We are all behavioral, more or less: A taxonomy of consumer decision making

Victor Stango and Jonathan Zinman*

August 2021

Abstract

We examine how 17 behavioral biases relate to each other, to three standard measures of risk and time preferences, to cognitive skills, personality, and demographics, and to outcomes in household finance, well-being, and health. Most consumers in our nationally representative panel data exhibit multiple biases, with substantial cross-person heterogeneity. Biases are positively correlated within person, especially after adjusting for measurement error. From that correlation structure, we reduce our 20 bias and standard preference measures to four behavioral common factors. Each BCF reflects a group of related biases re: beliefs, decision quality, discounting, or risk/uncertainty attitudes. The first two BCFs also strongly correlate with each other (positively) and cognitive skills (negatively). The first three BCFs and cognitive skills strongly correlate with various outcomes in the expected directions. Our results support processing-based models where basic limitations in cognition and/or attention produce multiple biases, and have several other implications for theory and practice.

Keywords: behavioral economics; multiple biases; factor analysis; measurement error instrumental variables; risky choice; intertemporal choice

JEL codes: C36, C81, D90, E70

---

* 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. Stango: UC Davis Graduate School of Management, vstango@ucdavis.edu; Zinman: Dartmouth College, IPA, J-PAL, and NBER, jzinman@dartmouth.edu. 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 and Dmitry Taubinsky 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.

"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 observed 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 related to each other and to other decision inputs like cognitive skills within-consumer, and what do those relationships imply for theories of consumer decision making? And how do groups of related biases and other decision inputs map into outputs—into outcomes in various domains? By analogy to a century of research in physics on the "Particle Zoo": What are economics' prospects for taxonomic classification of what might be called the "Cognitive/Behavioral Zoo"?[1]

We address such questions with evidence on the prevalence, correlation structure and cross-sectional explanatory power of multiple behavioral biases and other decision inputs. For each of 800+ adults from nationally representative panel data in the U.S., at two points in time three years apart, we measure a large set of decision inputs and outputs. For decision inputs, we measure 17 oft-cited biases using standard lab-style elicitations. We also measure classical decision inputs—patience and risk aversion using standard measures that are not designed to capture any biases, cognitive skills, and personality traits—and demographics. For decision outputs, we measure 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 set of findings describe bias prevalence and heterogeneity. Biases are more norm than idiosyncrasy. The median consumer exhibits 10 of 17 biases. No one exhibits all 17, but almost everyone exhibits multiple biases; e.g., the 5th percentile is 6. That said, cross-consumer

[1] Other instructive parallels include the century of painstaking research in various social sciences that has reduced the seemingly countless intelligences of Galton and Spearman's eras 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; e.g., Scerri (2011) states: "Fortunately, the periodic table allows chemists to function by mastering the properties of a handful of typical elements…."

heterogeneity in biases is substantial. The standard deviation of the number of biases exhibited is about 20% of its mean.

Our second findings describe correlations among biases. Biases positively correlate with each other within-consumer, especially after accounting for measurement error following Gillen et al. (2019).[2] Across all biases, 68% have positive pairwise correlations, with an average of 0.13, and 22% (50%) having p-values < 0.001 (<0.100). Correlations are stronger within six theoretically-related groups of biases that we define ex-ante: present-biased discounting, inconsistent and/or dominated choices, risk/uncertainty biases, overconfidence, math/statistical biases re: perceptions of risk and intertemporal prices, and limited attention/memory. Within those groupings the average pairwise correlation is 0.25 and 29% have p < 0.001.[3] We also find evidence of several cross-group correlations predicted by various theories; e.g., present-biased discounting being correlated with preference for certainty (Chakraborty, Halevy, and Saito 2020) and with limited memory and attention (Ericson 2017; Gabaix 2019).

Our third findings describe correlations between biases and other consumer characteristics. The average pairwise correlation between cognitive skills and biases is -0.25. Cognitive skills strongly negatively correlate with most biases, but positively correlate with loss aversion and ambiguity aversion.[4] Other classical inputs—patience, risk aversion, and personality traits—are relatively weakly correlated with biases, except for a few expected links between biases and our standard measures of time and risk preferences: patience is correlated with present biases, and risk aversion is correlated with ambiguity aversion, loss aversion, and math/statistical biases that can lead to undervaluation of returns to risk-taking.[5]

[2] The measurement error instrumental variables strategy we use here is feasible because biases are temporally stable within-consumer (Stango and Zinman 2021).

[3] See also Dean and Ortoleva (2019) and Chapman et al. (2020) on relationships among biases re: risk attitudes, and Chapman et al. (2020) on relationships among overconfidence varieties.

[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] See also, e.g., Benjamin et al. (2013), Chapman et al. (2020), 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.

Next, we conduct exploratory factor analysis (EFA) on the measurement error-adjusted correlation matrix of biases and standard preferences. EFA groups variables (factors) by estimating linear combinations of them and "lets the data speak" in identifying any latent variables (common factors) underlying patterns in the correlation matrix.

Fewer common factors imply greater prospects for dimension reduction, and our fourth finding is that four BCFs capture over 50% of the cross-sectional variation in our 20 measures of biases/preferences. This is the first hint that reality is perhaps lower-dimensional than our ex-ante groupings, with 4 BCFs vs. 6 groups. The first two BCFs alone explain 34% of the variance.

Our fifth set of findings examines empirical relationships within and across our four BCFs. One striking pattern is that our ex-ante groupings of theoretically-related biases are clearly reflected in BCF loadings.[6] The first BCF loads most strongly on the math/statistical biases and overconfidence, and as such we label it the "beliefs" BCF. The second BCF, "decision quality," loads most clearly on inconsistent choice biases and limited memory/attention. The third BCF, "discounting," loads on present-biased discounting and impatience and little else. The fourth BCF, "risk attitudes," loads on two of our three biases re: choice under risk and uncertainty, on our two standard measures of risk aversion, and on little else. BCFs 3 and 4 are consistent with some primitive role for preferences over time and risk/uncertainty.

But two other empirical patterns clarify how our ex-ante groupings do not tell the full story. One is that each BCF, and especially the first two, loads on biases outside their namesake group(s). E.g., present-biased discounting and preference for certainty both load fairly strongly on the decision quality BCF (as foreshadowed by the correlation matrix). Moreover, the first three BCFs, and especially the first two, are strongly correlated with each other. These are additional hints that decision inputs may be amenable to modeling with relatively few parameters. In any case, these results suggest that it is important for models to account for strong empirical relationships across biases that may seem conceptually distinct at first glance.

If biases do not live in conceptually neat isolation from each other, what explains their relationships? Another key hint comes from the fact that the first two BCFs are very strongly linked

[6] Loadings are the weights associated with each input to factor analysis—in our case, the 20 bias and standard preference measures—for each retained BCF. Please see Section 4-B for details.

to cognitive skills, in both correlation and fit; e.g., our cognitive skills measures explain 46% (30%) of the variance in the beliefs (decision quality) BCF. The discounting common factor is also tied to cognitive skills, albeit less strongly. BCFs exhibit a similar pattern of relationships with education and income. As we discuss in Section 5, these strong links between biases, cognitive skills and their offshoots support what might be called "behavioral microfoundation" theories where an underlying cognitive process can drive many biases, even seemingly disparate ones (e.g., Bordalo, Gennaioli, and Shleifer 2020; Enke and Graeber 2021; Gabaix 2019).

We also examine empirical relationships between BCFs and six outcomes: objective and subjective financial condition, and four standard measures of life satisfaction, happiness, and health. Results from various specifications suggest that the first three BCFs are negatively correlated with outcomes, and especially strongly so with financial condition. In contrast, we find that the fourth BCF is uncorrelated with outcomes, and that cognitive skills are strongly positively so.

Altogether our results suggest that behavioral biases are commonplace, heterogeneous in the cross-section of consumers, amenable to dimension reduction in various ways, and linked to household financial outcomes through a set of behavioral common factors and cognitive skills. 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 5 for references). On relationships among various biases, we add to two other papers examining relationships among a relatively large number of biases and other decision inputs.[8]

We discuss key takeaways in Section 5 and summarize them here. Our results support processing-based theories and highlight directions for future research on which process(es) are

[7] For reviews see e.g., the Handbook of Behavioral Economics (D. 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. (2020) 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.

most fundamental. Basic constructs used to organize the study of behavioral economics, as we do with our ex-ante groupings (present-biases, inconsistent choices, risky choice, etc.), are clearly useful but incomplete. Related, preference-based theories receive mixed support in our data. Even more broadly, our results make a strong case that accounting for behavioral agents is important for economic modeling, but that doing so often requires some richness: researchers should be wary of homogeneity and separability assumptions.[9] Our results and measurement tools provide a reference for assessing such assumptions and can help guide the development of more portable and powerful theoretical and empirical models.

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 6-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.[10]

[9] Many empirical and theoretical exercises in behavioral economics consider biases one or two at a time, implicitly assuming separability from other 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). 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] 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 2021) 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 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 biases and other consumer characteristics, we held in mind 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.[11] 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. Measuring biases, preferences, cognitive skills and personality traits

We measure 17 behavioral biases, three standard measures of patience and risk aversion, four standard measures of cognitive skills, and standard measures of the "Big Five" personality traits. Table 1 summarizes these 29 variables and their measures. The Data Appendix provides additional details on the data, elicitation methods, and their theoretical and methodological antecedents. We define behavioral biases as deviations from a clear normative classical benchmark (e.g., time-consistent discounting, accurate beliefs about one's own skills).

