From Proliferation to Parsimony in Behavioral Economics:

New Elicitation Methods, Summary Statistics, and Links to Real-World Outcomes

Victor Stango, Joanne Yoong, and Jonathan Zinman*

First draft: February 2016
This draft: July 2016

Abstract

Decades of research has distilled some aspects of consumer decision making into parsimonious constructs (e.g., fluid and crystalized intelligence, the Big Five). Behavioral economics, a younger field, is still characterized by proliferation: a wide set of relevant behavioral factors, without a portable framework to tie them together or represent them parsimoniously. We work toward empirically-informed parsimony by developing low-cost methods for directly eliciting data on 17 potentially behavioral factors from each person in a large, representative sample. Nearly all consumers are behavioral on one or more dimensions, even if one only counts large deviations from classical norms as behavioral. Behavioral biases are positively correlated, within-consumer. New statistics summarizing a consumer's behavioral tendencies across multiple factors—the "B-count" and "B-tile"—strongly and negatively correlate with financial condition/well-being and other outcomes, controlling for demographics, classical risk attitudes and patience, and cognitive skills.

---

* Previous title: "We are all behavioral…". 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 Russell Sage Foundation, the Roybal Center (grant # 3P30AG024962), and the National University of Singapore for funding and patience, 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, Dartmouth, the Federal Reserve Bank of Philadelphia, National University of Singapore, NYU/Stern Finance, UC-Davis, UCSD-Rady, DIW-Berlin, and the Aspen Consumer Decision Making Conference for comments.

Introduction

Human decision making is complex. One hundred years of research on some of its aspects has distilled what first appeared to be countless factors/traits into parsimonious constructs (e.g., fluid and crystalized intelligence, the Big Five personality traits).

Behavioral economics (BE) is a younger line of inquiry that studies deviations from classical economic specifications of preferences, decision-making rules and beliefs. BE is arguably still in the countless factors phase of its development (e.g., Fudenberg 2006; Chetty 2015): evidence is mounting that behavioral deviations are manifold and plentiful, but thus far behavioral factors largely have been studied in isolation from each other. Yet a given consumer could be behavioral in multiple ways, creating confounds in single factor studies, raising questions about whether different behavioral factors having reinforcing or countervailing influences, and highlighting the need for more parsimony in measurement and application.

Great parsimony in BE would generate more portable, predictive models, and clarify key inputs into policy and business strategy across many domains.[1] Policymakers invoke behavioral factors as a basis for a specific rulemaking (as done recently by, e.g., the Department of Energy). More generally, many agencies formulate high-level strategy based on assumptions about how behavioral factors *as a whole* impact decision making (e.g., the Consumer Financial Protection Bureau, the Securities and Exchange Commission, the Financial Conduct Authority, the World Bank). "Nudge units" and other centers of applied behavioral social sciences operate on similar premises. Basic research will continue to shape future applications; for example, the National Institutes for Health recently issued a call for research proposals that included "identification and measurement of appropriate economic phenotypes in population-based studies, based on approaches honed in behavioral and experimental studies."[2]

Our work informs all of these fronts. We develop relatively low-cost methods for directly eliciting data on *multiple* potentially behavioral factors (17 to be exact), per consumer, from a large, representative sample. Our methods also leave ample time for collecting data on cognitive skills, classical preferences, and real-world economic choices/outcomes. The resulting data allow

[1] For other work and discussions re: interactions among behavioral factors and other challenges in behavioral modeling, see, e.g., Benjamin, Raymond, and Rabin (forthcoming), Dean and Ortoleva (2015), Ericson (2014), Heidhues et al. (2015), and O'Donoghue and Rabin (1999).

[2] http://grants.nih.gov/grants/guide/rfa-files/RFA-AG-16-010.html

us to examine the prevalence (and in some cases distributions) of different behavioral factors (parameters), to estimate how behavioral tendencies are correlated within-consumer, to explore links between behavioral factors and real-world behavior, and to construct summary statistics for behavioral tendencies that aggregate across factors to the consumer level. Our methods and results have particular implications for an emerging class of behavioral sufficient statistic models (e.g., Chetty, Looney, and Kroft 2009; Mullainathan, Schwartzstein, and Congdon 2012; Allcott and Taubinsky 2015; Farhi and Gabaix 2015; Gabaix 2015), and broader implications for various other literatures that estimate and/or apply behavioral parameters.

We start with standard elicitation methods from recent high-profile papers and modify them for suitability in studies of modest length/budgets by shortening, simplifying and combining tasks.[3] 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. In all we elicit 17 behavioral factors ("B-factors"). Some relate to preferences: present-biased discounting (Read and van Leeuwen 1998; Andreoni and Sprenger 2012), 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). Others capture biased beliefs, biased perceptions, and behavioral decision rules: three varieties of 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).[4]

Even with such 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

[3] 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). But this prior 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).

[4] 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.

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.

RAND Corporation administered our survey across two separate modules fielded in 2014 and 2015, as part of its nationally representative the American Life Panel (ALP). The two modules span about 60 minutes of total survey time and have been completed by over 1,000 panelists. The ALP captures response times at the respondent-question level, which we use to control flexibly for survey effort at the level of B-factors, behavioral summary statistics, and the survey as a whole. We also assess whether survey effort is correlated with behavioral tendencies and/or with financial condition, and implement a series of corrections for any possible spurious results driven by links between survey effort and responses.

Our data appear to be of comparable quality to those in previous and more intensive single-factor studies. Some of our elicitations contain standard built-in data quality checks, and to a surprising extent the patterns suggest that our methods produce data of comparable quality to standard implementations that use more (in-person) instruction and task repetition.

We find that some behavioral tendencies are nearly ubiquitous. For example, 77% of our sample exhibits some preference for certainty, and 73% some ambiguity aversion. Meanwhile other behavioral tendencies are relatively rare; e.g., only 11% exhibit any under-confidence, and only 7% exhibit future-bias with respect to consumption. In cases where our elicitation is rich enough to permit estimation of the distribution of a behavioral parameter, we find substantial heterogeneity across consumers. On balance, our estimates of prevalence and variance are similar to comparable prior studies.

Next we estimate empirical links between an individual-level index of overall financial condition[5] and each B-factor, conditional on a rich set of controls: demographics, cognitive skills including financial literacy, risk/time preferences, and survey time. The overall pattern of results shows that even considered singly, many behavioral tendencies are substantively negatively

[5] 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.

correlated with financial condition. Even the correlations that lack statistical significance are almost exclusively negative. Controlling for survey effort does not change the results.

Further evidence that our behavioral measures capture something distinct from demographics, cognitive ability, and classical preferences comes from regressing each B-factor on the complete set of controls. For twelve of the 17 factors the adjusted r-squareds from such models are below 0.10, and in only one case is an r-squared above 0.20. In many of the models, one cannot reject the hypotheses that cognitive skills, demographics, and survey time are uncorrelated with the B-factor.

We also find that behavioral factors are positively correlated with each other within-person: being behavioral on one dimension implies that one is more likely to be behavioral on other dimensions.[6] These correlations vary in strength and significance, implying that single-factor studies can yield misleading inferences about the relationships between factors and real-world outcomes.[7]

To aggregate information in all of those factors to the consumer level, we then construct two behavioral summary statistics ("B-stats"): a "B-count" measuring the number of factors on which an individual exhibits any deviation from the classical norm, and a "B-tile" that also captures the degree of bias across all factors, in normalized percentile units.[8]

At this summary level nearly everyone in our data is behavioral, exhibiting one or more B-factor indicators. That finding is not an artifact of measuring so many factors that someone is bound to exhibit one: individuals at the 10th percentile have a B-count of six. Nor is it an artifact of survey response noise combined with a weak standard for what counts as behavioral: we also

---

[6] Dean and Ortoleva (2015) and Gillen et al. (2015) find similar evidence from student samples. Several papers have reported correlations among a smaller number of behavioral factors; see, e.g., Li et al. (2013). We focus here on the behavioral economics literature while noting that there are related literatures in other disciplines; e.g., on "deviations from rational thinking". Those literatures are also only beginning to grapple with correlations among behavioral factors (e.g., Stanovich 2016).

[7] The only other paper we know of that conducts predictive analysis, in a nationally representative sample, with measures of multiple behavioral factors is Goda et al. (2015), which does so for present-biased money discounting and exponential growth bias. Bruine de Bruin, Parker, and Fischoff (2007) and Li et al. (2015) also consider a relatively small set of behavioral factors, in convenience samples, as do von Gaudecker et al. (2011) in a representative Dutch sample without exploring links to real-world behavior. 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.

[8] Note that this is a relative, not absolute, measure of the size of behavioral deviations. One could take a similar approach for estimating absolute deviations, e.g., by using parameter estimates for each factor. We do not take this approach because about half of our elicitations are too coarse to yield parameter estimates.

infer that nearly everyone exhibits one or more behavioral tendencies even if we count only large deviations from neoclassical norms, and the distributions of B-counts are identical across deciles of survey time spent on the B-factor questions (though short response times are strongly related to non-response, i.e. skipping questions).

Perhaps most usefully, heterogeneity in B-stats explains cross-sectional variation in financial condition and other outcomes. Both B-counts and B-tiles are robustly negatively correlated with our rich summary index of financial condition, which captures "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. The results hold conditional on our controls: standard demographics such as income and education, measures of classical patience and risk aversion, measures of fluid intelligence, numeracy, financial literacy and executive attention, state of residence, and survey time. A one standard deviation change in a B-count has a conditional correlation with financial condition that is larger and/or more robust than the ones for cognitive skills and many other variables commonly thought to be important correlates of financial decisions and outcomes (like gender, education, standard measures of risk attitudes, and patience). If instead we model education or income as dependent variables, we find that B-stats strongly correlate with these outcomes as well.

Auxiliary findings push against interpretations of B-stats capturing something different than behavioral characteristics. 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 over-confidence) 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. Omitting control variables most likely to introduce omitted variable bias if they themselves are incomplete or measured with error—such as cognitive skills, education, patience, risk aversion and so on—has no discernible effects on (raising) the B-stat coefficients. Parsing 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), yields no evidence that variation in "math bias" drives our results. Controlling flexibly for survey-taking effort (using question-by-question response times) and item non-response has no effect on the results. We lack measures of non-cognitive skills like personality traits, but *a priori* this should not be a concern given that prior evidence finds weak if any correlations between B-factors and non-cognitive skills.[9] Our B-stats are negatively correlated with self-assessed financial well-being considered separately from the more welfare-ambiguous outcomes like savings and wealth, pushing against the interpretation that B-stats simply capture classical preferences. These results all push against hypotheses that B-stats measure noise, "mistakes," math ability, classical inputs, and/or other non-behavioral sources of heterogeneity across consumers.

Methodologically, our results suggest that adapting standard direct elicitation techniques for the purposes of measuring multiple behavioral factors is a fruitful approach. Our approach has yielded data with several interesting patterns. First, the prevalence of different behavioral factors varies greatly. Second, there is substantial heterogeneity in individual behavioral factors, and this heterogeneity helps predict behavior (although the results can be noisy). Third, behavioral indicators are positively correlated within-person, suggesting that single-factor studies should consider those correlations when interpreting results, and also that different behavioral dimensions could be subsumed by lower-dimensional factors. Fourth, one can summarize how "behavioral" people are with statistics—our B-counts and B-tiles. Fifth, these summary statistics indicate that nearly everyone is behavioral, more or less. Sixth, cross-sectional heterogeneity in these more-parsimonious parameters strongly conditionally correlates with financial, labor market, and education outcomes.

Besides expanding the direct elicitation toolkit, our results inform several literatures. The behavioral sufficient statistic approach to modeling discussed above relies on assumptions that have yet to be tested for the most part. Our findings support some of the key assumptions,[10] while casting doubt on others.[11] We also broaden the scope for empirically applying such models with consumer-level behavioral summary statistics that are suitable for domain-level

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

[10] E.g., our results strongly support the assumptions of positive within-consumer correlation among biases (e.g., Chetty 2015), and of consumer-level bias that is nonnegative, positive for some, and not mean-zero in the aggregate (Allcott and Taubinsky 2015, p. 2510).

[11] 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).

applications.[12] Going forward we expect that our methods will further complement sufficient statistic modeling 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.

Our findings also 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): many and perhaps most individuals are behavioral in some sense, and heterogeneity in behavioral tendencies helps explain behavior.[13] 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 widespread and economically important correlates of financial well-being.

1. 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).

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 aged 18 and older.

---

[12] Empirical implementation thus far 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 (Baicker, Mullainathan, and Schwartzstein 2015), and lightbulbs (Allcott and Taubinsky 2015). A narrow focus is understandable and in many cases desirable, given the presumed importance of context in mediating the effect of behavioral tendencies.

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

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. Over 400 surveys have been administered in the ALP, and data becomes 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. 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 thirty-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 aged 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. 1,515 individuals responded to at least one of our questions in module 315, and those 1,515 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. 1,427 individuals responded in part or whole to that second module.

Taken together, the two modules yielded a high retention rate (1427/1515 = 94%), low item non-response rate, and high response quality (see below, and Data Appendix)—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 1,515 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 directly eliciting useful measures of behavioral factors without breaking the bank, we prioritize elicitation methods that have been featured recently in top journals, did not seek to capture social preferences (we had to draw the line somewhere given our constraints), 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).

We conduct elicitations of 17 potentially behavioral factors, 16 of which produce one or more cardinal measures of deviations from classical norms, 15 of which produce some measure of the intensity of deviation, and 8 of which produce data sufficiently rich to permit structural parameter estimation. Table 1 summarizes our list of factors, elicitation methods and their key antecedents. Details are in the Data Appendix.

Deviations from classical norms may be uni-directional, 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 bi-directional. For example, in the case of discounting one can be either present-biased or future-biased relative to being time-consistent (unbiased). 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 under-estimate than over-estimate the effects of compounding on future values, and so we count under-estimation as the standard bias and over-estimation as non-standard. Our empirics allow for the possibility that standard vs. non-standard biases have different links to behavior.

The Data Appendix provides further details on i) motivation for trying to measure that factor; ii) our elicitation method and its key antecedents; iii) data quality indicators, including item non-response; iv) sample size (as it compares to that for other factors); and v) descriptions of the *magnitude* and *heterogeneity* of behavioral deviations, including descriptions of the distribution and—where the data permits—estimates of key parameters used in behavioral models.

D. Measuring Control Variables: Demographics, Cognitive Ability, Risk Attitudes, and Patience

Our modules also elicit 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. Table 2 describes these variables and how we specify them in the empirics.

We measure aspects of cognitive skills using 4 standard tests. 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. A second test is comprised of 2 "numeracy" questions,[14] labeled as such and popularized in economics since their

[14] "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.

deployment in the 2002 English Longitudinal Study of Ageing.[15] Our mean number correct is 1.7, with a standard deviation of 0.6. A third test is a 3-question "financial literacy" quiz developed and popularized by Lusardi and Mitchell (2014).[16] The median respondent gets all 3 correct, with a mean of 2 and a SD of 0.93. We also measure executive function—including working memory and the regulation of attention—using a two-minute Stroop task (MacLeod 1991).[17] Each time the subject chooses an answer that action completes what we refer to as a "round."[18] 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 an SD of 24. Mean (median) proportion correct is 0.91 (0.99), with an SD of 0.19. These four test scores are strongly correlated with each other (Appendix Table 1).

