

NBER WORKING PAPER SERIES

QUICKSAND OR BEDROCK FOR BEHAVIORAL ECONOMICS? ASSESSING  
FOUNDATIONAL EMPIRICAL QUESTIONS

Victor Stango  
Joanne Yoong  
Jonathan Zinman

Working Paper 23625  
<http://www.nber.org/papers/w23625>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
July 2017

This paper expands on, and supersedes, one line of inquiry in the working paper “The Quest for Parsimony in Behavioral Economics: New Methods and Evidence on Three Fronts.” Stango: UC Davis Graduate School of Management, [vstango@ucdavis.edu](mailto: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](mailto:joanne.yoong@gmail.com); Zinman: Dartmouth College, IPA, J-PAL, and NBER, [jzinman@dartmouth.edu](mailto:jzinman@dartmouth.edu). Thanks to Hannah Trachtman and Sucitro Dwijayana Sidharta for outstanding research assistance, and to the Sloan/Sage Working Group on Behavioral Economics and Consumer Finance, the Roybal Center (grant # 3P30AG024962), and the National University of Singapore for funding and patience. We thank Shachar Kariv and Dan Silverman for helping us implement their (with Choi and Muller) interface for measuring choice consistency, Charlie Sprenger for help with choosing the certainty premium elicitation tasks and with adapting the convex time budget tasks, Georg Weizsacker for help in adapting one of the questions we use to measure narrow bracketing, Julian Jamison for advice on measuring ambiguity aversion, Josh Schwartzstein for many conversations, and audiences at the Sage/Sloan Foundation, UCSD-Rady, DIW-Berlin, and the Aspen Consumer Decision Making Conference for comments on survey design. For comments on the paper we thank Stefano DellaVigna, Xavier Gabaix, Michael Haliassos, Paul Heidhues, Theresa Kuchler, Gautam Rao, Doug Staiger, Johannes Stroebel, Dmitry Taubinsky, and seminar and conference participants at Berkeley/Haas, the Boulder Conference on Consumer Financial Decision Making, the CEPR Network in Household Finance, CFPB, the CFPB Research Conference, Columbia GSB, Dartmouth, the Federal Reserve Bank of Philadelphia, National University of Singapore, NBER Law and Economics, NYU/Stern, the Research in Behavioral Finance Conference in Amsterdam, and UC-Davis. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Victor Stango, Joanne Yoong, and Jonathan Zinman. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Quicksand or Bedrock for Behavioral Economics? Assessing Foundational Empirical Questions  
Victor Stango, Joanne Yoong, and Jonathan Zinman  
NBER Working Paper No. 23625  
July 2017  
JEL No. D03,D14,D8,D9,E03,G02

### ABSTRACT

Behavioral economics lacks empirical evidence on some foundational empirical questions. We adapt standard elicitation methods to measure multiple behavioral factors per person in a representative U.S. sample, along with financial condition, cognitive skills, financial literacy, classical preferences and demographics. Individually, B-factors are prevalent, distinct from other decision inputs, and correlate negatively with financial outcomes in richly-conditioned regressions. Conditioning further on other B-factors does not change the results, validating common practice of modeling B-factors separately. Corrections for low task/survey effort modestly strengthen the results. Our findings provide bedrock empirical foundations for behavioral economics, and offer methodological guidance for research designs.

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

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

Joanne Yoong  
National University of Singapore  
21 Lower Kent Ridge Road  
Singapore 119077  
joanne\_yoong@nuhs.edu.sg

## **Quicksand or Bedrock for Behavioral Economics? Assessing Foundational Empirical Questions**

Victor Stango, Joanne Yoong, and Jonathan Zinman \*

July 2017

Behavioral economics lacks empirical evidence on some foundational empirical questions. We adapt standard elicitation methods to measure multiple behavioral factors per person in a representative U.S. sample, along with financial condition, cognitive skills, financial literacy, classical preferences and demographics. Individually, B-factors are prevalent, distinct from other decision inputs, and correlate negatively with financial outcomes in richly-conditioned regressions. Conditioning further on other B-factors does not change the results, validating common practice of modeling B-factors separately. Corrections for low task/survey effort modestly strengthen the results. Our findings provide bedrock empirical foundations for behavioral economics, and offer methodological guidance for research designs.

JEL Nos. D03, D14, D6, D8, D9, E03, G02

---

\* This paper expands on, and supersedes, one line of inquiry in the working paper “The Quest for Parsimony in Behavioral Economics: New Methods and Evidence on Three Fronts.” Stango: UC Davis Graduate School of Management, [vstango@ucdavis.edu](mailto: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](mailto:joanne.yoong@gmail.com); Zinman: Dartmouth College, IPA, J-PAL, and NBER, [jzinman@dartmouth.edu](mailto:jzinman@dartmouth.edu). Thanks to Hannah Trachtman and Sucitro Dwijayana Sidharta for outstanding research assistance, and to the Sloan/Sage Working Group on Behavioral Economics and Consumer Finance, the Roybal Center (grant # 3P30AG024962), and the National University of Singapore for funding and patience. We thank Shachar Kariv and Dan Silverman for helping us implement their (with Choi and Muller) interface for measuring choice consistency, Charlie Sprenger for help with choosing the certainty premium elicitation tasks and with adapting the convex time budget tasks, Georg Weizsacker for help in adapting one of the questions we use to measure narrow bracketing, Julian Jamison for advice on measuring ambiguity aversion, Josh Schwartzstein for many conversations, and audiences at the Sage/Sloan Foundation, UCSD-Rady, DIW-Berlin, and the Aspen Consumer Decision Making Conference for comments on survey design. For comments on the paper we thank Stefano DellaVigna, Xavier Gabaix, Michael Haliassos, Paul Heidhues, Theresa Kuchler, Gautam Rao, Doug Staiger, Johannes Stroebel, Dmitry Taubinsky, and seminar and conference participants at Berkeley/Haas, the Boulder Conference on Consumer Financial Decision Making, the CEPR Network in Household Finance, CFPB, the CFPB Research Conference, Columbia GSB, Dartmouth, the Federal Reserve Bank of Philadelphia, National University of Singapore, NBER Law and Economics, NYU/Stern, the Research in Behavioral Finance Conference in Amsterdam, and UC-Davis.

Behavioral economics (BE) is in the zeitgeist. Researchers have identified a panoply of behavioral biases that could affect decisions and outcomes, typically starting with lab-based direct elicitation methods that measure a bias using stylized tasks. Micro- and macroeconomists alike have incorporated behavioral specifications of preferences, expectations, and problem-solving approaches into models. Policymakers now invoke BE as a basis for specific rulemakings and broader regulatory authority. “Nudge units” and other centers of applied behavioral social sciences are multiplying.<sup>1</sup>

Yet many economists remain skeptical of the economic importance and usefulness of BE.<sup>2</sup> This skepticism is warranted by the lack of nationally representative empirical evidence on some foundational questions.

First, are behavioral biases prevalent enough in broad populations to merit thorough and continued examination? Perhaps not, if biases are truly anomalous. Yet our literature reviews of direct elicitation work on 17 prominently-researched behavioral factors (“B-factors”) indicates that many biases have only been measured in non-representative samples. 10 of the 17 B-factors we consider lack even a single prevalence estimate from nationally representative U.S. data.

Second, does cross-sectional heterogeneity in B-factors shows meaningful links to individual decisions and outcomes? Perhaps not, if learning, market forces, delegation, and/or policy protections neutralize behavioral tendencies. Yet our literature reviews again find scant prior evidence on the question: 14 of 17 B-factors lack even a single estimate, from U.S. representative data, of links between field outcomes and a behavioral bias conditional on cognitive skills and other “classical factors” such as risk aversion, patience, age, gender, income, education and others.

