Research Dialogue | Issue no. 147
July 2018

TIAA Institute

We are all behavioral, more or less: New consumer-level summary statistics for multiple behavioral factors

Victor Stango,
University of California, Davis

Joanne Yoong,
University of
Southern California

Jonathan Zinman,
Dartmouth College

Abstract

Measuring the individual-level prevalence, heterogeneity, and predictive power of behavioral factors—deviations from classical assumptions about consumer choice—is critical for theory, empirics, and policy. We develop low-cost, informative techniques to elicit a rich suite of behavioral factors from a large, representative sample of individuals. Some behavioral *factors* are less widespread than in previous studies, but nearly all *individuals* are behavioral along one or more dimensions. Individual-level statistics summarizing the extensive and intensive margins of "being behavioral" across multiple factors strongly and negatively correlate with financial condition/well-being and other outcomes, controlling for demographics, risk attitudes, patience, cognitive ability and other correlates.

Stango: UC Davis Graduate School of Management, vstango@ucdavis.edu; Yoong: Center for Economic and Social Research, University of Southern California, National University of Singapore and the National University Hospital System and the London School of Hygiene and Tropical Medicine, joanne.yoong@gmail.com; Zinman: Dartmouth College, IPA, J-PAL, and NBER, jzinman@dartmouth.edu. Thanks to Hannah Trachtman and Sucitro Dwijayana Sidharta for outstanding research assistance, the Michigan Retirement Research Center (MRRC), TIAA and the Pension Research Council (PRC), Russell Sage Foundation, the Roybal Center (grant #3P30AG024962), and the National University of Singapore for funding, Shachar Kariv and Dan Silverman for helping us implement the Choi, Kariv, Muller, and Silverman user 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, and seminar and conference participants at Columbia GSB, NBER, Dartmouth, the Federal Reserve Bank of Philadelphia, National University of Singapore, UC-Davis, UCSD-Rady, DIW- Berlin, and the Aspen Consumer Decision Making Conference for comments.

1. Introduction

Over the last few decades, research at the intersection of economics and psychology has documented and modeled a rich taxonomy of “behavioral factors”—deviations from the classical economic specifications of preferences, decision-making rules and beliefs—that may help explain a wide variety of economic decisions and outcomes.[1] That work includes studies of preferences such as present-biased discounting (Read and van Leeuwen, 1998; Andreoni and Sprenger, 2012a), loss aversion (Fehr and Goette, 2007), preference for certainty (Callen, et al., 2014), ambiguity aversion (Dimmock, et al., forthcoming), and choice inconsistency (Choi, et al., 2014). It also includes studies of biased beliefs, perceptions, and decision rules such as overconfidence (Moore and Healy, 2008), narrow bracketing (Rabin and Weizsäcker, 2009), exponential growth bias (Stango and Zinman, 2009; Levy and Tasoff, forthcoming), statistical fallacies (D. Benjamin, Moore, and Rabin, 2013; D. Benjamin, Rabin, and Raymond, forthcoming), and limited attention/memory (K. M. M. Ericson, 2011).[2] Those studies have become important inputs for economists and policymakers in domains ranging from household finance to energy policy to health economics.[3]

What that body of work lacks to date, and what we develop in this paper, are tractable methods for eliciting empirically useful measures of *multiple* behavioral factors from large samples of individuals. Measuring multiple behavioral tendencies at the level of the individual is important for two broad reasons. First, one might be interested in discerning relationships and correlations between factors, within-individual. Most extant empirical studies measure just one or a few factors, due to budget or methodological constraints,[4] but different factors can produce observably similar effects on behavior while having markedly different policy implications. Single-factor studies face potentially serious omitted variable and identification problems if behavioral factors are correlated with each other.[5]

A second broad reason for measuring multiple behavioral factors, and the focus of the present paper, is to inform the development of models using a behavioral summary statistic meant to capture multiple, reinforcing behavioral factors on consumer decision making (Mullainathan, Schwartzstein, and Congdon, 2012; Allcott and Taubinsky, 2015; Chetty, 2015; Farhi and Gabaix, 2015). Empirical implementation of such models has focused on public finance applications for specific decisions and product markets, e.g., sales taxes on food and alcohol (Chetty, Looney, and Kroft, 2009), health insurance

1 We use behavioral “factor” instead of, e.g., “anomaly” for two reasons: 1) our work suggests that behavioral tendencies are universal, not anomalous; 2) “factor” evokes the crucial point that there are many (behavioral) inputs to decision making.

2 This paragraph cites the papers that had the greatest influence on our methods—which rely on direct elicitation for measuring behavioral factors. We emphasize that the cites here are not meant to be exhaustive: they do not cover all of the important work on each behavioral factor, nor do they cover the complete set of potential behavioral biases or the set of methods for identifying behavioral biases and their effects. Below we discuss how our direct elicitation methods can complement other methods.

3 Some examples are the proliferation of “nudge units” and other centers of applied behavioral social sciences in both government agencies and private sector companies, the invoking of behavioral economics as a basis for rulemaking by agencies ranging from the Consumer Financial Protection Bureau (CFPB) to the Department of Energy, and a recent request for applications by the National Institutes for Health for work in the “identification and measurement of appropriate economic phenotypes in population-based studies, based on approaches honed in behavioral and experimental studies.”

4 Goda, et al. (2015), is an important exception and the paper most similar to ours in examining the prevalence and predictive power of multiple behavioral factors in national samples. They do so for a much smaller number of behavioral factors, as do Bruine de Bruin, Parker, and Fischoff (2007) and Li, et al. (2015), on convenience samples. Tanaka, et al. (2010), do lab-style elicitations for estimating loss aversion, present-bias, and probability-weighting for 181 Vietnamese villagers, and link those elicitations to survey data (on income, etc.), but they consider each behavioral factor independently.

5 Our data shows that behavioral factors are indeed inter-correlated (Stango, Yoong, and Zinman, in progress), with a strong slant toward the positive correlation that is key for applying the tools in behavioral public finance discussed below. See also Dean and Ortoleva (2015) and Gillen, et al. (2015), for evidence from student samples. For other work and discussions regarding interactions among behavioral factors, identification issues, and other challenges in behavioral modeling, see, e.g., Benjamin, Raymond, and Rabin (forthcoming); Ericson (2014); Fudenberg (2006); Heidhues, et al. (2015); and O’Donoghue and Rabin (1999).

(Baicker, Mullainathan, and Schwartzstein, 2015), and lightbulbs (Allcott and Taubinsky, 2015).[6] A narrow focus is understandable and in many cases desirable, given the presumed importance of context in mediating the effect of behavioral tendencies, but the market-by-market approach to applying behavioral insights stops short of testing the sort of domain-level assumptions that motivate high-level policy decisions. Taking household finance as an example, many agencies—the CFPB, the SEC, the Financial Conduct Authority in the UK, etc.—formulate strategy based on assumptions about how behavioral factors impact decision making, and even owe their existence to the assumption that consumers tend to make systematic behavioral mistakes in household finance.

We build a bridge from behavioral sufficient statistic modeling techniques to domain-level applications by creating consumer-level behavioral summary statistics that are useful for evaluating model assumptions and for predicting behavior. We do so by addressing several key and heretofore unanswered questions: Is it possible to obtain useful measures of particular behavioral factors using relatively cheap and quick methods? Can one use such methods to measure a relatively comprehensive suite of factors for each individual in a nationally representative sample? Once measured, do the data indicate that behavioral tendencies are prevalent? Does measuring heterogeneity in behavioral tendencies, across people, improve predictions about economic outcomes that motivate policy interventions—like household financial condition or labor market outcomes—conditional on standard covariates?

We address these questions by developing two new online survey instruments and administering them to a nationally representative sample of more than 1,000 individuals. We start with standard elicitation methods from recent high-profile behavioral studies and modify them for suitability in studies of modest length/budgets by shortening, simplifying and combining tasks.[7] Our modified elicitations are low-touch, not incentivized (with one exception) and therefore, at least by the standards of previous survey and experimental work, not prohibitively expensive.

Altogether, our instruments elicit measures of 16 behavioral factors. Even with this breadth, we acknowledge that we do not capture each and every margin on which individuals might be behavioral (e.g., we do not have a measure of projection bias, or any measures of social preferences), but our survey does collect extensive data on other inputs to decision-making—demographics, financial literacy and cognitive ability, standard measures of risk attitudes and patience, etc.—and on outcomes that might be affected by behavioral factors, particularly in the financial domain. The entire exercise takes roughly 60 minutes of online survey time, spread out over two 30-minute modules, and was fielded in late 2014 and early 2015 to respondents from the American Life Panel (ALP), a nationally representative panel administered by the RAND Corporation.

With data providing a relatively comprehensive picture of behavioral factors in hand, we provide the first broad-based evidence on the prevalence and heterogeneity of such factors at the person-level, and also construct a rich and empirically useful "behavioral parameter" at the individual level. Specifically, we define and measure for each person a "B-count," measuring the number of dimensions on which an individual is behavioral. It turns out that in our data, nearly everyone is behavioral, exhibiting one or more behavioral factors. That finding is not an artifact of measuring so many factors that someone is bound to exhibit one deviation among the 16 (individuals at the 10th percentile of the count of behavioral indicators exhibit a "B-count" of six), nor does the inference that we are all behavioral depend on observing a greater incidence of "being behavioral"

6 Much behavioral work in household finance is similar in focusing on a particular product market; see, e.g., Bertrand, et al. (2010), Stango and Zinman (2011), and Stango and Zinman (2014).

7 In this sense, we follow in the footsteps of prior work on modifying lab-type elicitation methods for use in nationally representative surveys, including Barsky, et al. (1997), Dohmen, et al. (2010; 2011), and Falk, et al. (2015; 2015). Unlike ours, that work does not focus on measuring behavioral factors. We also build on work in developing countries, using local samples, that modifies lab-type methods for measuring behavioral factors—albeit a small number of them—in surveys, including Ashraf, et al. (2006), Callen, et al. (2014), and Gine, et al. (2015).

relative to previous work. In fact, we tend to find *lower* prevalence factor-by-factor compared to prior work.[8] Rather, it is the aggregation of individual factors—capturing a heretofore unseen picture of what is potentially behavioral about a person—that renders "being behavioral" closer to universal than anomalous.

We also find substantial cross-sectional heterogeneity in the prevalence and intensity of behavioral indicators; this is the "more or less" qualifier to "we are all behavioral." That heterogeneity exists across factors because some are more common than others. It also exists across people, with some exhibiting a small number of our 16 factors and some individuals being behavioral on nearly every dimension. Individuals vary in both the "B-count," measuring the extensive margin of being behavioral across multiple factors, and a "B-tile," measuring the intensive margin, where an individual lies, across all factors, in the distribution of being behavioral.

Perhaps most usefully, we show that B-counts and B-tiles explain cross-sectional variation in financial condition and other outcomes. Both B-counts and B-tiles are robustly negatively correlated with a rich summary index of financial condition, capturing both "hard" outcomes like wealth, savings and stock market participation, as well as "soft" self-assessed outcomes like financial distress and self-evaluated retirement savings adequacy.[9] The results hold conditional on an unusually rich set of covariates, some of which we elicit as part of our survey instrument, covering not only standard demographics such as income and education, but also measuring preference parameters such as risk aversion and patience, human capital/cognitive ability metrics such as numeracy, financial literacy and executive attention, and other standard correlates. In all of our empirical models, a one standard deviation change in a B-count has a conditional correlation with financial condition that is larger than the ones for cognitive ability, and a more robust correlation than many variables commonly thought to be important correlates of financial decisions and outcomes (like gender, education, standard measures of risk attitudes, and patience). We also show that B-counts are meaningfully linked to both education and income outcomes in the cross-section.

We consider and find little support for the possibilities that our B-count simply measures variation in mathematical ability or survey-taking effort—confounding factors that if also correlated with financial condition could explain our findings. We parse our behavioral factors into those with a "right answer" in mathematical terms (like understanding compounding) and those without any one correct answer (like the degree of present-bias in discounting), but find no evidence that variation in "math bias" answers drives our results. Controlling flexibly for survey- taking effort (using question-by-question response times) has no effect on the results.

Further corroborative evidence comes from results that exploit the directionality of certain biases—such as time-inconsistent discounting, which can be either present-biased or future- biased. Our results show that the directionality of bias matters in the way predicted by most existing work: "standard biases" (such as present-bias, or under-estimating the effects of compounding, or overconfidence) are significantly negatively correlated with financial condition, while similar-magnitude "non-standard biases" (such as future-bias, or over- estimating exponential growth, or under-confidence) have no significant correlation with financial condition. This pushes harder against a story that our B-counts measure noise, or "mistakes," or math ability, or other sources of variation that would more likely predict symmetric negative effects in either direction.

8 Another companion paper focuses on the prevalence, heterogeneity, and predictive power of individual behavioral factors (Stango, Yoong, and Zinman, in progress).

9 Our financial outcome measurement is a contribution in its own right, in the sense that we show how it captures signals from inter-correlated measures of wealth, assets, recent (dis)saving, self-assessed financial condition, and severe financial distress.

Although we stop short of making inferences about welfare, our results belie an alternative interpretation that we are simply measuring—or measuring more finely—classical preference parameters, or omitted but "not-behavioral" variables. Importantly, our B-count is negatively correlated with self-assessed financial well-being considered separately from the more welfare-ambiguous outcomes like savings and wealth. Indeed, the pattern of results suggests that we are measuring something distinct from classical preferences like patience, or human capital metrics such as cognitive ability. The pattern is consistent with the view that behavioral tendencies can reduce economic welfare. Having said that, we stop short of inferring that the empirical links between our behavioral summary statistics and outcomes are causal. We plan to explore causal links—and the possibility of reverse causality between outcomes and B-counts—in future drafts and papers.

Altogether, our results suggest that adapting standard direct elicitation techniques for the purposes of measuring multiple behavioral factors is a fruitful approach. Our first pass at implementing this approach has yielded data with several interesting patterns. First, we are all behavioral, more or less. Second, one can summarize how "behavioral" people are with statistics—our B-counts and B-tiles—that require few ancillary assumptions to construct. Third, cross-sectional heterogeneity in these summary statistics is strongly conditionally correlated with financial, labor market, and education outcomes.

Our findings and methods inform several literatures. Our findings support assumptions and inferences that are fundamental to many strands of the behavioral social sciences, from micro (DellaVigna, 2009; Kőszegi, 2014) to macro (Akerlof, 2002; Driscoll and Holden, 2014)[10]: many and perhaps most individuals are behavioral in some sense, and heterogeneity in behavioral tendencies helps explain behavior. We thereby add to the extensive literature on the cross- sectional correlates of wealth accumulation and other measures of household financial condition (Poterba, 2014; Campbell, forthcoming) by showing that behavioral factors are important, not the least because omitting them can confound inferences about other correlates.

Our findings also support many key assumptions currently embedded in the emerging behavioral public finance toolkit,[11] while casting doubt on others.[12] Going forward, we expect that our methods will complement work in behavioral public finance by expanding the set of tools for testing and refining assumptions, for identifying key parameters like the prevalence of behavioral agents, the number of behavioral agents on a given margin, and the extent of their biases (Mullainathan, Schwartzstein, and Congdon, 2012), for identifying differences between experienced utility and decision utility in the large samples required to accommodate heterogeneity in behavioral biases, and for testing and refining predictions by using behavioral "typing" to, e.g., target/tag or estimate heterogeneous treatment effects. More broadly, we expand the direct elicitation toolkit by demonstrating that methods adapted for lower-touch channels, tight research budgets, and/or wider-ranging surveys can still produce useful data.

2. Research design: data, sample and how we measure behavioral factors

In this section we describe our sample, research design—including elicitation methods used to measure behavioral factors—and data (including outcome variables and control variables).

[10] See also Heathcote, Storesletten, and Violante (2009) for a survey of the increasingly important role of consumer heterogeneity in macroeconomic modeling.