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.[19] 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):

---

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

16 "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.

17 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).

18 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 her own. After practice ends, the task lasts for two minutes.

19 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).

"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).[20] The third and fourth are the switch points on the two multiple price lists we use to elicit the certainty premium (Data Appendix 1-D). 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 2). We use the first principal component of the four risk aversion measures in our regressions below.[21]

We elicit patience from the average savings rate across the 24 choices in our version of the Convex Time Budget task.

We also track and record survey response time, question-by-question from "click to click," as a measure of survey effort. We aggregate total response time spent for each B-factor, for each individual in the survey, and in the empirics below control for time spent as a measure of survey effort. We also present below some summary evidence examining links between survey effort, B-factors, and outcomes.

Finally, our other source of control variables is the ALP's standard set of demographic variables (such as age, income, education, etc.), which are collected when a panelist first registers, then refreshed quarterly and merged onto each new module. Those data also include state of residence, which we include in our empirics as state fixed effects.

E. Measuring Financial Outcomes

Finally, we designed our instrument to elicit rich data on financial outcomes for use as a dependent variable or variables in our empirics below linking B-factors to financial well-being. 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 elicit information on net worth, financial assets, recent savings

[20] We also elicit Dohmen et al.'s general risk taking scale, which is correlated 0.68 with the financial scale.
[21] The eigenvalue of the 1st principal component is 1.7, and none of the other principal components have eigenvalues greater than 1.

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.

Table 3 shows these financial outcomes, means and correlations between then. In each case "1" indicates plausibly better financial condition (greater wealth, more financial security, better "financial health", etc.). Our indicators include both stocks and flows. They include "hard" quantitative measures such as positive net worth, positive retirement assets, owning stocks, spending less than income in the last 12 months and not experiencing any of four specific indicators of severe financial distress in the last 12 months. They also include "soft" subjective and self-assessed measures of financial well-being such as financial satisfaction, financial stress, and viewing retirement or non-retirement savings as "adequate" or better.

1,508 of our 1,515 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 3 shows, these indicators are strongly correlated with each other.

Our main outcome takes 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 exhibits 4 of our 9 indicators of good financial condition.

2. Individual B-factors: prevalence and correlations with outcomes, controls and other B-factors

This section lays out the summary data on our individual B-factors. We then examine empirical links between those factors and financial condition. In part to allay concerns that our B-factors measure things other than behavioral tendencies, we then examine empirical links between those factors and our other controls. We conclude (and lead into the B-count and B-tile analysis) by discussing correlations between B-factors.

A. Summary Statistics on Individual B-Factors

Table 4 presents summary data on individual behavioral factors in our sample. There are several noteworthy patterns.[22]

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

First, estimated prevalence, defined as exhibiting any deviation from classical norms, varies substantially across B-factors. 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 re: consumption, gambler's fallacies, and overconfidence. The Data Appendix compares these findings to prior work; if anything, we find slightly lower prevalence than those other studies.

Third, despite the relatively modest prevalence in our data, many B-factors are quite prevalent, with some deviation indicated by at least 50% of the sample for 12 of the 16 factors for which we can estimate prevalence. Fourth, the "standard" directional bias emphasized by prior literature is indeed more prevalent in our data, in 7 of the 8 B-factors for which we can capture bi-directional biases.

Fifth, we can measure the absolute magnitude of deviations from neoclassical norms for 8 of our B-factors (9 if one counts our ambiguity aversion elicitation), and in these cases we see evidence of substantial deviations (see the Data Appendix). All but one of these factors exhibits large average deviations, and all of them include a substantial proportion of consumers exhibiting large deviations. Relatedly, both the magnitude and prevalence estimates suggest substantial heterogeneity across consumers, within B-factor.

B. B-factors and financial condition: estimation and econometric issues

We now turn to the primary empirical exercise, which we conduct first for individual B-factors and later for B-counts and B-tiles: regressing our summary measure of financial condition on B-factors and our rich set of controls, to estimate the conditional correlation between each factor and financial condition. Our main specification is:

$$FinCond_i = \alpha + \beta 1 \quad b\_factor\_s_i + \beta_2 b\_factor\_ns_i + \beta_3 b\_factor\_miss_i + \gamma X_i + \varepsilon_i,$$

where *i* indexes individuals. *FinCond* is individual-level financial condition on a [0, 1] scale. The next three variables are indicators for standard bias, non-standard bias (if applicable) and bias missing. The vector *X* contains the full set of control variables. That vector includes the full set of variables described in Table 2. In some cases responses for a particular variable (like patience) are missing, in which case we include a "missing" dummy for that observation and variable. In all we have 105 controls including categorical variables, derived from 17 different underlying control variables.

Econometric Issues

The empirical concerns plaguing nearly every non-experimental cross-sectional regression are unobserved heterogeneity and measurement error. In this case, our results could be spurious if other variables are correlated with our B-factors, and also correlated with financial condition, but omitted from the empirical model. Similarly, if other (included) variables are measured with error, and B-factors are measured with error and correlated with those included variables, then B-factor coefficients could reflect the influence of those other variables.

Primary candidates for such omitted variables, in our view, are “smarts” on some dimension: financial literacy, or cognitive ability, or math skill. Others could include classical preference parameters. While our set of controls is (unusually) rich and measures financial literacy, numeracy, cognitive ability, and both risk and patience, we nonetheless admit the possibility that imperfect or incomplete measures of those variables could create confounding influences on the B-factor variables—assuming, of course, that those other variables are correlated with our measured B-factors. While we plan to pursue more sophisticated approaches in later versions of the paper, here we address the omitted variable problem simply by estimating correlations between B-factors and *measured* financial literacy, numeracy, cognitive ability risk/patience, and so on. We then ask whether those correlations between observables are strong enough to warrant serious concern about the unobservable components of those same variables (or other related variables).

A second area of concern is survey effort. The idea there is that consumers who don’t care much about attending carefully to our (admittedly difficult) questions might give slapdash answers, inflating the prevalence of B-factors so long as behavioral answers are easier to give than non-behavioral answers. That would most likely be true in cases involving mathematics or statistics, because quick answers are less accurate and inaccuracy may be interpreted as being behavioral. It’s less clear how that would happen in questions about preferences (say, on present-bias for snacks vs. healthy food). Regardless, low survey effort by itself would not contaminate the findings—except to bias our B-factor coefficients *downward* since they would be noisier.

The more serious concern is that low survey effort is correlated with some other behavioral issue that would lead to truly worse real-world financial outcomes. We take three empirical approaches to addressing that latter concern. First, we control quite flexibly for survey effort in the empirics, by calculating B-factor-level survey response time and including decile indicators for survey time as controls.[23]

Second, we rely on the bi-directionality of some biases. Take exponential growth bias on the loan side, where consumers must infer an APR from a stream of payments and loan principal. If low survey effort leads to errors, and low survey effort is associated with poorer financial condition, then errors *in both directions* should be correlated with poorer financial condition. In many cases, however, the "theory" behind the behavioral bias suggests that the bias should have stronger effects in one direction—in the APR case, if consumers underestimate APRs (leading to too much borrowing) rather than over-estimate APRs (which would encourage savings). We therefore attend closely to the pattern of coefficients on bi-directional biases: are they symmetric, as the low survey effort story implies, or asymmetric, as behavioral theory in many cases implies?

And third, in some cases we stratify the sample or discard observations that seemingly reflect low survey effort (this primarily involves the very short response times, but can occasionally involve those at the other tail), and look at whether the results change or go away.

C. B-factors and financial condition: results

Table 5 shows results of these models, factor by factor in columns. We suppress coefficients on most of the controls, such as demographics and state effects. Because they are the focus of our econometric concerns above, we do show coefficients for the risk/patience, cognitive ability, time spent variables and education. Reading within a column, the "standard bias" refers to the one denoted as standard in Table 1. In some cases the "bias missing" variable is dropped due to collinearity.

The results generally support the view that individual behavioral factors are negatively correlated with financial condition. Five of the sixteen "standard" bias coefficients are negative

[23] The concern would also arise independent of true financial condition if "worse" answers were easier to give than "better" answers. Such a pattern is not likely given how most of our questions are laid out visually (see Appendix Table 5).

and statistically significant. Furthermore, of the remaining eleven, ten are negative and one is positive; a sign test rejects the hypothesis that these coefficients collectively are centered on zero rather than negative.

Magnitudes vary but one can easily interpret them since the RHS variables are dummies, the LHS variable is scaled on [0, 1] and the mean of the LHS is 0.43. For example, in the first column, present-bias on money leads to a reduction in financial condition of 0.041, about a 10% decline from the base. The most impactful biases, by this measure, are exponential growth bias (on the asset side) and limited attention. Collectively, the biases can have quite influential links to financial condition; in the extreme, turning all factors "on" and simply adding the coefficients leads to a decline in financial condition of 0.49, which moves someone from the 75th percentile to below the 25th in our data. Of course, few consumers are biased on every factor, and biases may not add up linearly. We address that latter point below when we include summary B-counts in similar models.

The results are robust to the inclusion of controls for survey time, education, risk aversion, patience, fluid intelligence, numeracy, financial literacy and executive attention. Of those, financial literacy appears most impactful and in the expected direction, with higher literacy associated with better financial condition.

Looking at the directional results, the pattern is more consistent with interpretation that our factors measure behavioral preferences and decision inputs rather than simply math or guesses. Two "nonstandard" coefficients are statistically significant, but one is positive and one is negative, more overall are positive than for the standard biases, and in seven out of eight instances the "standard" coefficient is *more negative* than the non-standard coefficient. Of course, the standard errors on these estimates are large relative to the point estimates, so this should be taken with a grain of salt.

C. B-Factors and links to other controls

As we said above, one way to think about the prospects for measurement error or omitted variable bias stemming from things related to other observables is to examine conditional correlations between B-factors and those observables. Table 6 does that, regressing each B-factor on the entire set of controls from Table 5.

The central takeaway is that B-factors are not simply functions of other covariates. Setting aside the general lack of statistical significance in the top rows (with the occasional exception of time spent, which enters consistent with greater survey effort being connected to a lower incidence of B-factors), the real keys lie in the columns showing r-squared and adjusted r-squared. The former range from 0.00-0.15 in all cases except two, generally lying around 0.10, and never are more than 0.31. By themselves fits around 0.10 in a cross-sectional regression aren't disastrous, but recall that our model is heavily and flexibly specified, perhaps overly so. That is borne out by the adjusted r-squared figures, which almost always lie in the 0.00-0.10 range, reaching above that in only three instances. Furthermore, the next three rows show fairly conclusively that even when the model fits some small amount of variation in the data, that explanatory power almost never comes from demographics, risk/patience/cognitive skill, or survey time spent. We show this by listing the change in adjusted r-squared one observes from adding each set of those variables to an otherwise fully specified (by us) model.[24]

In sum, the B-factors we observe are not clearly linked to any other observable variables. While it is possible that some omitted components of "smarts" or preferences or effort might affect the coefficients on B-factors in Table 5, those would have to be essentially orthogonal to the observable measures of survey effort, risk aversion, patience, cognitive skill, and demographics to comport with the overall pattern of results. Given that those same controls are generally viewed as key explanatory variables in other studies (and not others capturing behavior in the same domain), we view such a story as fairly implausible.

D. Correlations between B-factors

We next explore whether behavioral indicators are correlated within-consumer. Table 7 shows tetrachoric correlations among 15 B-factor indicators. We include 15 instead of 17 B-factors in this analysis because: a) the two choice inconsistency factors are strongly positively correlated by construction; b) we lack a cardinal measure of overconfidence in placement (see Appendix for details). For B-factors with bi-directional biases we include only the standard directional indicator. We find far more statistically significant positive correlations than one would expect to find by chance, and many fewer significant negative correlations.

[24] "Demographics" here include age, education, household size, gender, and ethnicity. They exclude income and fixed state effects (which are the baseline and account for any fit unexplained by the other groupings).

These simple results have several implications.[25] First, they hint at a key assumption of behavioral sufficient statistic models: behavioral biases are indeed positively correlated within-consumer (e.g., Chetty 2015). Second, they suggest that making inferences about (conditional) correlations between a B-factor and behavior could well be confounded by omitted B-factors. Third, they suggest that combining information across B-factors and within-consumer could help with predicting behavior, as we explore below.

4. B-counts and B-tiles: distributions, relationships with "standard" demographics, and correlation with financial condition

Here we define and examine the key constructs in the paper: summary individual-level statistics measuring how "behavioral" a person is, overall.

A. *The "B-count" and "B-tile": Person-level Summary Statistics for Behavioral Tendencies*

Our simplest summary measure is the "B-count": the sum of how many B-factor indicators an individual displays from Table 4. It is a count of extensive margins of behavioral deviations from classical benchmarks. Our primary B-count counts *any* deviation as behavioral, although we also consider other definitions.[26] The maximum possible B-count is 16 under most of our definitions.[27]

Our more nuanced "B-tile" aggregates factor-level measures of *how behavioral* an individual is, across B-factors, to the consumer-level. 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. For discrete-response and uni-directional outcomes like loss aversion, the –tiles take on fewer values but still measure the degree of deviation from classical benchmarks in useful ways.[28] Our

---

[25] There is clearly more to discover about the relationships among behavioral factors, and we are working on a companion paper devoted entirely to this subject [cite].

[26] A previous working paper version varied the thresholds at which a factor was "behavioral," with little change in the primary results.

[27] We have 24 indicators across 16 behavioral factors, but factors with bi-directional deviations allow for a maximum of one deviation per individual—bi-directional deviations are mutually exclusive within-person. Overconfidence in relative performance (the 17th factor) can only be measured in relative rather than absolute terms.

[28] 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.

approach 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, we calculate percentiles separately in each direction relative to the classical benchmark.

Specifically, the B-tile is the sum of all of the factor-level percentiles, rescaled for comparability to the B-count. We normalize the classical benchmark at 0 (rather than 1), to comport with the B-count: 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) 17. The B-tile has the advantage of being measurable each of our 17 factors, and it adds information beyond the factor-level indicator(s) for 15 of our 17 factors.

Table 8 shows that the median B-count, considering all possible deviations, is 9 with a standard deviation of 2.5. Nearly every consumer exhibits at least one deviation (100% with rounding), and the 10th percentile is 6th. Counting only standard-direction deviations among the bi-directional B-factors produces only slightly lower B-counts than counting any deviation (Column 1 vs. Column 3).

As with the B-counts, the B-tiles suggest that most consumers have substantial behavioral tendencies, although the B-tile distribution is shifted leftward of the B-count distribution, suggesting that small deviations are more common than large ones. The B-tiles also exhibit substantial heterogeneity across consumers, although the distribution is compressed compared to the B-counts because few people are out on the tails systematically across factors.