We measure biases that have featured prominently in behavioral economics, are amenable to compact elicitation in online surveys, and fall into one of six *related bias* groups that we construct ex-ante and use as one lens for interpreting our empirical estimates (Table 1 Column 7). Two biases 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 measured two ways (Choi

[11] 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."

et al. 2014), and narrow bracketing (Rabin and Weizsäcker 2009).[12] Another three relate to preferences or attitudes toward *risk and 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/statistical 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). Two relate to *limited attention and limited memory*, in ways that are motivated by behavioral models of inattention (Ericson 2011; Bronchetti et al. 2020).[13] We also aggregate these six groups to three broader ones (Table 1 Column 8): *preferences* (the discounting and risk biases), *beliefs* (overconfidence and math/statistical), or poor *decision quality* (choice inconsistency and limited attention/memory).[14]

Many bias sources are potentially bi-directional: one can be present- or future-biased, one can underestimate or overestimate future values, and so on. One might be interested in whether those different directions are important empirically, but we have found that they are not, at least for the correlation-based focus here. Practically speaking, bi-directional biases are mechanically (negatively) correlated with their flipsides, and if, as an example, present-bias appears in a common factor, future-bias does as well, with essentially the opposite weight. We therefore focus on just one direction for each bias, choosing the one that is more prominent in theory or more common empirically.[15]

In Section 2, we describe consumers' biases on an easy-to measure extensive margin: biased or not. Measuring the extensive margin does not require an elaborate elicitation or granular

[12] 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).
[13] Following a common delineation in behavioral economics, we do not measure social preferences. See Dean and Ortoleva (2019) and Chapman et al. (2020) for evidence on relationships between behavioral biases and social preferences.
[14] We classify the discounting and risk biases as preferences because they map most closely into time and risk preference parameters. A caveat is that such preferences may not be primitives *per se* but instead emerge from interactions of primitives with beliefs and/or cognitive limitations. See Sections 5-A and 5-B for related discussion.
[15] We examine bi-directional biases in more detail in Stango and Zinman (2021).

response data. The extensive margin is unit-free in cross-bias comparisons and maps well into models with a mix of biased and unbiased agents.

Mostly we rely on a more granular measure of intensity for 15 of the 17 biases (Table 1 Column 4): 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 takes on 11 distinct rank values, with others ranging from continuous (e.g., present-bias on money) to binary (e.g., limited memory).

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),[18] 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 refer to these as measures of "standard" or "classical" preferences/attitudes over risky and intertemporal choice because they are not designed to identify biases *per se*. But they do have theoretical and empirical links to biases that our analysis below takes into account.

We measure cognitive skills using standard tests for general/fluid intelligence (McArdle, Fisher, and Kadlec 2007), numeracy (Banks and Oldfield 2007), cognitive control/executive function (MacLeod 1991; Miyake and Friedman 2012), and financial literacy (Lusardi and Mitchell 2014). The first three measure the types of cognitive processes at the heart of many processing-based theories of decision making and biases that we consider in Section 5-A, while financial literacy can be conceptualized as crystalized intelligence for financial decision making and has been strongly linked to the financial decisions and outcomes we consider in Section 6-C.[19] In our second round of surveying, we add elicitations of the Big Five personality traits

---

[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] These two risk aversion measures have an estimated within-person correlation of 0.29 after adjusting for measurement error.

[19] Consistent with prior work on generalized intelligence (e.g., Jensen 1998), our four cognitive skills measures are very strongly correlated with each other (Appendix Table 1 Panel A) and have a single principal component or common factor.

(extraversion, agreeableness, neuroticism, openness and conscientiousness) to the end of our second module (Rammstedt and John 2007).[20]

D. Demographics, response times

We also consider standard demographic variables in some of our analysis (see Appendix Table 2 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. 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 as covariates in some specifications of our outcome regressions (Section 6).

E. Data availability

The data analyzed in this paper is available at [link TBD], together with annotated code for cleaning the data, creating the variables we use, and merging with other ALP surveys.

2. Prevalence and heterogeneity of multiple behavioral biases

Is exhibiting multiple behavioral biases common or anomalous? How biased are consumers, and 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-count".

Table 2 Panel A shows that the exhibiting many biases is the rule, not the exception. The mean B-count 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

[20] We initially decided against eliciting personality measures, given our resource constraints and the lack of prior evidence of correlations between them and behavioral biases; e.g., Becker et al.'s (2012) review article does not mention biases.

interpretation of B-count variation.[21] Column 2 suggests that cross-consumer heterogeneity is the number of biases exhibited is substantial. The B-count standard deviation is about 20% of its mean.

The key findings from Table 2 are that almost everyone exhibits multiple biases, with substantial heterogeneity across people in how many.

3. Correlations 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 and standard measures of risk and time preferences described in Section 1-C. These correlations are interesting descriptively, and they also serve as inputs to the common factor analyses starting in Section 4.

Measurement error could bias these estimated correlations and complicate identification of 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 2021). Specifically, we use univariate Obviously Related Instrumental Variables (ORIV) regressions to estimate pairwise correlations, following Gillen et al. (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.[22]

For the rest of our analyses, we scale each variable as a percentile rank,[23] because rank offers

[21] On average, only about one out of the maximum possible 17 biases is missing due to non-response (Table 2 Column 1), with a standard deviation of about 1.5. Every panelist completes at least five of the bias elicitations (Column 5).

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

[23] 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. 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 37th percentile, and 13% of responses fall into that category. Summing, anyone in the next category is in the 50th(=37th+13th), and so on.

both conceptual and practical advantages relative to alternative parametrizations.[24] 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 (recall from Table 1 Column 4 that most of our decision input elicitations produce at least somewhat granular measures).[25] 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 ORIV estimates of pairwise correlations among the 20 variables we include in our factor analysis in Section 4: our 17 biases, as well as patience and our two standard measures of risk aversion.

Table 4 summarizes several key patterns from the ORIV correlation matrix. Most biases are positively correlated with other biases: 68% of the pairwise correlations are > 0.[26] This is consistent with the hypothesis that many biases could emerge from a common underlying cognitive process or processes. Many cross-bias correlations are statistically significant at conventional cutoffs: 50%, 29%, and 22% of pairwise correlations have p-values <0.10, <0.01, and <0.001. This suggests that theorists and empiricists alike should be attuned to potential interactions among biases, with Table 3 serving as a guide to which interactions could be most important. Correlations are stronger within the ex-ante conceptually-related bias groupings described in Table 1 and Section 1-C. E.g., while the average correlation across all biases is 0.13, the average correlation within each of our six related bias groups is 0.25. Within the three broader groupings, the average correlation among preference biases is 0.17, among belief biases is 0.21, and among decision quality biases is 0.40. This suggests that many of the basic constructs underpinning the study of behavioral biases are indeed useful guides for interpreting empirical relationships.

---

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

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

[26] Many of the negative correlations involve risk biases, with loss aversion accounting for largest magnitudes and the only negative correlations with p <=0.001. We consider how to interpret this, and other rich patterns of relationships involving our risk- and uncertainty-related variables, below.

Several other correlations span our ex-ante groupings and evoke prior work theorizing or documenting nuanced connections between seemingly different biases. Starting with work showing that present-bias could emerge separately from time-inconsistent preferences *per se*, our present biases are strongly positively correlated with GARP violations, which is consistent with Gabaix and Laibson's (2017) theory of how present bias can arise from cognitive noise (note also the negative correlations between present bias and cognitive skills, summarized in Table 4 and discussed in the next sub-section). We also find that present-bias (over money at least) is very strongly positively correlated with preference for certainty, echoing Halevy's work (mostly recently with Chakraborty and Saito (2020)). Present-bias over money is also strongly correlated with distorted perceptions of probability (NBLLN and Gambler's Fallacy in our case), echoing its correlation with probability weighting in Epper et al.'s (2011) student sample. Present-bias (over consumption at least) is also strongly positively correlated with limited attention and memory, as suggested by e.g., Ericson (2017) and Gabaix (2019). And it is correlated with the exponential growth biases, as suggested by, e.g., Ericson and Laibson (2019). Also noteworthy: Ambiguity aversion seems at least somewhat related to biased beliefs, as is in e.g., Michelacci and Paciello (2020); limited attention and memory are positively correlated with a statistical bias (NBLLN in our case), which is reminiscent of Enke and Zimmerman's (2019) finding that attention treatments reduce correlation neglect; GARP violations are strongly correlated with various belief biases and limited memory, suggesting that they stem at least in part from some underlying cognitive limitation and reminiscent of e.g., Abaluck and Gruber's (forthcoming) inferences about drivers of inconsistencies in health insurance plan choice.

These nuanced patterns motivate statistical analysis that complements ex-ante classifications—like our related bias groups—by "letting the data speak". Factor analysis does that, and we undertake it in Section 4.

C. Correlations between biases, cognitive skills, and other decision inputs

Table 4 also summarizes correlations between biases and other decision inputs.

Starting with patience and risk aversion, which we include in our factor analysis in the next section, we find that patience is negatively correlated with the discounting biases (see also e.g., Dean and Ortoleva 2019) but mostly uncorrelated with other biases. Risk aversion is positively correlated with loss aversion (e.g., Chapman et al. 2020; Mrkva et al. 2020) and ambiguity aversion (e.g., Dean and Ortoleva 2019), and with math biases that distort perceived returns to risk-taking (e.g., D. Benjamin, Rabin, and Raymond 2016; Rabin and Vayanos 2010; Levy and Tasoff 2016), although none of these correlations are larger 0.22. Interestingly, preference for certainty is not strongly correlated with our risk aversion measures, further buttressing the conjecture that PFC is more strongly linked to intertemporal choice than to risky choice *per se*. We explore this further in our factor analysis below.

Turning to other inputs, personality traits are weakly correlated with biases, if at all.[27] Cognitive skills are strongly negatively correlated with most biases (see footnote 5 for references to prior work considering relationships with one or two biases at a time). But they are positively correlated with loss aversion (as in Chapman, et al. 2019) and with ambiguity aversion. Overall, the average pairwise correlation between cognitive skills and biases is the same in absolute value (0.25) as that between the biases themselves.[28] Section 4-C further explores empirical connections between cognitive skills and groups of biases, and Section 5-A discusses implications of our results for processing-based theories where biases emerge from cognitive limitations.