[11] Our results support a view that person-level bias is nonnegative, positive for some, and not mean-zero in the aggregate (Allcott and Taubinsky, 2015, p. 2510).

[12] Our results caution against assuming the homogeneity in person-level bias required to use Chetty, Looney,and Kroft's (2009) equivalent price metric to identify the average marginal bias distribution that is a key input to welfare analysis (Allcott and Taubinsky, 2015; Mullainathan, Schwartzstein, and Congdon, 2012).

A. The American Life Panel

Our data come from 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 age 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.

Panel members are regularly offered opportunities to participate in surveys, the purposes of which range from basic research to political polling. More than 400 surveys have been administered in the ALP, and all data is publicly available (after a period of initial embargo). This opens up great opportunities for future work linking our data to other modules.

B. Our research design and sample

Speaking broadly, our goal is to design readily applicable elicitation methods that robustly yield data on the widest possible range of behavioral factors at a reasonable cost. We chose a goal of keeping total elicitation time to an hour. This is a round figure that needn't overwhelm a research budget. We also sought to use elicitation methods that could be employed online rather than in-person (given that in-person elicitation typically comes at higher cost).

In consultation with ALP staff, we divided our elicitations and other survey questions into two 30-minute modules. This strategy adheres to ALP standard practice of avoiding long surveys (based on staff findings that shorter surveys improve both response rates and quality), and allows us to evenly disburse the more difficult tasks across the two modules.

All but one of our elicitations are unincentivized on the margin. Again, this helps manage elicitation costs. There is prior evidence that unpaid tasks do not necessarily change inferences about behavioral factors in large representative samples (Von Gaudecker, Van Soest, and Wengström, 2011; Gneezy, Imas, and List, 2015). Unpaid tasks (with hypothetical rewards) may even offer some conceptual advantages (e.g., Montiel Olea and Strzalecki, 2014).

After extensive piloting, the ALP fielded the first part of our instrument as ALP module 315, sending standard invitations to panel participants ages 18-60 in November 2014. Given our target of 1,500 respondents, the ALP sent 2,103 initial invitations. The invitation remained open until March 2015, but most respondents submitted completed surveys during the first few weeks after the initial invitation, as is typical in the ALP. 1,511 individuals responded to at least one of our questions in module 315, and those 1,511 comprise the sample for our study and the sample frame for part two of our instrument.

The ALP fielded the second part of our instrument as ALP module 352, sending invitations to everyone who responded to module 315 starting in January 2015 (to avoid the holidays), with a minimum of two weeks in between surveys. We kept that invitation open until July 2015. 1,407 individuals responded in part or whole to that second module.

Taken together, the two modules yielded a high retention rate (1407/1511 = 93%), low item non-response rate, and high response quality—all features that suggest promise for applying our methods in other contexts. We end up with usable data on a large number of behavioral factors for nearly all 1,511 participants: the respondent-level mean count of measurable behavioral factors is 14 out of a maximum of 16, with a median of 15 and a standard deviation of 2.9. We explore below the possibility that the individual-level degree of missingness in behavioral factors is itself informative in explaining outcomes.

Module 352 also included an invitation to complete a short follow-up survey (module 354) the next day. We use responses to the invitation and actual next-day behavior to measure limited memory as described at the end of the next sub-section.

C. Measuring behavioral factors: Elicitation methods and key antecedents

Given our goals of robustly eliciting behavioral factors without breaking the bank, we prioritized elicitation methods that had been featured recently in top journals and were short and simple enough (or could be so modified) to fit into modules that would also allocate substantial survey time to measuring control variables (Section 2-D) and outcome variables (Section 2-E).

Looking ahead to the data elicited, for each factor we measure whether an individual deviates from the classical norm. Deviations can occur on the extensive margin—i.e., is there any deviation? Does an individual have present-biased preferences? In that case, we use a 0/1 indicator to classify whether an individual is behavioral. We use those indicators to construct our "B-count," which is simply the sum of how many factors an individual displays on the extensive margin.

For most factors, the degree of deviation varies, as well, allowing us to measure *how behavioral* an individual is for that factor. We quantify that intensive margin by calculating each person's percentile ranking for each factor compared to others in the sample. Some of our factors are continuous, permitting percentiles to take on the full range of values from 1 to 100. Others are discrete, which creates lumpiness in the distributions but still allows calculations of percentiles. We then sum all of the factor-level percentiles to calculate each person's "B-tile," which is a quantitative measure that captures degrees of bias relative to others in the sample.

Deviations from classical norms may be unidirectional, as in the case of choice inconsistency: someone either chooses consistently with the General Axiom of Reveal Preference, or does not. For other factors, deviations from classical norms are bidirectional. For example, in the case of discounting, one can be either present-biased or future-biased (relative to being unbiased). We have eight unidirectional factors and eight bidirectional factors, yielding a total of 16 "behavioral indicators."

For bi-directional B-factors, we define in each case a "standard" direction based on what has been more commonly observed or cited in prior work. For example, work on Exponential Growth Bias (EGB) more commonly finds that people underestimate than overestimate the effects of compounding on future values, and so we count underestimation as the standard bias and overestimation as non-standard. Directionality provides a potentially useful avenue empirically, as in some cases the evidence is stronger that "standard" biases are the welfare- reducing ones—we test and find evidence in support of that view below.

We discuss definitions, elicitation methods and other particulars factor-by-factor below.[13]
Tables 1 and 2 provide summaries.

Present- or future-biased discounting with 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).

We measure discounting bias with respect to money using the Convex Time Budgets (CTB) method created by Andreoni and Sprenger (2012a). In our version, subjects make 24 decisions, allocating 100 hypothetical tokens each between (weakly) smaller-sooner and larger-later amounts. The 24 decisions are spread across 4 different screens with 6 decisions each. Each screen varies the start date (today or 5 weeks from today) and delay length (5 weeks or 9 weeks); each decision within a screen offers a different yield on saving.

[13] In defining behavioral factors, we impose minimal assumptions (as opposed, to say, the complementary exercise of using the data to estimate the parameters of a particular model).

We calculate biased discounting for each individual by subtracting the savings rate when the sooner payment date is five weeks from today from the savings 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.

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[14] and could actually lead to more wealth accumulation.

Present- or future-biased discounting with consumption

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?"* A respondent exhibits present bias by choosing *(consume treat today, plan to eat healthy in the future)* and future bias by choosing *(consume healthy today, plan to eat treat in the future)*.[15] These are discrete indicators of bias in either direction. As with money discounting, we denote present-bias as "standard."

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 real-world 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 to 11 decisions.[16] 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.

Following Choi, et al., we average across these 11 decisions 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 respect to first-order stochastic dominance (FOSD).[17] Hence, again following Choi, et al., our second measure captures inconsistency with both GARP and FOSD.[18] Choice inconsistency is unidirectional: we classify an individual as consistent or inconsistent.

14 Although see Koszegi and Szeidl (2013) for a theory of future-biased discounting.

15 If we limit the sample to those who did not receive snack-related informational/debiasing treatment about self-control in ALP module 212 (Barcellos and Carvalho, 2014), we find 15% with present bias and 8% with future bias (N=749).

16 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 decision quality 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.

17 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.

18 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.

Risk attitude toward certainty vs. gambles

Behavioral researchers have long noted a seemingly disproportionate preference for certainty (PFC) vs. gambles, and posited various theories to explain it, including Disappointment Aversion (Bell, 1985; Loomes and Sugden, 1986; Gul, 1991) and u-v preferences (Neilson, 1992; Schmidt, 1998; Diecidue, Schmidt, and Wakker, 2004). PFC may help to explain extremely risk-averse behavior, such as not participating in the stock market.

We use Callen, et al.'s (2014) two-task method for measuring a subject's certainty premium (CP).[19] In one task, subjects make 10 choices between two lotteries, one a $(p, 1\text{-}p)$ gamble over X and $Y > X$, $(p; X, Y)$, the other a $(q, 1\text{-}q)$ gamble over Y and 0, $(q; Y, 0)$. Both we and Callen, et al., 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. 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). Of these subjects, 1,049 choose consistently with monotonic utility and switch on both tasks, as is required to estimate the CP.[20]

We estimate the 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 large outcome, Y, such that one can refer to certainty of X being worth a specific percent chance of Y relative to its uncertain value," and the sign of CP 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. CP<0 indicates a cumulative prospect theory (CPT) type. We denote PFC as the standard bias, simply because CP>0 is far more common than CP<0 in our data (Table 2).

Loss aversion/small-stakes risk aversion

Loss aversion refers to placing higher weight on losses than gains, in utility terms. 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) (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.[21] Small-stakes risk aversion is also often classified as behavioral because it is incompatible with expected utility theory (Rabin, 2000).

Our indicator of loss-aversion/small-stakes risk aversion equals one if the respondent rejects either lottery. We also order sets of deviations to indicate greater degrees of loss aversion based on whether the individual

[19] Callen, et al., describe the method as "a field-ready, two-[task] modification of the uncertainty equivalent presented in Andreoni and Sprenger (2012b)."

[20] Eleven percent of our sample 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.). Fourteen percent of our subjects switch too soon for monotonic utility in the two-lottery (compared to 13% in Callen, et al.).

[21] A related point is that there is no known "model-free" method of eliciting loss aversion (Dean and Ortoleva, 2015).

respondent rejects the compound but not the single lottery, rejects the single but not the compound lottery, or rejects both.[22] These are unidirectional indicators; we either classify someone as loss-averse/small-stakes risk averse, or not.

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 Weizsacker (2009). Each task instructs the subject to make two decisions (i.e., two tasks each with two decisions). The two decisions are each between a certain payoff and a gamble, appear on the same screen, and are accompanied by instructions to consider the decisions jointly.

Our first task follows Rabin and Weizsacker'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.[23] As Rabin and Weizsacker 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. Our second task reproduces Rabin and Weizsacker'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 (i.e., who chooses A and D) violates dominance, here with an expected loss of $75 relative to BC. 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 variance.

Putting the two tasks together to create summary indicators of NBDC, our 0/1 indicator captures not broad-bracketing on both tasks, and we then order other responses as indicating greater deviations: narrow-bracketing on either task, narrow-bracketing on the second task, and narrow-bracketing on both tasks. These are unidirectional indicators: we either classify someone as narrow-bracketing, or not.

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. (forthcoming), find that it is indeed conditionally correlated with lower stockholdings and worse diversification in their ALP sample (see footnote 21, and also Dimmock, Kouwenberg, and Wakker (forthcoming)).

We elicit ambiguity aversion using just one or two questions about a hypothetical game in which the respondent chooses from a bag with green and yellow balls, winning $500 if the ball is green. The first question asks which is preferred: Bag One with 45 green and 55 yellow balls, or Bag Two in which the distribution is unknown. Those who choose the 45-55 bag are ambiguity-averse. The survey then asks, among those who are ambiguity-averse, what number of green balls would make the known distribution less attractive than

[22] Our companion paper explores whether subjects playing the single but not the compound lottery misunderstood the questions, but finds only limited support for that hypothesis (Stango, Yoong, and Zinman, in progress).

[23] Given the puzzling result in Rabin and Weizsacker that their 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.

the unknown distribution, allowing us to measure a more-continuous degree of ambiguity aversion.[24] Our measure of ambiguity aversion is unidirectional because it does not allow ambiguity-seeking.

Overconfidence

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 two distinct measures of overconfidence, following Larrick, et al. (2007), and Moore and Healy (2008). The first measure comes from a question that follows questions on simple numeracy and future value: *"How many of the last 3 questions (the ones on the disease, the lottery and the savings account) do you think you got correct?"* Overestimating the number of correct answers is a measure of overconfidence, and underestimating a measure of under-confidence. This variable, therefore, is bidirectional, with overconfidence the "standard" and indeed more common bias. We code these biases the same under all thresholds because few self-assessed scores deviate from the actual score by more than one. The second variable measures overconfidence in precision, as indicated by responding "100%" on sets of questions about likelihoods (of different possible numeracy quiz scores or of future income increases). We combine answers to these two precision questions and code being overconfident on at least one question as the 0/1 measure of overconfidence, but also code those who are overconfident on both questions as being more overconfident.

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 (aka NBLLN) can act as an "enabling bias" for other biases like overconfidence and loss aversion (D. Benjamin, Rabin, and Raymond, forthcoming).

Following Benjamin, Moore, and Rabin (2013; see also Kahneman and Tversky, 1972), we measure NBLLN using responses to the following question:

> *… say the computer flips the coin 1,000 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. We measure NBLLN using the distance between the subject's answer for the [481, 519] range and 78, counting any deviation as non-belief, and greater deviations as greater underestimation or overestimation. Deviations can be bidirectional, but underestimation is far more common in theory and practice, and so we label under-convergence to LLN as the "standard" bias.

The gambler's fallacy

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

We use a Benjamin, Moore, and Rabin (2013) elicitation for the gambler's fallacy (GF):

[24] We code as missing the 165 respondents who exhibit ambiguity aversion on the first question and respond with >45 green balls on the second question.

"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?"

A classic GF (which we label the "standard" deviation) implies a response < 50%, while the "hot hand" fallacy implies a response > 50%. Responding with something other than 50% triggers the 0/1 code for bias, and we also code greater deviations as indicating greater bias.

Exponential growth bias

Exponential Growth Bias (EGB) is a systematic tendency to underestimate the effects of compounding on costs of debt and benefits of saving. It has been shown to affect a broad range of financial outcomes (Levy and Tasoff, forthcoming; Stango and Zinman, 2009).

Our first measure of EGB follows 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?" We infer 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 APR under-estimators, overestimators, unbiased, and unknown bias.[25] Among those with known bias, we can then order individuals by the degree of under- or overestimation. Those who underestimate the loan APR demonstrate the "standard" bias.

Our second measure of EGB comes from a question popularized by Banks, et al. (2007), as part of a series designed to assess numeracy: *"Let's say you have $200 in a savings account. The account earns 10% interest per year. You don't withdraw any money for two years. How much would you have in the account at the end of two years?"* We calculate "asset-side EGB" by comparing the difference between the correct future value ($242), and the future value supplied by the same individual.[26] Those who underestimate display the "standard" direction of bias, although overestimation also occurs (to a much lesser extent).[27]

Limited attention/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., forthcoming; Stango and Zinman, 2014). Behavioral inattention is a very active line of theory inquiry, as well (e.g., Bordalo, Gennaioli, and Shleifer, 2015; Kőszegi and Szeidl, 2013; Schwartzstein, 2014).

In the absence of widely used methods for directly measuring behavioral limited attention, we create our own, using four simple questions. 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 take into account the opportunity cost of attention (Appendix Table 1, Panel A), and we define being behaviorally inattentive as: *"Yes, and I/we often regret not paying greater attention."* (In contrast, we do *not* classify someone as behavioral if they respond: "*Yes, but paying more attention would require too much time/effort.")* A fourth measure of limited attention is based on answers to *"Do you believe that you could improve the prices/terms your household typically receives on financial products/services by shopping more?"*[28] We classify those responding *"Yes, and I/we often regret not shopping*

25 Non-response is relatively small, as only 4% of the sample does not respond to both questions. Most of those we label as unknown-bias give responses that imply or state a 0% APR. Seven percent 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.

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

27 We label as unknown the 9% of the sample answering with future value < present value, the 4% of the sample answering with a future value > 2x the correct future value, and the 1% of the sample who skip this question.

28 This question is motivated by evidence that shopping behavior strongly predicts borrowing costs (Stango and Zinman, forthcoming).

more" as behaviorally inattentive.[29] An affirmative response to any of the four questions represents bias, and we can also order the degree of bias based on the number (1/2/3/4) of affirmative responses.[30] These are unidirectional measures.

