survey of Forces: "To what extent, if any, are finances a source of stress in your life?"; responses on a 100-point scale (respondents can input using slider or text box).

These four index components correlate strongly and positively with each other: the pairwise correlations range from 0.31 to 0.53, each with p-values < 0.001. This index is correlated 0.57 with the objective financial condition index.

*C. Life satisfaction, happiness, and health status*

Except for one elicitation of life satisfaction, each of these other elicitations comes from modules other than ours, in periods roughly coincident with our study period.[23]

*Life satisfaction* is measured using one of three minor variants on the standard "… how satisfied are you with your life as a whole these days?" asked in many surveys worldwide. For the other-module measure, we take the within-panelist average of non-missing responses to this question across the six ALP modules in which it has appeared subsequent to our round 1 modules, as of this writing. Of the 809/845 panelists with at least one non-missing response, 640 have at least two.

*Happiness* is measured by taking the within-panelist average of responses to two standard questions on happiness in general and in the last 30 days. These are asked in five other ALP modules after our Round 1 modules, with 787 of our 845 panelists completing at least one of these happiness questions and 397 completing both the 30-day version and the in-general-

[23] In deciding which measures to merge in from other modules, we define "study period" as post-our Round 1 (we could not find any relevant measure post-our Round 2), and select questions that have: a) been used in other studies; b) measure highly rated "aspects" of subjective well-being in the marginal utility sense per Benjamin, Heffetz, Kimball, and Szembrot (2014); c) are answered at least once by at least 2/3 of our sample.

version. Happiness last 30 days is measured using the standard "During the past 30 days, how much of the time have you been a happy person?" asked in many surveys worldwide. We take the within-panelist average of non-missing responses to this question across the four ALP modules in which it has appeared after our round 1 modules, as of this writing. Of the 509/845 panelists with at least one non-missing response, 474 have at least two. Happiness in general is measured using the standard "Taking all things together, I am generally happy" question asked in many surveys worldwide, including ALP module 425.

*Health status* is from the standard question: “Would you say your health is excellent, very good, good, fair, or poor?”. We take the within-panelist average across eight different modules in which this question has appeared after our Round 1 modules. Of the 840/845 panelists completing at least one of these, 780 complete more than one.

The pairwise correlations between our measures of life satisfaction, happiness, and health status range from 0.32 to 0.65. These measures are also strongly positively correlated with our indexes of subjective financial condition (the range is 0.29 to 0.50) and objective financial condition (from 0.29 to 0.35).

## References for Data Appendix


Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2011. "Reference Points and Effort Provision." *American Economic Review* 101 (2): 470–92.

Afriat, S. N. 1972. "Efficiency Estimation of Production Functions." *International Economic Review* 13 (3): 568.

Andreoni, James, and Charles Sprenger. 2012. "Estimating Time Preferences from Convex Budgets." *The American Economic Review* 102 (7): 3333–56.

———. 2016. "Prospect Theory Revisited: Unconfounded Experimental Tests of Probability Weighting."

Augenblick, Ned, Muriel Niederle, and Charles Sprenger. 2015. "Working over Time: Dynamic Inconsistency in Real Effort Tasks." *The Quarterly Journal of Economics* 130 (3): 1067–1115.

Banks, J., and Z. Oldfield. 2007. "Understanding Pensions: Cognitive Function, Numerical Ability, and Retirement Saving." *Fiscal Studies* 28 (2): 143–70.

Banks, James, Cormac O'Dea, and Zoë Oldfield. 2010. "Cognitive Function, Numeracy and Retirement Saving Trajectories." *The Economic Journal* 120 (548): F381–410.

Barber, Brad, and Terrence Odean. 2008. "All That Glitters: The Effect of Attention on the Buying Behavior of Individual and Institutional Investors." *Review of Financial Studies* 21 (2): 785–818.

Barberis, Nicholas C. 2013. "Thirty Years of Prospect Theory in Economics: A Review and Assessment." *Journal of Economic Perspectives* 27 (1): 173–96.

Barcellos, Silvia, and Leandro Carvalho. 2014. "Information about Self-Control and Intertemporal Choices."

Barsky, Robert B, F. Thomas Juster, Miles S Kimball, and Matthew D Shapiro. 1997. "Preference Parameters and Behavioral Heterogeneity; An Experimental Approach in the Health and Retirement Study." *Quarterly Journal of Economics* 112 (2): 537–79.

Bell, David E. 1985. "Disappointment in Decision Making under Uncertainty." *Operations Research* 33 (1): 1–27.

Benartzi, S., and R. H. Thaler. 1995. "Myopic Loss Aversion and the Equity Premium Puzzle." *The Quarterly Journal of Economics* 110 (1): 73–92.