D. Summing up the results so far

Biases tend to be positively correlated with each other, and negatively correlated with cognitive skills, with loss aversion and ambiguity aversion exceptions to those patterns. Patience is strongly negatively correlated with present bias, as expected. Risk aversion is positively correlated with loss aversion, ambiguity aversion, and math biases that distort perceived returns to risk-taking, although all these correlations are modest in magnitude. Personality traits are essentially uncorrelated with biases. Among our 17 biases, correlations are indeed stronger within subsets of the ex-ante groupings we formed based on constructs. But there are also several strong correlations that span those groupings and are consistent with nuanced theories of how seemingly different cognitive and psychological processes interact and emerge. These motivate taking a statistical

[27] Table 4 shows only the average pairwise correlation between each bias and our five personality measures, but among the 85 pairwise correlations, none exceed |0.21| and only 8 have a p-value<0.01. A caveat is that we do not have repeated measures of personality and consequently cannot adjust for measurement error therein, although personality measurement may be refined enough at this point, after a century of research, that measurement error is a relatively minor concern (Anusic and Schimmack 2016).

[28] And the average of the absolute values of the pairwise correlations is larger between cognitive skills and biases (0.32) than among biases (0.27).

approach to identifying relationships among biases, and other decision inputs, that "lets the data speak." We do that next.

4. Behavioral common factors: Identification and interpretation

A. Identifying behavioral common factors

With the correlations from the previous section in hand, we now ask whether the data identify one or more behavioral common factors (BCFs): linear combinations of our 17 biases and 3 standard preferences that might reflect more fundamental latent characteristics or processes.[29] The goal of factor analysis, and other dimension-reducing statistical methods, is to let the data choose a relatively small set of latent variables that best span variation in the data.

Mechanically, we use our correlation matrix in Table 3 to estimate common factors that best explain the collective variation in our matrix of 20 biases and preferences. The steps involve first estimating a full set of BCFs that collectively explain the data. Each BCF is a linear combination of our 20 variables, and factor analysis ranks the BCFs by eigenvalues—by the proportion of total variance explained by each factor. Appendix Table 3 shows the first step in this process, listing eigenvalues for 18 BCFs that collectively explain 100 percent of the variation in our 20 variables.

The second step is choosing a set of BCFs to "retain," or to focus on for interpretation—and in our case, also for our second-stage models correlating BCFs with outcomes. The selection of factors to retain often focuses on eigenvalues, in one of three ways: by setting a cutoff value for retention, by looking for "breaks" where eigenvalues drop sharply, or by choosing a set of eigenvalues that collectively explain a threshold share of variation in the data (say, 50% or 70%).[30] In our case we retain the first four BCFs, with eigenvalues ranging from 4.92 to 1.49, since eigenvalues above 1.00 indicate dimension reduction, we have no obvious breaks beyond the difference between the first and second eigenvalues, and the first four BCFs collectively explain 50%+ of variation in the data.

[29] In a prior version of the paper, we included cognitive skills and personality traits in the set of variables that could enter the common factors. We thank our Editor and referees for suggesting the more pared-down approach here, which allows us to more clearly identify and interpret: (a) relationships between behavioral common factors and cognitive skills; (b) relationships between behavioral common factors and outcomes, *conditional* on cognitive skills.

[30] See https://www.stata.com/manuals13/mvfactor.pdf as a useful reference, and compendium of other good sources.

B. Interpreting behavioral common factors: Loadings

The next step is to examine common factor loadings: the weights associated with each bias/preference in each retained BCF, scaled from -1 to 1. These weights are "rotated" to facilitate interpretation and we present them in Table 5.[31]

One lens for interpreting these loadings is our ex-ante "related bias" groupings, which are based on relatively traditional constructs (Section 1-C). For each of these groups, Table 5 shades the BCF in which it appears with the highest weights. We also shade the BCFs in which our standard measures of patience and risk aversion appear most prominently.

Factor 1 loads most strongly on biases from two related bias groupings: overconfidence and math/statistical biases. In terms of our broader beliefs vs. preferences vs. decision quality groupings (Table 1 Column 8), these loadings map well into the former and as such we label this BCF "beliefs." Factor 2 loads most strongly on the choice inconsistency and limited attention/memory groupings. Since each of the biases in these groupings indicate lower decision quality given available information—by misapplying it (GARP and dominance violations), considering it isolation (narrow bracketing), or ignoring or forgetting it (attention/memory)—we label this BCF "decision quality."[32] BCF 3 is clearly a "discounting" factor, loading most strongly on the two present biases and patience. BCF 4 is clearly a "risk/uncertainty" factor. It loads most strongly on two of the three risk biases, and on the two standard risk aversion measures. As discussed further in Sections 5-A and 5-B, one interpretation is that BCFs 1 and 2 capture something like "processing", while BCFs 3 and 4 capture something closer to "preferences".

Although our ex-ante bias groupings do well ordinally, in the sense of mapping the relative strength of bias loadings, they clearly miss some important relationships. BCFs 1 and 2, which

---

[31] We use a promax rotation; this is from the class of oblique rotations, which allow for correlation between common factors. We have used other rotations, including orthogonal rotations such as varimax, with indiscernible effects on the results except the orthogonal rotations fail to identify the strong correlations among the common factors evident in Table 5. Many online sources concisely describe rotations and their interpretation. See, e.g., https://www.theanalysisfactor.com/rotations-factor-analysis/ and the references listed there.

[32] It makes sense to us to consider our limited attention and memory measures as mapping into low decision quality, rather than constrained-optimal quality as in models of rational attention, because: (a) our limited attention measure captures an element of regret; (b) forgetting is quite costly in our memory task (see Bronchetti et al. (2020) for a related discussion of stakes in distinguishing between rational and behavioral attention models).

together explain 34% of the variance in our 20 decision inputs, each load on multiple other biases and bias groupings as well. Indeed, together BCFs 1 and 2 span all biases except perhaps present-biased money discounting. Also, as foreshadowed by the correlation matrix, preference for certainty seems far more tightly linked to processing biases (BCFs 1 and 2) and discounting (BCF 3) than to risk and uncertainty *per se* (BCF 4). Moreover, the first three BCFs are strongly positively correlated with each other (see Table 5), with BCFs 1 and 2 particularly strongly so at 0.71. Meanwhile, the risk-related BCF is largely unrelated to the other three, although the risk biases themselves each load nontrivially on at least one of the other BCFs.

These empirical relationships across our ex-ante groupings suggest another lens for interpreting the BCF loadings: newer theories that examine which cognitive or psychological processes might be most fundamental and hence powerful for explaining the “Cognitive Zoo”. We consider such theories in Sections 5-A and 5-B, after examining related evidence on relationships between BCFs and other individual characteristics including cognitive skills.

C. BCF scores and links between behavioral common factors and other consumer characteristics

Examining statistical relationships between the BCFs and other consumer characteristics further illuminates how to interpret the BCFs. Table 6 takes a first step by showing pairwise ORIV correlations between each BCF and our other decision inputs (cognitive skills and personality) and demographics, and also r-squared values from OLS regressions of each BCF on those inputs. It does the same for the B-count, for completeness.

More precisely, this exercise requires that we calculate a score for each BCF for each observation, to get an individual- and time-specific measure of each BCF for each person. For each behavioral common factor j, the score *BCFS* sums the values of our 20 bias and standard preference measures indexed by k, c_{ikt} (recall that each is scaled as a percentile rank), each weighted by its estimated loading ω_{jk}, for each panelist and survey date:

$$BCFS_{it}^{j} = \sum_{jk} \omega_{jk} c_{ikt}$$

The most striking pattern in this table is the strong relationship between cognitive skills and the processing *BCFS*s—BCFs 1 and 2. BCF 1, the beliefs BCF, has correlations with our four cognitive skills variables ranging from -0.32 to -0.60.[33] Together the four cognitive skills variables alone explain 46% of the variance in the beliefs BCF. 46% is notably high given that our fit estimates do not adjust for measurement error.[34] BCF 2, the decision quality BCF, has correlations with our four cognitive skills variables ranging from -0.19 to -0.47. Together the four cognitive skills variables alone explain 30% of the variance in the decision quality BCF. As we discuss in Section 5-A, these results are broadly consistent with processing-based theories where multiple biases emerge from a smaller set of fundamental cognitive processes.

The other BCFs have weaker if any relationships with cognitive skills. BCF 3 (discounting), has weaker correlations with cognitive skills than BCFs 1 and 2, and cognitive skills explain only 7% of the variance in BCF 3. BCF 4 (risk/uncertainty) is basically uncorrelated with cognitive skills, and cognitive skills only explain 1% of its variance. The further strengthens the inference that BCFs 3 and 4 capture preferences that operate relatively independently from cognitive skills, although not completely so in case of discounting.

Turning to other characteristics, the patterns for education and income mirror those for cognitive skills—unsurprising given that cognitive skills drive both of those things. More surprisingly, age is only weakly correlated with our BCFs, perhaps because our sample does not

[33] Similarly, Almenberg and Gerdes (2012) and Goda et al. (2019) find negative correlations between exponential growth bias and cognitive skills. And low cognitive skills has been linked to overconfidence in several studies (largely outside of economics, to our knowledge; see, e.g., Ehrlinger et al. 2008). We worried that the link between overconfidence and low cognitive skills could be mechanical in our data, but robustness checks suggest otherwise. 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. The level overconfidence measure is based on self-assessed performance on the 3-question numeracy questions, and the relative overconfidence bias is based on self-assessed performance on the 15-question number series questions. A negative correlation between cognitive skills and overconfidence could arise in either case because of ceiling effects: someone with the highest score on either test could never be overconfident, so overconfidence could only raise among those with lower (estimated) cognitive skills. To examine the ceiling effect, we re-calculated the ORIV correlations only for those getting 1+ numeracy questions wrong, or 2+ number series questions wrong. The correlation between level overconfidence and the numeracy score went from -0.89 to -0.69. The correlation between relative overconfidence and the number series score went from -0.99 to -0.97. All other correlations between level/relative overconfidence and cognitive skills change from an average of -0.49 to an average of -0.35/-0.45 excluding the "ceiling observations." We take this as evidence that ceiling effects may exert some downward bias on correlations, but do not fully explain the negative correlations.

[34] Our ORIV regressions employ instrumental variables, and as such standard fit measures such as r-squared are not useful.

include older adults who could be experiencing serious cognitive decline.[35] And personality traits are largely unrelated to BCFs, with the 20 pairwise correlations ranging from -0.07 to 0.10, and the four r-squareds from 0.01 to 0.03.