Third, are B-factors distinct from each other, in terms of their relationships to field behavior and outcomes? Not necessarily, if bias on one dimension is correlated with bias on others, within-person.<sup>3</sup> Yet most prior work in BE, both theory and empirics, analyzes one or two B-

---

<sup>1</sup> For reviews of behavioral economics research and/or applications see, e.g., Akerlof (2002), Chetty (2015), DellaVigna (2009), Driscoll and Holden (2014), Kahneman (2003), Koszegi (2014), Rabin (1998), and the Behavioral Science and Policy Association.

<sup>2</sup> For critiques of behavioral economics see, e.g., Fudenberg (2006) and Levine (2012).

<sup>3</sup> We do find within-person bias correlations in our sample (Stango, Yoong, and Zinman 2017), as do Dean and Ortoleva (2016) in a sample of university students.

factors in isolation from the many others. The validity of that practice, and the extent to which one can rely on findings from prior work, hinges on B-factor separability assumptions that have yet to be tested in representative data.<sup>4</sup>

And fourth, are B-factors distinct from other decision inputs? There are many dimensions to this question, and here we are primarily interested in two. One relates to research design: which other decision inputs should researchers consider when trying to identify relationships between B-factors and field outcomes? The second relates to the underpinnings of behavioral biases themselves: how well fit are they by other individual characteristics? Specifically, prior work often groups or links B-factors with other individual characteristics like cognitive skills, standard preferences, and demographics, but those approaches have questionable merit if behavioral factors capture information that measures of classical factors do not.

We provide new evidence on these foundational questions for 17 B-factors. One set of B-factors relates 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. 2016), and choice inconsistency (Choi et al. 2014). Other B-factors capture biased beliefs, biased perceptions, and behavioral decision rules: three varieties of overconfidence (Moore and Healy 2008), narrow bracketing (Matthew Rabin and Weizsäcker 2009), exponential growth biases (Stango and Zinman 2009; Levy and Tasoff 2016), statistical fallacies (Dohmen et al. 2009; Benjamin, Moore, and Rabin 2013; Benjamin, Rabin, and Raymond 2016), and limited attention/memory (Ericson 2011). We selected B-factors by drawing on recent direct elicitation papers in top economics and finance journals, consulting with seminar and conference audiences during the design phase of the project, and making some allowances for tractability.<sup>5</sup>

---

<sup>4</sup> For work and discussions re: interactions among behavioral factors and other challenges in behavioral modeling, see, e.g., the cites in footnotes 1-3, and also Benjamin et al. (2016), Ericson (forthcoming), Heidhues et al. (2016), and O'Donoghue and Rabin (1999).

<sup>5</sup> This paragraph cites the papers that had the greatest influence on our elicitation designs. With respect to drawing the line on what we did and did not seek to measure, some examples of tractability considerations are that we could not devise methods for eliciting projection bias or confirmation bias that seemed feasible given our budget constraint and other measurement priorities. If unconstrained we also would have elicited social preferences; given constraints this seemed like a natural line of demarcation, as in e.g., Gabaix (2014, 2017).

We can elicit a large set of B-factors because we streamline standard direct elicitation methods by shortening, simplifying and combining tasks/questions. Streamlining elicitations saves costs/time and allows us to construct an unusually rich, person-level dataset capturing behavioral tendencies, other decision inputs, and field choices/outcomes. Specifically, we measure 17 B-factors per person, together with classical factors— patience and risk attitudes, demographics that capture human capital and life-cycle inputs, cognitive skills including financial literacy— and financial choices/outcomes. Altogether we measure a broad set of B-factors, other decision inputs, and outcomes.

We implement our elicitations as part of two online survey modules administered to a nationally representative U.S. sample of 1400+ participants in RAND’s American Life Panel (ALP) in 2014-15. The two modules take about 60 minutes per respondent in total. Respondents are compensated for completing each module, per the ALP’s standard practice, and we offered additional compensation for only one of our many B-factor and cognitive tasks.

Measurement error concerns have shaped every step of this project, with no fewer than a dozen features of our research design and data helping us deal with classical or non-classical measurement error. Ex-ante, our design choices were influenced by findings that the presence or absence of marginal financial incentives need not change inferences about B-factors,<sup>6</sup> and by the success of previous work adapting lab elicitations for administration in surveys.<sup>7</sup> Additionally, our survey formatting discourages the production of mechanical relationships between B-factor measures and financial condition measures. Ex-post, we find nearly all respondents reporting that the elicitations are interesting, that standard internal consistency checks produce results comparable to prior work, and that B-factor prevalence does not vary with the amount of time respondents spend on elicitation tasks. Nor do we find unusually high B-factor prevalence, in cases where prior work provides comparable estimates. These patterns assuage the concern that lack of respondent effort will spuriously inflate B-factor prevalence. We moreover control for

---

<sup>6</sup> For recent evidence on whether financial incentives change behavioral parameter estimates see, e.g., Von Gaudecker et al. (2011) and Gneezy et al. (2015). Unpaid tasks (with hypothetical rewards) may even offer some conceptual advantages (Cohen et al. 2016), as may simpler/shorter tasks (Chuang and Schechter 2015).

<sup>7</sup> On behavioral biases see, e.g., Ashraf et al. (2006), Callen et al. (2014), Choi et al. (2014), and Gine et al. (forthcoming); on classical preferences, expectations, and decision rules see, e.g., Barsky et al. (1997), Dohmen et al. (2010, 2011) and Falk et al. (2015; 2015).

respondent survey effort—measured using time spent responding to B-factor elicitations—or downweight low-effort respondents in some specifications. Theory helps as well, by generating testable predictions—that certain directional biases (“standard” ones like over-confidence, and present-bias) should be more prevalent and predictive than others (“non-standard” biases like under-confidence, and future-bias). Those predictions find support in our data and are inconsistent with alternative interpretations of B-factors as two-sided “mistakes” or noise. And finally, our results linking B-factors to outcomes provide “proof is in the pudding” that classical measurement error does not fully obscure valuable signals from our B-factor data.

Further previewing our key results, on the first foundational question we find that most B-factors are indeed quite prevalent, with some deviation from the classical norm exhibited by at least 50% of the sample for 11 of the 17 B-factors. High prevalence is not simply an artifact of “trembles,” or low survey effort; we find substantial prevalence even when applying stricter standards for what counts as “behavioral,” or when discarding low-effort responses. Nor is high prevalence an artifact of our streamlined elicitations: we actually classify fewer people as behavioral than prior studies using comparable elicitation methods on representative U.S. samples, for 5 of the 7 B-factors with prior comparable studies.<sup>8</sup> Standard directional biases emphasized by prior literature (e.g., under-estimating exponential growth instead of over-estimating it) are indeed more prevalent in our data, for 6 of the 7 B-factors where we can capture bi-directional biases and there is a clear standard. Our main takeaways on the first foundational question are that B-factors are widespread enough in the general population to motivate continued scrutiny, and that our streamlined methods are useful for eliciting them.