Benjamin, Daniel, Don Moore, and Matthew Rabin. 2017. "Biased Beliefs about Random Samples: Evidence from Two Integrated Experiments."

Benjamin, Daniel, Matthew Rabin, and Collin Raymond. 2016. "A Model of Nonbelief in the Law of Large Numbers." *Journal of the European Economic Association* 14 (2): 515–44.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. forthcoming. "Memory, Attention, and Choice." *Quarterly Journal of Economics*.

Bradford, David, Charles Courtemanche, Garth Heutel, Patrick McAlvanah, and Christopher Ruhm. 2017. "Time Preferences and Consumer Behavior." *Journal of Risk and Uncertainty* 55 (2–3): 119–45.

Butler, Jeffrey V., Luigi Guiso, and Tullio Jappelli. 2014. "The Role of Intuition and Reasoning in Driving Aversion to Risk and Ambiguity." *Theory and Decision* 77 (4): 455–84.

Callen, Michael, Mohammad Isaqzadeh, James D Long, and Charles Sprenger. 2014. "Violence and Risk Preference: Experimental Evidence from Afghanistan." *The American Economic Review* 104 (1): 123–48.

Carvalho, Leandro, Stephan Meier, and Stephanie Wang. 2016. "Poverty and Economic Decision-Making: Evidence from Changes in Financial Resources at Payday." *American Economic Review* 106 (2): 260–84.
Carvalho, Leandro, Arna Olafsson, and Dan Silverman. 2019. "Misfortune and Mistake: The Financial Conditions and Decision-Making Ability of High-Cost Loan Borrowers."
Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman. 2014. "Who Is (More) Rational?" *American Economic Review* 104 (6): 1518–50.
Daniel, Kent, and David Hirshleifer. 2015. "Overconfident Investors, Predictable Returns, and Excessive Trading." *Journal of Economic Perspectives* 29 (4): 61–88.
Dean, Mark, and Pietro Ortoleva. 2019. "The Empirical Relationship between Non-Standard Economic Behaviors." *Proceedings of the National Academy of Science* 116 (33): 16262–67.
DellaVigna, Stefano, and Joshua M Pollet. 2009. "Investor Inattention and Friday Earnings Announcements." *The Journal of Finance* 64 (2): 709–49.
Diecidue, Enrico, Ulrich Schmidt, and Peter P Wakker. 2004. "The Utility of Gambling Reconsidered." *Journal of Risk and Uncertainty* 29 (3): 241–59.
Dimmock, Stephen, Roy Kouwenberg, Olivia S. Mitchell, and Kim Peijnenburg. 2016. "Ambiguity Aversion and Household Portfolio Choice Puzzles: Empirical Evidence." *Journal of Financial Economics* 119 (3): 559–77.
Dimmock, Stephen, Roy Kouwenberg, and Peter P Wakker. 2016. "Ambiguity Attitudes in a Large Representative Sample." *Management Science* 62 (5): 1363–80.
Dohmen, Thomas, Armin Falk, David Huffman, Felix Marklein, and Uwe Sunde. 2009. "Biased Probability Judgment: Evidence of Incidence and Relationship to Economic Outcomes from a Representative Sample." *Journal of Economic Behavior & Organization* 72 (3): 903–15.
Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2010. "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review* 100 (3): 1238–60.
Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association* 9 (3): 522–50.
Ericson, Keith. 2011. "Forgetting We Forget: Overconfidence and Memory." *Journal of the European Economic Association* 9 (1): 43–60.
Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences." *The Quarterly Journal of Economics* 133 (4): 1645–92.
Fehr, Ernst, and Lorenz Goette. 2007. "Do Workers Work More If Wages Are High? Evidence from a Randomized Field Experiment." *American Economic Review* 97 (1): 298–317.
Giné, Xavier, Jessica Goldberg, Dan Silverman, and Dean Yang. 2018. "Revising Commitments: Field Evidence on the Adjustment of Prior Choices." *The Economic Journal* 128 (608): 159–88.
Gneezy, Uri, Alex Imas, and John List. 2015. "Estimating Individual Ambiguity Aversion: A Simple Approach."
Goda, Gopi Shah, Matthew R Levy, Colleen Flaherty Manchester, Aaron Sojourner, and Joshua Tasoff. 2019. "Predicting Retirement Savings Using Survey Measures of Exponential-Growth Bias and Present Bias." *Economic Inquiry* 57 (3): 1636–58.