All told, it seems that we have two BCFs that are relatively distinct from other consumer characteristics, and two that are strongly intertwined with other consumer characteristics and especially with cognitive skills. The fact that these latter two BCFs are quantitatively the most important ones for capturing multiple biases—they span nearly all our biases in terms of loadings, and together explain more than more than a third of the variation in our 20 bias and standard preference measures—makes the links between cognitive skills and biases especially noteworthy when considering implications for theory, as we do next.

5. Implications for theory and open questions for empirics

Before correlating our BCFs with outcomes, we discuss the implications of our results thus far for theory and future empirical work.

A. Processing-based theories

Starting with the strong links between biases and cognitive skills documented in the previous sub-section, our results support what might be called "behavioral microfoundation" theories where an underlying cognitive process can drive many biases, even seemingly disparate ones. Examples include Bordalo, Gennaioli and Shleifer (2012; 2020), Bushong et al. (2021), Enke and Graeber (2021), Gabaix (2014; 2019), and Koszegi and Szeidl (2013).

One key open empirical question is: What exactly *are* the key underlying cognitive processes? Some theories in the cognitive sciences privilege executive function (Miyake and Friedman 2012) or other closely related concepts of cognitive control. But our measure of that process (Stroop) is not as strongly correlated with our biases as our measures of fluid and crystalized intelligence (Appendix Table 1 Panel B), and Shenav et al.'s (2017) review of work in neuroscience concludes: "although overwhelming evidence suggests that [cognitive] control is costly, it is far from clear

[35] We report results for linear age but have explored other functional forms and found little evidence of such non-monotonicity. That does not exclude the possibility that individual biases are non-monotonically related to age, of course, and it is important to keep in mind that our sample includes only working age adults.

how those costs should be operationalized (i.e., what are their constituents) or, more importantly, how they should be measured" (p. 114).

Many of the economic theories noted above model attention and memory as the behavioral microfoundation, in ways that are at least partially supported by our results. E.g., the Bordalo et al. models seem at least partially reflected in our decision quality BCF.[36] We also find a strong link between present-bias and limited attention and memory, per Ericson (2017) and Gabaix (2019). Gabaix's limited attention microfoundation can also produce the various biases in our beliefs BCF (Gabaix 2019, Section 2.3); these predictions receive less support in our data because our attention/memory biases appear only in the decision quality BCF, not the beliefs BCF. Recall however that these two BCFs are correlated 0.71, suggesting empirical as well as theoretical potential for consolidation and further dimension reduction. And of course our measurements and results are merely a starting point; both Caplin's (2016) and Gabaix's (2019) recent reviews of work on limited attention emphasize the lack of standard measures and urgent need for more work on measurement.[37]

Other theories have what might be construed as a mix of attention, memory, and cognitive skills at their heart; these too are at least partially supported by our results. Lian (forthcoming) seems well-reflected in our decision quality BCF and its strong relationship to cognitive skills; in that model, bounded recall and selective memory (which are likely reflected in our limited attention and memory measures), together with noisy perception (which is probably reflected in our cognitive skills measures), produce narrow thinking (which is likely reflected in our

[36] In Bordalo, Gennaioli and Shleifer (2020) limited memory and inattention produce context-specific choices, as well as other behavioral biases in which preferences/choices seem inconsistent. We do not measure context-specific choices but do find strong links between limited attention and limited memory, between attention/memory and inconsistent choices, and between attention/memory, inconsistent choices, and cognitive skills (see especially the loadings on BCF 2 in Table 5, and the strong links between BCF 2 and cognitive skills in Table 6). In Bordalo, Gennaioli and Shleifer (2012) a form of limited attention produces inconsistent choices and risky choice biases in ways that are also broadly consistent with our decision quality BCF, and particularly so if, for subjects with limited memory and attention, the possible gain presented by our loss aversion elicitation is more salient than the possible loss. This, and Khaw, Li, and Woodford (forthcoming), motivates more refined measurement of loss aversion, small-stakes risk aversion, and reference points, together with more refined measurement of basic cognitive processes, in future work.

[37] For some recent work on measuring attention, see e.g., Bronchetti et al. (2020), Caplin et al. (2020), Abaluck and Adams-Pressl (2021). Chun et al.'s (2011) review suggests to us that measuring working memory will be crucial for understanding links between cognitive skills, attention, and memory.

Inconsistent/Dominated Choice measures). In Enke and Graeber (2021), cognitive uncertainty (which is probably reflected in some combination of our attention/memory and cognitive skills measures) produces various risk and belief biases, along the lines of our BCFs 1 and 2 and their strong relationships with cognitive skills.[38] Benjamin (2019, p. 74) states that the representativeness heuristic is "generally considered to be a unifying theory for many of the [probabilistic reasoning and judgment] biases," suggesting that future work would do well to test that hypothesis. For now, it bears noting that the statistical biases we do measure are indeed strongly correlated with each other, and with cognitive skills.

A related open empirical question is: How well do those basic cognitive processes function? Our results here and in Section 6 support the hypothesis that biases and "mental gaps" a la Handel and Schwartzstein (2018) are cumulatively important for outcomes and welfare.

A key takeaway is that future work will do well to focus on measuring the basic cognitive processes at the heart of these models, together with the specific biases predicted to emerge from those processes, in order to more precisely map the taxonomy of consumer decision making.

B. Preference-based theories

Preference-based theories receive mixed support in our results.

On the supportive side, our BCF 3 can be read as a relatively comprehensive reflection of both standard and behavioral time preferences, and BCF 4 as a relatively comprehensive reflection of standard and behavioral risk preferences, as discussed above in Sections 4-B and 4-C. Those BCFs are consistent with some primitive role for preferences over time and risk.

On the less supportive side, our results suggest that it is unlikely that preferences alone, or at least the ones that we measure, can unify the "Cognitive Zoo". BCFs 3 and 4 explain much less of the variance in our full portfolio of biases and standard preferences than the processing BCFs 1 and 2. BCF 3, our discounting BCF, is related to cognitive skills, suggesting that processing limitations may influence discounting *a la* Gabaix and Laibson (2017). Our risk biases also load meaningfully on every other BCF in addition to the risk/uncertainty BCF 4, in ways consistent with processing-based theories and Halevy and co-authors' belief-based theories of present bias

[38] See also Woodford (2020).

(e.g., Chakraborty, Halevy, and Saito 2020). Broadly speaking, our rich patterns of how risk biases relate to other biases, to standard measures of risk aversion, and to cognitive skills are not fully explained by any extant theory we know of.[39]

C. Other prospects for unification?

The lack of a central role, or even much of a role at all, for personality traits in our taxonomy is noteworthy,[40] as recent work in labor economics has emphasized a central role for personality traits and other non-cognitive skills in consumer choice, at least in terms of child development (Heckman, Kautz, and Jagelka, 2021). Our results potentially contrast with Jagelka's (2020) evidence on high school students, where both cognitive skills and three of the Big Five personality traits relate strongly to choice inconsistency and to risk and time preferences. Some future work would do well to focus on the potential tension between Jagelka's findings and ours by measuring the non-cognitive skills of greatest interest to labor economists, together with multiple behavioral biases, in adult samples.

D. Basic assumptions and practices in (behavioral) economics

Our results also inform several basic assumptions and practices in the study of consumer choice. We highlight several key implications in bold.

Reconsider identification assumptions. Our evidence of common factors underlying multiple biases, and even groups of "unelated" biases, should prompt researchers to reconsider whether empirical models correlating outcomes with a single bias, or a few biases, are well-posed or subject to omitted variable bias (see footnote 9 above). Our findings, along with those of Chapman et al. (2020) and Dean and Ortoleva (2019), can serve as guides for whether and when single-bias studies are likely to be empirically informative. And, when our findings or others identify how biases are linked empirically to each other and/or cognitive skills, they can guide research design.

Accounting for behavioral agents is important for economic modeling. This is hardly news for practicing behavioral economists, but we hope that our results—e.g., on the prevalence of multiple

[39] For related evidence on the complexity of risk attitudes see, e.g., Chapman et al. (2021; 2020); Chapman, Snowberg, Wang, and Camerer (2019); Dean and Ortoleva (2019); O'Donoghue and Somerville (2018).
[40] See also Cohen et al. (2020) for a review of work finding modest or no correlations between personality and patience measures, and Ericson and Laibson's (2019) discussion of conjectured links between conscientiousness and self-control.

behavioral biases (Section 2), on prospects for taming the "Cognitive Zoo" (Section 5-A), and on relationships between groups of biases and outcomes (Section 6)-- will help convince other economists that allowing for consumer decision making limitations is important and tractable. We have focused our discussions on micro models, but note that our results have implications for macro models as well. E.g., they cast doubt on each of the "… three crucial assumptions underlying many [classical] models" articulated by Hommes (2021): "(i) agents have rational expectations; (ii) agents behave optimally [i.e., successfully maximize]….; (iii) agents have an infinite horizon…." And they strongly motivate allowing for multiple frictions or biases in representative agent models where the average consumer drives equilibrium outcomes.

Accounting for cross-consumer heterogeneity in behavioral biases is important. It has long been understood that there is substantial heterogeneity in cognitive skills and other classical decision inputs, and our work documents that heterogeneity in behavioral biases is also substantial, and strongly related to heterogeneity in cognitive skills and outcomes. This helps motivate the use of heterogenous agent models for many applications, including macroeconomic ones.

Basic constructs developed to help organize the study of behavioral biases are informative. As we discuss extensively above, our ex-ante groupings of biases—into groups of related biases and into broader groups of preferences, beliefs, and decision quality—provide useful, albeit incomplete, frameworks for understanding how biases and other cognitive processes are related to each other.

E. Summing up implications thus far

Our results inform the development of powerful and parsimonious models of consumer decision making. They fit best with processing-based models of behavioral biases, validate widely-used constructs in the study of biases, and provide guidance on assessing several key assumptions that underpin consumer choice models of all stripes. They also suggest several specific promising lines of inquiry for further work, while developing measurement tools and econometric approaches for facilitating such work.