In an earlier working paper version, we asked whether these results were sensitive to setting more stringent definitions of "behavioral." In other words, are we simply quantifying "trembles" in responses to (some) admittedly difficult questions? The answer appears is no. Even setting thresholds that label a factor behavioral only if it deviates substantially from the classical benchmark, we still find that nearly all consumers in the sample are behavioral on several factors. Ours is not a knife-edge result.

These shapes of these simple summary statistics have important implications. First, they suggest that most consumers have behavioral tendencies to some meaningful extent, ratifying the focus of many policymakers and researchers on behavioral factors and how to treat them.

Second, they strongly support a key assumption in the literature using reduced-form behavioral sufficient statistics: consumer-level bias that is nonnegative, positive for some, and not mean-zero (i.e., classical on average) in the aggregate (Allcott and Taubinsky 2015, p. 2510). Third, they 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). Fourth, the fact that there does appear to be substantial heterogeneity in summary behavioral tendencies across individuals suggests that it is worth exploring whether B-stats help predict behavior.

The bottom panel of the table shows that missing-ness does not overly complicate the construction of B-stats, with the mean (median) respondent supplying data required to measure 14 (15) of the 16 behavioral factors. Nevertheless we control for missing-ness in our regressions below.

B. Who is more behavioral? Correlates of B-stats

Figures 1-4 shows our "standard" B-counts broken out by cognitive ability, gender, income and education. The B-count 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 to males, or to low-income or low-education individuals. In most cases the median level of B-count is similar across splits, and the most striking pattern here is that any cross-group differences are dwarfed by the within-group variation. We see a similar pattern for B-tiles. These patterns further enhance the prospect that heterogeneity in B-stats will help predict real-world behavior, conditional on a rich set of standard covariates including the ones considered in Figures 1-4.

B. Links Between B-counts and B-tiles and financial condition: estimation and econometric issues

In Table 9 we take our summary measure of financial condition and regress it on B-counts, B-tiles and the same rich set of controls detailed in the B-factor models. The count of missing B-factors enters as well. We include both information about B-counts/-tiles and the number of missing responses as separate regressors. We specify the B-count B-tile as entering the models linearly.[29] As with the B-factor regressions 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. Survey response time deciles enter as well, here aggregated across all B-factor questions.

Many of the same econometric issues discussed alongside the B-factor regressions arise here, of course. We discuss them in turn. We again show coefficients on selected variables: (Appendix Table 3) shows results for a more complete set of control variables.

Starting with columns (1) and (2), we find that B-counts and B-tiles are negatively correlated with financial condition in an economically meaningful way. The coefficients are essentially identical, and economically meaningful: increase the B-count or B-tile by four (i.e., adding 4 factors to an individual’s “behavioralness”) reduces financial condition by roughly 0.10, which is 25% on the mean of 0.43. Put another way, it’s equivalent to switching one of the nine financial well-being indicators from “on” to “off.”

Proceeding to the separation into standard and non-standard, we find the same pattern as with individual B-factors: “standard” deviations from classical benchmarks are more negative and impactful than “non-standard” deviations. This holds whether we measure deviations using B-counts or B-tiles. In both cases (-count and -tile) one can reject that standard and nonstandard coefficients are equal.

When we include both counts and tiles (Column 5), we find that counts seem a more powerful predictor of financial condition than tiles. This is useful to know since generally speaking our counts are easier to quantify and sample-independent.

Harking back to the individual B-factor models, recall that there we summed up all of the individual coefficients when “on,” and found that simple addition would suggest a reduction in condition of about 0.49. Here, moving from zero to sixteen factors reduces condition by 0.38 in point terms.[30] This sheds light on whether B-factors have additive/reinforcing effects or are

[29] Results are similar for alternative functional forms; they not reject linearity.
[30] Again, we cannot reject linearity in this relationship.

simply different metrics of similar underlying behaviors leading to the same outcomes. If the latter were true, higher B-counts might not lead to lower financial condition, past a certain point anyway, because they would be measuring "different parts of the elephant." These findings suggest that to the contrary, B-factors seem to have reinforcing or additive effects.[31]

In the last two columns we show evidence on robustness, and the overall explanatory power of our summary statistics. Column 5 excludes the risk, patience, cognitive ability and time spent variables. If those were collectively correlated with B-counts, and/or misspecified, then dropping them should significantly affect the B-count coefficient as it soaked up the now-omitted effects of those variables. But that doesn't happen: dropping them all leads to almost no increase in the B-count coefficient at all. Again, our construct appears to measure something distinct.

The final column asks how much cross-sectional variation in financial condition one can explain using *only* our B-counts and B-tiles, and finds that they explain 11% of variation in the cross-section.

In unreported results, we also show that the B-counts and B-tiles are also strongly conditionally correlated with outcomes in other domains: income and education.

If one conducts an exercise similar to that in Table 7, regressing B-counts on our other controls, the other controls collectively explain about 17% of cross-sectional variation in B-counts with an adjusted r-squared of 0.10. Results are similar if the dependent variable is the individual-level B-tile. We do find in these models that cognitive ability is conditionally correlated with B-counts (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).

E. Robustness

Recall that the primary empirical concerns are omitted variables, or measurement error, that would spuriously lead to significant links between our B-counts/-tiles and outcomes. To this point we have addressed these concerns in a number of ways by explicitly controlling for as

[31] One caveat to this line of thinking involves measurement error: if one views the B-count as eliciting *repeated observations* of similar but mis-measured underlying behavioral factors, then as Dean and Ortoleva and others point out, the B-count might simply be correcting the attenuation bias one would get by measuring a single factor. We are exploring this currently, but not that if that were true, one might expect that the sum of the coefficients from the individual B-factor models would be less than the 16-factor effect from the B-count model. The (relatively) low inter-correlations between our factors also belie this interpretation.

many confounds as possible, by exploiting the directionality of some biases, and by examining correlations between B-factors/-counts/-tiles and other controls. We now undertake some complementary analysis to shed further light on these issues.

One might wonder whether a single B-factor is driving the results. In unreported results we rerun the primary specification, removing indicator(s) for each behavioral factor, one-by-one. In every case the B-count coefficient remains negative and significant.

Nor is it that case that a single outcome indicator drives the results. Appendix Table 4 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, and financial stress).

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 include both counts as separate variables, and observe that both matter.

To shed more detail on survey effort, 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 5). 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.

In Table 10 we push further on the survey issue point by segmenting the sample by deciles of survey effort: 1-2, 3-8 and 9-10. The coefficients of interest remain roughly the same.

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 consideration of elicitation tasks, less measurement error, and hence fewer deviations from neoclassical norms.[32] 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.[33] Third, Appendix Table 4 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 17 behavioral factors, from over 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 "B-stats": summary statistics that capture the prevalence and heterogeneity of behavioral factors across people. For example, "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.

[32] 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 her financial condition to improve—and hence the marginal utility of a dollar to decline—over time.

[33] 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.

Perhaps most importantly, our B-stats 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. We find little evidence that B-stats proxy for cognitive skills or other omitted variables.

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.[34] This is turn is key to unpacking the direction and extent of any causality underlying conditional correlations between behavioral factors and outcomes.

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, labor, health and others.

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

REFERENCES

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. 2012. "Estimating Time Preferences from Convex Budgets." *The American Economic Review* 102 (7): 3333–56.

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.

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.

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.

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*.

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.
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.
Dimmock, Stephen, Roy Kouwenberg, Olivia S. Mitchell, and Kim Peijnenburg. forthcoming. “Ambiguity Aversion and Household Portfolio Choice Puzzles: Empirical Evidence.” *Journal of Financial Economics*.
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.

Gabaix, Xavier. 2015. “Behavioral Macroeconomics Via Sparse Dynamic Programming.”
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, DC.
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.”
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 Koszegi, and Philipp Strack. 2015. “Unrealistic Expectations and Misguided Learning.” SSRN Scholarly Paper. Rochester, NY. http://papers.ssrn.com/abstract=2703524.
Kőszegi, Botond. 2014. “Behavioral Contract Theory.” *Journal of Economic Literature* 52 (4): 1075–1118. doi:10.1257/jel.52.4.1075.
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.
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 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.
Mullainathan, Sendhil, Joshua Schwartzstein, and William J. Congdon. 2012. “A Reduced-Form Approach to Behavioral Public Finance.” *Annual Review of Economics* 4 (1): 511–40.
O’Donoghue, Ted, and 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, 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.
Shah, Anuj K, Sendhil Mullainathan, and Eldar Shafir. 2012. “Some Consequences of Having Too Little.” *Science* 338 (6107): 682–85.
Stango, Victor, and Jonathan Zinman. 2009. “Exponential Growth Bias and Household Finance.” *The Journal of Finance* 64 (6): 2807–49.
Stanovich, Keith E. 2016. “The Comprehensive Assessment of Rational Thinking.” *Educational Psychologist* 51 (1): 23–34. doi:10.1080/00461520.2015.1125787.
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.

Data Appendix

This Appendix details, for each of the 17 individual behavioral factors: i) some brief motivation for trying to measure that factor; ii) our elicitation method and its key antecedents; iii) data quality indicators, including item non-response; iv) sample size (as it compares to that for other factors); v) definitions and prevalence estimates of behavioral *indicators*, at different cutoffs for classifying a deviation from the neoclassical norm as behavioral; vi) descriptions of the *magnitude* and *heterogeneity* of behavioral deviations, including descriptions of the distribution and—where the data permits-- estimates of key parameters used in behavioral models; vii) estimates of conditional correlations between measures of the behavioral factor and financial outcomes.

Table 1 summarizes our elicitation methods and their key antecedents. Table 2 presents descriptive statistics from our sample (see below for comparisons to prior work). Table 5 summarizes conditional correlations between each behavioral factor and financial outcomes. Table 6 correlates b-factors with other controls.

A. Present- or Future-Biased Discounting (Money)

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

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

The CTB already has been implemented successfully in field contexts in the U.S. (Barcellos and Carvalho 2014; Carvalho, Meier, and Wang forthcoming) and elsewhere (Gine et al. 2015). In exploring data quality and prevalence below we focus on comparisons to Andreoni and

Sprenger (2012a),and Barcellos and Carvalho (2014).[1] AS draw their sample from university students. BC's sample is drawn from the ALP, like ours (module 212 in their case), but they use a different adaptation of the CTB [(Appendix Figure 2).

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

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

Indicators of behavioral deviations here are bi-directional: we label someone as present-biased (future-biased) if the average difference is >0 (<0). We deem present-bias the "standard" direction, since future-bias is relatively poorly understood[4] and could actually lead to more wealth accumulation. Counting any deviation from time-consistent discounting as biased, 26% of our sample is present-biased and 36% is future-biased. As Table [2 shows, these prevalence estimates fall substantially if we set a higher threshold for classifying someone as behavioral; e.g., if we count only deviations > |20|pp, then only 3% of the sample is present-biased and 5% future-biased. Compared to prior prevalence estimates, our zero-threshold ones are in the middle of the range. E.g., Barcellos and Carvalho's CTB elicitation in the ALP shows 29% with any

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

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

[3] Bradford et al (2014) do find present-bias on average in their Qualtrics sample, classifying >50% as present-biased and 26% as future-biased.

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

present-bias, and 37% with any future-bias. Goda et al use a different elicitation method— a "time-staircase" multiple price list (Falk et al. 2015)—and classify 55% of their nationally representative sample (from the ALP and another online panel) as present-biased. In Andreoni and Sprenger's (AS') study sample 14% exhibit any present-bias and 12% any future-bias.

Interestingly, if we follow AS and use the CTB data to structurally estimate discounting-bias parameter values for each individual, we find that 90% of our subjects with no monotonicity violations lie within the interval [0.93, 1.07] (Appendix Table [see ctb_beta_i.xls], Columns 11-13).[5] This is noteworthy because behavioral macro papers sometimes assume representative agents with present bias that lies strictly below our 5th percentile (see, e.g., (İmrohoroğlu, İmrohoroğlu, and Joines 2003; Graham and Snower 2013; Pérez Kakabadse and Palacios Huerta 2013). As Harris and Laibson (2013) state: "the short-run discount factor… is typically thought to lie between ½ and 1." Our estimates should give researchers pause before choosing a value much below 1.

Table [finsumm_ctb.xls] presents estimates, from different specifications, of conditional correlations between discounting biases and our summary measure of financial condition. Our first two sets of specifications allow for the possibility that present-bias and future-bias have similar correlations with behavior. And indeed we find some evidence of directional similar effects, although only present-bias is statistically significant. [discuss specs details, emphasizing the controls for b-summ vars. Discuss magnitudes of present bias corrs].

Table [finsumm_ctb.xls] differs in several respects from previous studies that estimate relationships between directly elicited discounting biases and outcomes in broad samples (Bradford et al. 2014; Eisenhauer and Ventura 2006; Goda et al. 2015).[6] We use CTBs rather than Multiple Price Lists, test more flexible functional forms, and control for a much richer set of (behavioral) factors that could be correlated with both discounting and outcomes.[7]

[5] The 5th to 95th percentile interval AS' sample is [0.91, 1.11], as reported in their Table 3.

[6] Other papers have explored links between discounting biases and real-world behavior using direct elicitations on narrower samples, with narrower sets of covariates; see e.g., Chabris et al (2008), Meier and Sprenger (2010), and Li et al (2015).

[7] Other key differences include Bradford et al (2014) lacking controls for cognitive skills, and Eisenhauer and Ventura (2006) only controlling for income.

B. Present- or Future-Biased Discounting (Food)

In light of evidence that discounting can differ within-subject across domains (e.g., Augenblick, Niederle, and Sprenger 2015), we also obtain a coarse measure of discounting biases for consumption per se, by asking two questions that follow Read and van Leeuwen (1998) : *"Now imagine that you are given the choice of receiving one of two snacks for free, [right now/five weeks from now]. One snack is more delicious but less healthy, while the other is healthier but less delicious. Which would you rather have [right now/five weeks from now]: a delicious snack that is not good for your health, or a snack that is less delicious but good for your health?* We fielded these questions in ALP module 352, the second of our two surveys.

Of the 1427 taking our second survey, 1423 answer one of the two snack questions, and 1404 respond to both. 61% choose the healthy snack for today, while 68% choose it for five weeks in the future, with 15% exhibiting present bias (consume treat today, plan to eat healthy in the future) and 7% future bias (consume healthy today, plan to eat treat in the future).[8] Barcellos and Carvalho's ALP subjects answered similar questions in their baseline survey, albeit with only a one-week instead of a five-week delay, with 6% exhibiting present-bias and 9% future-bias. Read and van Leeuwen (1998) offer actual snacks to a convenience sample of employees in Amsterdam but do not calculate individual-level measures of bias. They do find substantial present-bias on average.

Table [finsumm_snack.xls] present the first estimates we know of correlating measures of consumption discounting biases with real-world outcomes. [describe predictive results]

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

Our third and fourth behavioral factors follow Choi et al (2014), which measures choice inconsistency with standard economic rationality. Choice inconsistency could indicate a tendency to make poor (costly) decisions in 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.

---

[8] If we limit the sample to those who did not receive the informational/debiasing treatment about self-control in ALP module 212 (Barcellos and Carvalho), we find 15% with present bias and 8% with future bias (N=748).