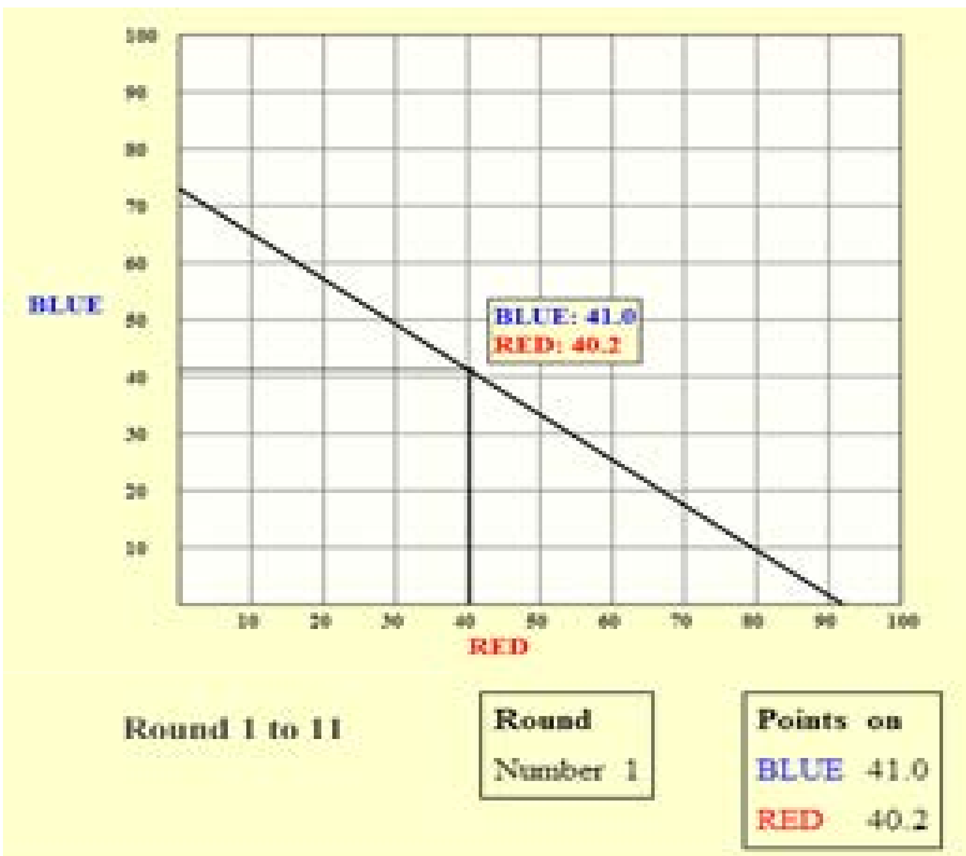
Grubb, Michael D. 2015. "Overconfident Consumers in the Marketplace." *Journal of Economic Perspectives* 29 (4): 9–36.
Gul, Faruk. 1991. "A Theory of Disappointment Aversion." *Econometrica: Journal of the Econometric Society* 59 (3): 667–86.
Gustman, Alan L, Thomas L Steinmeier, and Nahid Tabatabai. 2012. "Financial Knowledge and Financial Literacy at the Household Level." *American Economic Review* 102 (3): 309–13.
Hwang, In Do. 2016. "Prospect Theory and Insurance Demand."
Imai, Taisuke, Tom Rutter, and Colin F. Camerer. 2020. "Meta-Analysis of Present-Bias Estimation Using Convex Time Budgets."
Kahneman, D, and A Tversky. 1972. "Subjective Probability: A Judgement of Representativeness." *Cognitive Psychology* 3: 430–54.
Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47 (2): 263–91.
Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman. 2016. "Getting to the Top of Mind: How Reminders Increase Saving." *Management Science* 62 (12): 3393–3411.
Kartashova, Katya. 2014. "Private Equity Premium Puzzle Revisited." *American Economic Review* 104 (10): 3297–3334.
Kőszegi, Botond, and Matthew Rabin. 2009. "Reference-Dependent Consumption Plans." *American Economic Review* 99 (3): 909–36.
Kőszegi, Botond, and Adam Szeidl. 2013. "A Model of Focusing in Economic Choice." *The Quarterly Journal of Economics* 128 (1): 53–104.
Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112 (2): 443–77.
Larrick, Richard P, Katherine A Burson, and Jack B Soll. 2007. "Social Comparison and Confidence: When Thinking You're Better than Average Predicts Overconfidence (and When It Does Not)." *Organizational Behavior and Human Decision Processes* 102 (1): 76–94.
Levy, Matthew, and Joshua Tasoff. 2016. "Exponential-Growth Bias and Lifecycle Consumption." *Journal of the European Economic Association* 14 (3): 545–83.
Loomes, Graham, and Robert Sugden. 1986. "Disappointment and Dynamic Consistency in Choice under Uncertainty." *The Review of Economic Studies* 53 (2): 271–82.
Mata, Rui, Renato Frey, David Richter, Jürgen Schupp, and Ralph Hertwig. 2018. "Risk Preference: A View from Psychology." *Journal of Economic Perspectives* 32 (2): 155–72.
Meier, Stephan, and Charles Sprenger. 2010. "Present-Biased Preferences and Credit Card Borrowing." *American Economic Journal: Applied Economics* 2 (1): 193–210.
Moore, Don A., and Paul J. Healy. 2008. "The Trouble with Overconfidence." *Psychological Review* 115 (2): 502–17.
Moskowitz, T.J., and A. Vissing-Jorgensen. 2002. "The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle." *American Economic Review* 92 (4): 745–78.
Neilson, William S. 1992. "Some Mixed Results on Boundary Effects." *Economics Letters* 39 (3): 275–78.
Rabin, Matthew. 2000. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica* 68 (5): 1281–92.