We also measure limited prospective memory (e.g., K. M. M. Ericson, 2011), using an incentivized task offered 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 1,352 (out of the 1,407 respondents to Module 352). Among these 1,352, we classify individuals who do not complete the short survey as having limited memory. This is a unidirectional measure taking on only the values 0/1.

D. Measuring control variables: Demographics, cognitive ability, risk attitudes, and patience

Our modules also elicit unusually rich measures of cognitive skills, risk attitudes, and patience—measures of human capital and preference parameters that plausibly affect decisions and outcomes in classical models. These serve—among other purposes—as control variables in our outcome regressions linking behavioral indicators to financial outcomes (Section 4).

We assess general/fluid intelligence with a standard, 15-question "number series" test (McArdle, Fisher, and Kadlec, 2007) that is non-adaptive (i.e., everyone gets the same questions). The mean and median number of correct responses in our sample is 11, with a standard deviation of 3. Another is 2 "numeracy" questions,[31] labeled as such and popularized in economics since their deployment in the 2002 English Longitudinal Study of Ageing.[32] Our mean number correct is 1.7, with a standard deviation of 0.6. Another is a 3-question "financial literacy" quiz developed and popularized by Lusardi and Mitchell (2014).[33] The median respondent gets all 3 correct, with a mean of 2 and a standard deviation of 0.93. We also measure executive function—including working memory and the regulation of attention—using a two-minute Stroop task (MacLeod, 1991).[34]

29 Inattention indicators are strongly but not perfectly correlated across the four questions (Appendix Table 1, Panels B and C).

30 These behavioral limited attention indicator definitions impose a possibly unrealistic homogeneity assumption on the non-behavioral group, namely that individuals who say they do not have limited attention ("No, my household finances are set up so that they don't require much attention" or "No, my household is already very attentive to these matters") are identical, for the purposes of conditionally predicting behavior, to individuals who respond "Yes, but paying more attention would require too much time/effort." Indeed, it may be that the latter responses (and their analog for the shopping question) provide useful signals of time costs that can help control, e.g., for rational inattention. But in practice, more flexible parameterizations do not change the results.

31 "If 5 people split lottery winnings of two million dollars ($2,000,000) into 5 equal shares, how much will each of them get?"; "If the chance of getting a disease is 10 percent, how many people out of 1,000 would be expected to get the disease?" Response options are open-ended.

32 Banks and Oldfield (2007) interpret these as numeracy measures, and many other studies use them as measures of financial literacy (Lusardi and Mitchell, 2014).

33 "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?"; "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 for each of the three questions.

34 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 with "Simon Says," when an American crosses the street in England, etc.) is sometimes referred to as a "Stroop Mistake" (Camerer, 2007).

Each time the subject chooses an answer, that action completes what we refer to as a “round.”[35] The task is self-paced in the sense that the computer only displays another round after the subject completes a round by selecting a response. Subjects completed 71 rounds on average (both mean and median) within the two minutes, with a standard deviation of 21. Mean (median) number correct is 65 (68), with a standard deviation of 24. Mean (median) proportion correct is 0.91 (0.99), with a standard deviation of 0.19. These various measures of cognitive skills are strongly correlated with each other (Appendix Table 2), so in some cases we extract the first principal component of these four test scores to serve as a measure of cognitive ability in the regressions below (and thereby avoid potential collinearity problems).[36]

We also elicit four standard measures of risk attitudes/preferences. The first comes from the adaptive lifetime income gamble task developed by Barsky, et al. (1997), and adopted by the Health and Retirement Study and other surveys.[37] We use this to construct an integer scale from 1 (most risk tolerant) to 6 (most risk averse). The second is from Dohmen, et al. (2010; 2011): “How do you see yourself: Are you generally a person who is fully prepared to take financial risks?” (100 point scale, we transform so that higher values indicate greater risk aversion).[38] The third and fourth are the switch points on the two multiple price lists we use to elicit the certainty premium (Section 2-C). Each of the four measures is an ordinal scale, but we parameterize them linearly for the sake of concisely illustrating that they are strongly correlated with each other (Appendix Table 3). We use the first principal component of the four risk aversion measures in our regressions below.[39]

We elicit patience from the average savings rate across the 24 choices in our version of the Convex Time Budget task (Section 2-C).

Our other source of control variables is the ALP’s standard set of demographic variables, which are collected when a panelist first registers, then refreshed quarterly and merged onto each new module. Our regression tables and notes list and define our demographic control variables.

Finally, we also track and record survey response time, question by question from “click to click.” We aggregate total response time spent for each factor, for each individual in the survey, and in some empirics below control for time spent as a measure of survey effort.

E. Measuring financial outcomes

Finally, we designed our instrument to elicit rich data on financial outcomes for use in predictive analysis (Section 4). We chose nine indicators of financial condition that we construct from 15 survey questions, 14 of which are in module 315 (the question on non-retirement savings adequacy is in module 352). We drew the content and wording for these questions from other American Life Panel modules and other surveys (including the National Longitudinal Surveys, the Survey of Consumer Finances, the National Survey of American Families, the Survey of Forces, and the World Values Survey). The questions

35 Before starting the task, the computer shows demonstrations of two rounds (movie-style)—one with a correct response and one with an incorrect response—and then gives the subject the opportunity to practice two rounds on their own. After practice ends, the task lasts for two minutes.

36 In practice, results are unchanged if we control for the four test scores separately instead of for their first principal component (Section 4-C). The eigenvalue of the 1st principal component is 2.2, and none of the other principal components have eigenvalues greater than 1.

37 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).

38 We also elicit Dohmen, et al.’s general risk taking scale, which is correlated 0.68 with the financial scale.

39 The eigenvalue of the 1st principal component is 1.7, and none of the other principal components have eigenvalues greater than 1.

elicit information on net worth, financial assets, recent savings behavior, severe distress (missed housing utility payments, forced moves, postponed medical care, hunger), and summary self-assessments of savings adequacy, financial satisfaction and financial stress. Each indicator is scaled such that a 1 signals higher wealth or financial security. We describe these data in more detail below when we correlate our behavioral indicators with financial outcomes.

F. Definitions and distinctions: What is "behavioral"?

Some natural questions of interpretation arise with the data above in hand. First, what differentiates a "behavioral" factor from a non-behavioral one? Definitions can vary, but for practical purposes here we think of behavioral factors as those that can lead to welfare-reducing decisions and outcomes. For example, present-bias leads to borrowing decisions that a borrower later regrets: "over-borrowing" that leads to lower utility than forbearance would have yielded. In contrast, impatience—which we also measure but view as classical—leads to greater borrowing, but as a consequence of utility *maximization*. An impatient borrower neither regrets his decision nor views forbearance as being the right move ex post. Similarly, inattention can be rational and welfare-maximizing because of time costs and cognitive limitations—but our measure of behavioral inattention distinguishes that rational inattention from the type that leads to regret. Low levels of numeracy might lead to different decisions, but a person aware of their numeracy will not necessarily attach numeracy to greater or lesser financial well-being, or systematically "under-save" in a way that causes regret—in the way, for example, that someone with Exponential Growth Bias in the standard direction would.

The upshot of our taxonomy is that we try to distinguish classical preferences and problem- solving abilities, which can have ambiguous or neutral effects on financial well-being, from those that will lead both to the "standard" hard metrics of welfare-reducing decisions—lower savings, lower wealth accumulation conditional on income, and so on—and to lower self-assessed financial condition.

One might also wonder how our measured behavioral factors are correlated with measured variables (such as education) or omitted variables (such as a component of numeracy not captured by our survey questions), or simply measure survey effort. We consider these possibilities in detail below, after presenting the primary empirical results.

3. Are we all behavioral? Summary evidence

In this section, we present three complementary answers to the "are we all behavioral?" question. We first show prevalence estimates for our individual behavioral factors, based on the elicitation methods discussed above in Section 2-C. We then discuss construction, prevalence, and heterogeneity of our summary "B-count" and "B-tile" statistics, aggregating behavioral factors to the level of the individual. We also show how B-counts vary within groups segmented by cognitive ability, income, education and gender.

A. Summary statistics on individual behavioral factors

Table 2 presents summary data on the frequencies of individual behavioral factors in our sample. For each factor, we show prevalence relative to the classical benchmark. We also list the share of individuals for whom data are missing due to non-response or nonsensical answers; we treat such instances as possibly informative in our predictive analyses below.

Prevalence varies, with some factors being fairly common, and others less so.[40] The most common B-factors are inconsistency with GARP (and dominance avoidance), non-belief in the law of large numbers, limited memory, and preference for certainty. The least common are discounting biases regarding consumption, gambler's fallacies, and overconfidence. One of our companion papers compares these findings with those in prior work (Stango, Yoong, and Zinman, in progress). In brief, we tend to find weakly less prevalence of behavioral factors than other studies, including those with nationally representative data.

[40] Results are basically unchanged if we use the ALP's population weights.

All that said, most behavioral indicators are far from uncommon at the individual level, and many are seemingly widespread.

B. The "B-count" and "B-tile": Summary behavioral parameters at the person-level

Table 3 aggregates the indicators in Table 2 in two ways that yield individual-level measures of being "behavioral." We first construct a "B-count" summing the 0/1 behavioral indicators measuring deviation from the classical benchmark for each factor. The maximum possible B- count is 16.[41] We show counts including deviations in both directions, and also just including "standard" directions.

Similarly, we calculate for each person a "B-tile," summing across all factors the percentiles describing where each individual lies in the distribution of deviations from the classical benchmark. In many cases this is completely straightforward because deviations are uni directional and continuous, as in the case of deviations from GARP. We normalize the classical benchmark at 0 (rather than 1) to comport with the B-count—i.e., so that someone who meets every classical benchmark has a B-count and a B-tile both equal to zero. If a person were to be the *most* biased person in the sample on every factor, that person would have a B-tile of (close to) 16. For discrete-response and unidirectional outcomes like loss aversion, the tiles take on fewer values but still measure the degree of deviation from classical benchmarks in useful ways.[42] This approach is valuable because it confers comparability across discretely coded factors based on how many people are in each "bin" relative to the benchmark. For bi directional deviations, whether discrete or continuous, the only difference is that we calculate percentiles separately in each direction relative to the classical benchmark.

The B-counts and B-tiles show that nearly all individuals are "behavioral" on one dimension or more. This is not simply a function of small "trembles" away from classical benchmarks. In a previous version of the paper, we imposed larger thresholds for calling something a deviation and still found that 98% of individuals have a B-count of 1 or more. Indeed, most individuals are behavioral on several factors.

That said, the degree to which individuals are behavioral varies quite a bit in the cross- section at each threshold. The median B-count is 9 (8 on "standard" directions), with a standard deviation of 2.5. Missing responses are not a big issue, with the mean (median) respondent supplying data required to measure 14 (15) of the 16 behavioral factors. The B-tile displays similar variation, with some compression because few people are out on the tails systematically.

Although our main focus below is on how cross-sectional variation in B-counts/tiles correlates with financial condition and other outcomes, the raw prevalence exhibited in Table 3 is striking. On the extensive margin, essentially everyone is "behavioral," with a typical individual exhibiting half of the behavioral indicators elicited here.

C. Who is behavioral? B-counts and demographics

A natural question is how our B-count relates to other measurable individual-level characteristics. From a policy perspective, the question might be framed a bit differently: is being "behavioral" more or less widespread in, say, low-income or low-education populations? Many policies now explicitly note a goal of combating the incentives of firms to cater to behavioral biases. Such policies also cite disproportionate effects of such catering on "disadvantaged" populations or sub-groups.

[41] Recall that we have 24 indicators across 16 behavioral factors, but that factors with bidirectional deviations allow for a maximum of one deviation per individual—bidirectional deviations are mutually exclusive within-person.

[42] For example, loss aversion takes on four values: unbiased, and then three ordered responses (whether the individual respondent rejects the compound but not the single lottery, rejects the single but not the 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.

With this in mind, Figures 1-4 show standard B-counts broken out by cognitive ability, gender, income and education. The latter three are collected by the ALP as a matter of course. The bottom line of these splits is that our B-count measure varies substantially within all of the sub-groups we examine. That is to say, being "behavioral" is not confined to those with low cognitive ability, or by gender, or to low-income or low-education individuals. In most cases, the median level of B-count is similar across splits, and any differences are swamped by the within- group variation.

Table 4 shows some related results, regressing B-counts on a rich set of demographics and measures of cognitive skills, standard risk attitudes, and patience. Results are similar if the dependent variable is the individual-level B-tile. We do this with a control for the count of missing behavioral factors (even-numbered columns) and without. Several key patterns emerge. First, many demographic variables have strong (statistically speaking) conditional correlations with B-counts (e.g., gender, age). Second, cognitive ability is also conditionally correlated with B-counts, in the expected (negative) direction (D. J. Benjamin, Brown, and Shapiro, 2013; Burks, et al., 2009; Frederick, 2005; although see also Cesarini, et al., 2012; and Li, et al., 2013). Third, despite these strong correlations, it appears that B-counts are quite far from fully explained by standard factors. One can see this in the magnitudes of the correlations, e.g., a one standard deviation increase in cognitive skills is associated with a 4 to 8 percent decrease in B-count, which is nontrivial but hardly enormous. One can also see this in the R-squareds: Our complete set of covariates, not including the count of behavioral factors with missing data (which mechanically reduces B-counts), explains at most 42% of the variation in a B-count.

Of course, heterogeneity in B-counts could reflect noise rather than signal. We address this in the next section by examining conditional correlations between B-counts/tiles and outcomes, particularly in the financial domain.

4. Do B-counts/tiles help explain financial condition and other outcomes?

In this section, we ask whether our B-counts and B-tiles help explain individual-level financial condition and other outcomes. The central findings are that B-counts/tiles are meaningfully and negatively correlated with overall financial condition, and are also meaningfully negatively correlated with income and education (which some might consider related outcomes).

A. Measuring financial condition

Recall that in Section 2-E we mentioned eliciting a set of indicators for financial condition. There are nine, and Table 5 details the measures, definitions, frequencies in our data, and pairwise correlations. In each case, "1" indicates plausibly better financial condition (greater wealth, more financial security, better "financial health," etc.):

- Positive net worth
- Positive retirement assets
- Owning stocks
- Spending less than income in the last 12 months
- Financial satisfaction (above the median in our data)
- Self-assessing retirement savings as "adequate" or better
- Self-assessing non-retirement savings as "adequate" or better
- Not experiencing severe financial distress in the last 12 months
- Having self-assessed financial stress below the sample median

1,508 of our 1,511 respondents provide data we can use to construct one or more of the nine indicators. The median respondent supplies the full nine, with a mean of 8.8 and standard deviation of 0.64. As Table 6 shows, these indicators are strongly correlated with each other. Each of the 36 pairwise correlations are positive, ranging from 0.02 to 0.82, and 34 have p- values of 0.01 or less.

To measure individual-level financial condition, we take the individual-level mean of these nine indicator variables. In our sample, the average value of this summary measure is 0.43, meaning that the average respondent affirms 4 of our 9 indicators of good financial condition.

As we hinted above, this measure of financial condition includes "hard" outcomes like savings and net worth—which are more concrete but less tightly linked to welfare or financial well-being in theoretical terms—and "soft" self-assessments of financial well-being, which one might view as being more strongly correlated with individual-level welfare. While we go no further than that observation, and stop short of declaring that our metric decisively captures utility or financial welfare, we do conduct some empirical exercises below linking B-counts/tiles to each individual component of our overall metric. The bottom line is that B-counts/tiles seem to be equally strongly correlated with almost all of the individual components above, in contrast to classical preference parameters and decision inputs, which seem more strongly linked to the "hard" outcomes and less strongly linked to the "soft" outcomes.