We use the same task and user interface as in Choi et al (2014) but abbreviate it from 25 decisions to 11.[9] Each decision confronts respondents with a linear budget constraint under risk ([Appendix Figure 3): subjects choose a point on the line, and then the computer randomly chooses whether to pay the point value of the x-axis or the y-axis. 1,270 of the 1,427 individuals taking our second survey make all 11 decisions, and comprise our sample for measuring *choice inconsistency*.[10]

Following Choi et al, we average across these 11 decisions, within-consumer, to benchmark choices against two different standards of rationality. One benchmark is a complete and transitive preference ordering adhering to the General Axiom of Revealed Preference (GARP), as captured by the Afriat (1972) Critical Cost Efficiency Index. 1-CCEI can be interpreted as the subject's degree of choice inconsistency: the percentage points of potential earnings "wasted" per the GARP standard. But as Choi et al. discuss, consistency with GARP is not necessarily the most appealing measure of decision quality because it allows for violations of monotonicity with respect to first-order stochastic dominance (FOSD).[11] Hence, again following Choi et al., our second measure captures inconsistency with both GARP and FOSD.[12] Note that these measures of inconsistency are unidirectional: there is no way of classifying someone as being *overly* consistent.

Our distribution of individual-level CCEI estimates is nearly identical to Choi et al's— if we use only the first 11 rounds of choices from Choi et al to maximize comparability to our setup. Our median (1-CCEI) is 0.002, suggesting nearly complete consistency with GARP. The mean is 0.05. The median (1-combined-CCEI), capturing FOSD violations as well, is 0.10, with a mean of 0.16. Interestingly, choice inconsistency is substantially higher when using the full 25 rounds

[9] 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.

[10] 1424 individuals view at least one of the instruction screens, and 1,311 are recorded as completing at least one round of the task, and 1,270 are recorded as completing each of the 11 rounds.

[11] 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.

[12] 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.

in both our pilot data and Choi et al (e.g., mean CCEI of 0.12 in both samples), suggesting that task performance deteriorates in later rounds—more than offsetting any learning-by-doing—perhaps due to fatigue and/or declining attention.

In terms of prevalence in our data, 53% of subjects exhibit any inconsistency with GARP, and 96% exhibit any inconsistency with GARP or FOSD (Table 2). If we set a 20pp threshold for classifying someone as inconsistent, only 7% are inconsistent with GARP, and 31% are inconsistent with GARP or FOSD. Looking more directly at heterogeneity, we see standard deviations of 0.08 and 0.18, and 10-90 ranges of 0.16 and 0.41.

Choi et al finds that choice inconsistency is conditionally correlated with lower net worth. But our data shows little evidence of a conditional correlation between our broader measure of financial condition and various measures of choice inconsistency constructed from our data.

D. Risk attitude re: certainty

Behavioral researchers have long noted a seeming disproportionate preference for certainty (PFC) and posited various theories to explain it: Cumulative Prospect Theory (Daniel Kahneman and Tversky 1979; Amos Tversky and Kahneman 1992), 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 seemingly extreme risk averse behavior.

We use Callen et al's (2014) two-task method[13] for measuring a subject's *certainty premium* (CP).[14] Similar to Holt and Laury tasks, in one of the Callen et al tasks subjects make 10 choices between two lotteries, one a (p, 1-p) gamble over X and Y > X , (p; X, Y), the other a (q, 1-q) gamble over Y and 0, (q; Y, 0). Both Callen et al and we fix Y and X at 450 and 150 (hypothetical dollars in our case, hypothetical Afghanis in theirs), fix p at 0.5, and have q range from 0.1 to 1.0 in increments of 0.1. In the other task, p = 1, so the subject chooses between a lottery and a certain option. Our two tasks are identical to Callen et al's except for the currency units, but our settings, implementation, and use of the elicited data are very different. Callen et al

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

[14] The Callen et al tasks also elicit non-parametric measures of classical risk aversion: a higher switch point indicates greater risk aversion. We use the two switch points as two of our four measures of classical risk aversion..

administer the tasks in-person, using trained surveyors, at polling centers and homes in Afghanistan. They use the data to examine the effects of violence on risk preferences.

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

We estimate CP for each respondent i by imputing the likelihoods q* at which i expresses indifference as the midpoint of the q interval at which i switches, and then using the two likelihoods to estimate the indirect utility components of the CP formula. As Callen et al detail, the CP “is defined in probability units of the high outcome, Y, such that one can refer to certainty of X being worth a specific percent chance of Y relative to its uncertain value.” We estimate a mean CP of 0.16 in our sample (SD=0.24, median =0.15), compared to 0.37 (SD=0.15) in Callen et al. Callen et al’s findings suggest that much of the difference could be explained by greater exposure to violence in their sample.

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

E. Loss aversion/small-stakes risk aversion

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

We measure loss aversion using the two choices developed by Fehr and Goette (2007) in their study of the labor supply of bike messengers (see Abeler et al (2011) for a similar elicitation method). Choice 1 is between a lottery with a 50% chance of winning \$80 and a 50% chance of losing \$50, and zero dollars. Choice two is between playing the lottery in Choice 1 six times, and zero dollars. As Fehr and Goette (FG) show, if subjects have reference-dependent preferences, then subjects who reject lottery 1 have a higher level of loss aversion than subjects who accept lottery 1, and subjects who reject both lotteries have a higher level of loss aversion than subjects who reject only lottery 1. In addition, if subjects' loss aversion is consistent across the two lotteries, then any individual who rejects lottery 2 should also reject lottery 1 because a rejection of lottery 2 implies a higher level of loss aversion than a rejection of only lottery 1. Other researchers have noted that, even in the absence of loss aversion, choosing Option B is compatible with small-stakes risk aversion.[15] We acknowledge this but use "loss aversion" instead of "loss aversion and/or small-stakes risk aversion" as shorthand. Small-stakes risk aversion is also often classified as behavioral because it is incompatible with expected utility theory (Rabin 2000).

Response rates suggest a high level of comfort with these questions; only two of our 1,515 subjects skip, and only two more who answer the first question do not answer the second. 37% of our 1,511 respondents reject both lotteries, consistent with relatively extreme loss aversion, compared to 45% of FG's 42 subjects. Another 36% of our subjects accept both lotteries, consistent with standard neoclassical behavior, compared to 33% in FG. The remaining 27% of our subjects (and 21% of FG's) exhibit moderate loss aversion, playing one lottery but not the

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

other, with our main difference from FG being that 14% of our subjects (vs. only 2% of theirs) exhibit the puzzling behavior of playing lottery 1 but not lottery 2. Although one wonders whether these 14% misunderstood the questions, we find only a bit of evidence in support of that interpretation: those playing the single but not compound lottery have slightly lower cognitive skills than other loss averters, conditional on our rich set of covariates, but actually have higher cognitive skills than the most-neoclassical group. And playing the single but not the compound lottery is uncorrelated with our measure of ambiguity aversion, pushing against the interpretation that the compound lottery is sufficiently complicated as to appear effectively ambiguous (Dean and Ortoleva 2015).

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

Despite the massive amount of work on loss aversion, research exploring links between directly elicited measures of loss aversion and real-world behavior is only beginning. von Gaudecker et al (2011) do not explore links between loss aversion and real-world behavior. Dimmock and Kouwenberg (2010) do, like von Gaudecker et al using CentERdata, but lack many important covariates. Fehr and Goette (2007) find that loss aversion moderates the effect of a wage increase, but their sample includes only bike messengers and lacks measures of many other potentially moderating factors. Abeler et al (2011) find that loss aversion is strongly correlated with effort choices in the lab among their student sample, but again they lack data on many covariates of interest. Hwang (2016) uses the Abeler et al measures to infer a strong correlation between loss aversion and insurance holdings in an earlier ALP module, but lacks many important covariates and the only other behavioral factor considered is an interaction between loss aversion and a measure of the Gambler's Fallacy (labelled "Heuristics" in the Hwang paper).[16]

F. Narrow Bracketing and Dominated Choice

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

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

Our first task follows RW's Example 2, with Decision 1 between winning \$100 vs. a 50-50 chance of losing \$300 or winning \$700, and Decision 2 between losing \$400 vs. a 50-50 chance of losing \$900 or winning \$100.[17] As RW show, someone who is loss averse and risk-seeking in losses will, in isolation (narrow bracketing) tend to choose A over B, and D over C. But the combination AD is dominated with an expected loss of \$50 relative to BC. Hence a broad-bracketer will never choose AD. 29% of our subjects choose AD, compared to 53% in the most similar presentation in RW.

Our second task reproduces RW's Example 4, with Decision 1 between winning \$850 vs. a 50-50 chance of winning \$100 or winning \$1,600, and Decision 2 between losing \$650 vs. a 50-50 chance of losing \$1,550 or winning \$100. As in task one, a decision maker who rejects the risk in the first decision but accepts it in the second decision (A and D) violates dominance, here with an expected loss of \$75 relative to BC. 23% of our subjects choose AD, compared to 36% in the most similar presentation in RW. A new feature of task two is that AD sacrifices expected

[16] Hwang (2016) also discusses the potential (mediating) role of narrow framing/bracketing but lacks a directly elicited measure of such.

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

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

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

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

Research linking directly-elicited measures of NBDC to real-world outcomes is just beginning. The only paper we know of in this vein, Gottlieb and Mitchell (2015), uses a different method for measuring narrow bracketing—one that does not allow for dominated choice—the Tversky and Kahneman (1981) "sensitivity to framing" questions regarding the policy response to an epidemic. 30% of subjects in the Health and Retirement Study choose different policy options under the two different frames, an indicator of framing sensitivity, and this indicator is negatively correlated with the holding of long-term care insurance, conditional on a rich set of covariates include a measure loss aversion.

G. Ambiguity Aversion

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

We elicit a coarse measure of ambiguity aversion using just one or two questions about a game that pays $500 if you select a green ball. The first question offers the choice between a Bag One with 45 green and 55 yellow balls vs. a Bag Two of unknown composition. 1,397 subjects respond to this question (out of 1,427 who answer at least one of our questions on ALP module 352). 73% choose the 45-55 bag, and we label them ambiguity averse. The survey then asks these subjects how many green balls would need to be in Bag One to induce them to switch. We subtract this amount from 50, dropping the 99 subjects whose response to the second question is >45 (and the 10 subjects who do not respond), to obtain a continuous measure of ambiguity aversion that ranges from 0 (not averse in the first question) to 50 (most averse=== the three subjects who respond "zero" to the second question). The continuous measure (N=1,288) has a mean of 14 (median=10), and a SD of 13. If we impose a large-deviation threshold of 10 (20% of the max) for labeling someone as ambiguity averse, 50% of our sample exceeds this threshold and another 16% are at the threshold. Our elicitation does not distinguish between ambiguity-neutral and ambiguity-seeking choices (for more comprehensive but still tractable methods see, e.g., Dimmock, Kouwenberg et al (forthcoming), Dimmock, Kouwenberg, and Wakker (forthcoming), Gneezy et al (2015)), and so our measure of deviations from ambiguity-neutrality is one-sided.

Despite the coarseness of our elicitation, comparisons to other work suggests that it produces reliable data. Our ambiguity aversion indicator correlates strongly with one constructed from Dimmock et al's elicitation in the ALP (0.14, p-value 0.0001, N=789), despite the elicitations taking place roughly 3 years apart. Prevalence at our 10pp large-deviation cutoff nearly matches that from Dimmock, Kouwenberg et al's (forthcoming) ALP sample and Butler et al's (2014) Unicredit Clients' Survey sample from Italy, and the prevalence of any ambiguity aversion (0% cutoff) nearly matches Dimmock, Kouwenberg, and Wakker's (forthcoming) from the Dutch version of the ALP.

Our predictive analysis builds on the papers by Dimmock and co-authors cited above, which measure conditional correlations between ambiguity aversion and financial market behavior, by broadening the set of both outcomes and control variables (especially other b-factors).[18]

[18] The other paper we know of examining correlations between ambiguity attitudes and real-world behavior is Sutter et al's (2013) study of adolescents in Austria.

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

The first measures it in level/absolute terms, by following our three question numeracy quiz (Section 2-D) with the question: *"How many of the last 3 questions (the ones on the disease, the lottery and the savings account) do you think you got correct?"* We then subtract the respondent's assessment from her actual score. 39% of 1,366 subjects are overconfident ("overestimation" per Moore and Healy) by this measure (with 32% overestimating by one question), while only 11% are underconfident (with 10% underestimating by one question). Larrick et al (2007), Moore and Healy, and other studies use this method for measuring overestimation, but we are not aware of any that report individual-level prevalence estimates (they instead focus on task-level data, sample-level summary statistics, and/or correlates of cross-sectional heterogeneity in estimation patterns).

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

The third measures confidence in placement (relative performance), using a self-ranking elicited before taking our number series test (Section 2-D): *"We would like to know what you think about your intelligence as it would be measured by a standard test. How do you think your performance would rank, relative to all of the other ALP members who have taken the test?..."* We find a better-than-average effect in the sample as a whole (70% report a percentile>median) that disappears when we ask the same question immediately post-test, still not having revealed any scores (50% report a percentile>median). We also construct an individual-level measure of confidence in placement by subtracting the subject's actual ranking from his prettest self-ranking

(N=1,395). This measure is useful for capturing individual-level heterogeneity ordinally, but not for measuring prevalence because the actual ranking is based on a 15-question test and hence its percentiles are much coarser than the self-ranking.

Among the three sets of pairwise correlations at the individual level, only placement and precision are not strongly correlated with each other. The other two pairs have correlations ranging from 0.10 to 0.19 depending on functional form, with p-values <0.001.

We are not aware of any prior work exploring conditional correlations between the sorts of overconfidence measures describe above and real-world outcomes.

I. Non-belief in the Law of Large Numbers

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

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

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

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

1,375 subjects (out of the 1,427 who answer at least one of our questions in Module 352) respond,[19] with mean (SD) responses of 27 (18), 42 (24), and 31 (20). We measure NBLLN using the distance between the subject's answer for the [481, 519] range and 78. Only one subject gets it exactly right. 87% underestimate; coupled with prior work, this results leads us to designate underestimation as the "standard" directional bias. The modal underestimator responds

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

with 50 (18% of the sample). The other most-frequent responses are 25 (10%), 30 (9%), 33 (8%), and 40 (7%). Few underestimators—only 4% of the sample-- are within 10pp of 78, and their mean distance is 43, with an SD of 17. 9% of the sample underestimates by 20pp or less. 13% overestimate relative to 78, with 5% of the sample quite close to correct at 80, and another 5% at 100. Benjamin, Moore, and Rabin (2013) do not calculate individual-level measures of underestimation or overestimation in their convenience sample, but do report that the sample means are 35%, 36%, and 29% for the three bins. The comparable figures in our data our 27%, 42%, and 31%.

We are not aware of any prior work exploring conditional correlations between directly-elicited NBLLN and real-world outcomes.