6. Conditional correlations between behavioral common factors and outcomes

We now examine empirical correlations between our behavioral common factors, cognitive skills, and various outcomes: objective financial condition, subjective financial well-being, overall life satisfaction, happiness, and health status.[41] These results shed light on whether BCFs tend to matter for outcomes, and provide additional empirical guidance on which factors researchers should consider when modeling relationships between decision inputs and outputs for a given outcome domain.

A. Specification and variable construction

The general form of our empirical specification is:

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

where *Outcome* is one of the measures detailed below in Section 6-C for each panelist *i* and survey round *t*. In various specifications, we correlate each outcome with one or more factor scores $BCFS_{it}^{j}$ detailed in Section 4-C, where j indexes our four behavioral common factors. Some specifications also include additional covariates X_{it}; e.g., a common factor capturing cognitive skills,[42] the Big 5 personality traits, relatively exogenous demographics (age, gender, race and ethnicity), and more endogenous variables including income and education (see Appendix Table 2 for a complete list of covariates and their definitions).

B. Econometric issues

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

The first is measurement error, because the *BCFS* use not only the loadings from Table 5 but also the values c_{it} of each characteristic *k*—which are themselves measured with error. To address this we once again 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.[43] That

[41] 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. Such convergence issues are common; see, e.g., https://www.stata.com/manuals/semintro12.pdf.
[42] We estimate the cognitive skills common factor from the four variables listed in Table 1. The data suggest one common factor underlying all four variables, including all four with loadings between 0.4 and 0.7 (see also Appendix Table 1 Panel A). We include that common factor in the models of Table 7 Panel C and Appendix Table 4 because it is easier to interpret, particularly when comparing magnitudes of coefficients to those on the BCFs. Including the four cognitive skills variables separately does not materially affect the coefficients on the BCFs but makes those comparisons murkier, and also causes collinearity problems that complicate interpreting coefficients on the cognitive skills variables.
[43] Exceptions are our five personality traits, which we elicit only once.

allows us to again use ORIV to estimate the models, using both the first elicitation to instrument for the second and the second elicitation to instrument for the first, clustering standard errors by panelist. ORIV will produce an unbiased estimate of the univariate correlation between a common factor and outcome and *Y* if the measurement errors in the common factors are uncorrelated across rounds.[44]

A second potential obstacle is that while we would like to accurately measure partial correlations between BCFs and outcomes, it is not clear what should be the appropriate set of controls. Simply correlating a BCF with an outcome without including any controls could overstate the strength of the true empirical link between the two, if for example both are correlated in the expected directions with omitted cognitive skills. Relatedly, a partial correlation between a BCF and outcome, estimated conditional on endogenous covariates that are causally and negatively affected by the BCF (e.g., income and education, perhaps), could be statistically biased due to over-controlling. We therefore present results for both univariate and different multivariate specifications, to shed light on which BCF-outcome links are more robust to different specifications.

C. Results

Table 7 reports results for three specifications, one per panel, for each of six outcomes. We normalize every factor score, so the coefficients are the marginal effect of a one-standard deviation 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 and sample means presented in the penultimate row of Table 7. The first outcome (Column 1) is an index of five 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 measures of subjective financial condition: financial satisfaction, savings adequacy for retirement and other purposes, and lack of financial stress. The third and

[44] 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.

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.[45]

Panel A shows univariate correlations—one regression per cell. The first three BCFs are strongly correlated with nearly every outcome, and all 18 of these coefficients are negative: more bias is associated with worse outcomes. Magnitudes are large for the financial condition variables, and especially so for objective financial condition. E.g., the coefficient of -0.146 on the belief BCF implies that a one-standard deviation increase in that BCF, in the cross-section of consumers, is associated with a reduction in objective financial condition of roughly 28% on its mean. Magnitudes are much smaller for the life satisfaction, happiness, and health status outcomes, on the order of 3 to 5% changes per SD change in BCF, but still statistically distinguishable from zero at conventional thresholds for nearly all the 18 outcome-BCF pairs for the first three BCFs. The fourth BCF (risk/uncertainty) has modest if any links with outcomes—the confidence intervals exclude changes larger than |8%|. One can interpret this as additional evidence that the risk/uncertainty BCF captures preferences *per se* and that our outcome measures are indeed utility proxies: consumers who implement their risk/uncertainty preferences, even behavioral ones, need not experience utility reductions.

The Panel B specification includes all four BCFs in each specification, without any other covariates—one regression per column. Standard errors on the first two BCFs increase substantially, as one would expect given their 0.71 correlation with each other. Coefficient magnitudes tend to fall substantially on the first three BCFs, which again is unsurprising given their inter-correlations (Table 5). Nevertheless 17 of the 18 coefficients remain negative, many remain statistically different from zero at conventional levels, and several imply economically large magnitudes as well. Estimates for the risk/uncertainty BCF are basically unchanged, as one expects given the lack of correlation between this BCF and the others (Table 5).

[45] 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 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.

The Panel C specification adds cognitive skills, Big 5 personality traits, and demographics that are relatively exogenous: gender, age, race, and ethnicity.[46] Standard errors on the BCFs are basically unchanged, suggesting that adding relatively exogenous covariates does not exacerbate the multicollinearity problem. 15 of the 18 coefficients on the first three BCFs remain negative. But their magnitudes tend to decrease, while the cognitive skills common factor is positively correlated with each outcome. This is consistent with cognitive skills playing a more robustly fundamental role than behavioral biases in driving non-financial outcomes, although the identification challenges discussed above caution against drawing firm inferences. For financial outcomes, on the other hand, the coefficient on one or more of the first three BCFs implies a larger relationship than that for cognitive skills.

Appendix Table 4 further explores robustness by estimating several additional covariate specifications, focusing on the subjective financial condition outcome. We focus on this outcome because it is arguably our best-measured aspect of well-being, and because it is interesting to include objective financial condition as an additional covariate in our richest specification; i.e., we estimate partial correlations between subjective financial condition and BCFs and cognitive skills, conditional on other covariates including objective financial condition (in Column 6). Reading across columns, covariate specifications start as relatively parsimonious; e.g., Column 1 includes only the BCFs and relatively exogenous demographics. Each subsequent column adds covariates, with Columns 5 and 6 most likely to be overcontrolled, as Column 5 includes all demographics including education and income, cognitive skills, and personality traits, and Column 6 adds objective financial condition. Column 4 is the same specification we use in Table 7, Panel C. The overall pattern suggests robustness for moderately-sized correlations between subjective financial condition and BCFs 2 and 3, and for a lack of correlation with BCF 4. Inferences for BCF 1 and for cognitive skills are more sensitive to covariate specification.

We see the main takeaways from this section as twofold: first, omitted variable bias and/or over-controlling could present challenges to inference when linking biases and other decision inputs to outcomes such as financial condition and subjective well-being. Second, despite those challenges the links between our BCFs and financial condition are robust. Correlations between

[46] By way of contrast, income and education are much more likely to be endogenous, and we present specifications including them below in Appendix Table 4.

the fourth BCF (risk/uncertainty) and outcomes are rather precisely estimated zeros, regardless of outcome and covariate specification. Correlations between the first three BCFs and outcomes are almost universally negative, and between cognitive skills and outcomes almost universally positive, regardless of outcome and covariate specification. These correlations are economically meaningful in many specifications, and the correlations between subjective financial condition and BCF 2 (decision quality) and BCF 3 (discounting) are particularly robust.

Overall it seems reasonable to infer that cognitive skills are positively implicated in many outcomes, while some groups of behavioral biases are negatively so. Which if any decision inputs are more fundamental in driving outcomes is a vital question for future research.

7. Conclusion

We investigate the taxonomy of consumer decision making. We find that behavioral biases are commonplace, heterogeneous in the cross-section of consumers, amenable to dimension reduction in various ways, and linked to each other and to outcomes through a set of behavioral common factors and cognitive skills. Lower cognitive skills are associated with more behavioral biases, especially strongly so for biased beliefs and choice inconsistencies, and robustly positively correlated with outcomes.

Our results support processing-based theories where basic cognitive functions like intelligence, attention, and/or memory are fundamental and provide a microfoundation for various behavioral biases, including seemingly disparate ones. Closely related, basic constructs used to organize the study of behavioral economics, as we do with our ex-ante groupings (present-biases, inconsistent choices, risky choice, etc.), are clearly useful but incomplete. And preference-based theories receive mixed support in our data.

Even more broadly, our results make a strong case that accounting for behavioral agents is important for economic modeling, but that doing so often requires some richness: researchers should be wary of homogeneity and separability assumptions. Our results can aid in assessing such assumptions, and our approaches to measurement and inference provide tools for collecting and interpreting new data as-needed.

Our work here also highlights many promising directions for future research. Conducting similar analysis with additional and/or different biases, including social preferences, should be fruitful. It would be especially impactful to identify which are the key fundamental cognitive processes, figure out the best ways to measure them, and assess how well they function in the cross-section of consumers. Such data, together with data on the specific biases predicted to emerge from processing-based models, would facilitate further development of such models and sharper testing of them.

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

Abaluck, Jason, and Abi Adams-Prassl. 2021. “What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses.” *The Quarterly Journal of Economics* 136 (3): 1611–63.

Abaluck, Jason, and Jonathan Gruber. forthcoming. “When Less Is More: Improving Choices in Health Insurance Markets.” *The Review of Economic Studies*.

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.

Anusic, Ivana, and Ulrich Schimmack. 2016. “Stability and Change of Personality Traits, Self-Esteem, and Well-Being: Introducing the Meta-Analytic Stability and Change Model of Retest Correlations.” *Journal of Personality and Social Psychology* 110 (5): 766–81.

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.” *European Economic Review* 127 (August): 103493.

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. 2019. “Errors in Probabilistic Reasoning and Judgment Biases.” In *Handbook of Behavioral Economics: Applications and Foundations 1*, edited by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, 2:69–186. North-Holland.

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.