B. Do B-counts help explain financial condition, income and education?

In Table 6, we take our summary measure of financial condition and regress it on B-counts, B-tiles and a rich set of controls to estimate the conditional correlation between B-counts/tiles and financial condition. Because not all respondents answer the full set of B-factor questions, we include both information about B-counts/tiles and the number of missing responses as separate regressors. Our main specification is:

$$Outcome_i = \alpha + \beta_1 Bfactor_i + \beta_2 Bcount\_Miss_i + \gamma X_i + \varepsilon_i$$

Here, i indexes individuals, *Outcome* is an individual-level economic outcome (such as financial condition), the B-factor variable includes some specification of B-counts, B-tiles or both, and missing B-factor counts are also included as a control.[43] The vector X is the full set of control variables.

We vary the inclusion of b-factors in two main ways. First, we alternately include B-counts and B-tiles to see whether both explain outcomes and whether one seems more impactful. Second, we separate both counts and tiles into "standard" and "non-standard" components, as we detailed in Section 2. Doing so is informative for a few reasons—not least to the extent that behavioral theories predict stronger (negative) effects on financial condition of "standard" biases, these specifications allow corroboration and inference about whether our counts measure "behavioral factors" versus other omitted variables.

Results are reported in Table 6. Other covariates i nclude gender, age, income, education, state of residency, risk attitudes, patience, marital status, household size and employment status, and a full vector of variables measuring cognitive ability. We also include time spent on survey questions in order to control generally for survey effort. Table 6 shows coefficients on a subset of control variables for the purposes of comparing their magnitudes to those of B-count correlations. Appendix Table 4 shows results for a more complete set of control variables.

Starting with columns (2) and (4), we find that both B-counts and B-tiles are negatively correlated with financial condition in an economically meaningful way. The p-values in both cases are less than 0.01 and robust to the inclusion of our full set of controls. Table 6 shows that the coefficient on the count of missing B-factors is not significant once we include the full set of controls (Columns 2+). Column (1) shows how the exclusion of our cognitive ability measures affects (or rather, does not affect) the coefficient on the B-count.

Proceeding to the separation into standard and non-standard, we find that "standard" directions of deviation from classical benchmarks are more impactful than "non-standard" deviations. This holds whether we measure deviations using B-counts or B-tiles.

The size of the coefficients implies that increasing the B-count or B-tile by 4 indicators reduces the level

[43] Results are similar for alternative functional forms; they do not reject linearity.

of summary financial condition by roughly 0.10 (a 25% decline from the mean of 0.43). A one standard deviation increase in the B-count (2.46) is associated with a decline in financial condition that is similar to the difference between someone age 55-60 vs. age 18-35, all else equal. It is also interesting to note that B-counts are more significant correlates of financial outcomes than many variables commonly thought to be important correlates, like cognitive skills, gender, standard measures of risk attitudes, and patience.

In the last column, we include *both* B-counts and B-tiles in order to assess whether, conditional on "being behavioral" on the more extensive margin, the *degree* of being behavioral matters incrementally. The evidence suggests that it does, though it appears that including either B-counts or B-tiles by themselves usefully captures the variation in how behavioral are different individuals. From these results, a key takeaway is that measuring *any* deviations from neoclassical norms—measuring the extensive margins of behavioral deviations at the level of individual factors—can be quite informative. This is important to keep in mind for future applications because it suggests that simple approaches to elicitation can be quite useful.

Tables 7 and 8 show that the B-counts and B-tiles are also strongly conditionally correlated with outcomes in other domains: income (Table 7) and education (Table 8).

C. Robustness and interpretation

Here we consider several issues of robustness and interpretation.

One might wonder whether a single B-factor is driving the results. Table 9 shows that B- counts do in fact capture the contributions of multiple behavioral factors: they are not driven by any one behavioral factor in particular. We show this by rerunning a primary specification (Table 6, Column 2), removing indicator(s) for each behavioral factor one-by-one. For example, the second row of results in Table 9 shows the B-count coefficient where we replace the B-count with one that excludes the indicators for present-biased and future-biased money discounting. Altogether, the results in Table 9 show essentially no difference as we drop these factors one-by- one. No one factor makes an outsized contribution to the significance of the B-count with the possible exception of our limited attention indicator (Section 2-C). Of course, one might expect to find one outlier among 16 factors, simply by statistical chance.[44]

Nor is it the case that a single outcome indicator drives the results. Appendix Table 5 shows results for each of our nine indicators of financial condition (compare to Table 6 Column 2). Our behavioral summary statistics are correlated with both quantitative measures (net worth, retirement assets, savings rate) and qualitative measures (self-assessed savings adequacy, financial satisfaction, financial stress).

A natural concern is that our B-factors do not measure behavioral factors per se, but rather capture unmeasured cognitive ability. We see little evidence that this is true. First, as we noted above, the first two columns in Table 6 present results from regressions with and without our controls for cognitive skills, and the coefficients on the B-counts are stable in the spirit of Altonji, et al. (2005). Conversely, coefficients on cognitive skills are sensitive to the exclusion of behavioral factors, at least for financial outcomes [add comparison to Appendix Table 6].

Table 10 provides additional evidence that B-count results are not driven by a conflation of behavioralness with (math) ability. Here, we segment our B-factors into two categories: those that reflect preferences or decision rules, and a set of "math biases" for which the neoclassical benchmark is a clear correct answer. The math bias category includes EG biases, the gambler's fallacies, and non-belief in the law of large numbers. We then omit the math factors from our B-count measures and re-estimate the models. The results are unchanged in any qualitative or qualitative sense (compare to

[44] Nevertheless, these results do motivate closer scrutiny of our attention variables, and we are undertaking such analysis in our companion papers. See also footnote 27.

Table 6) and are also stable if we include the count of "math bias" factors as a regressor (even-numbered columns in Table 10). Moreover, the coefficients on cognitive ability and education are not significantly affected by the inclusion or exclusion of the "math bias" count, suggesting that even these variables are not correlated with education or cognitive skills in a way that substantively affects the results.

Yet another concern is that the correlations between B-counts and outcomes somehow reflect survey-taking behavior rather than actual behavior. For example, perhaps—for whatever reason—people who exhibit low effort on surveys have worse financial outcomes, and we classify those low-effort people as behavioral due to measurement error. But if this were the case, then large deviations from neoclassical norms should be more negatively correlated with outcomes than small deviations, since large deviations are the strongest indicators of low effort. Yet (in unreported results) we find the opposite result. A complete theory of confounding survey effort would require a positive correlation between survey effort and financial condition. Yet close scrutiny of our task design and user interfaces yields little cause for concern that a low- effort (and hence erroneously behavioral) respondent would be more likely to respond in a way that would erroneously indicate poor financial condition; e.g., it seems no easier, effort-wise, to indicate poor condition than good condition (Appendix Table 7). Empirically, we do not see a strong pattern of extreme responses; e.g., only 9% of our sample exhibits 0 of the 9 indicators of sound financial condition (Table 5). If anything, more behavioral respondents seem to be more "positive perceivers" than "negative nellies," as suggested by the weakly positive conditional correlations between B-count and responses to questions about expected financial condition a year from now (Appendix Table 8).[45] This suggests that any mechanical or artificial relationship between B-counts and self-reported outcomes may actually push *against* our findings of negative correlations. Perhaps most to the point, controlling (flexibly) for time spent on our behavioral elicitations does not change the estimated conditional correlation between the B-count and financial outcomes in our main specification.

A similar finding emerges if we consider the directional nature of the "standard" vs. "non- standard" bias distinction. This again argues against an interpretation of our B-counts as capturing noise or math/cognitive ability that is also negatively correlated with financial condition. If that were true, one might expect symmetric negative correlations regardless of the direction of bias. Our results, to the contrary, suggest that a bias leading to "under-saving" is associated with reduced financial condition, while the direction associated with "over-saving" is not, as one might expect given the received wisdom that the latter is less problematic than the former.[46]

We also consider whether our B-variables might be capturing effects/correlations of omitted non-cognitive skills. A priori there is little reason for concern, as we are not aware of any evidence that even suggests a strong correlation between behavioral factors and non-cognitive skills.[47] We also have checked robustness by adding to our regressions a measure of non-cognitive skills elicited in a later model, and find that the results discussed here do not change.

A final issue of interpretation is whether the B-count correlations reflect reverse causality. Reverse causality would be a novel finding—it would indicate not just instability in behavioral factors (within-subject over time), but a particular cause of instability that would affect how theorists and empiricists model relationships between behavioral factors and decisions—circumstantial evidence casts doubt on its importance for our results. First, in theory, reverse causality could just as easily push in the opposite direction of our results, with worse financial condition leading to *more* deliberate

45 The level of optimism in response to these questions is quite striking, as well, and further pushes against any intuition that certain (erroneously labelled as behavioral) people self-report relatively negatively.

46 Without going into too much detail, the penalty function for under-saving and over-borrowing can involve bankruptcy or foreclosure, which are large, discrete negative events.

47 We reached this conclusion after reading Becker, et al.'s (2012) review article and doing a Google Scholar search of papers that cite it.

consideration of elicitation tasks, less measurement error, and hence fewer deviations from neoclassical norms.[48] Second, the limited empirical evidence on instability in elicited behavioral factors suggests that it is due to measurement error rather than to marginal changes in financial condition or other life circumstances, although disastrous events may play a role.[49] Third, Appendix Table 5 shows that our B-counts are just as strongly correlated with outcomes that are relatively sticky and objectively-measured (e.g., a stock variable like our indicator for positive net worth) as they are with outcomes that are probably relatively unstable and subjectively-measured (e.g., our indicator for whether someone feels stressed by their finances).

5. Conclusion

We directly elicit measures of 16 behavioral factors from more than 1,000 individuals participating in a nationally representative U.S. online panel survey, using low-cost, low-touch, and short adaptations of standard methods. We use the resulting data to construct new summary statistics that capture the prevalence and heterogeneity of behavioral factors across people. These "B-counts"—counts of the number of factors for which an individual indicates a behavioral tendency—show that behavioral factors are closer to universal than anomalous. Nearly all of our sample exhibits at least one behavioral indicator, and most people exhibit several. We also find substantial heterogeneity in behavioral indicators across individuals: the standard deviation of the B-count is 2.5.

Perhaps most importantly, our B-counts and B-tiles are strongly conditionally correlated with outcomes. This inference holds for financial outcomes, for which we have a rich set of data showing that 9 indicators of "different" measures of financial condition/security/wealth are in fact strongly positively correlated with each other. It also holds for income and education, for which we have standard measures. We find little evidence that B-counts/tiles proxy for cognitive skills (for which we have several, strongly inter-correlated measures), and some evidence that omitting behavioral variables can lead to mistaken inferences about the strength of the conditional correlation between cognitive ability and financial outcomes.

This paper only begins to tap the potential of the new survey instrument and dataset described herein. In companion papers, we are exploring the absolute and relative predictive power of different behavioral factors, inter-correlations among behavioral factors, and the common factor structure. There are many possibilities for exploiting the panel, multi-topic architecture of the ALP to explore relationships between our behavioral variables, covariates, and outcomes in yet more domains. Our behavioral data—some of it at least—is also well suited for structurally estimating parameters. Our methods are suitable for collecting data in other settings and should be helpful in adding to the small but growing body of evidence on the reliability (intertemporal stability) of directly elicited behavioral factors. For example, comparing the reliability of individual behavioral factors to that of B-counts—in the ALP and elsewhere— could shed light on the extent to which measurement error drives any instability.[50] This in turn is key to unpacking the direction and extent of any causality underlying conditional correlations between behavioral factors and outcomes.

48 The only exception we know of is present-biased discounting with respect to money, which should, in theory, increase under financial distress if the subject expects their financial condition to improve—and hence the marginal utility of a dollar to decline—over time.

49 Meier and Sprenger (2015) find moderate (in)stability in present-biased money discounting over a two-year period. This instability is uncorrelated with observables (in level or changes), which is consistent with measurement error but not environmental factors (including those that could generate reverse causality) playing an important role. Callen, et al. (2014), find that exposure to violent conflict increases preference for certainty. Li, et al. (2013), find moderate (in)stability in present-biased money discounting and in loss aversion over several months. Carvalho, et al. (forthcoming), find small changes in present-biased money discounting around payday in a low-income sample and no changes in choice inconsistency (or in cognitive skills, *contra*, e.g., Shah, et al. (2012), and Mani, et al. (2013)). There is a larger body of evidence on the reliability of non-behavioral measures of time and risk preferences; see Meier and Sprenger (2015) and Chuang and Schechter (2015) for recent reviews.

50 Chuang and Schechter (2015) speculate that simpler elicitations may produce better reliability by reducing noise.

Pushing further to map links between the multitude of behavioral factors, decisions, and outcomes will improve understanding about individual economic behavior, market functioning, and policy design across the many domains in which behavioral economics has taken hold—energy, household finance, health and others.

References

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.

Akerlof, George A. 2002. "Behavioral Macroeconomics and Macroeconomic Behavior." *American Economic Review* 92 (3): 411–33.

Allcott, Hunt, and Dmitry Taubinsky. 2015. "Evaluating Behaviorally Motivated Policy: Experimental Evidence from the Lightbulb Market." *American Economic Review* 105 (8): 2501–38.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113 (1): 151–84.

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

———. 2012b. "Uncertainty Equivalents: Testing the Limits of the Independence Axiom."

Ashraf, Nava, Dean Karlan, and Wesley Yin. 2006. "Tying Odysseus to the Mast: Evidence from a

Commitment Savings Product in the Philippines." *Quarterly Journal of Economics* 121 (2): 673– 97.

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.

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein. 2015. "Behavioral Hazard in Health Insurance." *The Quarterly Journal of Economics* 130 (4): 1623–67.

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

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.

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.

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

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

Benjamin, Daniel, Don Moore, and Matthew Rabin. 2013. "Misconceptions of Chance: Evidence from an Integrated Experiment."

Benjamin, Daniel, Matthew Rabin, and Collin Raymond. forthcoming. “A Model of Nonbelief in the Law of Large Numbers.” *Journal of the European Economic Association*.

Bertrand, Marianne, Dean Karlan, Sendhil Mullainathan, Eldar Shafir, and Jonathan Zinman. 2010. “What’s Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment.” *Quarterly Journal of Economics* 125 (1): 263–305.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2015. “Memory, Attention, and Choice.”

Bruine de Bruin, Wändi, Andrew M. Parker, and Baruch Fischhoff. 2007. “Individual Differences in Adult Decision-Making Competence.” *Journal of Personality and Social Psychology* 92 (5): 938–56. doi:10.1037/0022-3514.92.5.938.

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

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

Camerer, Colin F. 2007. “Neuroeconomics: Using Neuroscience to Make Economic Predictions.” *The Economic Journal* 117 (519): C26–42. doi:10.1111/j.1468-0297.2007.02033.x.

Campbell, John. forthcoming. “Restoring Rational Choice: The Challenge of Consumer Financial Regulation.” *American Economic Review*.

Carvalho, Leandro, Stephan Meier, and Stephanie Wang. forthcoming. “Poverty and Economic Decision-Making: Evidence from Changes in Financial Resources at Payday.” *American Economic Review*.

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

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

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. “Salience and Taxation: Theory and Evidence.” *The American Economic Review* 99 (4): 1145–77.

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

Chuang, Yating, and Laura Schechter. 2015. “Stability of Experimental and Survey Measures of Risk, Time, and Social Preferences: A Review and Some New Results.” *Journal of Development Economics* 117 (November): 151–70.

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. 2015. “Is It All Connected? A Testing Ground for Unified Theories of Behavioral Economics Phenomena.”

DellaVigna, Stefano. 2009. “Psychology and Economics: Evidence from the Field.” *Journal of Economic Literature* 47 (2): 315–72.

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. forthcoming. “Ambiguity Aversion and Household Portfolio Choice Puzzles: Empirical Evidence.” *Journal of Financial Economics*.

Dimmock, Stephen, Roy Kouwenberg, and Peter P. Wakker. forthcoming. “Ambiguity Attitudes in a Large Representative Sample.” *Management Science*.