J. 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 take a slice of Benjamin, Moore, and Rabin's (2013) elicitation for the gambler's fallacy (GF):

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

1,392 subjects respond, out of the 1,427 respondents to module 352. The standard GF implies a response < 50%, while the "hot hand" fallacy implies a response > 50%. Our mean response is 45% (SD=25), which is consistent with the GF but substantially above the 32% in Benjamin, Moore, and Rabin. Another indication that we find less evidence of the GF is that, while they infer that "at the individual level, the gambler's fallacy appears to be the predominant pattern of belief" (2013, p. 16), we find only 26% answering < "50". 14% of our sample responds with >"50" (over half of these responses are at "90" or "100"). So 60% of our sample answers correctly. Nearly everyone who responds with something other than "50" errs by a substantial amount—e.g., only 2 % of the sample is [30, 50) or (50, 70]. Sixteen percent of our sample answers "10",[20] which Benjamin, Moore, and Rabin speculates is an indicator of miscomprehension; we find that while subjects with this indicator do have significantly lower cognitive skills than the unbiased group, they actually have higher cognitive skills than the rest of subjects exhibiting a gambler's fallacy.

We are not aware of any prior work exploring conditional correlations between directly-elicited GF and real-world outcomes.

K. Exponential Growth Bias

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

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

We infer an individual-level measure of "debt-side EGB" by comparing the difference between the APR i*mplied* 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 underestimators (the standard bias), overestimators, unbiased, and unknown (37% of the sample).[21] The median level difference between the correct and stated value is 500bp, with a

---

[20] 34% of the sample in Benjamin, Moore, and Raymond respond "10%"on one or more of their ten questions.

[21] 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. 7% state payment amounts that imply a negative APR, even after being prompted to reconsider their answer. We also classify the 4% of respondents with implied APRs >=100% as having unknown bias.

mean of 1,042bp and SD of 1,879bp. Among those with known bias, we count as biased 70%, 64%, 47%, and 34% under error tolerance of zero, 100bp, 500bp, and 1000bp. Under these tolerances we count 3%, 13%, 41%, and 61% as unbiased, and 27%, 22%, 10%, and 3% as negatively biased. This is substantially less EGB than Stango and Zinman (2009; 2011) infer from similar questions in the 1983 Survey of Consumer Finances, where 98% of the sample underestimates, and the mean bias is 1,800bp or 3,800bp depending on the benchmark.

Stango and Zinman (2009; 2011) find that debt-side EGB is strongly conditionally correlated with loan terms, debt allocation, savings, and wealth. But those papers lack controls for cognitive skills and other behavioral factors.

Our second measure of EGB comes from a question popularized by Banks and Oldfield (2007) as part of a series designed to measure basic numeracy: "Let's say you have $200 in a savings account. The account earns 10 percent interest per year. You don't withdraw any money for two years. How much would you have in the account at the end of two years?" 1,389 subjects answer this question (out of the 1,427 respondents to Module 352), and we infer an individual-level measure of "asset-side EGB" by comparing the difference between the correct future value ($242), and the future value supplied by the same individual.[22] We again bin individuals into underestimators (the standard bias), overestimoators, unbiased, and unknown (14% of the sample).[23] Among those with known bias (N=1,222), the median bias is $0, with a mean of $2 and SD of $14.[24] 44% of our sample provides the correct FV. 47% of our sample underestimates by some amount, with most underestimators (29% of the sample) providing the linearized (uncompounded) answer of $240. Nearly all other underestimates provide an answer that fails to account for even simple interest; the most common reply in this range is "$220". 9% of our sample overestimates the FV, with small mass points at 244, 250, 400, and 440.

---

[22] 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.

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

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

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

Goda et al (2015) and Levy and Tasoff (forthcoming) measure asset-side EGB using more difficult questions in their representative samples. They find that 9% and 11% overestimate FVs, while 69% and 85% underestimate.

The only prior paper we know looking directly at links between a measure of asset-side EGB and real-world outcomes is Goda et al, which has data on fewer behavioral factors.

L. 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 measuring limited attention and/or memory, we create our own, using five simple questions and tasks.

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

Responses are strongly but not perfectly correlated (ranging 0.56 to 0.69 among pairwise expressions of regret). A fourth measure limited attention is also strongly correlated with the others, based on the question: "Do you believe that you could improve the prices/terms your household typically receives on financial products/services by shopping more?"[25] 18% respond "Yes, and I/we often regret not shopping more", and the likelihood of this response is correlated 0.25 with each of the regret measures above. 1,483 subjects answer all four questions, out of the 1,515 respondents to Module 315. Summing the four indicators of attentional regret, we find that 49% of subjects have one or more (earning a classification of behavioral inattention), 29% have two or more, 19% three or more, and only 6% have all four.

We also seek to measure limited prospective memory (e.g., 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 1,358. Only 14% actually complete the short survey. Failure to complete is weakly positively correlated with our indicators of attentional regret, with correlations ranging from 0.02 to 0.04, and p-values from 0.20 to 0.50.

Our indicator of behavioral limited memory—(do not complete | intended to complete)—is of course a coarse and noisy one. We suspect that most of the noise is introduced because our elicitation makes it costless to express an intention to complete (in future research we plan to

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

explore charging a small “sign up” fee), thereby including in the indicator’s sample frame some (many?) subjects who rationally do not complete the task. Relatedly, although we set the payoff for task completion to be sufficiently high to dominate any attention/memory/time costs in *marginal* terms for most subjects (the effective hourly wage is in the hundreds of dollars), it may well be the case that the *fixed* cost exceeds $10 for some (many?) respondents.

References for Data Appendix

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

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

Andreoni, James, and Charles Sprenger. 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."

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

Banks, James, Cormac O'Dea, and Zoë Oldfield. 2010. "Cognitive Function, Numeracy and Retirement Saving Trajectories." *The Economic Journal* 120 (548): F381–410. doi:10.1111/j.1468-0297.2010.02395.x.

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."

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

Benartzi, S., and R. H. Thaler. 1995. "Myopic Loss Aversion and the Equity Premium Puzzle." *The Quarterly Journal of Economics* 110 (1): 73–92. doi:10.2307/2118511.

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*.

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

Bradford, David, Charles Courtemanche, Garth Heutel, Patrick McAlvanah, and Christopher Ruhm. 2014. "Time Preferences and Consumer Behavior." National Bureau of Economic Research.

Butler, Jeffrey V., Luigi Guiso, and Tullio Jappelli. 2014. "The Role of Intuition and Reasoning in Driving Aversion to Risk and Ambiguity." *Theory and Decision* 77 (4): 455–84. doi:10.1007/s11238-013-9407-y.

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

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

Chabris, Christopher F., David Laibson, Carrie L. Morris, Jonathon P. Schuldt, and Dmitry Taubinsky. 2008. "Individual Laboratory-Measured Discount Rates Predict Field

Behavior." *Journal of Risk and Uncertainty* 37 (2–3): 237–69. doi:10.1007/s11166-008-9053-x.
Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman. 2014. "Who Is (More) Rational?" *American Economic Review* 104 (6): 1518–50.
Daniel, Kent, and David Hirshleifer. 2015. "Overconfident Investors, Predictable Returns, and Excessive Trading." *Journal of Economic Perspectives* 29 (4): 61–88.
Dean, Mark, and Pietro Ortoleva. 2015. "Is It All Connected? A Testing Ground for Unified Theories of Behavioral Economics Phenomena."
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 G., and Roy Kouwenberg. 2010. "Loss-Aversion and Household Portfolio Choice." *Journal of Empirical Finance* 17 (3): 441–59. doi:10.1016/j.jempfin.2009.11.005.
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*.
Eisenhauer, Joseph G, and Luigi Ventura. 2006. "The Prevalence of Hyperbolic Discounting: Some European Evidence." *Applied Economics* 38 (11): 1223–34.
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, David Huffman, and Uwe Sunde. 2015. "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences."
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.
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, DC.
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."
Gottlieb, Daniel, and Olivia S. Mitchell. 2015. "Narrow Framing and Long-Term Care Insurance."
Graham, Liam, and Dennis J Snower. 2013. "Hyperbolic Discounting and Positive Optimal Inflation." *Macroeconomic Dynamics* 17 (3): 591–620.
Grubb, Michael D. 2015. "Overconfident Consumers in the Marketplace." *Journal of Economic Perspectives* 29 (4): 9–36.
Gul, Faruk. 1991. "A Theory of Disappointment Aversion." *Econometrica: Journal of the Econometric Society* 59 (3): 667–86.

Gustman, Alan L, Thomas L Steinmeier, and Nahid Tabatabai. 2012. "Financial Knowledge and Financial Literacy at the Household Level." *American Economic Review* 102 (3): 309–13. doi:10.1257/aer.102.3.309.

Harris, Christopher, and David Laibson. 2013. "Instantaneous Gratification." *The Quarterly Journal of Economics* 128 (1): 205–48.

Hwang, In Do. 2016. "Prospect Theory and Insurance Demand."

İmrohoroğlu, Ayşe, Selahattin İmrohoroğlu, and Douglas H. Joines. 2003. "Time-Inconsistent Preferences and Social Security." *The Quarterly Journal of Economics* 118 (2): 745–84.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47 (2): 263–91.

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, 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, 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.

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

Moore, Don A., and Paul J. Healy. 2008. "The Trouble with Overconfidence." *Psychological Review* 115 (2): 502–17.

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

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

Pérez Kakabadse, Alonso, and Ignacio Palacios Huerta. 2013. "Consumption and Portfolio Rules with Stochastic Hyperbolic Discounting."

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.

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.

Sutter, Matthias, Martin G Kocher, Daniela Glätzle-Rützler, and Stefan T Trautmann. 2013. “Impatience and Uncertainty: Experimental Decisions Predict Adolescents’ Field Behavior.” *American Economic Review* 103 (1): 510–31. doi:10.1257/aer.103.1.510.

Toubia, Olivier, Eric Johnson, Theodoros Evgeniou, and Philippe Delquié. 2013. “Dynamic Experiments for Estimating Preferences: An Adaptive Method of Eliciting Time and Risk Parameters.” *Management Science* 59 (3): 613–40. doi:10.1287/mnsc.1120.1570.

Tversky, A., and D. Kahneman. 1981. “The Framing of Decisions and the Psychology of Choice.” *Science* 211 (4481): 453–58. doi:10.1126/science.7455683.

Tversky, Amos, and Daniel Kahneman. 1992. “Advances in Prospect Theory: Cumulative Representation of Uncertainty.” *Journal of Risk and Uncertainty* 5 (4): 297–323.

Von Gaudecker, Hans-Martin, Arthur Van Soest, and Erik Wengström. 2011. “Heterogeneity in Risky Choice Behavior in a Broad Population.” *The American Economic Review* 101 (2): 664–94.