Turning to the second foundational question, we find that cross-sectional heterogeneity in B-factors does in fact correlate with outcomes. As a threshold matter, we start by showing that cross-sectional heterogeneity exists, to a substantial degree. In cases where comparable prior estimates of B-factor (parameter) distributions exist, we find similar distributions. We also document that, for each B-factor, there is at least some prior theory or empirical evidence

---

<sup>8</sup> By our accounting, the 7 B-factors with a nationally representative prevalence estimate for the U.S. in prior work are: money discounting biases (Bradford et al. 2014; Goda et al. 2017), snack discounting biases (Barcellos and Carvalho 2014), loss aversion (Hwang 2016), narrow bracketing (Gottlieb and Mitchell 2015; Rabin and Weizsäcker 2009), ambiguity aversion (Dimmock et al. 2016), debt-side exponential growth bias (Stango and Zinman 2009, 2011), and asset-side EGB (Levy and Tasoff 2016; Goda et al. 2017). See Section 2-A and the Data Appendix for more details.

suggesting links between that B-factor and worse financial condition. We then regress a rich index of financial condition on each B-factor, conditional on measures of classical risk aversion and patience; cognitive skills including financial literacy; income, age, gender, education, and other demographics; and survey effort. As noted above, prior work lacks this sort of richly conditioned test, in nationally representative U.S. data, for 14 of the 17 B-factors.<sup>9</sup> Overall, we find that B-factors negatively correlate with financial condition, with economically significant magnitudes. And, as with prevalence, bi-directional biases exhibit the pattern predicted by theory: standard directional biases exhibit negative correlations, while non-standard biases do not. Down-weighting observations reflecting low task/survey effort strengthens the results, albeit modestly. Our main takeaway here is that B-factors do have economically substantial conditional correlations with field outcomes.

On the third foundational question, we provide new tests of whether B-factors are empirically distinct from each other for each of the 17 B-factors.<sup>10</sup> The findings are striking: links between single B-factors and financial condition are essentially invariant to whether one controls for (even the full set of) other B-factors or not. These findings suggest that theorists are on firm ground when modeling the influence of a single bias on (financial) decisions while abstracting away from other potential biases. And empiricists interested in the relationship between a particular bias and (financial) outcomes need not be overly concerned about omitting other B-

---

<sup>9</sup> By our accounting, the 3 B-factors with comparable estimates in prior work are: biased money discounting and asset-side exponential growth bias in Goda et al. (2017), and ambiguity aversion in Dimmock et al. (2016). See Section 3-B and the Data Appendix for more details.

<sup>10</sup> The only other paper we know of that estimates comparable conditional relationships between field outcomes and multiple behavioral factors, in a broadly nationally representative sample, is Goda et al. (2017), which does so for present-biased money discounting and exponential growth bias (i.e., for 2 of our 17 B-factors). But Goda et al. always include both B-factors in their regressions and hence does not address the question of whether inferences change with single vs. multiple B-factors in the model. Bruine de Bruin, Parker, and Fischhoff (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 field behavior. Gottlieb and Mitchell (2015) estimates the relationship between holding long-term care insurance and narrow bracketing in a representative sample of older Americans, and checks robustness to adding controls for loss aversion in tandem with adding many other variables. Hwang (2016) estimates relationships between insurance holdings and loss aversion (and the interaction between loss aversion and the Gambler's Fallacy), in a nationally representative sample, but does not control for cognitive skills and does not examine whether the loss aversion estimate is sensitive to the inclusion of a Gambler's Fallacy main effect. Tanaka et al. (2010) does lab-style elicitations for estimating loss aversion, present-bias, and probability-weighting for 181 Vietnamese villagers, and links those elicitations to survey data (on income, etc.), but considers each B-factor separately.



factors. Our main takeaways here are strong empirical support for the hypothesis that behavioral biases have distinct relationships with choices/outcomes, and pushback against criticism that behavioral economics is overly siloed.

On the fourth foundational question, various results suggest that our B-factor measures are distinct from classical factors. Removing classical factors from our regressions does not substantively affect point estimates on B-factors (this exercise also serves as a specification test *a la* Altonji et al. (2005)). B-factors are poorly fit by even our full set of covariates, with adjusted R-squareds below 0.10 in most cases, even after correcting for measurement error. We also briefly examine correlations between B-factors and other covariates, and find that strong correlations are the exception rather than the rule.<sup>11</sup> Our main takeaways here inform research design: researchers may not actually need a rich set of control variables to identify conditional correlations between field outcomes and behavioral biases. Conversely, measures of classical factors are likely poor proxies for measures of B-factors.

Overall, our results suggest that BE rests on empirical bedrock. B-factors are prevalent, conditionally correlated with outcomes as predicted by theory, empirically distinct from each other in ways that validate modeling one or few biases at a time, and largely distinct from other decision inputs. Our use of response times to diagnose and correct for measurement error could have broad applicability to survey and lab data. Our new direct elicitation methods open the door to collecting much more data, at lower cost, for probing these questions and otherwise describing behavioral biases, shaping models, and testing predictions.<sup>12</sup> Our findings on B-factor separability open the door to collecting such data one factor at a time, for example by adding B-factor elicitations, perhaps on a rotating basis, to large, established representative household surveys that should offer more statistical power for testing links between outcomes and B-factors. The trajectories of older literatures on consumer decision making inputs (e.g., on

---

<sup>11</sup> Among correlations between B-factors and other decision inputs, cognitive skills have been a particular focus of prior work. Our findings are in line with this prior work: while we tend to find some negative correlations between behavioral tendencies and fluid intelligence (Benjamin, Brown, and Shapiro 2013; Burks et al. 2009; Frederick 2005), those correlations are modestly-sized and do not hold across all B-factors (Cesarini et al. 2012; Li et al. 2013). See also Dohmen et al. (2010) on correlations between measures of classical preferences/attitudes and intelligence. We lack measures of non-cognitive skills like personality traits, but prior evidence finds weak if any correlations between those skills and B-factors (see e.g., Becker et al.'s (2012) review article and subsequent papers citing it). Hence *a priori* there is little concern that omitting non-cognitive skills will affect inferences about B-factors

<sup>12</sup> For some other uses of our existing data see Stango, Yoong, and Zinman (2017).

cognitive skills, on personality) suggest that such data will key for advancing basic and applied research in BE.

## **1. Research design**

In this section we describe our sample, survey design and elicitation methods, and empirical strategies, including those for diagnosing and dealing with measurement error.

### *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.

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 become publicly available after a period of initial embargo. This opens up opportunities for future work linking our data to other modules.

### *B. Our research design and sample*

Speaking broadly, our goal was to design 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 adhered to ALP standard practice of avoiding long surveys (based on staff findings that shorter surveys improve both response rates and quality), and allowed us to evenly disburse the more demanding tasks across the two modules. Per

standard ALP practice, we paid panelists \$10-\$20 per completed module. We discuss task-based incentives, and lack thereof, in Section 1-G.

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 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 (we measure two of our B-factors using the same elicitation, and so the max here is 16 instead of 17), with a median of 15 and a standard deviation of 2.9.

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 in the Data Appendix (Section L).

### *C. Measuring and describing behavioral factors: Elicitation methods and key antecedents*

We conduct elicitations of 17 potentially behavioral factors, 15 of which produce some measure of the intensity of deviation, and six of which produce data sufficiently rich to permit structural parameter estimation (such as an individual-level present- (or future-) biased discounting parameter in a beta-delta model). Given our goals of directly eliciting useful measures of B-factors without breaking the bank, we prioritize elicitation methods that have been

featured recently in top journals and were short and simple enough (or could be so modified) to fit into modules that would also allocate substantial survey time to measuring other decision inputs (Section 1-D) and outcome variables (Section 1-E).