Rabin, Matthew, and Dimitri Vayanos. 2010. “The Gambler’s and Hot-Hand Fallacies: Theory and Applications.” *Review of Economic Studies* 77 (2): 730–78.
Rabin, Matthew, and Georg Weizsäcker. 2009. “Narrow Bracketing and Dominated Choices.” *American Economic Review* 99 (4): 1508–43.
Rammstedt, Beatrice, and Oliver P. John. 2007. “Measuring Personality in One Minute or Less: A 10-Item Short Version of the Big Five Inventory in English and German.” *Journal of Research in Personality* 41 (1): 203–12.
Read, Daniel, and Barbara van Leeuwen. 1998. “Predicting Hunger: The Effects of Appetite and Delay on Choice.” *Organizational Behavior and Human Decision Processes* 76 (2): 189–205.
Schmidt, Ulrich. 1998. “A Measurement of the Certainty Effect.” *Journal of Mathematical Psychology* 42 (1): 32–47.
Schwartzstein, Joshua. 2014. “Selective Attention and Learning.” *Journal of the European Economic Association* 12 (6): 1423–52.
Stango, Victor, and Jonathan Zinman. 2009. “Exponential Growth Bias and Household Finance.” *The Journal of Finance* 64 (6): 2807–49.
———. 2011. “Fuzzy Math, Disclosure Regulation, and Credit Market Outcomes: Evidence from Truth-in-Lending Reform.” *Review of Financial Studies* 24 (2): 506–34.
———. 2014. “Limited and Varying Consumer Attention: Evidence from Shocks to the Salience of Bank Overdraft Fees.” *Review of Financial Studies* 27 (4): 990–1030.
———. 2016. “Borrowing High versus Borrowing Higher: Price Dispersion and Shopping Behavior in the U.S. Credit Card Market.” *The Review of Financial Studies* 29 (4): 979–1006. https://doi.org/10.1093/rfs/hhv072.
Toubia, Olivier, Eric Johnson, Theodoros Evgeniou, and Philippe Delquié. 2013. “Dynamic Experiments for Estimating Preferences: An Adaptive Method of Eliciting Time and Risk Parameters.” *Management Science* 59 (3): 613–40.
Tversky, Amos, and Daniel Kahneman. 1992. “Advances in Prospect Theory: Cumulative Representation of Uncertainty.” *Journal of Risk and Uncertainty* 5 (4): 297–323.

Allocate 100 tokens between **5 weeks from today** and **14 weeks from today**

| | Token value 5 weeks from today | Token value 14 weeks from today | Decision: How many of the 100 tokens would you like to allocate to the **sooner** payment **5 weeks from today?** | Tokens received 5 weeks from today | Tokens remaining 14 weeks from today | Total payment 5 weeks from today | Total payment 14 weeks from today |
|---|---|---|---|---|---|---|---|
| 1 | $1 | $1 | 0 out of 100 tokens | 0 | 100 | $0.00 | $100.00 |
| 2 | $1 | $1.02 | 0 out of 100 tokens | 0 | 100 | $0.00 | $102.00 |
| 3 | $1 | $1.04 | 0 out of 100 tokens | 0 | 100 | $0.00 | $104.00 |
| 4 | $1 | $1.07 | 0 out of 100 tokens | 0 | 100 | $0.00 | $107.00 |
| 5 | $1 | $1.11 | 0 out of 100 tokens | 0 | 100 | $0.00 | $111.00 |
| 6 | $1 | $1.17 | 0 out of 100 tokens | 0 | 100 | $0.00 | $117.00 |

**Data Appendix Figure 1. Discounting choices, screenshot**
(1 of 4 screens, 6 choices per screen)




**Data Appendix Figure 2. Consistency with GARP choices, screenshot**
(1 of 11 rounds, 1 choice per round).