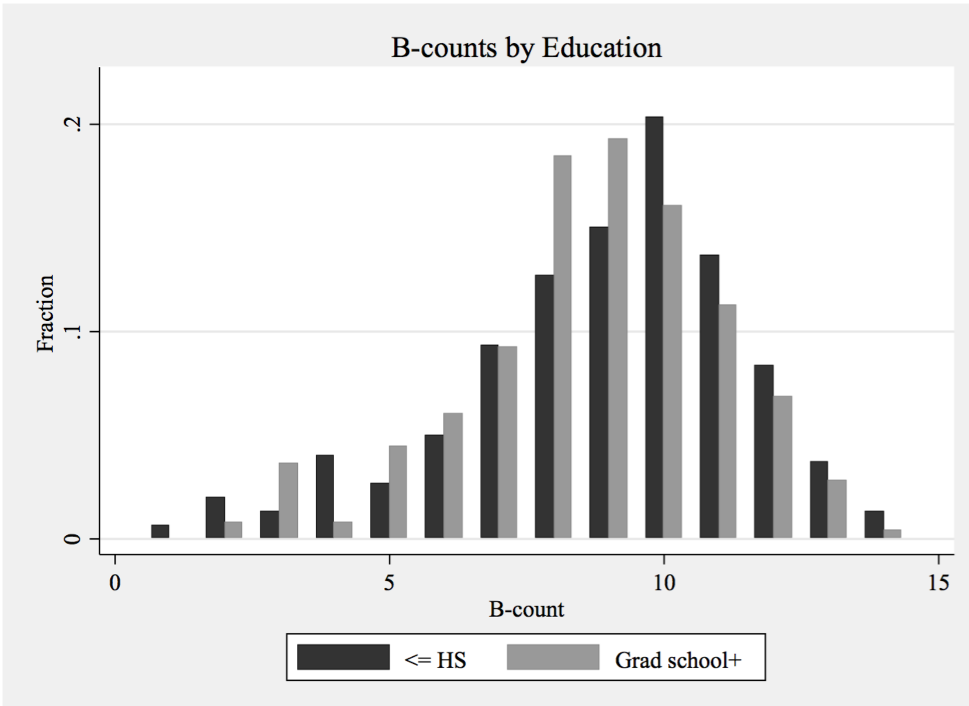

Figure 1

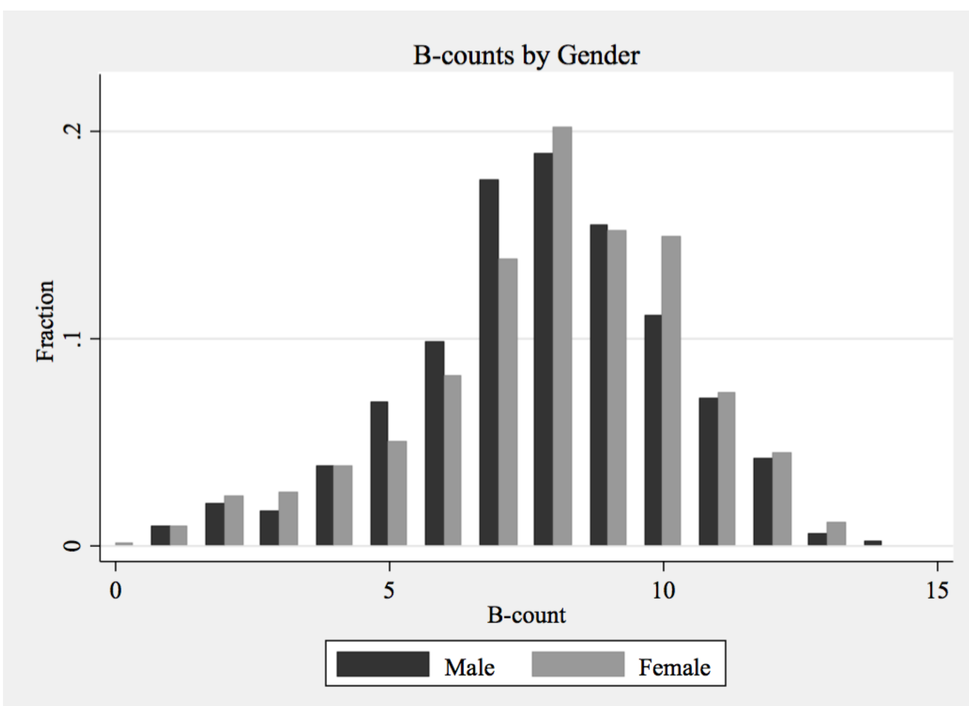

Figure 2

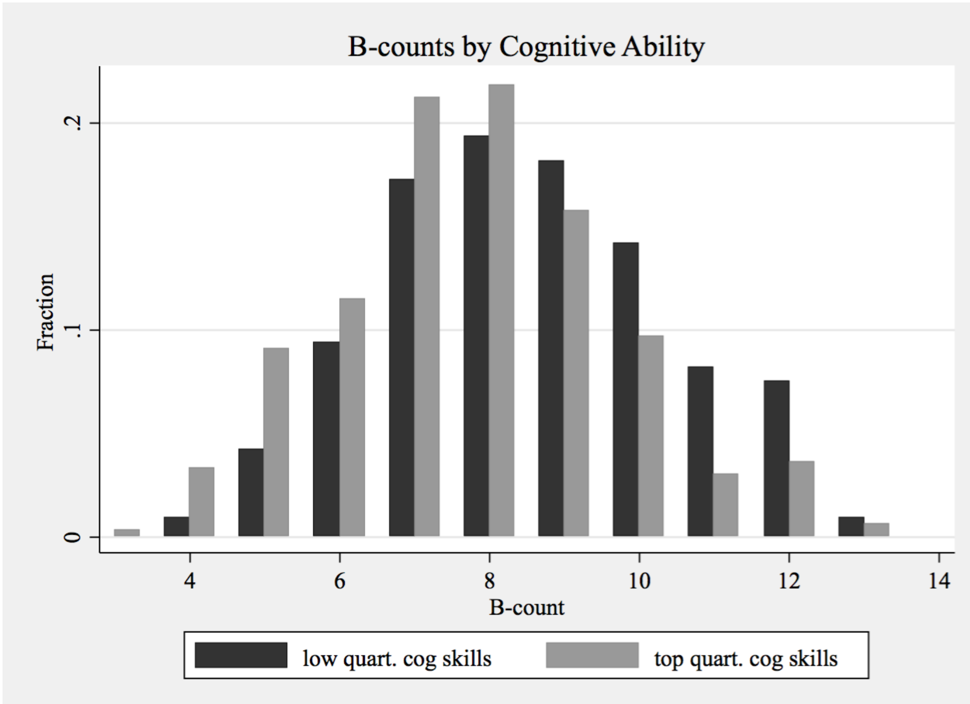

Figure 3

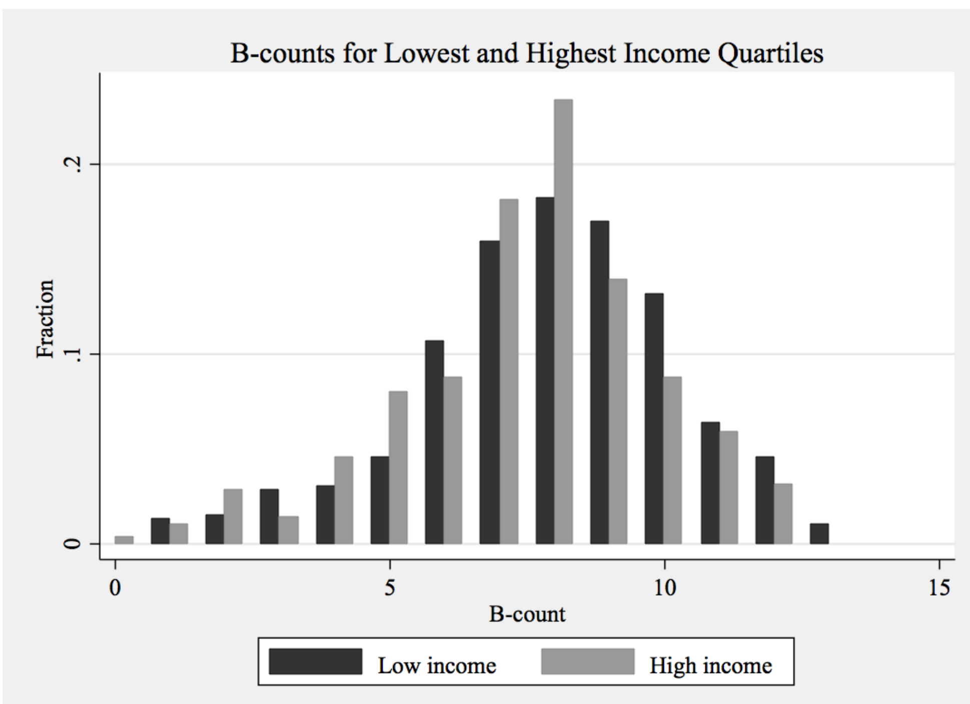

Figure 4

Table 1. Research design: eliciting data on multiple behavioral factors, and defining biases

Factor name: *key antecedents*	Elicitation method description	Behavioral indicator(s), "standard" deviation direction 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 tdy
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 tdy, healthy for 5 weeks from now** Future-biased: choose healthy for tdy, 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, and 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 and 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 cetainty: certainty premium (CP) >0** Cumulative prospect theory: 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: making a choice that is dominated given 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: indicated if prefers 45 green balls to uncertain mix, increases as number of green balls declines**
(Over-)confidence in 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** Underconfidence: self-assessment < actual score
(Over-)confidence in relative performance: *Larrick et al (2007), Moore & Healy (2008)*	"… what you think about your intelligence as it would be measured by a standard test. How do you think your performance would rank, relative to all of the other ALP members who have taken the test?"	**Greater diff between self-assessed and actual rank indicates more overconfidence. But only ordinal, no cardinal, measures here, because test has < questions than percentiles**
Overconfidence in precision: *Larrick et al (2007), Moore & Healy (2008)*	Questions about about likelihoods of different numeracy quiz scores and 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, 1000]. 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**
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 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 given the opportunity cost of attention, with questions varying the types of decisions: day-to-day, medium-run, long-run, or choosing financial	**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**

The Data Appendix provides additional details on measuring individual behavioral factors. "pp" = percentage points. "CCEI" = Critical Cost Efficiency Index. "Standard" bias accounting applies only to factors with bi-directional biases.

Table 2. Controls variable definitions/specifications.

Control	Definition/specification in empirics
Demographics:	
Gender	*Indicator, "1" for female.*
Age	*Four categories: 18-34, 35-45, 46-54, 55+*
Education	*Four categories: HS or less, some college/associates, BA, graduate*
Income	*The ALP's 17 categories*
Race/ethnicity	*Three categories: White, Black, or Other; seperate indicator for Hispanic*
Marital status	*Three categories: married/co-habitating; separated/divorced/widowed; never married*
Household size	*Five categories for count of other members: 0, 1, 2, 3, 4+*
Employment status	*Five categories: working, self-employed, not working, disabled, missing*
Immigrated to USA	*Indicator, "1" for immigrant*
State of residence	*Fixed effects*
Risk, patience:	
Risk aversion	*1st principal component of four measures, standardized*
Patience	*Average savings rate across the 24 Convex Time Budget decisions, standardized*
Cognitive skills:	
Fluid intelligence	*# correct on standard 15-question, non-adaptive number series quiz*
Numeracy	*# correct on Banks and Oldfield questions re: division and %*
Financial literacy	*# correct on Lusardi and Mitchell "Big Three" questions re: interest, inflation, and diversification*
Executive attention	*# correct on 2-minute Stroop test; respondents instructed to answer as many q's correctly as they can*
Survey effort (time spent)	*Measured for each B-factor (and other variable), included as decile indicators relative to other consumers*

Notes: See Section 2 of the text for detail on elicitations and construction.

Table 3. 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 months	fin stress < median
Net worth>0	0.44	1								
Retirement assets>0	0.53	***0.33***	1							
Owns stocks	0.49	***0.34***	***0.82***	1						
Spent < income last 12 months	0.36	***0.28***	***0.21***	***0.20***	1					
Financial satisfaction > median	0.46	***0.23***	***0.23***	***0.22***	***0.31***	1				
Retirement saving adequate	0.26	***0.23***	***0.19***	***0.18***	***0.27***	***0.30***	1			
Non-ret saving adequate	0.25	***0.12***	0.02	**0.05**	***0.18***	***0.17***	***0.31***	1		
No financial hardship in last 12 mos.	0.56	***0.30***	*0.29*	***0.29***	***0.32***	***0.34***	***0.30***	***0.15***	1	
Self-assessed financial stress < median	0.51	***0.26***	***0.15***	***0.17***	***0.29***	***0.33***	***0.29***	***0.16***	***0.32***	1
Mean of all indicators	0.43									

Unconditional pairwise correlations. Bold/italics indicate significance at p < 0.01.
Pairwise sample sizes range from 1391 to 1508.

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 401ks/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 an 100 point scale in response to: "To what extent, if any, are finances a source of stress in your life?"

Table 4. Prevalence and missing values for individual B-factors.

			Missing detail		
Factor and bias	Share biased	Share "missing"	survey nonresponse	item nonresponse	responded, not usable
Discounting money: Present-biased	**0.26**	0.06	0.00	0.06	0.00
Discounting money: Future-biased	0.36				
Discounting snacks: Present-biased	**0.15**	0.07	0.06	0.02	0.00
Discounting snacks: Future-biased	0.07				
Violates GARP (based on CCEI)	**0.53**	0.16	0.06	0.10	0.00
Loses by violating GARP or dominance violations	**0.96**	0.16	0.06	0.10	0.00
Preference for certainty type (positive certainty premium)	**0.77**	0.31	0.00	0.03	0.18
Cumulative prospect theory type (negative certainty premium)	0.23				
Loss-averse: prefers certain zero payoff	**0.63**	0.00	0.00	0.00	0.00
Narrow-brackets	**0.59**	0.02	0.00	0.02	0.00
Ambiguity-averse	**0.73**	0.08	0.06	0.03	0.07
Overconfident in level performance	**0.38**	0.10	0.06	0.03	0.07
Underconfident in level performance	0.11				
Overconfident in precision	**0.44**	0.11	0.06	0.04	0.00
Overconfident in relative performance	n/a		0.06	0.04	0.00
Non-belief in the law of large numbers: under-estimates convergence	**0.87**	0.09	0.06	0.03	0.00
Non-belief in the law of large numbers: over-estimates convergence	0.13				
Gambler's fallacy	**0.26**	0.08	0.06	0.02	0.00
Hot hand fallacy	0.14				
Exponential growth bias, loan-side: underestimates APR	**0.70**	0.37	0.00	0.05	0.32
Exponential growth bias, loan-side: over-estimates APR	0.27				
Exponential growth bias, asset-side: underestimates future value	**0.47**	0.19	0.06	0.03	0.11
Exponential growth bias, asset-side: over-estimates future value	0.09				
Limited attention with regret	**0.49**	0.02	0.00	0.02	0.00
Limited prospective memory	**0.86**	0.10	0.06	0.02	0.02

Section 2-C provides some details on measuring individual behavioral factors; see the text and Data Appendix for additional details."Share biased" is conditional on non-missing values. "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. "Survey nonresponse" generates missing values when the consumer failed to take the second survye module. "Item nonresponse" can occur on either module. "Unusable" reaponses are those that are illogical or internally inconsistent (e.g., negative loan APRs on the EGB questions). Most later analyses use standard and non-standard bias indicators; others use only "standard" indicators

Table 5. Individual B-factor indicators and financial condition.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	RHS bias variable shown as column header. LHS variable in all models is summary financial condition.															
Variable	Discounting money	Discounting snack	GARP	GARP FOSD	Pref for certainty	Loss averse	Narrow bracket	Ambig averse	OC perf	OC precision	NBLLN	Gambler's fallacy	EGB loan	EGB asset	Limited attention	Limited memory
Standard bias indicator	**-0.041****	-0.018	-0.010	-0.021	-0.011	-0.011	-0.001	**-0.039***	-0.015	-0.015	-0.091	-0.013	0.017	**-0.059*****	**-0.111*****	**-0.043****
	(0.016)	(0.018)	(0.013)	(0.034)	(0.034)	(0.013)	(0.013)	**(0.020)**	(0.015)	(0.014)	(0.235)	(0.016)	(0.044)	**(0.017)**	**(0.012)**	**(0.018)**
Non-standard bias indicator	-0.007	-0.009			-0.003				**0.037***		-0.064	**-0.033***	0.021	-0.037		
	(0.016)	(0.025)			(0.037)				**(0.021)**		(0.235)	**(0.020)**	(0.046)	(0.026)		
bias missing	-0.008	0.034	-0.011				0.022	0.025	0.058	**0.070***	0.051	-0.077	0.024	-0.022	-0.071	0.010
	(0.028)	(0.050)	(0.023)				(0.049)	(0.025)	(0.109)	**(0.037)**	(0.240)	(0.096)	(0.044)	(0.025)	(0.052)	(0.042)
1st PC risk attitudes (stdized)	-0.008	-0.009	-0.009	-0.009	-0.009	-0.008	-0.009	-0.008	-0.010	-0.009	-0.008	-0.009	-0.009	**-0.010***	**-0.011***	-0.009
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	**(0.006)**	**(0.006)**	(0.006)
patience (stdized)	0.004	0.006	0.007	0.007	0.006	0.006	0.006	0.005	0.007	0.005	0.006	0.005	0.005	0.006	0.004	0.007
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
fluid intell # correct	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.000	0.002	-0.001	0.001	0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
numeracy # correct	0.001	0.002	0.002	0.003	0.001	0.001	0.002	0.003	-0.006	0.006	0.003	0.004	0.003	0.003	0.001	0.000
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
financial literacy # correct	**0.029*****	**0.030*****	**0.029*****	**0.029*****	**0.030*****	**0.029*****	**0.028*****	**0.028*****	**0.030*****	**0.032*****	**0.027*****	**0.028*****	**0.029*****	**0.023****	**0.027*****	**0.028*****
	(0.009)	**(0.009)**	**(0.009)**	**(0.009)**	**(0.009)**	**(0.009)**	**(0.009)**	**(0.009)**	**(0.009)**	**(0.009)**	**(0.009)**	**(0.009)**	**(0.009)**	**(0.009)**	**(0.009)**	**(0.009)**
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	0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Time spent decile 2	0.000	-0.048*	0.014	0.014	-0.026	0.019	0.001	-0.001	0.007	-0.021	0.036	-0.029	-0.031	-0.006	**-0.046***	0.008
	(0.027)	(0.027)	(0.028)	(0.028)	(0.028)	(0.026)	(0.028)	(0.028)	(0.024)	(0.026)	(0.028)	(0.028)	(0.027)	(0.027)	**(0.026)**	(0.026)
Time spent decile 3	-0.007	-0.044	0.013	0.014	-0.003	0.021	-0.002	0.039	0.039	0.023	0.037	-0.002	-0.040	0.038	-0.037	0.007
	(0.027)	(0.027)	(0.029)	(0.029)	(0.028)	(0.026)	(0.027)	(0.029)	(0.027)	(0.027)	(0.029)	(0.026)	(0.028)	(0.027)	(0.027)	(0.028)
Time spent decile 4	0.015	-0.023	-0.002	-0.001	**-0.051***	**0.061****	0.013	0.009	-0.032	0.020	0.026	0.020	-0.044	0.011	-0.013	-0.033
	(0.028)	(0.028)	(0.029)	(0.029)	**(0.028)**	**(0.027)**	(0.027)	(0.031)	(0.029)	(0.027)	(0.029)	(0.028)	(0.028)	(0.029)	(0.027)	(0.026)
Time spent decile 5	-0.004	-0.025	0.035	0.036	0.019	0.044	-0.001	0.042	0.019	-0.004	0.045	0.036	**-0.049***	0.021	-0.021	-0.022
	(0.027)	(0.026)	(0.030)	(0.030)	(0.028)	(0.027)	(0.027)	(0.032)	(0.030)	(0.027)	(0.029)	(0.028)	**(0.029)**	(0.027)	(0.027)	(0.029)
Time spent decile 6	0.013	**-0.054****	-0.019	-0.018	**-0.049***	0.036	-0.001	0.027	0.005	0.026	0.048	0.028	-0.022	-0.004	0.010	0.039
	(0.028)	**(0.027)**	(0.029)	(0.029)	**(0.028)**	(0.027)	(0.027)	(0.033)	(0.031)	(0.028)	(0.030)	(0.027)	(0.029)	(0.028)	(0.027)	(0.029)
Time spent decile 7	0.003	-0.028	0.005	0.006	-0.026	0.004	0.018	-0.002	-0.031	-0.010	0.006	-0.028	-0.028	**0.050***	-0.039	0.032
	(0.028)	(0.028)	(0.030)	(0.030)	(0.028)	(0.028)	(0.028)	(0.032)	(0.028)	(0.028)	(0.029)	(0.027)	(0.029)	**(0.029)**	(0.026)	(0.027)
Time spent decile 8	0.009	-0.024	0.011	0.012	-0.033	0.018	0.019	0.031	0.029	-0.003	0.030	-0.017	0.004	-0.002	-0.028	-0.030
	(0.028)	(0.028)	(0.030)	(0.029)	(0.029)	(0.027)	(0.027)	(0.033)	(0.029)	(0.028)	(0.029)	(0.028)	(0.029)	(0.028)	(0.027)	(0.027)
Time spent decile 9	-0.019	-0.022	0.018	0.019	0.005	0.009	0.007	0.009	-0.028	0.033	0.023	-0.010	-0.030	0.009	-0.020	0.023
	(0.028)	(0.028)	(0.030)	(0.030)	(0.028)	(0.027)	(0.028)	(0.033)	(0.028)	(0.028)	(0.030)	(0.028)	(0.029)	(0.028)	(0.027)	(0.028)
Time spent decile 10	0.003	-0.015	-0.024	-0.023	-0.041	0.016	0.020	0.030	0.052*	-0.011	**0.052***	-0.001	-0.003	0.016	-0.015	0.015
	(0.028)	(0.028)	(0.029)	(0.029)	(0.028)	(0.027)	(0.028)	(0.032)	(0.029)	(0.028)	**(0.029)**	(0.027)	(0.029)	(0.028)	(0.027)	(0.028)
R-squared	0.40	0.40	0.40	0.40	0.40	0.39	0.39	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.43	0.40
Observations	1505	1505	1505	1505	1505	1502	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505