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.
Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2012. “Salience Theory of Choice Under Risk.” *The Quarterly Journal of Economics* 127 (3): 1243–85.
———. 2020. “Memory, Attention, and Choice.” *Quarterly Journal of Economics* 135 (3): 1399–1442.
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 Field.”
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.
Bushong, Benjamin, Matthew Rabin, and Joshua Schwartzstein. 2021. “A Model of Relative Thinking.” *The Review of Economic Studies* 88 (1): 162–91.
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.
Caplin, Andrew. 2016. “Measuring and Modeling Attention.” *Annual Review of Economics* 8 (1): 379–403.
Caplin, Andrew, Dániel Csaba, John Leahy, and Oded Nov. 2020. “Rational Inattention, Competitive Supply, and Psychometrics.” *The Quarterly Journal of Economics* 135 (3): 1681–1724.
Chakraborty, Anujit, Yoram Halevy, and Kota Saito. 2020. “The Relation between Behavior under Risk and over Time.” *American Economic Review: Insights* 2 (1): 1–16.
Chapman, Jonathan, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer. 2020. “Econographics.”
———. 2021. “On the Relation between Willingness to Accept and Willingness to Pay.”
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.
Chun, Marvin M., Julie D. Golomb, and Nicholas B. Turk-Browne. 2011. “A Taxonomy of External and Internal Attention.” *Annual Review of Psychology* 62: 73–101.
Cohen, Jonathan D., Keith Ericson, David Laibson, and John Myles White. 2020. “Measuring Time Preferences.” *Journal of Economic Literature* 58 (2): 299–347.
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.

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.

Enke, Benjamin, and Thomas Graeber. 2021. “Cognitive Uncertainty.”

Enke, Benjamin, and Florian Zimmermann. 2019. “Correlation Neglect in Belief Formation.” *The Review of Economic Studies* 86 (1): 313–32.

Epper, Thomas, Helga Fehr-Duda, and Adrian Bruhin. 2011. “Viewing the Future through a Warped Lens: Why Uncertainty Generates Hyperbolic Discounting.” *Journal of Risk and Uncertainty* 43 (3): 169–203.

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

———. 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.

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, 2:261–343. Elsevier.

Gabaix, Xavier, and David Laibson. 2017. “Myopia and Discounting.”

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.
Handel, Benjamin, and Joshua Schwartzstein. 2018. "Frictions or Mental Gaps: What's Behind the Information We (Don't) Use and When Do We Care?" *Journal of Economic Perspectives* 32 (1): 155–78.
Heckman, James J, Tim Kautz, and Tomáš Jagelka,. 2021. "Some Contributions of Economics to the Study of Personality." In *Handbook of Personality: Theory and Research*, edited by Oliver P. John and Richard W. Robins, Fourth. Guilford Press.
Heidhues, Paul, Botond Kőszegi, and Philipp Strack. 2018. "Unrealistic Expectations and Misguided Learning." *Econometrica* 86 (4): 1159–1214.
Hommes, Cars. 2021. "Behavioral and Experimental Macroeconomics and Policy Analysis: A Complex Systems Approach." *Journal of Economic Literature* 59 (1): 149–219.
Jagelka, Tomas. 2020. "Are Economists' Preferences Psychologists' Personality Traits? A Structural Approach."
Jensen, Arthur Robert. 1998. *The g Factor: The Science of Mental Ability*. Praeger.
Khaw, Mel Win, Ziang Li, and Michael Woodford. forthcoming. "Cognitive Imprecision and Small-Stakes Risk Aversion." *The Review of Economic Studies*. https://doi.org/10.1093/restud/rdaa044.
Kőszegi, Botond. 2014. "Behavioral Contract Theory." *Journal of Economic Literature* 52 (4): 1075–1118.
Kőszegi, Botond, and Adam Szeidl. 2013. "A Model of Focusing in Economic Choice." *The Quarterly Journal of Economics* 128 (1): 53–104.
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.
Lian, Chen. forthcoming. "A Theory of Narrow Thinking." *The Review of Economic Studies*.
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.
Michelacci, Claudio, and Luigi Paciello. 2020. "Ambiguity Aversion and Heterogeneity in Households' Beliefs."

Miyake, Akira, and Naomi P. Friedman. 2012. "The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions." *Current Directions in Psychological Science* 21 (1): 8–14.
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.
Shenhav, Amitai, Sebastian Musslick, Falk Lieder, Wouter Kool, Thomas L. Griffiths, Jonathan D. Cohen, and Matthew M. Botvinick. 2017. "Toward a Rational and Mechanistic Account of Mental Effort." *Annual Review of Neuroscience* 40 (1): 99–124.
Stango, Victor, and Jonathan Zinman. 2009. "Exponential Growth Bias and Household Finance." *The Journal of Finance* 64 (6): 2807–49.
———. 2021. "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.
Woodford, Michael. 2020. "Modeling Imprecision in Perception, Valuation, and Choice." *Annual Review of Economics* 12 (1): 579–601.

Table 1. Measuring decision inputs: Behavioral biases, standard preferences, cognitive skills, and personality traits

	Definition	Unique values	Non-missing obs		Ex-ante groupings	
			Round 1	Round 2	Related biases	Broader
	(1)	(4)	(5)	(6)	(7)	(8)
Behavioral biases						
Present-bias: Money discounting	Discounts more when sooner date is today vs. 5 weeks from today, across 24 choices	28	803	819	Present-bias	Preferences
Present-bias: Consumption discounting	Chooses less healthy snack today and more healthy snack for 5 weeks from now	3	835	829	Present-bias	Preferences
Violations of General Axiom of Revealed Preference	Inconsistent across 11 choices under uncertainty subject to budget constraint	54	774	799	Inconsist/dom choices	Decision quality
Violations of GARP and dominance avoidance	See above, also counts dominated choices as inconsistent	97	774	799	Inconsist/dom choices	Decision quality
Narrow bracketing	Makes dominated choice(s) given implications of an earlier decision	5	827	837	Inconsist/dom choices	Decision quality
Preference for certainty	Certainty premium > 0, from 20 MPL choices between certain payoffs and lotteries	20	620	620	Risk/uncertainty	Preferences
Loss aversion	Chooses certain $0 over pos expected value gamble(s) with potential for small loss	4	843	845	Risk/uncertainty	Preferences
Ambiguity aversion	Prefers lower but known expected payoff to unknown payoff	17	784	751	Risk/uncertainty	Preferences
Overconfidence in level performance	Self-assessment > actual score on numeracy quiz	4	829	813	Overconfidence	Beliefs
Overconfidence in precision	Indicates 100% certainty about quiz performance and/or future income change	2	793	775	Overconfidence	Beliefs
Overconfidence in relative performance	Greater diff between self-assessed and actual intelligence test rank relative to others	78	844	818	Overconfidence	Beliefs
Non-belief in the law of large numbers	Overestimates variance in sample of 1000 coin flips	21	833	819	Math/statistical	Beliefs
Gambler's fallacy	After 10 straight "heads," thinks prob. "tails" >50%	10	842	817	Math/statistical	Beliefs
Exponential growth bias, loan-side	Underestimates loan APR given other terms	50	778	783	Math/statistical	Beliefs
Exponential growth bias, asset-side	Underestimates future value given other terms	11	761	735	Math/statistical	Beliefs
Limited attention	Regrets paying too little attention to finances, taking opportunity cost into account	5	832	829	Ltd attention/memory	Decision quality
Limited memory	Says will complete short survey for $10 tomorrow but does not complete	2	825	803	Ltd attention/memory	Decision quality
Standard measures of time and risk preferences						
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		

For details on elicitation methods and their antecedents please see Section 1-C and Data Appendix Sections 1 and 2. Our data consist of two survey rounds, of two modules each, conducted 3 years apart. Unless noted otherwise, tables include only those panelists who took all four modules across both rounds (N=845).

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-count						
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-count						
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. Here 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-count components and definitions are summarized in Table 1 and Section 1-C; please see Data Appendix Section 1 for details. The proportion in Column 6 Panel A is: (B-count)/(Number of nonmissing inputs to B-count). In Column 6 Panel B it is: (Count missing inputs)/17 .

Table 3. ORIV correlations among behavioral biases, and between biases and standard preference measures

	PB money	PB cons'n	Viols. GARP	Viols. GARP+	Narr. brack.	Pref. cert.	Loss aver.	Ambig. aver.	OC level	OC prec.	OC rel.	NBLLN	Gambler Fallacy	EGB APR	EGB asset	Lim. att.	Lim. mem.	Patience	Risk av. fin	Risk av. large
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 and uncertainty 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/statistical 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								
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			
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

Pairwise rank correlation point estimates, each estimated using the ORIV method described in Section 3-A. Standard errors are clustered by panelist but not shown; please see Table 4 for a summary of p-values. 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. ORIV correlations of behavioral biases with other biases and decision inputs: summary