Table 1 summarizes our list of B-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 B-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 each bi-directional B-factor we define a “standard” direction based on what has been more commonly observed or cited in prior work; e.g., present-biased discounting (with future-bias classified as non-standard and time-consistent as unbiased), over-confidence in performance (with under-confidence as non-standard and accurate assessment of one’s own performance as unbiased), and underestimating exponential growth (with over-estimating as non-standard and accurate estimation as unbiased).<sup>13</sup>

Table 2 shows that missing values from question-skipping/non-completion (“item non-response”) or nonsense answers (“responded, not usable”) are uncommon: 11 of 17 B-factors we have usable responses from more than 90% of the sample. Only two B-factors have usable data for less than 80% of the sample—with one of those cases due in part to an inherent limitation of the certainty premium elicitation rather than true item non-response.<sup>14</sup> Standard internal consistency checks for respondent understanding and diligence produce results comparable to prior work, as detailed in the Data Appendix.

---

<sup>13</sup> We use “B-factor” terminology in tandem with “behavioral bias” for two reasons. First, “factor” evokes an input to decision making, and much of our analysis considers behavioral factors alongside other factors. Second, as discussed above, many B-factors have two, mutually exclusive directional biases. The Data Appendix provides further details, for each B-factor, on: i) motivation for trying to measure it; 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*, with background on the distinctions between standard vs. non-standard directional biases where applicable, at different cutoffs for classifying a deviation from the classical norm as behavioral; vi) descriptions of the *magnitude* and *heterogeneity* of behavioral deviations, including descriptions of the distribution and—where the data permit—estimates of key parameters used in behavioral models; and vii) estimates of conditional correlations with financial outcomes, including particular components of our financial condition index that have particularly strong links to a given B-factor per theory.

<sup>14</sup> Only those subjects who switch at some point on both multiple price lists identify a certainty premium or discount. See the Data Appendix Section D for details.

Using our data to estimate B-factor prevalence, distributions, and structural model parameters is straightforward, and we undertake those exercises in Sections 2, 3-A, and the Data Appendix.

#### *D. Measuring classical factors and survey effort*

Our modules also elicit rich measures of cognitive skills, demographics, patience, and risk attitudes. We refer to these as “classical” factors because they are measures of human capital, life cycle considerations and preference parameters that plausibly affect decisions and outcomes in any model. These factors serve as control variables in regressions of outcomes on B-factors. Measures of survey effort serve the same purpose, and also help correct for measurement error (Sections 1-G and 3-C). Table 3 summarizes the classical factors and survey effort measures, and this sub-section provides additional details.

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,<sup>15</sup> labeled as such and popularized in economics since their deployment in the 2002 English Longitudinal Study of Ageing.<sup>16</sup> 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).<sup>17</sup> 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

---

<sup>15</sup> “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.

<sup>16</sup> Banks and Oldfield (2007) interpret these as numeracy measures, and many other studies use them as measures of financial literacy (Lusardi and Mitchell 2014).

<sup>17</sup> “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.

1991).<sup>18</sup> Each time the subject chooses an answer that action completes what we refer to as a “round.”<sup>19</sup> 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.

The four test scores—fluid intelligence, numeracy, financial literacy, and Stroop—have pairwise correlations ranging from 0.19 to 0.45. Some (most) of our respondents had taken the number series test (financial literacy quiz) in a prior module; we have confirmed robustness to incorporating information from prior modules.

We also elicit four standard measures of classical risk attitudes/preferences, although we end up using only two of them in the empirics. 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.<sup>20</sup> We use this to construct an integer scale from 1 (most risk tolerant) to 6 (most risk averse). The second is from Dohmen et al. (2010, 2011): “How do you see yourself: Are you generally a person who is fully prepared to take financial risks” (100 point scale, we

---

<sup>18</sup> Our version displays the name of a color on the screen (red, blue, green, or yellow) and asks the subject to click on the button corresponding to the color the word is printed in (red, blue, green, or yellow; not necessarily corresponding to the color name). Answering correctly tends to require using conscious effort to override the tendency (automatic response) to select the name rather than the color. The Stroop task is sufficiently classic that the generic failure to overcome automated behavior (in the game “Simon Says,” when an American crosses the street in England, etc.) is sometimes referred to as a “Stroop Mistake” (Camerer 2007).

<sup>19</sup> 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.

<sup>20</sup> 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).

transform so that higher values indicate greater risk aversion).<sup>21</sup> Those first two scales are correlated with each other (0.38) and we use them as covariates in our regressions. The third and fourth measures are the switch points on the two multiple price lists we use to elicit the certainty premium (Data Appendix Section D). Those latter two measures are correlated with each other and with the first two, but we find empirically that they add no explanatory power to models explaining either B-factors or financial outcomes. We have also used the first principal component of the four risk aversion measures in our empirics, but find that the Dohmen et al. and Barsky et al. measures included separately have more explanatory power.

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

The ALP tracks and record survey response time, screen-to-screen, and we use this to construct measures of survey effort. Specifically, for each respondent we measure time spent on each B-factor, on all B-factor elicitation, and on the entirety of the two modules.

Our other source of control variables is the ALP's standard set of demographic variables (such as gender, age, income, education, household size, etc.), which are collected when a panelist first registers, then refreshed quarterly and merged onto each new module.

#### *E. Measuring financial choices and outcomes: A new index of financial condition*

Our instruments also elicit rich data on financial choices and outcomes. We construct nine indicators of financial condition from 15 survey questions, 14 of which are in module 315. The questions elicit information on net worth, financial assets, recent savings behavior, household distress as measured by recent events (missed housing utility payments, forced moves, postponed medical care, hunger), and summary self-assessments of savings adequacy, financial satisfaction and financial stress. 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). 1,508 of our 1,515 respondents provide data we can use to construct one

---

<sup>21</sup> We also elicit Dohmen et al.'s general risk-taking scale, which is correlated 0.68 with the financial scale.

or more of the nine indicators. The median respondent supplies the full nine, with a mean of 8.8 and standard deviation of 0.6.

Table 4 shows the financial indicators, their sample proportions, and correlations between them. In each case “1” indicates plausibly better financial condition. Our indicators include both stocks and flows. They include five “hard” quantitative measures: positive net worth, positive retirement assets, owning stocks, spending less than income in the last 12 months, and not experiencing any of four specific objective indicators of severe financial distress in the last 12 months. They also include four “soft” subjective and self-assessed measures of financial condition: financial satisfaction (1=above median), financial stress (1=below median), and viewing retirement or non-retirement savings as “adequate” or better.

As Table 4 shows, these nine indicators are positively correlated with each other, which is consistent with the hypothesis that they are each signals of some underlying construct like “financial health,” or “financial security,” although we use the more purely descriptive label of “financial condition.” Our main outcome takes the individual-level mean of these nine indicator variables. The sample mean of this summary measure is 0.43, meaning that the average respondent exhibits 4 of our 9 indicators of better financial condition.

#### *F. Regression specifications*

Our primary empirical exercise, beyond describing B-factor prevalence/distributions, is estimating how cross-individual variation in B-factors correlates with financial condition. Our main empirical specification does so by regressing (OLS), separately for each B-factor, financial condition on the B-factor and a full set of controls:

$$(1) \text{FinCond}_i = \alpha + \beta_1 \text{Bfactor}_{s_i} + \beta_2 \text{Bfactor}_{ns_i} + \beta_3 \text{Bfactor}_{miss_i} + \gamma X_i + \varepsilon_i$$

In these models  $i$  indexes individuals, and *FinCond* is the summary measure of financial condition described in Section 1-E. The next two variables are B-factor *indicators* for standard bias and non-standard bias (if applicable). The omitted category is unbiased/classical, unless noted otherwise. We start with indicators because they provide a straightforward way to summarize the data and check theoretical predictions re: standard vs. non-standard biases. We also use measures that capture information on the magnitudes of behavioral deviations. One approach standardizes the scale of deviations across B-factors, by calculating each individual’s



percentile location (for both the standard and non-standard directional biases, where applicable) in the distribution of each B-factor.  $X$  contains the full set of covariates for classical factors and survey effort described above and in Table 3. In cases where responses for a particular variable are missing we include a “missing” dummy for that observation and variable. In all we have over 100 control variables, many of them categorical, derived from up to 33 underlying measures.