* 0.10 ** 0.05 *** 0.01.Unit of observation is the individual. LHS variable is our summary statistic for financial condition: the proportion of positive financial indicators in Table 4. Each column includes a different behavioral factor bias measure, as denoted by column headers. Models are OLS and also include the full set of controls in Table 2: age categories, gender, education categories, four race/ethnicity categories, state of residence, immigrant indicator, 17 income categories, 3 marital status categories, 4 household size categories, 5 work status categories, fixed state effects, and dummies for missing values associated with each variable. Patience is the average savings rate across the 24 CTB decisions; we also include a dummy for missing this variable. See Section 2 and the Data Appendix for detail on variable construction.

Table 6. Are individual B-factor indicators well-explained by standard factors and other controls?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Dependent variable shown as column header, coded as 0/1 with "1" indicating standard bias and "0" indicating no or nonstandard bias															
Variable	Discounting money	Discounting snack	GARP	GARP FOSD	Pref for certainty	Loss averse	Narrow bracket	Ambig averse	OC perf	OC precision	NBLLN	Gambler's fallacy	EGB loan	EGB asset	Limited attention	Limited memory
highest ed: some college or associates	0.032	0.041	0.042	0.020	0.046	0.004	0.048	0.108***	0.068*	0.031	0.016	-0.046	-0.065	0.018	-0.025	0.028
	(0.036)	(0.030)	(0.043)	(0.017)	(0.043)	(0.036)	(0.039)	(0.039)	(0.038)	(0.039)	(0.027)	(0.035)	(0.047)	(0.040)	(0.039)	(0.029)
highest ed: bachelor's	0.030	0.001	-0.012	0.007	-0.021	0.041	0.026	0.127***	0.007	0.048	-0.023	-0.104**	-0.040	0.024	-0.111**	0.003
	(0.041)	(0.034)	(0.050)	(0.020)	(0.047)	(0.042)	(0.045)	(0.045)	(0.044)	(0.045)	(0.031)	(0.041)	(0.053)	(0.045)	(0.045)	(0.034)
highest ed: graduate	0.030	0.019	0.035	-0.002	-0.060	-0.036	0.066	0.147***	-0.027	-0.039	-0.054	-0.055	-0.041	-0.010	-0.159***	-0.007
	(0.047)	(0.039)	(0.057)	(0.023)	(0.052)	(0.048)	(0.052)	(0.052)	(0.051)	(0.052)	(0.036)	(0.047)	(0.060)	(0.052)	(0.053)	(0.039)
1st PC risk attitudes (stdized)	0.007	-0.016*	-0.002	-0.003	-0.038***	0.064***	-0.018	0.042***	-0.021*	0.001	0.015*	-0.009	-0.017	-0.015	-0.026*	-0.005
	(0.012)	(0.010)	(0.014)	(0.006)	(0.012)	(0.013)	(0.013)	(0.013)	(0.012)	(0.013)	(0.009)	(0.012)	(0.016)	(0.012)	(0.013)	(0.010)
patience (stdized)	-0.027**	0.006	-0.002	-0.004	0.009	-0.024*	0.011	-0.026*	-0.006	0.002	-0.010	0.002	0.002	-0.006	-0.016	0.004
	(0.012)	(0.010)	(0.015)	(0.006)	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.009)	(0.012)	(0.016)	(0.013)	(0.014)	(0.010)
fluid intell # correct	0.007	-0.004	-0.012*	-0.005**	0.006	0.009	-0.018***	0.003	0.002	0.019***	-0.018***	-0.019***	0.009	-0.043***	-0.003	-0.005
	(0.006)	(0.005)	(0.007)	(0.003)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.004)	(0.005)	(0.008)	(0.006)	(0.006)	(0.005)
numeracy # correct	-0.023	0.001	-0.028	0.012	-0.013	-0.066**	0.036	0.047*	-0.288***	0.077***	0.005	0.015	-0.039	0.060**	0.001	-0.018
	(0.027)	(0.021)	(0.031)	(0.012)	(0.035)	(0.027)	(0.029)	(0.028)	(0.027)	(0.029)	(0.019)	(0.025)	(0.036)	(0.028)	(0.029)	(0.021)
financial literacy # correct	-0.004	0.006	-0.024	-0.014*	-0.032	-0.010	-0.014	0.031	0.004	-0.022	-0.024*	-0.022	0.010	-0.092***	-0.029	-0.037***
	(0.018)	(0.014)	(0.021)	(0.008)	(0.021)	(0.018)	(0.019)	(0.019)	(0.018)	(0.019)	(0.013)	(0.017)	(0.024)	(0.020)	(0.020)	(0.014)
exec attention # correct	-0.000	-0.000	-0.001	-0.000	-0.001	0.002***	0.001	-0.001	-0.000	-0.001	0.001	-0.000	-0.000	-0.001**	-0.000	-0.000
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Time spent decile 2	0.106*	0.049	0.006	0.005	-0.149*	-0.042	-0.020	-0.004	-0.037	-0.105*	0.025	-0.009	-0.149*	-0.110*	-0.044	0.037
	(0.054)	(0.043)	(0.070)	(0.028)	(0.079)	(0.053)	(0.059)	(0.057)	(0.049)	(0.060)	(0.041)	(0.053)	(0.080)	(0.058)	(0.060)	(0.040)
Time spent decile 3	0.040	-0.001	-0.012	0.031	-0.227***	-0.032	-0.033	0.038	0.069	-0.159**	0.091**	-0.007	-0.188**	-0.098*	-0.073	-0.043
	(0.055)	(0.043)	(0.070)	(0.028)	(0.078)	(0.054)	(0.059)	(0.058)	(0.055)	(0.062)	(0.041)	(0.048)	(0.083)	(0.057)	(0.061)	(0.044)
Time spent decile 4	0.027	0.079*	-0.077	0.027	-0.193**	0.061	0.044	0.014	0.027	-0.306***	0.071*	0.041	-0.136*	-0.164***	-0.013	-0.023
	(0.056)	(0.045)	(0.072)	(0.029)	(0.076)	(0.055)	(0.059)	(0.058)	(0.059)	(0.060)	(0.041)	(0.052)	(0.079)	(0.061)	(0.060)	(0.041)
Time spent decile 5	0.001	-0.010	-0.089	0.022	-0.124	0.049	-0.019	0.050	0.041	-0.277***	0.036	-0.019	-0.170**	-0.189***	0.025	-0.037
	(0.055)	(0.042)	(0.071)	(0.028)	(0.077)	(0.055)	(0.058)	(0.057)	(0.060)	(0.061)	(0.041)	(0.052)	(0.083)	(0.057)	(0.061)	(0.045)
Time spent decile 6	0.109**	-0.011	-0.066	0.023	-0.194**	0.017	-0.047	0.054	0.016	-0.380***	0.050	0.047	-0.285***	-0.228***	-0.109*	0.031
	(0.055)	(0.043)	(0.071)	(0.028)	(0.077)	(0.055)	(0.058)	(0.056)	(0.062)	(0.061)	(0.043)	(0.050)	(0.080)	(0.059)	(0.061)	(0.046)
Time spent decile 7	0.123**	0.080*	-0.058	0.027	-0.196**	0.045	-0.009	0.009	0.056	-0.410***	0.077*	0.062	-0.291***	-0.321***	-0.034	-0.048
	(0.055)	(0.045)	(0.073)	(0.029)	(0.079)	(0.057)	(0.060)	(0.060)	(0.057)	(0.063)	(0.042)	(0.051)	(0.081)	(0.061)	(0.060)	(0.042)
Time spent decile 8	0.148***	0.003	-0.195***	-0.011	-0.187**	0.061	0.004	-0.032	0.057	-0.495***	0.077*	0.065	-0.290***	-0.265***	0.048	-0.058
	(0.057)	(0.045)	(0.071)	(0.028)	(0.078)	(0.055)	(0.060)	(0.059)	(0.058)	(0.062)	(0.042)	(0.053)	(0.082)	(0.059)	(0.061)	(0.043)
Time spent decile 9	0.107*	0.055	-0.151**	-0.011	-0.208***	0.131**	-0.017	0.098*	-0.016	-0.464***	0.148***	0.045	-0.232***	-0.397***	0.005	-0.087**
	(0.057)	(0.044)	(0.072)	(0.029)	(0.078)	(0.056)	(0.060)	(0.057)	(0.057)	(0.062)	(0.043)	(0.052)	(0.081)	(0.058)	(0.061)	(0.043)
Time spent decile 10	0.177***	0.110**	-0.114	-0.001	-0.173**	0.076	-0.061	-0.042	0.053	-0.482***	0.081*	0.120**	-0.248***	-0.366***	-0.014	-0.078*
	(0.056)	(0.044)	(0.072)	(0.029)	(0.078)	(0.056)	(0.060)	(0.059)	(0.060)	(0.063)	(0.042)	(0.051)	(0.080)	(0.061)	(0.061)	(0.043)
R-squared	0.08	0.09	0.11	0.11	0.12	0.14	0.09	0.10	0.21	0.21	0.15	0.14	0.12	0.31	0.10	0.11
Adjusted r-squared	0.01	0.01	0.03	0.03	0.02	0.07	0.02	0.02	0.15	0.14	0.08	0.08	0.01	0.25	0.03	0.03
Observations	1416	1399	1265	1265	1044	1505	1480	1284	1361	1340	1370	1387	952	1217	1477	1353
Contribution to adjusted r-squared:																
risk/patience and cognitive skills	0.00	0.07	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.02	0.00	0.08	0.01
demographics	0.00	0.03	0.01	0.00	0.00	0.01	0.01	0.01	0.00	0.02	0.03	0.01	0.02	0.00	0.01	0.00
time spent	0.01	0.03	0.00	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.10	0.00	0.01	0.00	0.00	0.01

* 0.10 ** 0.05 *** 0.01.Unit of observation is the individual. LHS variable is the b-factor shown in the column header. Models are OLS and also include the full set of controls in Table 2: age categories, gender, education categories, four race/ethnicity categories, state of residence, immigrant indicator, 17 income categories, 3 marital status categories, 4 household size categories, 5 work status categories, fixed state effects, and dummies for missing values associated with each variable. Sample size varies because missing b-factor observations are discarded. Patience is the average savings rate across the 24 CTB decisions; we also include a dummy for missing this variable. See Section 2 and the Data Appendix for detail on variable construction.

Table 7. Correlations between b-factors.

Factors	Pres bias $	Pres bias snack	Incons with GARP	Cert prem>0	Loss averse	Narr bracket	Ambig aver	Overcon level	Overcon precis	NBLLN	GF cold	EGB APR	EGB FV	Ltd Attent	Ltd Mem
Present-bias money:	1.00														
Present-bias snacks:	0.01	1.00													
Choice inconsistency with GARP:	0.02	0.09	1.00												
Preference for certainty:	0.05	-0.09	***0.13***	1.00											
Loss aversion/small-stakes risk aversion:	-0.01	-0.04	***-0.15***	0.03	1.00										
Narrow bracketing:	***0.13***	0.03	0.07	-0.06	***-0.21***	1.00									
Ambiguity aversion:	-0.02	0.00	-0.04	***-0.12***	0.03	-0.08	1.00								
(Over-)confidence in performance:	0.07	***0.15***	***0.17***	***0.11***	***-0.11***	0.04	-0.07	1.00							
Overconfidence in precision:	0.01	-0.01	***-0.09***	0.06	-0.01	***-0.09***	-0.03	***0.16***	1.00						
Non-belief in the law of large numbers:	0.01	-0.06	***0.22***	***0.12***	-0.01	***0.12***	-0.05	***0.16***	***-0.16***	1.00					
Gambler's fallacy:	0.02	0.04	***0.08***	***0.24***	-0.07	***0.11***	-0.06	***0.16***	-0.02	***0.34***	1.00				
Exponential growth bias, debt-side:	0.02	0.02	-0.02	0.01	-0.03	***-0.11***	-0.03	-0.07	0.02	0.00	-0.08	1.00			
Exponential growth bias, asset-side:	0.08	***0.10***	***0.17***	***0.12***	***-0.12***	***0.13***	***-0.08***	***0.75***	-0.03	***0.24***	***0.30***	-0.08	1.00		
Limited attention:	-0.01	0.06	0.02	-0.06	0.04	0.01	-0.01	0.02	-0.04	***0.10***	0.00	-0.02	***0.11***	1.00	
Limited prospective memory:	0.01	***0.19***	0.08	-0.05	0.02	0.03	-0.01	0.06	***-0.11***	***0.17***	0.06	-0.10	***0.13***	0.05	1.00

Tetrachoric correlations. Bold/italics indicate significance at 10% or better, and shaded cells are positive correlations: there are 27 such positive correlations and 11 negative correlations. Here we only include the "standard" directional biases. Relative to the full set of 17 behavioral factors in Table 1, here we exclude: 1) inconsistency w/r/t GARP + dominance avoidance, because it: a) is correlated with inconsistency w/r/t GARP by construction; b) is indicated by nearly everyone in our sample if we define our indicators based on any deviation from the neoclassical norm; 2) overconfidence in relative performance, since we do not have an absolute indicator of that factor, as detailed in the Data Appendix Section H.