		Average correlation						
		Biases			Other decision inputs			
Bias(es)	Related bias set(s)	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	0.01
Present-bias: Consumption discounting	Present-bias	0.14	0.47	0.12	-0.19	0.01	-0.13	-0.01
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	0.01
Preference for certainty	Risk/uncertainty	0.18	-0.13	0.23	-0.07	0.06	-0.30	0.04
Loss aversion	Risk/uncertainty	-0.14	0.10	-0.17	-0.09	0.22	0.25	0.02
Ambiguity aversion	Risk/uncertainty	-0.09	-0.07	-0.10	-0.10	0.13	0.19	0.02
Overconfidence in level performance	Overconfidence	0.22	0.32	0.21	0.02	0.07	-0.59	0.01
Overconfidence in precision	Overconfidence	-0.01	0.12	0.07	0.02	0.03	0.03	0.05
Overconfidence in relative performance	Overconfidence	0.19	0.26	0.16	0.00	0.05	-0.64	0.05
Non-belief in the law of large numbers	Math/statistical	0.25	0.25	0.24	0.05	0.16	-0.49	0.06
Gambler's Fallacy	Math/statistical	0.21	0.28	0.20	0.04	0.21	-0.42	0.04
Underestimates APR: Exponential growth bias, loan-side	Math/statistical	-0.06	0.00	-0.07	-0.01	0.01	0.01	0.01
Underestimates future value: Exponential growth bias, asset-side	Math/statistical	0.21	0.21	0.21	-0.01	0.15	-0.51	0.02
Limited attention	Limited attention/memory	0.06	0.22	0.04	-0.16	-0.01	-0.09	0.01
Limited memory	Limited attention/memory	0.14	0.22	0.14	0.17	-0.06	-0.23	0.01
Preferences	Present-bias, Risk/uncertainty	0.05	0.17	0.04	-0.20	0.10	-0.06	0.02
Beliefs	Overconfidence, Math/stat	0.14	0.21	0.15	0.02	0.10	-0.37	0.03
Decision quality	Inconsist choice; Ltd att/mem	0.18	0.40	0.14	-0.03	0.02	-0.28	0.03
Average correlation	All	0.13	0.25	0.11	-0.06	0.08	-0.25	0.02
|Average correlation|	All	0.22	0.27	0.21	0.10	0.13	0.32	0.07
share positive	All	0.68	0.71	0.68	0.41	0.12	0.21	0.65
share with p<0.10	All	0.50	0.53	0.50	0.06	0.38	0.72	0.25
share with p<0.01	All	0.29	0.47	0.26	0.00	0.25	0.57	0.08
share with p<0.001	All	0.22	0.29	0.22	0.00	0.18	0.49	0.02

Average ORIV correlations are unweighted for each group of variables. For example, the "cognitive skills" column shows, for "Present-bias: Money discounting", the average correlation between that bias and each of the four cognitive skills variables. Standard errors, which are not shown, are clustered on panelist.

Table 5. Common factors from a 4-factor model: loadings, intracorrelations, and eigenvalues

Inputs to factor analysis: biases and standard prefs	Ex-ante classification: Related groups	Common factors BCF 1	BCF 2	BCF 3	BCF 4
Present-bias: Money discounting	Present-bias; Preferences	0.22	0.11	0.88	0.10
Present-bias: Consumption discounting	Present-bias; Preferences	-0.14	0.52	0.49	-0.07
Patience	Standard intertemporal; Preferences	0.08	0.07	-0.78	0.01
Preference for certainty	Risk/uncertainty; Preferences	0.38	0.43	0.23	0.08
Loss aversion	Risk/uncertainty; Preferences	-0.28	-0.32	0.07	0.57
Ambiguity aversion	Risk/uncertainty; Preferences	-0.45	0.20	0.06	0.37
Risk aversion: financial	Standard risk; Preferences	-0.06	-0.09	0.16	0.75
Risk aversion: large stakes	Standard risk; Preferences	0.38	0.13	-0.11	0.65
Overconfidence in level performance	Overconfidence; Beliefs	0.81	0.12	-0.06	-0.04
Overconfidence in precision	Overconfidence; Beliefs	0.20	-0.09	-0.27	0.11
Overconfidence in relative performance	Overconfidence; Beliefs	0.60	0.29	0.12	-0.09
Non-belief in the law of large numbers	Math/stat; Beliefs	0.61	0.38	0.11	0.17
Gambler's Fallacy	Math/stat; Beliefs	0.54	0.35	0.10	0.27
Exponential growth bias, loan-side	Math/stat; Beliefs	0.11	-0.35	0.20	0.03
Exponential growth bias, asset-side	Math/stat; Beliefs	0.76	0.10	0.15	-0.01
Violations of General Axiom of Revealed Preference	Inconsist/dom choice; Decision qual	0.32	0.83	0.12	-0.08
Violations of GARP and dominance avoidance	Inconsist/dom choice; Decision qual	0.42	0.77	0.08	0.04
Narrow bracketing	Inconsist/dom choice; Decision qual	0.44	0.15	0.13	-0.21
Limited attention	Ltd attent/mem; Decision qual	-0.05	0.32	0.12	-0.12
Limited memory	Ltd attent/mem; Decision qual	-0.12	0.79	-0.11	-0.04
	Correlation with:				
	BCF 1	1.00	0.71	0.33	0.03
	BCF 2		1.00	0.39	-0.03
	BCF 3			1.00	0.10
	BCF 4				1.00
	Eigenvalue	4.92	1.97	1.86	1.49
	Variation explained	0.25	0.10	0.09	0.07
	Cumulative variation explained	0.25	0.34	0.44	0.51

Higher BCF values indicate more bias. We conduct factor analysis on the ORIV correlation matrix in Table 3, as detailed in Section 4-A. Factor loadings are detailed in Section 4-B; these are the weights associated with each input (i.e., each factor) in each retained common factor. These weights are "rotated" to facilitate interpretation. We use a promax rotation here, which unlike orthogonal rotations allows for correlations between common factors (hence the "Correlation with" results above). Shaded cells show the common factor in which each ex-ante grouping of related biases/characteristics has the highest |loadings|. See Appendix Table 3 for eigenvalues for, and variation explained by, additional common factors beyond the first four.

Table 6. Statistical relationships between behavioral common factors and other consumer characteristics

	B-count	Behavioral common factors from Table 5			
	from Table 2	BCF 1	BCF 2	BCF 3	BCF 4
Panel A. Pairwise ORIV correlations					
Cognitive skills:					
Fluid intelligence	-0.32	-0.60	-0.47	-0.26	-0.04
Numeracy	-0.23	-0.32	-0.19	-0.13	0.02
Financial literacy	-0.24	-0.47	-0.38	-0.21	-0.08
Executive function	-0.05	-0.32	-0.23	-0.09	-0.04
Key demographics:					
Age (years)	-0.01	0.10	0.06	-0.01	0.08
Education (years)	-0.12	-0.34	-0.23	-0.14	-0.07
Income	-0.12	-0.38	-0.25	-0.15	-0.09
Personality traits:					
Extraversion	0.10	0.10	0.08	0.02	-0.07
Agreeableness	0.07	0.04	0.03	0.01	0.03
Conscientiousness	0.13	0.02	-0.03	-0.01	0.10
Neuroticism	0.04	-0.02	0.02	0.08	0.09
Openness	0.05	-0.03	-0.02	-0.03	-0.01
Panel B. R-squareds					
Cognitive skills	0.18	0.46	0.30	0.07	0.01
Demographics (exogenous)	0.05	0.17	0.13	0.04	0.04
Demographics (all)	0.13	0.36	0.24	0.11	0.11
Personality	0.03	0.01	0.01	0.01	0.03
All variables above	0.27	0.53	0.37	0.14	0.15

BCFs here are the factor scores detailed in Section 4-C, with higher values indicating more bias. Correlations are pairwise and estimated using ORIV, with standard errors clustered on panelist. All correlations >|0.06| have p-values <=0.01. Each r-squared is from an OLS regression of the B-count or BCF listed in the column heading on the variables described in the row heading. Exogenous demographics are age, gender, race, and ethnicity. Other demographics are education, income, household size, marital/cohabitation status, employment status, and state of residence.

Table 7. ORIV correlations between outcomes and behavioral common factors: Three specifications

Outcome variable	*Objective fin. condition*	*Subjective fin. condition*	*Life satisfaction I*	*Life satisfaction II*	*Happiness*	*Health status*
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A. Univariate regressions of each outcome on each BCF					
BCF 1: Beliefs	-0.146	-0.071	-0.032	-0.023	-0.015	-0.036
	(0.011)	(0.009)	(0.008)	(0.008)	(0.010)	(0.009)
BCF 2: Decision quality	-0.142	-0.085	-0.033	-0.025	-0.022	-0.030
	(0.013)	(0.009)	(0.008)	(0.009)	(0.010)	(0.010)
BCF 3: Discounting	-0.105	-0.078	-0.035	-0.022	-0.021	-0.032
	(0.013)	(0.010)	(0.009)	(0.008)	(0.010)	(0.010)
BCF 4: Risk/uncertainty	-0.018	-0.014	0.003	-0.007	-0.006	-0.010
	(0.013)	(0.010)	(0.008)	(0.009)	(0.010)	(0.009)
	Panel B. Multivariate regressions of each outcome on all BCFs					
BCF 1: Beliefs	-0.116	-0.021	-0.022	-0.010	0.006	-0.038
	(0.018)	(0.014)	(0.013)	(0.014)	(0.016)	(0.015)
BCF 2: Decision quality	-0.025	-0.049	-0.003	-0.013	-0.023	0.011
	(0.021)	(0.014)	(0.014)	(0.014)	(0.017)	(0.016)
BCF 3: Discounting	-0.034	-0.041	-0.024	-0.010	-0.011	-0.019
	(0.014)	(0.011)	(0.010)	(0.009)	(0.011)	(0.011)
BCF 4: Risk/uncertainty	-0.013	-0.012	0.006	-0.006	-0.006	-0.007
	(0.012)	(0.010)	(0.008)	(0.009)	(0.010)	(0.009)
	Panel C: Multivariate regressions of each outcome on all BCFs and other covariates					
BCF 1: Beliefs	-0.088	-0.019	-0.016	-0.001	0.006	-0.024
	(0.019)	(0.014)	(0.014)	(0.014)	(0.017)	(0.015)
BCF 2: Decision quality	0.001	-0.039	0.004	-0.007	-0.019	0.014
	(0.020)	(0.014)	(0.014)	(0.014)	(0.017)	(0.016)
BCF 3: Discounting	-0.019	-0.030	-0.018	-0.004	-0.006	-0.015
	(0.013)	(0.010)	(0.010)	(0.009)	(0.011)	(0.011)
BCF 4: Risk/uncertainty	-0.021	-0.010	0.008	-0.006	-0.005	-0.008
	(0.011)	(0.009)	(0.008)	(0.009)	(0.010)	(0.009)
Cognitive skills common factor	0.082	0.024	0.031	0.029	0.029	0.034
	(0.014)	(0.009)	(0.010)	(0.011)	(0.011)	(0.011)
Other covariates?	Yes	Yes	Yes	Yes	Yes	Yes
Mean of LHS	0.53	0.50	0.68	0.64	0.70	0.61
Observations	1689	1689	1687	1617	1567	1673

Each cell presents results from an ORIV regression of the outcome variable described in the column heading on the variables described in its panel title. Standard errors are clustered on panelist. Higher outcome values indicate better outcomes. Common factor variables here are the factor scores detailed in Section 6-A, with higher BCF values indicating more bias and higher cognitive skills CF values indicating more skills. 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. "Other covariates" included in the Panel C regressions are cognitive skills (as shown in the table), personality traits, and plausibly exogenous demographics: age, gender, race, and ethnicity. See Appendix Table 2 for details on covariate variable definitions. Appendix Table 4 shows further variations in the specification for subjective financial condition, varying the sets of other covariates more finely. 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 3 for details on outcome measurement.