### *G. Measurement error and other econometric concerns*

Classical measurement error in B-factors could bias their coefficients in regressions like equation (1) above, and lead to underestimates of correlations between B-factors and classical factors. More complicated measurement error structures, including non-classical and/or correlated measurement error, could affect inferences about each of our foundational questions in various ways.<sup>22</sup> Many features of our research design and data help us deal with measurement error in B-factors and other variables. We summarize them here and flesh out many of them in subsequent sections.

Describing and stratifying on respondent task/survey effort helps diagnose and deal with various measurement error concerns. B-factor prevalence does not vary with the amount of time respondents spend on elicitation tasks (Section 2-B), obviating concerns that “trembles” or low task engagement leads to upward bias in B-factor prevalence. Short response times are correlated with non-response, so we control for non-response in the empirics. Our survey user interfaces do not make “bad” financial condition an easier response than “good” financial condition (Appendix Table 1), so low survey effort does not spuriously generate poor financial condition (Appendix Table 2), or a spurious correlation between B-factors and financial condition.

We use survey effort (response time) as a control in our empirics, nonetheless. In specifications like equation (1) above, we include the respondent’s decile of time spent responding to the elicitation for the included B-factor. We have also included deciles of total time spent on all B-factor questions, and all questions overall, moving the control from B-factor elicitation effort more narrowly to survey effort more broadly.

---

<sup>22</sup> Despite its potential importance, there has been surprisingly little work on diagnosing and correcting for measurement error in direct elicitation data, as Gillen et al. (2017) discuss. See also, e.g., Beauchamp et al. (forthcoming) and Dean and Ortoleva (2016). For a general reference see, e.g., Buonaccorsi (2010).

We also use survey effort measures to implement more direct measurement error corrections. Respondents with shorter response times exhibit relatively weak correlations between B-factors and other variables including outcomes (Appendix Table 3). In a series of corrections, we weight observations in the regressions of equation (1) by survey time spent, specified various ways for robustness. This weighting approach follows a measurement error literature about using auxiliary data on observation “reliability,” where an observation’s reliability is inversely related to the degree of regression error associated with that observation. Reliability-weighting is a preferred approach when one is concerned about measurement error in many variables, not just the variable(s) of interest, and is equivalent to correcting for a known form of heteroscedasticity (Fuller 2009). We find that such corrections modestly strengthen the statistical relationships between financial condition and B-factors, and improve the fit of the regressions overall.

Several other features of the data offer further reassurances that our elicitations produce useful information. Standard internal consistency checks for respondent understanding and diligence produce results comparable to prior work (Data Appendix). Nor do we find unusually high B-factor prevalence relative to prior work, in cases where prior work provides comparable estimates (Section 2-A). These patterns further assuage the concern that our streamlined elicitations spuriously identify B-factors. Our regression results linking B-factors to outcomes provide “proof in the pudding” that our elicitations provide useful information.

The data also reveal patterns predicted by theory that run counter to what one would expect if our B-factors capture only noise or unmeasured aspects of classical factors. Purely noisy B-factor measures should produce symmetric prevalence of standard and non-standard biases. Noisy B-factors also should have weak or zero conditional correlations with outcomes. In contrast, theories predict asymmetry: standard biases should be more prevalent, and more negatively associated with outcomes, than non-standard biases. We generally do find these asymmetries in the data (Sections 2-A and 3-A).

Some readers may be concerned that 16 of our 17 B-factor elicitations are not incentivized on the margin. We had ex-ante reasons for this design choice, as we faced an especially large opportunity cost of incentivizing, in the sense that our research questions focus on multiple B-factors and their links to outcomes. Paying task incentives would have meant measuring fewer B-factors, on a smaller sample, thereby sacrificing breadth and likely statistical power.

Moreover, prior evidence suggests that paid vs. unpaid tasks do not necessarily change inferences about behavioral factors in large representative samples; for recent evidence see, e.g., Von Gaudecker et al. (2011) and Gneezy et al. (2015). Unpaid tasks (with hypothetical rewards) may even offer some conceptual advantages (Cohen et al. 2016), as may simpler/shorter tasks (Chuang and Schechter 2015). Ex-post, it is reassuring that only 3% of respondents found our modules uninteresting.<sup>23</sup> This finding pushes against the hypothesis that paying financial incentives is important because subjects find direct elicitation tasks unpleasant.

Beyond those measurement error concerns, we deal with two other unobserved heterogeneity issues directly. One might wonder whether B-factors are distinct, as we treat them by initially estimating equation (1) separately for each B-factor. We examine that directly in Section 4, by asking how the results change if we include all B-factors simultaneously in a single model. To preview the results, we find that the “all B-factors” model yields results almost identical to those from the individual B-factor regressions.

Another unobserved heterogeneity issue relates to whether B-factors reflect omitted components of other classical factors, such as cognitive skills, classical preferences, education or others. Section 5 present some specification tests along the lines of Altonji et al. (2005), changing the set of covariates to see whether that significantly affects point estimates on the B-factor coefficients. We also regress the B-factors themselves on *all* of our covariates, to estimate how well-explained are B-factors by other observable characteristics. Both sets of results suggest that our B-factor measures capture distinct influences rather than unobserved heterogeneity in classical factors.

Finally, in Section 3-D we consider how reverse causality would affect the interpretation of equation (1). We view any reverse causality as interesting in its own right but unlikely to occur.

## **2. B-factors are prevalent**

Here we address the first of our four foundational empirical questions, whether B-factors are prevalent enough in a broad population to merit continued intensive study.

---

<sup>23</sup> Each ALP survey ends with “Could you tell us how interesting or uninteresting you found the questions in this interview?” and roughly 90% of our sample replies “Very interesting” or “Interesting,” with only 3% replying “Uninteresting” or “Very uninteresting.”

### *A. B-factors: Prevalence and comparisons to prior work*

Are B-factors prevalent in a broad population? As noted at the outset, our literature reviews turned up U.S. population estimates for only 7 of the 17 B-factors we consider. The main question here is qualitative: does the overall pattern of evidence suggest that behavioral tendencies are more than just isolated anomalies?

Table 2 shows that our B-factors are indeed prevalent.<sup>24</sup> Some deviation from classical norms is indicated by at least 50% of the sample for 11 of the 16 factors for which we can estimate prevalence. In the Data Appendix, we impose more-stringent thresholds on what indicates a behavioral bias, counting only economically large deviations (defining “large” in various ways) from classical norms, and find that most B-factors remain prevalent (e.g., exhibited by >20% of the population).

The “standard” directional bias emphasized by prior literature is indeed more prevalent in our data, in 6 of the 7 B-factors where we capture bi-directional biases and there is a clear standard.<sup>25</sup> This pattern is consistent with theory but inconsistent with measurement error stories. Purely noisy B-factor measures should show symmetry between standard and non-standard biases. As an example, if people are unable to perform exponential calculations in their heads, there is little a priori reason to believe that random math mistakes would lie unequally in positive and negative directions. But in fact, and consistent with prior work, we find that under-estimation of compound growth or decline is far more common than over-estimation.