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.

Driscoll, John C., and Steinar Holden. 2014. “Behavioral Economics and Macroeconomic Models.” *Journal of Macroeconomics* 41 (September): 133–47.

Ericson, Keith. 2014. “On the Interaction of Memory and Procrastination: Implications for Reminders.”

Ericson, Keith M. 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. 2015. “The Nature and Predictive Power of Preferences: Global Evidence.”

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2015. “The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences.”

Farhi, Emmanuel, and Xavier Gabaix. 2015. “Optimal Taxation with Behavioral Agents.”

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

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

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

Gillen, Ben, Erik Snowberg, and Leeat Yariv. 2015. “Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study.”

Gine, Xavier, Jessica Goldberg, Daniel Silverman, and Dean Yang. 2015. “Revising Commitments: Field Evidence on the Adjustment of Prior Choices.” Working paper. The World Bank, Washington, D.C.

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. 2015. “The Role of Time Preferences and Exponential-Growth Bias in Retirement Savings.”

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.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante. 2009. “Quantitative Macroeconomics with Heterogeneous Households.” *Annual Review of Economics* 1 (1): 319–54.

Heidhues, Paul, Botond Kőszegi, and Philipp Strack. 2015. “Unrealistic Expectations and Misguided Learning.” SSRN Scholarly Paper. Rochester, NY. http://papers.ssrn.com/abstract=2703524.

Kahneman, D., and A. Tversky. 1972. “Subjective Probability: A Judgement of Representativeness.” *Cognitive Psychology* 3: 430–54.

Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman. Forthcoming. “Getting to the Top of Mind: How Reminders Increase Saving.” *Management Science*.

Kartashova, Katya. 2014. “Private Equity Premium Puzzle Revisited.” *American Economic Review* 104 (10): 3297–3334.

Kőszegi, Botond. 2014. “Behavioral Contract Theory.” *Journal of Economic Literature* 52 (4): 1075–1118. doi:10.1257/jel.52.4.1075.

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. forthcoming. “Exponential-Growth Bias and Lifecycle Consumption.” *Journal of the European Economic Association*.

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.

Li, Ye, Jie Gao, A. Zeynep Enkavi, Lisa Zaval, Elke U. Weber, and Eric J. Johnson. 2015. “Sound Credit Scores and Financial Decisions despite Cognitive Aging.” *Proceedings of the National Academy of Sciences* 112 (1): 65–69.

Loomes, Graham, and Robert Sugden. 1986. “Disappointment and Dynamic Consistency in Choice under Uncertainty.” *The Review of Economic Studies* 53 (2): 271–82.

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.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao. 2013. “Poverty Impedes Cognitive Function.” *Science* 341 (6149): 976–80.

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.

Meier, Stephan, and Charles Sprenger. 2010. “Present-Biased Preferences and Credit Card Borrowing.” *American Economic Journal: Applied Economics* 2 (1): 193–210.

Meier, Stephan, and Charles D. Sprenger. 2015. “Temporal Stability of Time Preferences.” *Review of Economics and Statistics* 97 (2): 273–86.

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.

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.

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.

Neilson, William S. 1992. "Some Mixed Results on Boundary Effects." *Economics Letters* 39 (3): 275–78.

O'Donoghue, Ted, and Matthew Rabin. 1999. "Doing It Now or Later." *American Economic Review* 89 (1): 103–24.

Poterba, James M. 2014. "Retirement Security in an Aging Population." *American Economic Review* 104 (5): 1–30.

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.

Read, Daniel, and Barbara van Leeuwen. 1998. "Predicting Hunger: The Effects of Appetite and Delay on Choice." *Organizational Behavior and Human Decision Processes* 76 (2): 189–205.

Schmidt, Ulrich. 1998. "A Measurement of the Certainty Effect." *Journal of Mathematical Psychology* 42 (1): 32–47.

Schwartzstein, Joshua. 2014. "Selective Attention and Learning." *Journal of the European Economic Association* 12 (6): 1423–52.

Shah, Anuj K., Sendhil Mullainathan, and Eldar Shafir. 2012. "Some Consequences of Having Too Little."

Science 338 (6107): 682–85.

Stango, Victor, Joanne Yoong, and Jonathan Zinman. in progress. "'B': Measuring Common Factor(s) in Behavioral Decision Making."

———. in progress. "Which Behavioral Factors Matter? The Measurement and Empirical Importance of Distinct Behavioral Factors."

Stango, Victor, and Jonathan Zinman. forthcoming. "Borrowing High vs. Borrowing Higher: Price Dispersion and Shopping Behavior in the U.S. Credit Card Market." *Review of Financial Studies*.

———. 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.

Tanaka, Tomomi, Colin F. Camerer, and Quang Nguyen. 2010. "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam." *American Economic Review* 100 (1): 557–71.

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.

About the authors

Victor Stango is an economics professor at the Graduate School of Management at the University of California, Davis. His research focuses on household financial decision making over both short- and long-term time horizons. His current work examines how behavioral influences on consumer decision-making are related to each other, to cognitive abilities and other demographics, and to financial decisions and outcomes.

Stango's work has been featured in *The Wall Street Journal*, *The New York Times*, *The New Yorker*, *Business Week*, *Newsweek* and major online business news media. He has appeared on "Good Morning America," Fox News, CNBC, Bloomberg, and many other news programs to discuss his work and provide expert commentary. His research has appeared in the *American Economic Review*, *The Journal of Finance*, *The Review of Financial Studies* and other leading academic journals. Stango is an associate editor of The International Journal of Industrial Organization. He is also an affiliate expert with Cornerstone Research, occasionally providing consulting in matters related to the financial service industry.

Before joining the Graduate School of Management in 2008, Stango gained experience at the Tuck School of Business at Dartmouth College, the Federal Reserve Banks of Chicago and New York, and other academic institutions. He also served for several years on the board of Consumer Credit Research Foundation.

Stango holds a B.A. in economics and political science from the University of Pennsylvania, and a Ph.D. in economics from UC Davis.

Joanne Yoong is an applied micro-economist conducting research on behavioural economics, health and financial decision making, and economic development. Dr. Yoong is a Senior Economist at the University of Southern California, where she directs the research program for the Center for Economic and Social Research (East). Dr Yoong is also an honorary senior lecturer at the London School of Hygiene and Tropical Medicine, and an adjunct economist at the RAND Corporation. She was previously also jointly appointed as Associate Professor of Health Systems and Behavioral Sciences at the Saw Swee Hock School of Public Health and the Director of the Center for Health Services and Policy Research at the National University Hospital System. She is the previous Director of the Asia Pacific Regional Capacity-Building for Health Technology Assessment (ARCH) Initiative, an APEC-funded multi-country collaboration to promote health technology assessment among member economies, and Director of the RAND Behavioural Finance Forum 2012. Dr. Yoong's academic research has been published in leading economics and public health journals including the American Economic Review and has been funded by the WHO, OECD, NIH, DFID, World Bank and USAID. Dr Yoong received her Ph.D. in Economics at Stanford University as an FSI Starr Foundation Fellow and her AB summa cum laude in Economics and Applied and Computational Mathematics from Princeton University.

Jonathan Zinman is a Professor of Economics at Dartmouth College, an Academic Lead for the Global Financial Inclusion Initiative of Innovations for Poverty Action (IPA), and co-founder of IPA's U.S. Finance Initiative.

Professor Zinman's research focuses on household finance and behavioral economics. He has papers published in several top journals in economics, finance, law, and general-interest science, and his work has been featured extensively in popular and trade media as well.

Professor Zinman applies his research by working with policymakers and practitioners around the globe. He has served on the inaugural Consumer Advisory Board of the Consumer Financial Protection Bureau, as a Visiting Scholar at the Federal Reserve Bank of Philadelphia, and as a Community Development Research Advisory Council member for the Federal Reserve Bank of Boston. He also works directly with financial service providers, ranging from startups to nonprofit to publicly-traded companies, to develop and test innovations that are beneficial to both providers and their clients.

Figure 1. B-counts by education

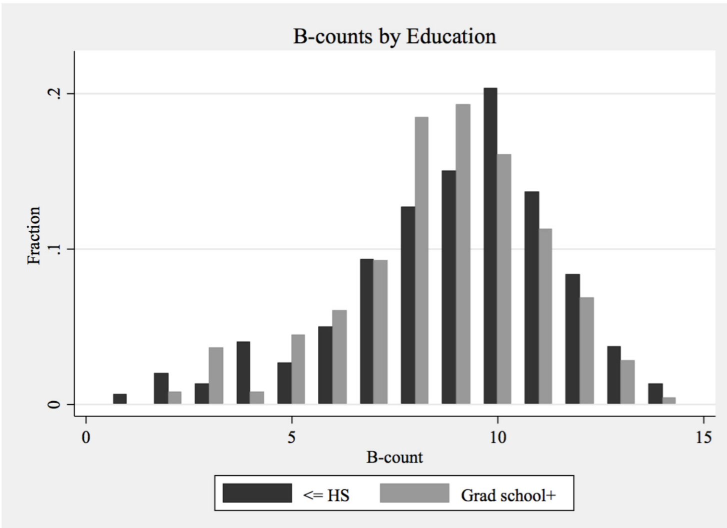

Figure 2. B-counts by gender

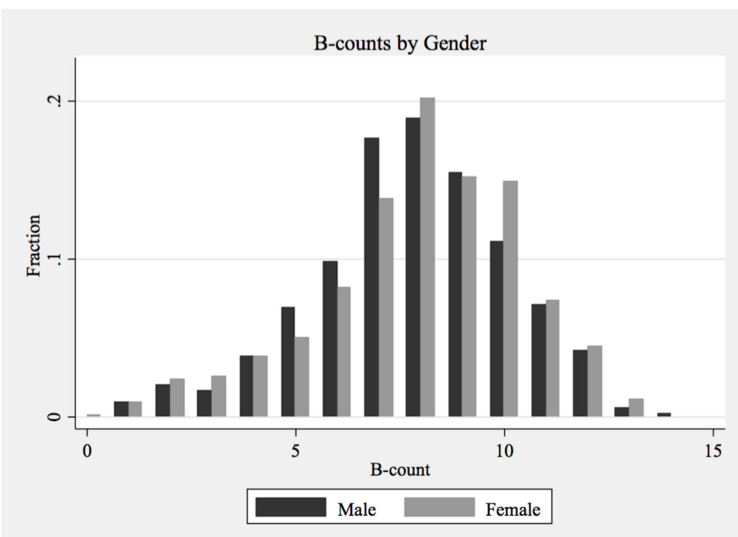

Figure 3. B-counts by cognitive ability

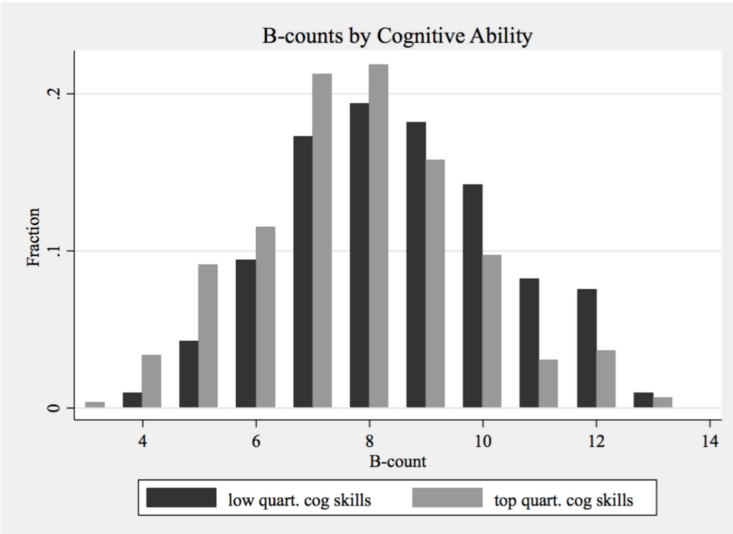

Figure 4. B-counts by lowest and highest income quartiles

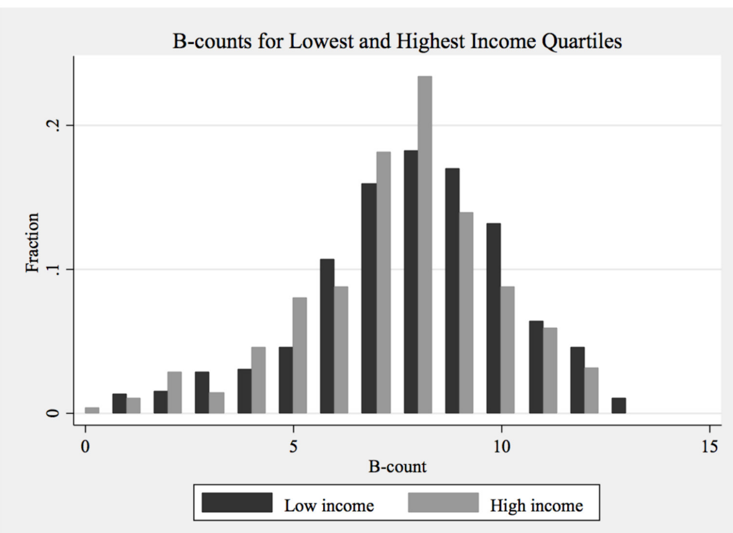

Table 1. Research design: eliciting multiple individual behavioral factors

Factor name: key antecedents	Elicitation method description	Bias measurement ("standard" bias directions in bold)
Discounting money: Andreoni & Sprenger (2012), Barcellos & Carvalho (2014)	Convex Time Budget. 24 decisions allocating 100 tokens each between smaller- sooner and larger-later amounts; decisions pose varying start dates, delay lengths & savings yields.	**Present-biased: discounts more when sooner date is today** Future-biased: discounts more when sooner date is 5 weeks from today
Discounting snacks: Read & van Leeuwen (1998), Barcellos & Carvalho (2014)	Two decisions between two snacks: healthier/less-delicious vs. less healthy/more delicious. Decision pose variation only in date snack is delivered: now, or 5 weeks from now.	**Present-biased: choose less healthy today, healthy for 5 weeks from now** Future-biased: choose healthy for today, less healthy for 5 weeks from now
Choice inconsistency with GARP (with dominance avoidance): Choi, et al. (2014)	Decisions from 11 different linear budget constraints under risk. Subjects choose a point on the line, then the computer randomly chooses whether to pay the point value of the x-axis or the y-axis.	**GARP only: percentage points of potential earnings wasted (CCEI) GARP or dominance avoidance: pp of potential earnings wasted (combined-CCEI)**
Preference for certainty: Callen, et al. (2014)	2 screens of 10 choices each 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). Y=\$450, X=\$150, q ϵ[0.1, 1.0], p=0.5 on one screen and 1.0 on the other.	**Preference for certainty bias: certainty premium (CP)>0** Cumulative prospect theory bias: certainty premium (CP)<0
Loss aversion/small-stakes risk aversion: Fehr & Goette (2007)	Two choices. Choice 1: between a 50-50 lottery (win \$80 or lose \$50) and \$0. Choice 2: between playing the lottery in Choice 1 six times and \$0.	**Loss aversion: choosing the certain \$0 payoff in one or more choices.**
Narrow bracketing: Rabin & Weizsacker (2009	Two tasks of two decisions each. 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.	**Narrow-bracketing: choosing one decision dominated by implications of an earlier decision, on one or both tasks.**
Ambiguity aversion: Dimmock, et al. (forthcoming)	Two questions re: a game where win \$500 if pick green ball. 1. Choose between bag with 45 green-55 yellow and bag with unknown mix. 2. If chose 45-55 bag, how many green balls in 45-55 bag would induce switch to bag with unknown mix.	**Aversion: exists if prefers 45 green balls to uncertain mix, increases as number of green balls declines**
(Over-)confidence performance: Larrick, et al. (2007), Moore & Healy (2008	"How many of the last 3 questions (the ones on the disease, the lottery and the savings account) do you think you got correct?"	**Overconfidence: self-assessment > actual score** Under-confidence: self-assessment < actual score
Overconfidence in precision: Larrick, et al. (2007), Moore & Healy (2008)	Questions about likelihoods of different numeracy quiz scores and/or future income increases.	**Overconfidence: responds 100% to one or both questions**
Non-belief in the law of large numbers: Benjamin, Moore, and Rabin (2013)	Question re: percent chances that, among 1,000 coin flips, the # of heads will fall in ranges [0, 480], [481, 519], and [520, 1,000]. NBLLN = distance between response for [481, 519] and 78.	**>78 (overestimate convergence to 50-50)** <78 (underestimate convergence to 50-50)
Gambler's or "hot hand" fallacy: Benjamin, Moore, and Rabin (2013)	"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?"	>50%: "hot hand" fallacy **<50%: gambler's fallacy**