Table 8. B-counts and B-tiles: summary statistics.

	Indicators	
	All	Standard
"B-count" = count of behavioral indicators (N=1511)		
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=1511)		
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=1511)		
median	1	1
mean	1.75	1.75
SD	2.60	2.60

Please see Table 1 and the Data Appendix for lists and descriptions of behavioral factors and indicators.

Table 9. Do B-counts and B-tiles help explain financial condition?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
B-count, all biases	-0.024***					-0.024***	-0.026***	
	(0.004)					(0.005)	(0.004)	
B-count, standard biases		-0.025***			-0.019***			-0.018**
		(0.004)			(0.007)			(0.008)
B-count, non-standard biases		-0.010						0.055
		(0.007)						(0.068)
B-tile, all biases			-0.025***					
			(0.005)					
B-tile, standard biases				-0.029***	-0.006			-0.041***
				(0.005)	(0.009)			(0.009)
B-tile, non-standard biases				-0.007				-0.098
				(0.008)				(0.076)
b-factor missing count	-0.008	-0.007	-0.003	-0.002	-0.004	-0.008	-0.019***	-0.036***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.004)	(0.004)
highest ed: some college or associates	-0.033*	-0.032*	-0.035**	-0.032*	-0.031*	-0.034*	-0.029*	
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.017)	
highest ed: bachelor's	0.008	0.010	0.007	0.009	0.010	0.008	0.016	
	(0.020)	(0.020)	(0.021)	(0.020)	(0.020)	(0.020)	(0.020)	
highest ed: graduate	0.022	0.021	0.023	0.022	0.022	0.022	0.024	
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	
1st PC risk attitudes (stdized)	-0.008	-0.009	-0.009	-0.009	-0.009	-0.008		
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)		
patience (stdized)	-0.002	-0.000	-0.001	0.001	0.001	-0.002		
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)		
fluid intell # correct	-0.000	-0.000	-0.003	-0.003	-0.001	-0.000		
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)		
numeracy # correct	-0.005	-0.005	-0.006	-0.006	-0.004	-0.005		
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)		
financial literacy # correct	0.022**	0.021**	0.023***	0.022**	0.022**	0.022**		
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)		
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)		
Time spent decile 2	-0.050	-0.048	-0.047	-0.043	-0.046	-0.050		
	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)		
Time spent decile 3	-0.091*	-0.087*	-0.091*	-0.085*	-0.085*	-0.091*		
	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)		
Time spent decile 4	-0.105**	-0.101*	-0.105**	-0.100*	-0.099*	-0.104**		
	(0.053)	(0.053)	(0.053)	(0.053)	(0.053)	(0.053)		
Time spent decile 5	-0.111**	-0.106**	-0.115**	-0.108**	-0.104*	-0.111**		
	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)		
Time spent decile 6	-0.084	-0.079	-0.084	-0.077	-0.076	-0.084		
	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)		
Time spent decile 7	-0.062	-0.058	-0.067	-0.061	-0.057	-0.062		
	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)		
Time spent decile 8	-0.095*	-0.093*	-0.098*	-0.094*	-0.093*	-0.095*		
	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)		
Time spent decile 9	-0.103*	-0.100*	-0.109*	-0.104*	-0.101*	-0.102*		
	(0.055)	(0.055)	(0.056)	(0.055)	(0.055)	(0.055)		
Time spent decile 10	-0.105*	-0.103*	-0.106*	-0.102*	-0.102*	-0.105*		
	(0.054)	(0.054)	(0.055)	(0.054)	(0.054)	(0.054)		
other controls?				yes, as detailed in notes				
R-squared	0.43	0.43	0.42	0.43	0.43	0.43	0.40	0.11
mean(LHS)				0.43				
Observations				1502				

* 0.10 ** 0.05 *** 0.01.Unit of observation is the individual. LHS variable is our summary statistic for financial condition: the proportion of positive financial indicators in Table 3. Models are OLS and also include the full set of controls in Table 2: age categories, gender, education categories, four race/ethnicity categories, state of residence, immigrant indicator, 17 income categories, 3 marital status categories, 4 household size categories, 5 work status categories, fixed state effects, and dummies for missing values associated with each variable. Patience is the average savings rate across the 24 CTB decisions; we also include a dummy for missing this variable. See Section 2 and the Data Appendix for detail on variable construction.

Table 10. Robustness.

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable or sub-sample:					
	Fin cond.	"Hard" outcomes	"Soft" outcomes	Time spent decs. 1-2	Time spent decs. 3-8	Time spent decs. 9-10
B-count, all biases		-0.022***	-0.027***	-0.024***	-0.025**	-0.022**
		(0.004)	(0.005)	(0.005)	(0.013)	(0.009)
B-count, math biases	-0.024***					
	(0.004)					
B-count, non-math biases	-0.021**					
	(0.009)					
b-factor missing count	-0.007	-0.015*	-0.024**	-0.008	-0.036	-0.006
	(0.009)	(0.009)	(0.012)	(0.011)	(0.026)	(0.020)
highest ed: some college or associates	-0.033*	-0.007	-0.003	-0.045**	-0.053	0.031
	(0.018)	(0.020)	(0.025)	(0.022)	(0.054)	(0.046)
highest ed: bachelor's	0.008	0.059***	0.081***	0.008	-0.094	0.047
	(0.020)	(0.023)	(0.029)	(0.026)	(0.069)	(0.053)
highest ed: graduate	0.022	0.055**	0.074**	0.015	-0.036	0.043
	(0.023)	(0.026)	(0.033)	(0.029)	(0.077)	(0.060)
1st PC risk attitudes (stdized)	-0.008	-0.005	-0.007	-0.001	0.022	-0.025*
	(0.006)	(0.007)	(0.009)	(0.007)	(0.026)	(0.015)
patience (stdized)	-0.002	0.000	-0.002	0.002	-0.023	-0.026
	(0.006)	(0.007)	(0.009)	(0.008)	(0.021)	(0.016)
fluid intell # correct	-0.000	0.000	-0.001	-0.002	0.011	-0.003
	(0.003)	(0.003)	(0.004)	(0.004)	(0.011)	(0.006)
numeracy # correct	-0.005	-0.007	-0.007	-0.013	-0.030	-0.007
	(0.013)	(0.015)	(0.019)	(0.017)	(0.046)	(0.030)
financial literacy # correct	0.022**	0.033***	0.036***	0.033***	0.022	-0.010
	(0.009)	(0.010)	(0.013)	(0.012)	(0.032)	(0.021)
exec attention # correct	0.000	0.000	0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Time spent decile 2	-0.049	-0.012	-0.019			
	(0.048)	(0.053)	(0.068)			
Time spent decile 3	-0.091*	-0.043	-0.053			
	(0.051)	(0.057)	(0.073)			
Time spent decile 4	-0.105**	-0.052	-0.085			
	(0.053)	(0.059)	(0.076)			
Time spent decile 5	-0.111**	-0.059	-0.073			
	(0.054)	(0.060)	(0.077)			
Time spent decile 6	-0.084	-0.040	-0.048			
	(0.054)	(0.060)	(0.077)			
Time spent decile 7	-0.062	0.001	0.001			
	(0.054)	(0.060)	(0.078)			
Time spent decile 8	-0.095*	-0.014	-0.017			
	(0.055)	(0.062)	(0.079)			
Time spent decile 9	-0.103*	-0.052	-0.062			
	(0.055)	(0.062)	(0.079)			
Time spent decile 10	-0.105*	-0.066	-0.085			
	(0.054)	(0.061)	(0.078)			
R-squared	0.43	0.51	0.50	0.45	0.71	0.58
Observations	1489	1489	1489	976	187	326

* 0.10 ** 0.05 *** 0.01. Unit of observation is the individual. LHS variable in Columns 1, 4, 5 and 6 is our summary statistic for financial condition: the proportion of positive financial indicators in Table 4. LHS in Column 2 calculates the proportion using "hard" financial outcomes: net worth, savings, stock market participation, lack of financial hardhip and retirement savings. LHS in Column 3 uses "soft" outcomes: self-assesed retirement savings adequacy, non-retirement savings adequacy, financial satisfaction and financial stress. Column 4 uses the subsample in the lowest 2 deciles by total survey time spent on the b-factor questions, Column 5 uses deciles 3-8, and Column 6 uses deciles 9-10

Models are OLS and also include the full set of controls in Table 2: age categories, gender, education categories, four race/ethnicity categories, state of residence, immigrant indicator, 17 income categories, 3 marital status categories, 4 household size categories, 5 work status categories, fixed state effects, and dummies for missing values associated with each variable. Patience is the average savings rate across the 24 CTB decisions; we also include a dummy for missing this variable. See Section 2 and the Data Appendix for detail on variable construction

Appendix Table 1. 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	
Executive attention	**0.36**	**0.19**	**0.22**	1
mean 65, SD 24,	*0*	*0*	*0*	
min 0, max 154	1352	1326	1355	1444

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 the 2 of the 6 questions popularized by Banks and Oldfield (2007).
Financial literacy measured using the three of the questions popularized by Lusardi and Mitchell (2014).
Executive attention measured using a two-minute Stroop test where respondents are instructed to answer as many questions correctly as they can.

Appendix Table 2. 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	1215	1133	1066	1222

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 3. Selected columns from Table 9, full results not showing state fixed effects.

	(1)	(2)
B-count	-0.024***	
	(0.004)	
B-tile		-0.029***
		(0.005)
b-factor missing count	-0.006	-0.013**
	(0.008)	(0.006)
female	-0.014	-0.014
	(0.013)	(0.013)
age 35-45	-0.009	-0.010
	(0.017)	(0.017)
age 46-54	0.008	0.009
	(0.019)	(0.019)
age>=55	0.052**	0.052**
	(0.021)	(0.021)
race=Black	-0.016	-0.013
	(0.020)	(0.020)
race=Other	-0.062***	-0.062***
	(0.020)	(0.020)
latino	-0.008	-0.011
	(0.018)	(0.018)
immigrant	0.063***	0.063***
	(0.021)	(0.021)
income 5k-7.5k	0.011	0.017
	(0.040)	(0.041)
income 7.5k-10k	0.015	0.022
	(0.038)	(0.038)
income 10k-12.5k	-0.008	-0.004
	(0.036)	(0.036)
income 12.5k-15k	-0.047	-0.040
	(0.038)	(0.039)
income 15k-20k	0.000	0.002
	(0.037)	(0.037)
income 20k-25k	0.033	0.040
	(0.035)	(0.035)
income 25k-30k	0.049	0.050
	(0.037)	(0.038)
income 30k-35k	0.043	0.050
	(0.034)	(0.035)
income 35k-40k	0.154***	0.158***
	(0.038)	(0.039)
income 40k-50k	0.111***	0.116***
	(0.035)	(0.035)
income 50k-60k	0.169***	0.172***
	(0.035)	(0.035)
income 60k-75k	0.199***	0.200***
	(0.035)	(0.035)
income 75k-100k	0.284***	0.289***
	(0.037)	(0.037)
income 100k-125k	0.287***	0.289***
	(0.038)	(0.038)
income 125k-200k	0.348***	0.353***
	(0.040)	(0.040)
income >=200k	0.402***	0.405***
	(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	1502	1502

Cont.'d

	(1)	(2)
marital status: separated/divorced/widowed	0.015	0.013
	(0.017)	(0.017)
marital status: never married	0.005	0.005
	(0.017)	(0.017)
# other hh members: 1	-0.027	-0.028*
	(0.017)	(0.017)
# other hh members: 2	-0.038**	-0.039**
	(0.018)	(0.018)
# other hh members: 3	-0.069***	-0.070***
	(0.020)	(0.021)
# other hh members: 4+	-0.031	-0.033
	(0.025)	(0.025)
highest ed: some college or associates	-0.031*	-0.034**
	(0.017)	(0.017)
highest ed: bachelor's	0.009	0.007
	(0.020)	(0.020)
highest ed: graduate	0.019	0.018
	(0.024)	(0.024)
work status: self-employed	-0.018	-0.015
	(0.023)	(0.023)
work status: not working	-0.028	-0.027
	(0.019)	(0.019)
work status: disabled	-0.083***	-0.086***
	(0.026)	(0.026)
work status: missing	-0.056	-0.059
	(0.057)	(0.057)
	-0.009	-0.009
1st PC risk attitudes (stdized)	(0.006)	(0.006)
	0.009	0.010
risk missing	(0.017)	(0.017)
	0.001	0.001
patience (stdized)	(0.006)	(0.006)
	-0.011	-0.017
patience missing	(0.027)	(0.027)
fluid intell # correct	-0.000	-0.000
	(0.003)	(0.003)
fluid intell score missing	-0.208**	-0.195**
	(0.083)	(0.081)
numeracy # correct	-0.004	-0.005
	(0.012)	(0.012)
numeracy score missing	-0.007	-0.006
	(0.043)	(0.043)
financial literacy # correct	0.023***	0.024***
	(0.008)	(0.008)
financial literacy score missing	0.093	0.100
	(0.072)	(0.071)
exec attention # correct	0.000	-0.000
	(0.000)	(0.000)
exec attention score missing	0.020	0.016
	(0.032)	(0.032)
10 quantiles of tsbfavg	0.001	0.001
	(0.002)	(0.002)
did not take module 352	0.009	0.004
	(0.046)	(0.046)

* 0.10 ** 0.05 *** 0.01.

Appendix Table 4. B-counts and individual components of financial well-being

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	net worth>0	retirement assets>0	owns stocks	spent < inc last 12 months	fin satis > median	adequate ret. Savings	adequate non-ret. Savings	no fin distress recent	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	1460	1475	1485	1496	1493	1492	1386	1497	1501

* 0.10 ** 0.05 *** 0.01.

Appendix Table 5. Outcome Measurement Survey Formatting

Variable	# of questions used	response options: # per q.	response options: orientation	response options: placement of one(s) indicating worse condition	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 bottom**	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	vertical	top	Spent more than income/Spent same as income/Spent less than income
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 3 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