Appendix Table 4 shows that our estimates of prevalence are in line with prior findings, for the nine B-factors for which we could find prior studies on nationally representative samples. Seven of these B-factors were estimated in prior work using comparable elicitations on U.S. samples, and in five of these cases we find lower prevalence (time-inconsistent money discounting, loss aversion, narrow bracketing, and both debt-side and asset-side exponential growth biases). Our streamlined elicitations need not inflate prevalence estimates.

---

<sup>24</sup> Results are basically unchanged if we use the ALP’s population weights.

<sup>25</sup> Both Gambler’s Fallacies—hot-hand and cold-hand—have attracted substantial researcher attention, and so for the purposes of estimating prevalence we do not think there is a clear standard directional bias. For the purposes of estimating links to financial condition, we focus more on the hot-hand bias, for reasons detailed in the Data Appendix.

*B. Does B-factor prevalence reflect low survey effort?*

The comparability of our B-factor prevalence with prior work notwithstanding, here we discuss possible relationships between elicitation measurement error and our measured prevalence. Do respondents rush through the survey with less care than one would hope, leading us to classify economically harmless “trembles” as behavioral biases?

As a threshold matter, if time spent responding to B-factor questions is a valid proxy for survey effort, it should be the case that the quickest times, at least, are strongly correlated with item non-response. Figure 1 and Appendix Table 5 confirm this relationship. Respondents in the bottom decile of time spent responding to B-factor elicitations supply substantially fewer usable B-factor responses, with an average proportion of missing B-factors of 0.29 compared to <0.10 in deciles 4 through 10.

However, among the usable data on B-factors, there is no clear relationship between B-factor prevalence and survey effort at the respondent level, as shown in Figure 2, which plots the average (Number of behavioral biases indicated)/(Number of non-missing B-factors) for each decile of survey effort. Appendix Table 6 breaks this out for each B-factor, and reveals that different B-factors have different and often non-monotonic relationships between prevalence and survey effort. Thus, among usable responses, quicker responses are *not* necessarily more likely to be classified as behavioral.

Putting the findings together, it seems that while non-response is a function of survey effort, prevalence conditional on providing a response does not vary systematically with survey effort. These findings are consistent with the abovementioned finding that 90% of respondents characterize our modules as “interesting” or “very interesting.” Our elicitations seem to produce attentive and thoughtful responses for all but a small number of very low effort respondents, many of whom simply fail to supply data instead of (spuriously) responding in a way that we would classify as behavioral.

Our regressions will control for B-factor missing data, and control for or weight by survey effort, to address measurement error concerns more formally.

### 3. B-factors exhibit important links to field choices and outcomes

Here we address the second of our four foundational empirical questions, whether cross-sectional heterogeneity in B-factors helps us understand field choices and outcomes. Theories of each B-factor predict negative relationships between each (standard direction) behavioral bias and field outcomes (see Data Appendix for details), yet these predictions have only been tested for 3 of our 17 B-factors in representative U.S. data. We perform tests for each of the 17 B-factors, and for B-factors and directional biases as a whole.

#### *A. B-factor cross-sectional heterogeneity is substantial*

As a threshold matter, we start by noting the substantial heterogeneity in behavioral biases across individuals. This is evident on both extensive and intensive margins. Our prevalence estimates show the extensive margin heterogeneity: for most B-factors, many individuals exhibit a bias, and many do not. The Data Appendix details the intensive margin heterogeneity, which also tends to be substantial, and congruent with distributions estimated in prior work. Six of our elicitations are rich enough to permit structural estimation of parameters from models used in prior work, and we find very similar parameter estimates for the closest available comparisons: time-inconsistent money discounting and GARP violations (Data Appendix Sections A and C).

#### *B. B-factor cross-sectional heterogeneity is conditionally correlated with financial condition*

We now turn to our main test of whether B-factors are meaningfully linked to field choices and outcomes: equation (1). As we detail in the Data Appendix, for each B-factor there is theory, and in some cases prior empirical work, motivating the hypothesis that the (standard) behavioral bias will induce sub-optimal choices and hence worse financial condition. Yet prior work on 14 of our 17 B-factors lacks even a single estimate of the relationship between a field outcome and a B-factor, conditional on classical factors including cognitive skills, in representative U.S. data.

Table 5 presents estimates of equation (1), regressing our rich measure of financial condition on B-factor indicators and our full set of controls from Table 3, with one regression per B-factor per column. We suppress coefficients on many of the 100+ control variables, focusing on variables of interest. Reading within a column, any applicable standard bias vs. non-standard bias

distinction is denoted in Table 1 and detailed in the Data Appendix.<sup>26</sup> The omitted B-factor category is unbiased except for Certainty Premium and NBLLN, which each have the non-standard bias as the omitted category because we lack unbiased responses.

Table 5 supports the hypothesis that B-factors are negatively conditionally correlated with financial condition, in ways predicted by behavioral theories. 6 (5) of the 17 standard bias coefficients have p-values less than 0.10 (0.05). Tests across all 17 coefficients, or the 11 coefficients with p-values less than 0.10, estimate an average standard bias coefficient of -0.11 to -0.28 depending on specification, with a p-value below 0.01 in 5 of 6 specifications (Table 6). In contrast, the nonstandard bias coefficient has a p-value below 0.10 in only one of the six cases where we can identify correlations on bi-directional biases, and that one case has a positive sign.

16 of the 17 standard bias coefficients have the expected negative sign, and joint sign tests confirm the hypothesis that the average coefficient is negative (Table 6). In contrast, we cannot reject that the non-standard bias coefficients are centered on zero. Moreover, the standard coefficient is *more negative* than the non-standard one in every case, although for the most part the estimates are too imprecise to reject equality. To the extent one can infer asymmetry between standard vs. non-standard biases, this pattern is consistent with theory and inconsistent with plausible measurement error structures. Purely noisy B-factor measures should produce symmetric (zero) conditional correlations between outcomes and standard vs. non-standard biases. Likewise, B-factor measures that capture unmeasured aspects of classical factors should produce symmetric (weakly negative) conditional correlations between outcomes and standard vs. non-standard biases.<sup>27</sup>

Focusing on specific standard biases, the ones with p-values less than 0.10 are: overconfidence in relative performance, limited attention, limited memory, present-biased money discounting, underestimating exponential growth on future values, and ambiguity aversion. The latter three are the only three among our 17 to have been considered in comparable prior work,

---

<sup>26</sup> Even though both Gambler's Fallacies have generated substantial research, to the point where we are a bit agnostic about which if either bias—hot hand vs. cold hand—should be considered standard, when forced to choose for the purposes of organizing tables we label the hot hand standard, as this directional bias has stronger conceptual links to arguably more harmful financial decisions (see the Data Appendix for discussion).

<sup>27</sup> Looking at the control variables themselves in Table 5, being near retirement age, having greater income and financial literacy, and financial risk aversion are all strongly conditionally correlated with financial condition in the expected directions. The fact that other covariates exhibit weaker statistical links in this specification may not mean much, given the possibility that we are over-controlling.

with that work finding statistically significant negative correlations with financial outcomes in all three cases (Dimmock et al. 2016; Goda et al. 2017).

Magnitudes vary but one can easily interpret them since the RHS variables are indicators, the LHS variable is scaled on  $[0, 1]$  and the mean of the LHS is 0.43. For example, in the first column, present-biased money discounting is associated with a reduction in financial condition of 0.038, about a 9% decline from the base. The other standard bias coefficients with p-values below 0.10 imply associated reductions in financial condition ranging from 6% to 25%.