Table 1 (continued). Research design: eliciting multiple individual behavioral factors

Factor name: key antecedents	Elicitation method description	Bias measurement ("standard" bias directions in bold)
Exponential growth bias, debt-side: Stango & Zinman (2009; 2011)	Survey first elicits monthly payment respondent would expect to pay on a $10,000, 48-month car loan. Survey then asks for APR implied by that payment. EGB = difference between actual implied APR and the subject's perceived implied APR.	**Underestimates** or overestimates APR
Exponential growth bias, asset-side: Banks, et al. (2007)	Elicits future value of $200, earning 10% annual, after two years. EGB = difference between the correct future value ($242), and the subject's perceived future value.	**Underestimates** or overestimates future value
Limited attention: Author-developed	Four questions re: whether subject's finances would improve with more attention, with questions varying the types of decisions: day-to-day, medium-run, long-run, or choosing financial products/services (see Appendix Table for scripts).	**One or more responses indicating regret about amount of attention paid**
Limited prospective memory: Ericson (2011)	"The ALP will offer you the opportunity to earn an extra $10...This special survey has just a few simple questions but will only be open for 24 hours, starting 24 hours from now...please tell us now whether you expect to do this special survey."	**Says will complete task but does not complete**

Section 2-C provides additional details on measuring individual behavioral factors; see Stango, Yoong, and Zinman (2016b) for additional details and results. "pp" = percentage points. "CCEI" = Critical Cost Efficiency Index. "Standard" bias accounting applies only to factors with bidirectional biases.

Table 2. Prevalence and missing values for B-factors

Factor and bias	Share biased	Share missing
Discounting money: Present-biased	**0.26**	**0.06**
Discounting money: Future-biased	0.36	
Discounting snacks: Present-biased	**0.15**	**0.07**
Discounting snacks: Future-biased	0.07	
Violates GARP (based on CCEI)	**0.53**	**0.16**
Loses by violating GARP or dominance violations	**0.96**	**0.16**
Preference for certainty type (positive certainty premium)	**0.77**	**0.31**
Cumulative prospect theory type (negative certainty premium)	0.23	
Loss-averse: prefers certain zero payoff	**0.63**	**0.00**
Narrow-brackets	**0.59**	**0.02**
Ambiguity-averse	**0.73**	**0.08**
Overconfident in level performance	**0.38**	**0.10**
Under-confident in level performance	0.11	
Overconfident in precision	**0.44**	**0.11**
Non-belief in the law of large numbers: underestimates convergence	**0.87**	**0.09**
Non-belief in the law of large numbers: overestimates convergence	0.13	
Gambler's fallacy	**0.26**	**0.08**
Hot hand fallacy	0.14	
Exponential growth bias, loan-side: underestimates APR	**0.70**	**0.37**
Exponential growth bias, loan-side: overestimates APR	0.27	
Exponential growth bias, asset-side: underestimates future value	**0.47**	**0.19**
Exponential growth bias, asset-side: overestimates future value	0.09	
Limited attention with regret	**0.49**	**0.02**
Limited prospective memory	**0.86**	**0.10**

Section 2-C provides some details on measuring individual behavioral factors; see Stango, Yoong, and Zinman (in progress) for additional details and results. "pp" = percentage points. "GARP" = General Axiom of Revealed Preference. "CCEI" = Critical Cost Efficiency Index. "Standard" bias classifications are those typically theorized/observed in prior work. Sample size "N" varies across factors due to differing response rates. Most later analyses use standard and non-standard bias indicators; others use only "standard" indicators. Appendix Table 4 provides prevalence estimates using ALP sampling weights designed to be representative of the U.S. population; the results are similar.

Table 3. Are we all behavioral? To what extent? Summary statistics at the level of the individual

	Indicators	
	All	Standard
"B-count" = count of behavioral indicators (N=1,511)		
proportion with any bias	1.00	1.00
mean	9.05	7.90
SD	2.46	2.40
10th percentile	6	5
25th percentile	8	7
50th percentile	9	8
75th percentile	11	10
90th percentile	12	11
"B-tile" = degree of bias for non-missing factors (N=1,511)		
mean	5.97	4.96
SD	1.47	1.44
10th percentile	1.09	3.16
25th percentile	4.97	3.95
50th percentile	5.94	4.92
75th percentile	6.95	5.88
90th percentile	7.86	6.91
Count of missing factors for measuring behavioral indicators (N=1,511)		
median	1	1
mean	1.75	1.75
SD	2.60	2.60

Please see Tables 1 and 2 for lists and descriptions of behavioral indicators, and Section 2-C for details on their definitions and measurement.

Table 4. Correlates of B-counts in the cross-section

	LHS=B-count	
	(1)	(2)
female	0.165	0.234**
	(0.109)	(0.091)
age 35-45	0.250	0.321***
	(0.153)	(0.122)
age 46-54	0.240	0.225*
	(0.152)	(0.126)
age≥55	0.007	0.089
	(0.172)	(0.140)
income 5k-7.5k	0.333	0.172
	(0.403)	(0.301)
income 7.5k-10k	0.032	-0.046
	(0.375)	(0.306)
income 10k-12.5k	0.396	0.019
	(0.369)	(0.277)
income 12.5k-15k	0.390	0.337
	(0.384)	(0.332)
income 15k-20k	0.315	0.480*
	(0.340)	(0.266)
income 20k-25k	0.142	0.221
	(0.367)	(0.282)
income 25k-30k	0.336	0.249
	(0.324)	(0.249)
income 30k-35k	0.372	0.361
	(0.325)	(0.255)
income 35k-40k	0.241	0.277
	(0.327)	(0.264)
income 40k-50k	0.065	0.125
	(0.320)	(0.256)
income 50k-60k	0.423	0.397*
	(0.314)	(0.239)
income 60k-75k	0.118	0.131
	(0.315)	(0.246)
income 75k-100k	-0.295	-0.164
	(0.320)	(0.253)
income 100k-125k	-0.072	-0.054
	(0.342)	(0.262)
income 125k-200k	-0.810**	-0.708**
	(0.345)	(0.285)
income≥200k	-0.282	0.020
	(0.402)	(0.322)
count missing b-factors		-0.663***
		(0.025)
constant	9.903***	10.240***
	(0.521)	(0.406)

	LHS=B-count	
	(1)	(2)
marital status: separated/divorced/widowed	0.087	0.080
	(0.138)	(0.119)
marital status: never married	-0.331**	-0.234**
	(0.147)	(0.118)
# other hh members: 1	-0.185	-0.066
	(0.139)	(0.118)
# other hh members: 2	0.007	0.095
	(0.152)	(0.121)
# other hh members: 3	0.097	0.173
	(0.179)	(0.147)
# other hh members: 4+	-0.070	0.095
	(0.212)	(0.176)
education attain: some college or associates	0.204	0.259**
	(0.151)	(0.119)
education attain: bachelor's	-0.152	-0.107
	(0.172)	(0.143)
education attain: graduate	-0.127	-0.115
	(0.195)	(0.160)
work status: self-employed	-0.464**	-0.425**
	(0.190)	(0.168)
work status: not working	-0.140	-0.139
	(0.167)	(0.136)
work status: disabled	-0.244	-0.030
	(0.214)	(0.187)
work status: missing	0.099	-0.190
	(0.475)	(0.385)
1st PC cog skills (stdized)	-0.327***	-0.426***
	(0.065)	(0.055)
cog. skills missing	-3.191***	0.307
	(0.226)	(0.206)
1st PC standard risk attitudes	0.018	-0.006
	(0.048)	(0.044)
risk missing	-1.593***	-0.254**
	(0.133)	(0.116)
patience	-0.217***	-0.220***
	(0.053)	(0.043)
patience missing	-1.038***	-0.067
	(0.225)	(0.176)
race=Black	0.209	0.167
	(0.181)	(0.144)
race=Other	-0.269	-0.156
	(0.179)	(0.145)
Latino	0.438**	0.365***
	(0.171)	(0.136)
immigrant	-0.034	-0.049
	(0.196)	(0.160)
R-squared	0.42	0.60
mean(LHS)	9.054	9.054
sd(LHS)	2.461	2.461
Observations	1,505	1,505

* 0.10 ** 0.05 *** 0.01. Each column presents results from an OLS regression of the LHS variable listed in the column heading on the variables listed down the rows, plus state of residence fixed effects.

Table 5. Measuring financial condition: Wealth/savings/distress indicators, prevalence (means) and pairwise correlations

Variable	Mean of indicator	net worth > 0	retirement assets > 0	owns stocks	spent < income last 12 months	financial satisfaction > median	retirement saving adequate	non-ret saving adequate	no severe distress last 12 mos	fin stress < median
net worth > 0	0.44	1								
retirement assets > 0	0.53	0.33	1							
		0.00								
owns stocks	0.49	0.34	0.82	1						
		0.00	*0.00*							
spent < income last 12 months	0.36	0.28	0.21	0.20	1					
		0.00	*0.00*	*0.00*						
financial satisfaction > median	0.46	0.23	0.23	0.22	0.31	1				
		0.00	*0.00*	*0.00*	*0.00*					
retirement saving adequate	0.26	0.23	0.19	0.18	0.27	0.30	1			
		0.00	*0.00*	*0.00*	*0.00*	*0.00*				
non-ret saving adequate	0.25	0.12	0.02	0.05	0.18	0.17	0.31	1		
		0.00	*0.39*	*0.09*	*0.00*	*0.00*	*0.00*			
no severe distress last 12 mos	0.56	0.30	0.29	0.29	0.32	0.34	0.30	0.15	1	
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*		
fin stress < median	0.51	0.26	0.15	0.17	0.29	0.33	0.29	0.16	0.32	1
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	
Mean of all indicators	**0.43**									
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Frequency of each count:	0.09	0.11	0.16	0.14	0.12	0.11	0.09	0.08	0.06	0.04

Unconditional pairwise correlations and their p-values (in italics).
Pairwise sample sizes range from 1,391 to 1,508. Sample size for frequency distribution is 1,511.
Variable definitions: net worth is from two summary questions–"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?" and "You stated that your household's [debts/assets] are worth more than your household's [assets/debts]. By how much?" Retirement assets is from questions on IRAs and workplace plans. Stockholding is from questions on stock mutual funds in IRAs, stock mutual funds in 401(k)s/other retirement accounts, and direct holdings. Spent < income is from a summary question on spending vs. saving over the past year, taken from the Survey of Consumer Finances. Financial satisfaction is based on a 100-point scale responding to "How satisfied are you with your household's overall economic situation?" Savings adequacy questions are placed one each in the two different modules to mitigate mechanical correlations, with response options framed to encourage people to recognize tradeoffs between saving and consumption. Indicators of severe financial distress are taken from the National Survey of American Families: late/missed payment 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. Financial stress is based on a 100-point scale in response to: "To what extent, if any, are finances a source of stress in your life?"

Table 6. Do B-counts and B-tiles help explain financial condition?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
B-count, all biases	-0.026***	-0.024***					-0.014***
	(0.004)	(0.004)					(0.006)
B-count, standard biases			-0.025***			-0.019**	
			(0.004)			(0.008)	
B-count, non-standard biases			-0.010			-0.005	
			(0.007)			(0.062)	
B-tile, all biases				-0.028***		-0.009	
				(0.005)		(0.011)	
B-tile, standard biases					-0.032***	-0.006	-0.016**
					(0.005)	(0.070)	(0.007)
B-tile, non-standard biases					-0.009	-0.009	
					(0.008)	(0.008)	
B-factor missing count	-0.022***	-0.006	-0.001	-0.005	-0.001	-0.005	-0.005
	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
female	-0.016	-0.014	-0.014	-0.015	-0.015	-0.015	-0.014
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
age 35-45	-0.002	-0.009	-0.010	-0.008	-0.008	-0.007	-0.007
	(0.017)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
age 46-54	0.018	0.008	0.009	0.010	0.012	0.011	0.011
	(0.018)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
age≥55	0.062***	0.052**	0.051**	0.055***	0.056***	0.055***	0.055***
	(0.020)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
highest ed: some college or associates	-0.027	-0.031*	-0.034*	-0.030*	-0.031*	-0.030*	-0.030*
	(0.017)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
highest ed: bachelor's	0.016	0.009	0.007	0.010	0.008	0.010	0.009
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
highest ed: graduate	0.025	0.019	0.018	0.018	0.017	0.018	0.018
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
1st PC risk attitudes (stdized)	-0.010	-0.009	-0.009	-0.010	-0.010	-0.010	-0.010
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
patience (stdized)	-0.000	0.001	0.001	0.002	0.003	0.002	0.002
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
fluid intell # correct		-0.000	0.000	0.000	0.000	-0.000	-0.000
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
fluid intell score missing		-0.208**	-0.201**	-0.207**	-0.195**	-0.204**	-0.200**
		(0.091)	(0.091)	(0.090)	(0.091)	(0.091)	(0.090)
numeracy # correct		-0.004	-0.005	-0.004	-0.005	-0.004	-0.005
		(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)

Table 6. Do B-counts and B-tiles help explain financial condition?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
numeracy score missing		-0.007	-0.009	-0.012	-0.016	-0.013	-0.013
		(0.051)	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)
financial literacy # correct		0.023^{***}	0.024^{***}	0.023^{**}	0.023^{***}	0.023^{**}	0.022^{**}
		(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
financial literacy score missing		0.093	0.101	0.087	0.090	0.086	0.087
		(0.078)	(0.078)	(0.078)	(0.078)	(0.078)	(0.078)
exec attention # correct		0.000	-0.000	0.000	-0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
exec attention score missing		0.020	0.015	0.020	0.015	0.019	0.018
		(0.034)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)
time spent on B-factor questions		0.001	0.001	0.001	0.001	0.001	0.001
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
other controls?				yes, as detailed in notes			
R-squared	0.41	0.41	0.41		0.41	0.41	0.41
mean(LHS)				0.43			
sd(LHS)				0.28			
Observations				1,502			

*0.10 ** 0.05 *** 0.01. LHS variable is our summary statistic for financial condition: the proportion of positive financial indicators in Table 6. Each column presents results from a single OLS regression, with Huber-White standard errors, of this LHS variable on the variables shown in the rows, and the following additional control variables: four race/ethnicity categories, state of residence, immigrant indicator, 17 income categories, 3 marital status categories, 4 household size categories, 5 work status categories, a dummy for missing the cognitive skills variable (see Appendix Table 2 and Section 2-D for details on measuring cognitive skills), a dummy for missing the first principal component of four measures of risk attitudes (see Appendix Table 3 and Section 2-D for details on measuring these components of non- behavioral risk attitudes). Patience is the average savings rate across the 24 CTB decisions; we also include a dummy for missing this variable. Unit of observation is the respondent.

Table 7. Do B-counts and B-tiles help explain income?