Appendix Table 1. Pairwise ORIV correlations among cognitive skills components, and between them and behavioral biases

Panel A. Among cognitive skills components	Fluid Int.	Num.	Fin. lit.	Exec. fcn.
Fluid intelligence	1.00			
Numeracy	0.75	1.00		
Financial literacy	0.62	0.77	1.00	
Executive function	0.51	0.52	0.29	1.00

Panel B. Cognitive skills components and biases	PB money	PB cons'n	Viols. GARP	Viols. GARP+	Narr. brack.	Pref. cert.	Loss aver.	Ambig. aver.	OC level	OC prec.	OC rel.	NBLLN	Gamb. Fallac.	EGB APR	EGB asset	Lim. att.	Lim. mem.
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
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
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
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

Pairwise rank correlation point estimates, each estimated using the ORIV method described in Section 3-A. Standard errors are clustered by panelist and not shown; coefficients >=|0.06| tend to have p-values<0.01 and see Table 4 for summary information on p-values. 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.

Appendix Table 2. Additional variables

Panel A. Outcomes used in Table 7

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 Section 3 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 Section 3 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 Section 3 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 Section 3 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 Section 3 for details on module coverage.

Panel B. Demographics and other additional covariates used in Table 6, Table 7 Panel C, and Appendix Table 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 3. Behavioral common factor (BCF) eigenvalues and explanatory power

	Eigenvalue		Variation Explained	
		Difference from BCF n+1	Proportion	Cumulative
BCF 1	4.92	2.95	0.25	0.25
BCF 2	1.97	0.10	0.10	0.34
BCF 3	1.86	0.37	0.09	0.44
BCF 4	1.49	0.15	0.07	0.51
BCF 5	1.34	0.06	0.07	0.58
BCF 6	1.28	0.12	0.06	0.64
BCF 7	1.16	0.10	0.06	0.70
BCF 8	1.06	0.23	0.05	0.75
BCF 9	0.83	0.07	0.04	0.80
BCF 10	0.76	0.10	0.04	0.83
BCF 11	0.66	0.02	0.03	0.87
BCF 12	0.64	0.02	0.03	0.90
BCF 13	0.62	0.15	0.03	0.93
BCF 14	0.47	0.14	0.02	0.95
BCF 15	0.34	0.04	0.02	0.97
BCF 16	0.29	0.08	0.01	0.98
BCF 17	0.22	0.13	0.01	1.00
BCF 18	0.09	0.09	0.00	1.00

Common factors estimated using the ORIV correlation matrix in Table 3. "Difference" shows the gap between the listed eigenvalue and the next highest. "Proportion" shows the share of variance in our 20 inputs to the factor analysis explained by that BCF. "Cumulative" shows the share of variance explained by all BCFs together, from the highest on the list to the listed factor.

Appendix Table 4. ORIV conditional correlations between behavioral common factors (BCFs) and subjective financial condition: varying specifications

Outcome variable	*Subjective financial condition*					
	(1)	(2)	(3)	(4)	(5)	(6)
BCF 1	-0.024	-0.012	-0.031	-0.019	0.008	0.016
	(0.014)	(0.014)	(0.013)	(0.014)	(0.015)	(0.013)
BCF 2	-0.055	-0.044	-0.038	-0.039	-0.039	-0.034
	(0.016)	(0.014)	(0.014)	(0.014)	(0.014)	(0.012)
BCF 3	-0.020	-0.032	-0.031	-0.030	-0.029	-0.021
	(0.013)	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)
BCF 4	-0.001	-0.012	-0.011	-0.010	0.001	0.002
	(0.013)	(0.009)	(0.010)	(0.009)	(0.009)	(0.008)
Cognitive skills common factor		0.026		0.024	0.007	-0.001
		(0.010)		(0.009)	(0.009)	(0.008)
Relatively exogenous demographics: age, gender, ethnicity, race?	Yes	Yes	Yes	Yes	Yes	Yes
Big 5 personality?	No	No	Yes	Yes	Yes	Yes
Income and education?	No	No	No	No	Yes	Yes
Control for objective financial condition?	No	No	No	No	No	Yes

Each column presents results from a single ORIV regression of subjective financial condition on the behavioral common factor scores and other variables described in the rows, with standard errors clustered on panelist. Higher values indicate better outcomes (see Data Appendix Section 3 for details on outcome measurement). Higher BCF values indicate more bias. All common factor scores are normalized such that a coefficient shows the marginal effect of a one-SD change in its factor score on an outcome. Column (4) here corresponds to the specification shown in Table 7 Panel C Column 2. Each regression here is estimated on a sample of 1689 observations from 845 panelists.

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 (forthcoming) 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 (2020) 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 for measuring a subject's *certainty premium* (CP).[11] 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 mean CP of 0.16 in our sample (SD=0.24, median =0.15), compared to 0.37 (SD=0.15) in Callen

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

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.[12] 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.

[12] 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.[13] 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

[13] 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.[14] We subtract this amount from 50, dropping the 99 subjects whose response to the second

[14] 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 (D. Benjamin, Rabin, and Raymond 2016).

Following Benjamin, Moore, and Rabin (see also D Kahneman and Tversky 1972; D. 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),[15] 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 and Hot Hand Fallacies

These fallacies involve falsely attributing statistical dependence to statistically independent events, in either the gambler's fallacy-- 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. These 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:

---

[15] 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 gambler's 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 gambler's fallacy but substantially above the 32% in Benjamin, Moore, and Rabin. Another indication that we find less evidence of the gambler's fallacy is that, while they infer that "at the individual level, the gambler's fallacy 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,"[16] 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 gambler's (compared to 26% in our sample), and 10% exhibit hot-hand (compared to 14% in our sample). Conditional on exhibiting gambler's, 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).

[16] 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).[17] 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

[17] 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.[18] We again bin individuals into underestimators (the standard bias), overestimators, unbiased, and unknown (14% of the sample).[19] Among those with known bias (N=1,222), the median bias is $0, with a mean of $2 and SD of $14.[20] 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.

[18] 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.

[19] 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.

[20] 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 (Gabaix 2019).

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?"[21] 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

[21] 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 Other Individual Characteristics

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.[22]

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-

[22] 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 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.

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. 2020. "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 Marzilli. 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.
Gabaix, Xavier. 2019. "Behavioral Inattention." In *Handbook of Behavioral Economics-Foundations and Applications 2, Volume 2*, edited by Douglas Bernheim, Stefano DellaVigna, and David Laibson. Elsevier.
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. forthcoming. “Meta-Analysis of Present-Bias Estimation Using Convex Time Budgets.” *Economic Journal*.

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.

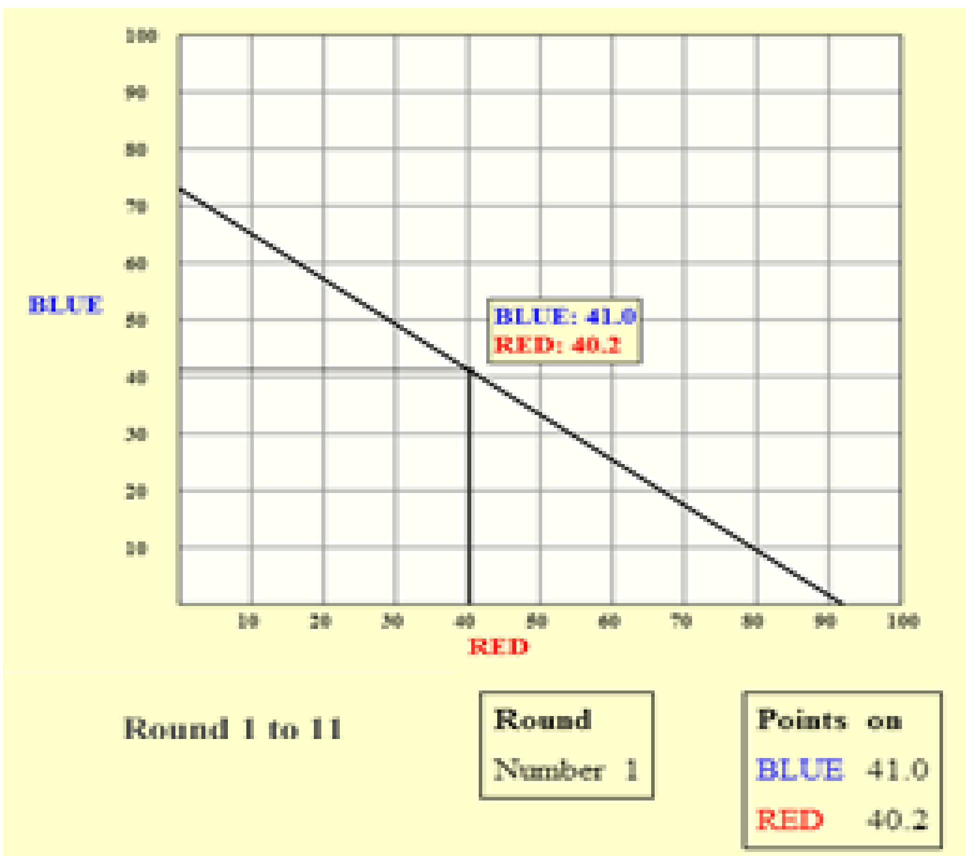
Moskowitz, T.J., and A. Vissing-Jorgensen. 2002. “The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle.” *American Economic Review* 92 (4): 745–78.

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.
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.
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).