On balance, we infer that behavioral biases are conditionally correlated with financial condition, to economically and statistically significant extents.<sup>28</sup> Results in Sections 4 and 5 suggest that these correlations are not biased by unobserved correlations (in measurement error) between a single B-factor and other B-factors, or between B-factors and (unmeasured aspects of) other covariates. Before detailing those results, we show that a more direct approach to dealing with measurement error sharpens inferences on B-factors, albeit modestly.

### *C. Using survey time/effort to correct for measurement error*

As outlined in Section 1-G, we correct for measurement error by weighting observations in the regressions of equation (1) by survey effort, with effort specified various ways for robustness. To motivate our approach, recall Appendix Table 3, which shows lower univariate correlations between B-factors and other key variables (financial condition, education, cognitive skills and income) in the first decile of survey effort relative to other deciles. Correlations are higher in the 2<sup>nd</sup>-5<sup>th</sup> deciles, higher still in the 6<sup>th</sup>-9<sup>th</sup> deciles, and then decline in three of the four cases in the highest decile (per discussions with ALP staff, we have learned that the highest decile likely includes respondents who take a break from the survey without logging out). We infer that measurement error affects the informativeness of some combination of B-factor measures and other key variables. Recall that prevalence does not seem affected by survey effort (Section 2-B). Putting the two results together, it seems that while “rushing” does not affect the level of B-factor prevalence, it may affect the signal-to-noise ratio in various measurements.

Table 6 shows how estimates of equation (1) change after weighting by different definitions and functional forms of survey effort. Each cell reports the coefficient on the B-factor indicator

---

<sup>28</sup> We find similar results when replacing the indicators for standard and non-standard biases with their linear percentiles (Appendix Table 7).



in the row label (the standard bias indicator where applicable), from a single regression. Column 1 reproduces our result on each B-factor (standard bias where applicable) from Table 5 row 1. Column 2 weights each observation by  $\log(\text{respondent's time spent responding across all B-factor elicitations})$ . Results are quite similar to Column 2. Column 3 weights by decile of time spent across all B-factor elicitations. Given extrema in both tails of time spent and the bit of nonlinearity in Appendix Table 3, we view this as a more reasonable functional form assumption. The coefficients tend to be a bit more negative, and the fit improves by two percentage points. Column 4 weights by the decile of B-factor-specific survey time spent; e.g., when time-inconsistent money discounting is the B-factor we weight by the decile of time spent reading instructions and responding to the money discounting elicitation. The results are similar to those from the other specifications, with perhaps the only moderately noteworthy difference being that two additional B-factors now have p-values below 0.10.

The overall pattern is that down-weighting very rapid survey responses modestly improves model fit, and also leads to modestly larger and more significant point estimates on the B-factors. Other weighting specifications produce similar results. What we take from this is twofold. First, measures of survey effort may be useful in diagnosing and correcting for the measurement error that afflicts elicitation and testing across a broad range of fields (not just behavioral economics). And second, our results are robust to these corrections, which (along with other results) encourages us about the validity of our elicitations.

#### *D. Reverse causality?*

We also consider whether B-factor correlations with financial condition 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 biases and decisions. In any case, we surmise that reverse causality is unlikely to drive our inferences. Theoretically, the sign of the bias introduced by any reverse causality is unclear; it could either understate or overstate the importance of behavioral biases. Empirically, the case for reverse causality depends on the existence of something that has not been found: substantial instability in B-factors, within-

person. Prior work suggests that measured instability is largely due to measurement error and not true changes in B-factors.<sup>29</sup>

#### **4. B-factors are distinct from each other re: relationships to choices and outcomes**

We now turn to our third foundational empirical question, whether B-factors have distinct relationships with field choice and outcomes. The standard practice of behavioral economics is to implicitly assume the answer to this question is “yes,” studying single or few B-factors in isolation from the many others. There are of course conceptual and practical reasons for this approach; focusing on the latter, theorists must simplify to generate tractable models, and empiricists have lacked cost-effective tools for measuring multiple B-factors in representative samples. But there are also reasons for questioning the validity of this practice (Fudenberg 2006). There has been little if any empirical evidence brought to bear on this debate, as we discuss in the Introduction, in part at least because of the lack of data linking multiple B-factors with field choices and outcomes in representative samples.

Our data is well-suited for a simple yet incisive test of B-factor separability. Our previous section documents patterns of conditional correlations that are consistent with theoretical predictions that single B-factors, in particular standard directional biases, have deleterious influences on financial condition. We now test whether these inferences change if we jointly consider our full set of behavioral biases rather than just one B-factor at a time.

Table 7 presents results from two specifications of this test. The first two columns compare the standard bias coefficients from equation (1) for each single B-factor (column 1 reproduces the coefficients from the first row of Table 5) to the standard bias coefficients from estimating equation (1) with the complete set of 25 behavioral bias indicators for all 17 B-factors (column

---

<sup>29</sup> 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 levels or changes), which is consistent with measurement error but not environmental factors (including those that could generate reverse causality) playing an important role. Li et al. (2013) find moderate (in)stability in present-biased money discounting and in loss aversion, over several months. Carvalho et al. (2016) 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)). The strongest evidence we know of for instability in a B-factor is the Callen et al. (2014) finding that exposure to violent conflict increases preference for certainty. There is a larger body of evidence on the reliability of non-behavioral measures of time and risk preferences; see, e.g., Meier and Sprenger (2015) and Chuang and Schechter (2015) for recent reviews.

2). The coefficients are similar for 16 of the 17 B-factors: for these 16 the biggest difference is a single percentage point, and none of these differences have  $t$ -statistics  $> 1$ . Columns 3 and 4 perform the same comparison, with the only specification difference being that we use our preferred measurement error correction: weighting by the decile of respondent survey effort across all B-factor elicitations. Again the coefficients are similar for 16 of the 17 B-factors, with the largest difference among these now being 1.2 percentage points.

These results also belie the concern that our B-factors are measured with too much error to be informative, as the vector of “all other B-factors” contains several that have strong stand-alone conditional correlations with outcomes in Table 5. If one interprets those stand-alone conditional correlations as evidence of meaningful signal in measurement, then including them in the “all other B-factors” vector should change inferences, *if* there is a true confounding correlation amongst multiple B-factors and outcomes. We find no evidence of such a confound.

All in all, Table 7 strongly supports B-factor separability assumptions, at least with respect to financial condition, with profound implications for research practice and scope. On a practical level, our results suggest that theorists and empiricists are on firm ground considering one or a few B-factors in isolation from the many other potential behavioral influences. This validates prior work and implies a relatively manageable set of technical and data requirements for making continued progress going forward. On a more fundamental level, our results suggest that the criticism of BE as studying separate B-factors in “silos” lacks empirical grounding. This stands in apparent contrast to kindred lines of inquiry on cognitive skills and personality, where decades of research and testing have distilled seemingly countless intelligences and traits to a few underlying constructs.<sup>30</sup>

## **5. B-factors are distinct from other decision inputs**

Lastly, we consider our fourth foundational empirical question, whether B-factors are distinct from other decision inputs. We explore aspects of this question that are focused on research design, and on the underpinnings of behavioral biases themselves. Which other factors does one need to control/account for to identify the relationships between B-factors and financial

---

<sup>30</sup> Stango, Yoong, and Zinman (2017) takes up the broader question of how the 17 B-factors are related to each other, within-person, and what the relationships imply for models, applications, and research directions. See also Dean and Ortoleva (2016).

condition documented above? And how much of the information in our B-factor measures is explained by classical factors (instead of behavioral influences *per se*), including those like demographics that tend to be more readily available from a data perspective?