	(1)	(2)	(3)	(4)	(5)	(6)
B-count, all biases	-0.052***	-0.032**				-0.018
	(0.014)	(0.014)				(0.021)
B-count, standard biases				-0.031**		
				(0.014)		
B-count, non-standard biases				-0.045*		
				(0.025)		
B-tile, all biases			-0.056***			
			(0.018)			
B-tile, standard biases					-0.055***	-0.024
					(0.018)	(0.025)
B-tile, non-standard biases					-0.059**	
					(0.027)	
B-factor missing count	-0.031	-0.077***	-0.082***	-0.078***	-0.082***	-0.076***
	(0.020)	(0.028)	(0.027)	(0.028)	(0.027)	(0.028)
female	-0.049	-0.014	-0.010	-0.013	-0.010	-0.014
	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)
age 35-45	0.119*	0.130**	0.133**	0.129**	0.134**	0.134**
	(0.063)	(0.063)	(0.062)	(0.063)	(0.062)	(0.063)
age 46-54	0.241***	0.244***	0.250***	0.243***	0.251***	0.251***
	(0.065)	(0.066)	(0.066)	(0.066)	(0.066)	(0.066)
age≥55	0.358***	0.397***	0.398***	0.396***	0.399***	0.404***
	(0.071)	(0.073)	(0.073)	(0.073)	(0.073)	(0.073)
highest ed: some college or associates	0.309***	0.213***	0.214***	0.213***	0.215***	0.217***
	(0.060)	(0.061)	(0.060)	(0.061)	(0.061)	(0.061)
highest ed: bachelor's	0.722***	0.584***	0.578***	0.583***	0.578***	0.585***
	(0.068)	(0.071)	(0.071)	(0.071)	(0.071)	(0.071)
highest ed: graduate	1.017***	0.844***	0.840***	0.845***	0.839***	0.842***
	(0.075)	(0.079)	(0.079)	(0.079)	(0.079)	(0.079)
1st PC risk attitudes (stdized)	-0.022	-0.013	-0.013	-0.013	-0.013	-0.014
	(0.022)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
patience (stdized)	-0.019	-0.015	-0.018	-0.017	-0.019	-0.014
	(0.023)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
fluid intell # correct		-0.008	-0.009	-0.008	-0.009	-0.008
		(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
fluid intell score missing		0.649**	0.666**	0.648**	0.666**	0.661**
		(0.295)	(0.296)	(0.296)	(0.296)	(0.295)
numeracy # correct		0.112**	0.105**	0.112**	0.105**	0.110**
		(0.046)	(0.046)	(0.046)	(0.046)	(0.046)

Table 7. Do B-counts and B-tiles help explain income?

	(1)	(2)	(3)	(4)	(5)	(6)
numeracy score missing		0.384*	0.377*	0.388**	0.378*	0.375*
		(0.196)	(0.195)	(0.196)	(0.195)	(0.197)
financial literacy # correct		0.180***	0.175***	0.180***	0.175***	0.178***
		(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
financial literacy score missing		0.264	0.261	0.270	0.263	0.255
		(0.236)	(0.234)	(0.237)	(0.235)	(0.235)
exec attention # correct		0.000	0.000	0.000	0.000	0.000
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
exec attention score missing		0.117	0.109	0.116	0.109	0.114
		(0.129)	(0.129)	(0.129)	(0.129)	(0.129)
time spent on B-factor questions		-0.038***	-0.038***	-0.038***	-0.038***	-0.03***
		(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
other controls?	yes, as detailed in notes					
R-squared	0.45	0.48	0.48	0.48	0.48	0.48
mean(LHS)	2.410					
Observations	1,509					

* 0.10 ** 0.05 *** 0.01. LHS variable is income quartile. Each column presents results from a single OLS regression, with Huber-White standard errors, of this LHS variable on the variables shown in the rows, and the following additional control variables: four race/ethnicity categories, state of residence, immigrant indicator, 3 marital status categories, 4 household size categories, 5 work status categories, a dummy for missing the cognitive skills variable (see Appendix Table 2 and Section 2-D for details on measuring cognitive skills), a dummy for missing the first principal component of four measures of risk attitudes (see Appendix Table 3 and Section 2-D for details on measuring these components of non-behavioral risk attitudes). Patience is the average savings rate across the 24 CTB decisions; we also include a dummy for missing this variable. Unit of observation is the respondent.

Table 8. Do B-counts and B-tiles help explain education?

	(1)	(2)	(3)	(4)	(5)	(6)
B-count, all biases	-0.030***	-0.020***				-0.008
	(0.007)	(0.007)				(0.011)
B-count, standard biases				-0.021***		
				(0.007)		
B-count, non-standard biases				-0.014		
				(0.013)		
B-tile, all biases			-0.029***			
			(0.010)			
B-tile, standard biases					-0.032***	-0.021
					(0.010)	(0.013)
B-tile, non-standard biases					-0.016	
					(0.015)	
B-factor missing count	-0.047***	-0.026*	-0.025*	-0.025*	-0.025*	-0.025*
	(0.010)	(0.015)	(0.014)	(0.015)	(0.014)	(0.015)
female	0.042*	0.054**	0.054**	0.053**	0.054**	0.054**
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
age 35-45	0.001	0.017	0.017	0.017	0.019	0.019
	(0.032)	(0.032)	(0.032)	(0.033)	(0.033)	(0.033)
age 46-54	-0.015	0.004	0.006	0.005	0.009	0.008
	(0.034)	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)
age≥55	-0.005	0.028	0.028	0.029	0.031	0.032
	(0.037)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)
1st PC risk attitudes (stdized)	-0.047***	-0.043***	-0.043***	-0.043***	-0.043***	-0.043***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
patience (stdized)	-0.021*	-0.018	-0.019	-0.018	-0.018	-0.017
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
fluid intell # correct		0.027***	0.026***	0.027***	0.026***	0.026***
		(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
fluid intell score missing		0.330**	0.338**	0.331**	0.342**	0.340**
		(0.138)	(0.141)	(0.138)	(0.140)	(0.139)
numeracy # correct		-0.010	-0.013	-0.010	-0.013	-0.012
		(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
numeracy score missing		-0.013	-0.015	-0.015	-0.020	-0.020
		(0.082)	(0.082)	(0.083)	(0.083)	(0.083)
financial literacy # correct		0.050***	0.049***	0.050***	0.049***	0.049***
		(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
financial literacy score missing		-0.048	-0.046	-0.051	-0.054	-0.056
		(0.119)	(0.122)	(0.119)	(0.122)	(0.121)

Table 8 (continued). Do B-counts and B-tiles help explain education?						
	(1)	(2)	(3)	(4)	(5)	(6)
exec attention # correct		0.000	0.000	0.000	0.000	0.000
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
exec attention score missing		-0.020	-0.024	-0.019	-0.024	-0.022
		(0.062)	(0.061)	(0.062)	(0.061)	(0.061)
time spent on B-factor questions		-0.008*	-0.008*	-0.008*	-0.008*	-0.008*
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
other controls?	yes, as detailed in notes					
R-squared	0.30	0.32	0.32	0.32	0.32	0.32
mean(LHS)	0.419					
Observations	1,509					

* 0.10 ** 0.05 *** 0.01. LHS variable is indicator for whether individual is a college graduate. Each column presents results from a single OLS regression, with Huber-White standard errors, of this LHS variable on the variables shown in the rows, and the following additional control variables: four race/ethnicity categories, state of residence, immigrant indicator, 17 income categories, 3 marital status categories, 4 household size categories, 5 work status categories, a dummy for missing the cognitive skills variable (see Appendix Table 2 and Section 2-D for details on measuring cognitive skills), a dummy for missing the first principal component of four measures of risk attitudes (see Appendix Table 3 and Section 2-D for details on measuring these components of non-behavioral risk attitudes). Patience is the average savings rate across the 24 CTB decisions; we also include a dummy for missing this variable. Unit of observation is the respondent.

Table 9. Robustness to excluding one behavioral factor at a time

Each cell presents results on a B-count coefficient from a single regression

remove indicator(s) of:	
nothing: full B-count result from Table 7 Column 2	0.025***
	(0.004)
biased discounting (money)	0.025***
	(0.004)
biased discounting (snacks)	-0.026***
	(0.004)
choice inconsistency with GARP	-0.027***
	(0.004)
choice inconsist w/GARP, dominance avoidance	-0.025***
	(0.004)
biased attitude re: certainty	-0.025***
	(0.004)
loss aversion/small-stakes risk aversion	-0.025***
	(0.004)
narrow bracketing	-0.027***
	(0.004)
ambiguity aversion	-0.025***
	(0.004)
over-/under-confidence in performance	-0.030***
	(0.004)
overconfidence in precision	-0.025***
	(0.004)
non-belief in the law of large numbers	-0.025***
	(0.004)
gambler's fallacy/hot hand	-0.027***
	(0.004)
exponential growth biases: debt-side	-0.024***
	(0.004)
exponential growth biases: asset-side	-0.024***
	(0.004)
limited attention	-0.016***
	(0.004)
limited prospective memory	-0.025***
	(0.004)
mean(LHS)	0.430
sd(LHS)	0.281
Observations	1,502

Each cell presents results from an OLS regression that is the same as the one in Table 6 Column 2: our financial condition summary outcome on a B-count of behavioral indicators capturing any deviation from standard assumptions, the count of missing B-factor variables, and the full set of controls. Each B-count excludes the indicator(s) described in the row label, to show whether it alone is driving the results.

Table 10. Links between B-counts and financial condition, excluding "math biases"

Compare to Table 7

	(1)	(2)
B-count excluding math biases	0.025***	0.025***
	(0.004)	(0.004)
B-factor missing count		-0.026***
		(0.008)
count of "math biases"	-0.012***	-0.020***
	(0.004)	(0.005)
female	-0.019	-0.014
	(0.013)	(0.013)
age 35-45	-0.004	-0.001
	(0.017)	(0.017)
age 46-54	0.017	0.021
	(0.018)	(0.018)
age≥55	0.065***	0.066***
	(0.021)	(0.021)
highest ed: some college or associates	-0.029*	-0.032*
	(0.017)	(0.017)
highest ed: bachelor's	0.014	0.008
	(0.020)	(0.020)
highest ed: graduate	0.021	0.016
	(0.023)	(0.024)
1st PC risk attitudes (stdized)	0.016**	0.011
	(0.007)	(0.007)
patience (stdized)	-0.008	-0.009
	(0.006)	(0.006)
1st PC cog skills (stdized)	0.001	0.000
	(0.006)	(0.006)
other controls?	yes, as detailed in Table 6	
R-squared	0.40	0.40
mean(LHS)	0.430	
sd(LHS)	0.281	
Observations	1,502	1,502

*0.10 ** 0.05 *** 0.01. There are 8 math bias indicators, capturing the bidirectional biases for: non-belief in the law of large numbers, gambler's fallacies, exponential growth bias (debt-side), and exponential growth bias (asset-side). See Table 1 and Section 2-C for descriptions.

Appendix

Appendix Table 1. Attention proxies

Panel A. Response tabs

		Time frame		
		day-to-day	med-run	long-run
Do you believe that your household's [horizon] finances...would improve if your household paid more attention to them?				
Yes, and I/we often regret not paying greater attention		0.26	0.23	0.35
Yes, but paying more attention would require too much time/effort		0.08	0.11	0.12
No, my household finances are set up so that they don't require much attention		0.15	0.16	0.13
No, my household is already very attentive to these matters		0.52	0.51	0.41
	N	1,488	1,486	1,487
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."				
long-run: "dealing with kids' college, retirement planning, allocation of savings/investments, etc."				
Do you believe that you could improve the prices/terms your household typically receives on financial products/services by shopping more?				
Yes, and I/we often regret not shopping more		0.18		
Yes, but shopping more would require too much time/effort		0.2		
No, my household already gets the best deals on most products/services		0.47		
No, my household wouldn't be able to get the best deal even with more shopping		0.14		
	N	1,491		

Panel B. Correlations between regret indicators

	day-to-day	med-run	long-run
day-to-day	1		
med-run	0.62	1	
long-run	0.51	0.57	1
shopping	0.23	0.25	0.25

Each pairwise correlation here has a p-value < 0.0001. Pairwise sample sizes range from 1481 to 1491.

Panel C. Correlations between too much time/effort indicators

	day-to-day	med-run	long-run
day-to-day	1		
med-run	0.43	1	
long-run	0.34	0.39	1
shopping	0.06	0.09	0.12

Each pairwise correlation here has a p-value < 0.01. Pairwise sample sizes range from 1481 to 1491.

Panel D. Proportion with each count of "regret" and "too much time/effort" responses

	regret	too much
0	0.52	0.67
1	0.20	0.22
2	0.10	0.07
3	0.13	0.04
4	0.06	0.01
N	1,480	1,480

Appendix Table 2. Pairwise correlations between measures of cognitive skills

	Fluid intelligence	Numeracy	Financial literacy	Executive attention
Fluid intelligence	1			
mean 10.6, SD 2.8,				
min 0, max 15	1403			
Numeracy	**0.44**	1		
mean 1.7, SD 0.6,	*0*			
min 0, max 2	1371	1372		
Financial literacy	**0.45**	**0.41**	1	
mean 2.2, SD 0.9,	*0*	*0*		
min 0, max 3	1399	1368	1406	1
Executive attention	**0.36**	**0.19**	**0.22**	
mean 65, SD 24,	*0*	*0*	*0*	
min 0, max 154	1,352	1,326	1,355	1,444

Results for each pair of variables show the correlation, p-value (in italics) and sample size.
Each cognitive skills measure is a count of correct responses.
Fluid intelligence measured using a standard 15-question, non-adaptive number series.
Numeracy measured using 2 of the 6 questions popularized by Banks and Oldfield (2007).
Financial literacy measured using 3 of the questions popularized by Lusardi and Mitchell (2014).
Executive attention measured using a 2-minute Stroop test where respondents are instructed to answer as many questions correctly as they can.

Appendix Table 3. Pairwise correlations among measures of risk aversion

	Lifetime income gamble	Financial risk-taking scale	Switch point two-lottery list	Switch point lottery vs. certain list
Lifetime income gamble	1			
mean 4.3, SD 1.3				
min 1, max 6	1503			
Financial risk-taking scale	**0.19**	1		
mean -43, SD 23	*0*			
min -100, max 0	1390	1403		
Switch point two-lottery list	**0.07**	**0.09**	1	
mean 7.6, SD 1.5	*0.02*	*0*		
min 2, max 10	1147	1068	1153	
Switch point lottery vs. certain list	**0.26**	**0.16**	**0.42**	1
mean 6.6, SD 1.8	*0*	*0*	*0*	
min 2, max 10	1,215	1,133	1,066	1,222

Results for each pair of variables show the correlation, p-value (in italics) and sample size.
Higher values indicate greater risk aversion. Each variable is an ordinal scale but parameterized linearly for convenience in summarizing the correlations.
Lifetime income gamble is from the Barsky, et al. (1997), task.
Financial risk-taking scale is from Dohmen, et al., (2010, 2011).
Switch points are from the two multiple price lists we use to measure the certainty premium. As Callen, et al. (2014), detail, these switch points provide non-parametric measures of risk aversion.

Appendix Table 4. Selected columns from Table 6, full results not showing state fixed effects

	(1)	(2)		(1)	(2)
B-count	-0.024***		marital status: separated/divorced/widowed	0.015	0.013
	(0.004)			(0.017)	(0.017)
B-tile		-0.029***	marital status: never married	0.005	0.005
		(0.005)		(0.017)	(0.017)
B-factor missing count	-0.006	-0.013**	# other hh members: 1	-0.027	-0.028*
	(0.008)	(0.006)		(0.017)	(0.017)
female	-0.014	-0.014	# other hh members: 2	-0.038**	-0.039**
	(0.013)	(0.013)		(0.018)	(0.018)
age 35-45	-0.009	-0.010	# other hh members: 3	-0.069***	-0.070***
	(0.017)	(0.017)		(0.020)	(0.021)
age 46-54	0.008	0.009	# other hh members: 4+	-0.031	-0.033
	(0.019)	(0.019)		(0.025)	(0.025)
age≥55	0.052**	0.052**	highest ed: some college or associates	-0.031*	-0.034**
	(0.021)	(0.021)		(0.017)	(0.017)
race=Black	-0.016	-0.013	highest ed: bachelor's	0.009	0.007
	(0.020)	(0.020)		(0.020)	(0.020)
race=Other	-0.062***	-0.062***	highest ed: graduate	0.019	0.018
	(0.020)	(0.020)		(0.024)	(0.024)
Latino	-0.008	-0.011	work status: self-employed	-0.018	-0.015
	(0.018)	(0.018)		(0.023)	(0.023)
immigrant	0.063***	0.063***	work status: not working	-0.028	-0.027
	(0.021)	(0.021)		(0.019)	(0.019)
income 5k-7.5k	0.011	0.017	work status: disabled	-0.083***	-0.086***
	(0.040)	(0.041)		(0.026)	(0.026)
income 7.5k-10k	0.015	0.022	work status: missing	-0.056	-0.059
	(0.038)	(0.038)		(0.057)	(0.057)
income 10k-12.5k	-0.008	-0.004		-0.009	-0.009
	(0.036)	(0.036)	1st PC risk attitudes (stdized)	(0.006)	(0.006)
income 12.5k-15k	-0.047	-0.040		0.009	0.010
	(0.038)	(0.039)	risk missing	(0.017)	(0.017)
income 15k-20k	0.000	0.002		0.001	0.001
	(0.037)	(0.037)	patience (stdized)	(0.006)	(0.006)
income 20k-25k	0.033	0.040		-0.011	-0.017
	(0.035)	(0.035)	patience missing	(0.027)	(0.027)
income 25k-30k	0.049	0.050	fluid intell # correct	-0.000	-0.000
	(0.037)	(0.038)		(0.003)	(0.003)
income 30k-35k	0.043	0.050	fluid intell score missing	-0.208**	-0.195**
	(0.034)	(0.035)		(0.083)	(0.081)
income 35k-40k	0.154***	0.158***	numeracy # correct	-0.004	-0.005
	(0.038)	(0.039)		(0.012)	(0.012)
income 40k-50k	0.111***	0.116***	numeracy score missing	-0.007	-0.006
	(0.035)	(0.035)		(0.043)	(0.043)
income 50k-60k	0.169***	0.172***	financial literacy # correct	0.023***	0.024***
	(0.035)	(0.035)		(0.008)	(0.008)
income 60k-75k	0.199***	0.200***	financial literacy score missing	0.093	0.100
	(0.035)	(0.035)		(0.072)	(0.071)
income 75k-100k	0.284***	0.289***	exec attention # correct	0.000	-0.000
	(0.037)	(0.037)		(0.000)	(0.000)
income 100k-125k	0.287***	0.289***	exec attention score missing	0.020	0.016
	(0.038)	(0.038)		(0.032)	(0.032)
income 125k-200k	0.348***	0.353***	10 quantiles of tsbfavg	0.001	0.001
	(0.040)	(0.040)		(0.002)	(0.002)

Appendix Table 4 (continued). Selected columns from Table 6, full results not showing state fixed effects

	(1)	(2)		(1)	(2)
income≥200k	0.402***	0.405***	did not take module 352	0.009	0.004
	(0.046)	(0.046)		**(0.046)**	**(0.046)**
Constant	0.600***	0.540***			
	(0.082)	(0.079)			
R-squared	0.41	0.41			
mean(LHS)	0.430	0.430			
sd(LHS)	0.281	0.281			
Observations	1,502	1,502			

* 0.10 ** 0.05 *** 0.01

Appendix Table 5. B-counts and individual components of financial well-being

	(1) net worth>0	(2) retirement assets>0	(3) owns stocks	(4) spent < inc last 12 months	(5) fin satis > median	(6) adequate ret. Savings	(7) adequate non-ret. Savings	(8) no fin distress recent	(9) no current fin stress
B-count	-0.026***	-0.021***	-0.015**	-0.022***	-0.023***	-0.039***	-0.022***	-0.029***	-0.029***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.008)
count missing B-factors	-0.027***	-0.022***	-0.018**	-0.014	-0.015	-0.033***	0.020	-0.018**	-0.032***
	(0.009)	(0.008)	(0.008)	(0.009)	(0.010)	(0.009)	(0.014)	(0.009)	(0.010)
female	0.023	0.016	-0.023	-0.041	-0.021	-0.040	-0.037	-0.026	0.008
	(0.025)	(0.023)	(0.024)	(0.027)	(0.027)	(0.025)	(0.026)	(0.026)	(0.028)
age 35-45	0.094***	0.067**	0.057*	-0.010	-0.090**	-0.055*	-0.029	-0.016	-0.024
	(0.034)	(0.032)	(0.032)	(0.035)	(0.037)	(0.032)	(0.035)	(0.035)	(0.038)
age 46-54	0.225***	0.105***	0.093***	0.032	-0.142***	-0.008	-0.017	-0.018	-0.066*
	(0.036)	(0.032)	(0.033)	(0.036)	(0.038)	(0.034)	(0.036)	(0.036)	(0.039)
age≥55	0.301***	0.116***	0.130***	0.018	-0.084**	0.052	0.066*	0.053	-0.048
	(0.039)	(0.036)	(0.037)	(0.040)	(0.042)	(0.039)	(0.039)	(0.039)	(0.042)
ed: some college or associates	0.034	-0.017	0.011	-0.025	-0.085**	-0.076**	-0.070**	-0.013	-0.046
	(0.035)	(0.032)	(0.032)	(0.034)	(0.037)	(0.033)	(0.035)	(0.037)	(0.038)
highest ed: bachelor's	0.037	0.116***	0.111***	0.015	-0.047	-0.038	-0.035	0.035	-0.113***
	(0.040)	(0.037)	(0.038)	(0.040)	(0.042)	(0.039)	(0.040)	(0.042)	(0.043)
highest ed: graduate	0.027	0.098**	0.102**	-0.083*	-0.095*	0.040	0.040	0.111**	-0.089*
	(0.045)	(0.042)	(0.042)	(0.047)	(0.049)	(0.046)	(0.048)	(0.046)	(0.050)
1st PC risk attitudes (stdized)	-0.016	0.002	-0.011	-0.013	-0.005	-0.014	-0.020*	-0.002	-0.003
	(0.012)	(0.011)	(0.011)	(0.012)	(0.012)	(0.011)	(0.011)	(0.012)	(0.013)
patience (stdized)	0.000	-0.003	-0.001	0.024*	-0.026**	0.001	0.009	-0.016	0.019
	(0.012)	(0.011)	(0.011)	(0.013)	(0.013)	(0.011)	(0.012)	(0.012)	(0.014)
R-squared	0.28	0.41	0.38	0.15	0.17	0.14	0.11	0.24	0.11
mean(LHS)	0.440	0.534	0.494	0.356	0.462	0.263	0.250	0.557	0.505
sd(LHS)	0.497	0.499	0.500	0.479	0.499	0.441	0.433	0.497	0.500
Observations	1,460	1,475	1,485	1,496	1,493	1,492	1,386	1,497	1,501

* 0.10 ** 0.05 *** 0.01.

Appendix Table 6. Outcome regressions without behavioral factors: control variable correlations and fit

LHS:	financial condition summary (1)	1=income top quintile (2)	1=college graduate (3)
female	-0.020	-0.036*	0.042*
	(0.013)	(0.020)	(0.024)
age 35-45	-0.009	-0.009	-0.009
	(0.017)	(0.025)	(0.031)
age 46-54	0.015	0.019	-0.009
	(0.018)	(0.026)	(0.033)
age≥55	0.063***	0.062**	0.006
	(0.021)	(0.031)	(0.036)
race=Black	-0.017	-0.021	-0.021
	(0.020)	(0.026)	(0.035)
race=Other	-0.058***	-0.040	0.026
	(0.020)	(0.026)	(0.037)
Latino	-0.017	-0.093***	0.008
	(0.019)	(0.026)	(0.035)
immigrant	0.062***	0.009	0.103**
	(0.021)	(0.031)	(0.042)
income 5k-7.5k	-0.003		-0.022
	(0.041)		(0.071)
income 7.5k-10k	0.005		-0.057
	(0.039)		(0.063)
income 10k-12.5k	-0.020		-0.002
	(0.035)		(0.059)
income 12.5k-15k	-0.074*		0.082
	(0.039)		(0.081)
income 15k-20k	-0.015		0.027
	(0.037)		(0.063)
income 20k-25k	0.017		-0.039
	(0.036)		(0.058)
income 25k-30k	0.030		0.069
	(0.038)		(0.060)
income 30k-35k	0.024		-0.007
	(0.035)		(0.057)
income 35k-40k	0.136***		0.177***
	(0.039)		(0.061)
income 40k-50k	0.098***		0.172***
	(0.035)		(0.060)
income 50k-60k	0.143***		0.283***
	(0.034)		(0.060)
income 60k-75k	0.186***		0.335***
	(0.035)		(0.061)
income 75k-100k	0.277***		0.329***
	(0.037)		(0.061)
income 100k-125k	0.278***		0.385***
	(0.038)		(0.062)
income 125k-200k	0.358***		0.554***
	(0.041)		(0.064)
income≥200k	0.389***		0.533***
	(0.046)		(0.077)

Appendix Table 6 (continued). Outcome regressions without behavioral factors: control variable correlations and fit

LHS:	financial condition summary (1)	1=income top quintile (2)	1=college graduate (3)
marital status: separated/divorced/widowed	0.011	-0.169***	0.003
	(0.017)	(0.023)	(0.031)
marital status: never married	0.013	-0.176***	0.089***
	(0.017)	(0.022)	(0.029)
# other hh members: 1	-0.027	0.030	-0.096***
	(0.017)	(0.026)	(0.031)
# other hh members: 2	-0.042**	0.011	-0.040
	(0.018)	(0.027)	(0.032)
# other hh members: 3	-0.078***	0.042	-0.172***
	(0.021)	(0.033)	(0.036)
# other hh members: 4+	-0.037	0.003	-0.138***
	(0.025)	(0.033)	(0.046)
education attain: some college or associates	-0.038**	0.048**	
	(0.017)	(0.020)	
education attain: bachelor's	0.011	0.162***	
	(0.020)	(0.027)	
education attain: graduate	0.019	0.274***	
	(0.023)	(0.034)	
work status: self-employed	-0.006	0.006	0.006
	(0.024)	(0.038)	(0.045)
work status: not working	-0.021	-0.009	-0.045
	(0.019)	(0.027)	(0.035)
work status: disabled	-0.082***	-0.038	-0.005
	(0.026)	(0.025)	(0.044)
work status: missing	-0.052	-0.071	0.035
	(0.059)	(0.078)	(0.098)
1st PC cog skills (stdized)	0.022***	0.039***	0.085***
	(0.007)	(0.010)	(0.013)
cog skills missing	-0.046**	0.028	-0.072**
	(0.019)	(0.029)	(0.033)
1st PC risk attitudes	-0.009	-0.026***	-0.042***
	(0.006)	(0.009)	(0.011)
risk missing	0.023	-0.013	-0.005
	(0.015)	(0.022)	(0.027)
Patience	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)
patience missing	0.030	-0.005	-0.062
	(0.028)	(0.035)	(0.044)
Constant	0.420***	0.280***	0.237**
	(0.062)	(0.103)	(0.120)
R-squared	0.39	0.24	0.30
mean(LHS)	0.430	0.191	0.415
sd(LHS)	0.281	0.393	0.493
Observations	1,502	1,601	1,601

* 0.10 ** 0.05 *** 0.01. Each column presents results from an OLS regression, with Huber-White standard errors, of the LHS variable listed in the column heading on the variables listed down the rows, plus state of residence fixed effects.

Appendix Table 7. Outcome measurement survey formatting

Variable	# of questions used	response options: # per q.	response options: orientation	response options: placements of one(s) indication worse conditions	response options: ordering details
net worth>0	1	3	vertical	middle	Yes (indicating assets>debts)/No in middle/ About the same
retirement assets>0	2	2	vertical	bottom	"Enter total amount: $[fill].00 OR "No one in my household (including myself) has any [other] retirement accounts
owns stocks	3	2	vertical	middle*	"About what percent of your household's [IRA/KEOGH; 401(k)/other retirement accounts] are invested in stocks or mutual funds (not including money market mutual funds)?" "[fill]%"
				or botttom**	Aside from anything you have already told us about, do you or another member of your household have any shares of stock or stock mutual funds? If you sold all those and paid off anything you owed on them, about how much would your household have? Enter total amount: $ [fill].00 OR "No one in my household (including myself) has any other shares..."
spent< income last 12 months	1	3			
financial satisfaction> median	1	slider horizontal left side of scale 0 to 100 point scale			
retirement saving adequate	1	5	vertical	top	see notes
non-ret saving adequate	1	5	vertical	bottom	see notes
no severe distress last 12 mos	4	2	vertical	top	Yes or No for each question, with Yes on top. (Only 3% of the sample says Yes to each of the 4.)
fin stress<median	1	slider horizontal right side of scale 0 to 100 point scale			

Please see Table 6 for additional details on variable definitions.
*middle: Someone could declare zero stockholdings by checking the two boxes for: "No one in my household… has any [other] retirement accounts"
** bottom: Someone could also declare zero stockholdings by entering zeros in the 3 fill boxes that specifically ask about stocks.

Retirement savings adequacy question and response options:
"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."

1 Not nearly enough: I/we should be saving much more and borrowing/spending much less

2 Not enough: I/we should be saving more and borrowing/spending less 3 Just about enough

4 More than enough: I/we should be saving less and borrowing/spending more

5 Much more than enough: I/we should be saving much less and borrowing/spending much more

Non-retirement savings adequacy question and response options:
"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:"

1	I wish my household saved a lot less and spent a lot more
2	I wish my household saved somewhat less and spent somewhat more
3	My household saving and spending levels are about right
4	I wish my household saved somewhat more and spent somewhat less
5	I wish my household saved a lot more and spent a lot less

Appendix Table 8. Higher B-count, more optimism about financial condition

LHS re: financial status next year	1=better off				1=more on-track			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
B-count	0.009	0.013	0.013	0.014*	0.009	0.011	0.011	0.009
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)
count missing B-factors	0.012	0.013	0.012	0.014	0.002	0.006	0.007	0.005
	(0.008)	(0.010)	(0.010)	(0.010)	(0.015)	(0.016)	(0.016)	(0.016)
1st PC cog skills (stdized)		0.041**	0.041**	0.041**		0.018	0.020	0.019
		(0.016)	(0.016)	(0.016)		(0.017)	(0.017)	(0.017)
current satis with fin situation higher is more			0.000				-0.000	
			(0.001)				(0.001)	
fin summ prop				0.019				-0.096
				(0.058)				(0.062)
R-squared	0.11	0.12	0.12	0.12	0.07	0.07	0.07	0.07
mean(LHS)	0.449				0.505			
Observations	1,499	1,499	1,490	1,499	1,387	1,387	1,379	1,387

* 0.10 ** 0.05 *** 0.01.

LHS=1 if responds “Will be better off” to: “Now looking ahead–do you think that a year from now your household will be better off financially, or worse off, or just about the same as now?” Response ordering and frequency:

Will be better off	0.45
About the same	0.46
Will be worse off	0.09

LHS=1 if responds "I will feel [much more/more] on-track" to: "How do you think you will feel about how your household is using money a year from now?" Response ordering and frequency:

I will feel much more off-track	0.03
I will feel more off-track	0.03
I will feel about the same	0.43
I will feel more on-track	0.39
I will feel much more on-track	0.12

Sample size is lower in Columns 5-8 because that LHS question was asked in our 2nd module (#352).