In addressing the distinctness question we focus on our covariates overall, or on groups of them. We do *not* interpret specific correlations between B-factor measures and other covariates, because the many correlations *among* the other covariates complicate that analysis and demand more thorough consideration than we have space for here. We leave this for future work, and for now provide some preliminary estimates of univariate and conditional correlations between behavioral biases and other covariates, for reference (Appendix Tables 3 and 8).

*A. Do estimates of links between financial condition and B-factors change based on the omission/inclusion of other covariates?*

One way to explore whether our B-factor measures capture distinct inputs to consumer decision making is to examine sensitivity to omitting other sets of covariates, following Altonji et al. (2005). If, the reasoning goes, our B-factor coefficients reflect omitted components of those covariates instead of behavioral influences *per se*, then dropping the observed components of those covariates should substantively change the B-factor coefficient estimates.

Table 8 show results from these exercises. The first column shows the standard-bias B-factor coefficients from Table 5 row 1. The second column drops our measures of cognitive skills from the specification in column 1. Comparing column 1 to column 2, none of the 17 B-factor coefficients are different with p-values below 0.10. Moreover the magnitude of any change in the point estimates is uniformly small, and in mixed directions. The third column drops our measures of classical preferences (risk and patience) from the specification in column 1. Again the B-factor coefficient estimates do not change. The fourth column drops age, education, and gender from the specification in column 1. Again the B-factor coefficient estimates do not change. The fifth drops our measures of survey effort from the specification in column 1. Again the B-factor coefficient estimates do not change. Across the 68 comparisons implicit in this table, none of them indicate statistically significant differences; in fact, the largest difference is only 1.1 percentage points, with a p-value of 0.69.

Overall there is little if any evidence here that our B-factor measures are simply proxies for other covariates. This suggests that the data requirements for estimating relationships between B-factors and field choices/outcomes need not be onerous, at least with respect to financial condition: one need not be overly concerned that omitting cognitive skills, classical preferences, some standard demographics, or even survey effort would substantively bias an estimated relationship between a B-factor and financial outcomes.

*B. Do other covariates explain B-factors in the cross section?*

Another way to explore whether our B-factor measures capture distinct inputs to consumer decision making is to estimate how much of their variance is explained by other covariates. To do this we simply take our main specification in (1) and move a measure of the (standard directional) behavioral bias to the left-hand-side:

$$(2) Bfactor_{s_i} = \alpha + cX_i + \varepsilon_i,$$

Table 9 reports four estimates of the adjusted R-squared for each B-factor. Columns 1 and 2 use standard bias indicators, without and with our preferred measurement error correction (weighting by the decile of time spent responding across all B-factor elicitations). Columns 3 and 4 use standard bias percentiles, again without and with the measurement error correction. Of these 68 estimates of fit, only 14 exceed 0.1, and 32 are less than 0.05. Weighting does tend to improve fit, but only modestly (by a percentage point or two in most cases). In short, our measures of behavioral biases are almost entirely unexplained by other covariates.

Several other results suggest that the lack of fit is *not* simply attributable to measurement error. Conceptually, including all of the other covariates is a powerful test, in the sense that even if one worries that a given measure fails to capture something important about that classical factor (e.g., a particular test score does not completely capture cognitive skills), one or more of the many other related measures (education, income, etc.) is likely to pick it up. Practically, both the B-factors together and the other covariates fit our other key variable, financial condition, relatively well. The fit of those models is over 0.40, with B-factors separately fitting 0.16 of variation across individuals and all other covariates as a group fitting 0.34. And various other findings above re: our behavioral bias measures suggest that they capture abundant signal (see Section 1-G for a summary).

## 6. Conclusion

Despite broad and growing influence, behavioral economics has lacked nationally representative evidence on basic empirical questions, and we provide methods and data for addressing four of them. Behavioral biases are prevalent, not anomalous. They are correlated with field choices and outcomes in ways predicted by theory, not neutralized by market forces, learning, or high stakes. They are separable in practice, not artificially siloed for modeling convenience. And they are distinct features of consumer decision making, not merely proxies for unmeasured aspects of classical inputs to decision making like cognitive skills, classical preferences, or life-cycle variables. No fewer than a dozen features of our research design, data, and results push against the hypothesis that measurement error drives our inferences.

This paper only begins to tap the potential of the new elicitation methods and dataset described herein. On the elicitation side, direct comparisons between our elicitations and standard ones would refine approaches to lowering the cost of measuring B-factors. Our methods are suitable for collecting data in a variety of settings and thus can be used to expand the evidence base on B-factors. In particular, our findings on the separability of behavioral biases imply that our streamlined elicitations could be integrated into established representative and large-sample household surveys, one B-factor at a time. This would increase power for estimating relationships between behavioral biases and field choices/outcomes. For the time being, we have another round of data collection from our sample underway that will add to the small body of evidence on the reliability (intertemporal stability) of directly elicited behavioral factors.<sup>31</sup> This is key to unpacking the direction and extent of any causality underlying conditional correlations between behavioral factors and outcomes.

In terms of the data used here, we are already at work exploring relationships among B-factors, and how to efficiently measure and summarize information on behavioral and other decision inputs. Beyond 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. That work could include more detailed consideration of behavioral theories, including structural models, than we undertake in this paper.

---

<sup>31</sup> Chuang and Schechter (2015) speculate that simpler elicitations may produce better reliability by reducing noise.

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

## REFERENCES

- Akerlof, George A. 2002. "Behavioral Macroeconomics and Macroeconomic Behavior." *American Economic Review* 92 (3): 411–33.
- 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.
- Banks, J., and Z. Oldfield. 2007. "Understanding Pensions: Cognitive Function, Numerical Ability, and Retirement Saving." *Fiscal Studies* 28 (2): 143–70.
- Barcellos, Silvia, and Leandro Carvalho. 2014. "Information about Self-Control and Intertemporal Choices."
- Barsky, Robert B, F. Thomas Juster, Miles S Kimball, and Matthew D Shapiro. 1997. "Preference Parameters and Behavioral Heterogeneity; An Experimental Approach in the Health and Retirement Study." *Quarterly Journal of Economics* 112 (2): 537–79.
- Beauchamp, Jonathan, David Cesarini, and Magnus Johannesson. forthcoming. "The Psychometric Properties of Measures of Economic Risk Preferences." *Journal of Risk and Uncertainty*.
- Becker, Anke, Thomas Deckers, Thomas Dohmen, Armin Falk, and Fabian Kosse. 2012. "The Relationship Between Economic Preferences and Psychological Personality Measures." *Annual Review of Economics* 4 (1): 453–78.
- Benjamin, Daniel, Sebastian Brown, and Jesse Shapiro. 2013. "Who Is 'Behavioral'? Cognitive Ability and Anomalous Preferences." *Journal of the European Economic Association* 11 (6): 1231–55.
- Benjamin, Daniel, Don Moore, and Matthew Rabin. 2013. "Misconceptions of Chance: Evidence from an Integrated Experiment."
- Benjamin, Daniel, Matthew Rabin, and Collin Raymond. 2016. "A Model of Nonbelief in the Law of Large Numbers." *Journal of the European Economic Association* 14 (2): 515–44.
- Bradford, David, Charles Courtemanche, Garth Heutel, Patrick McAlvanah, and Christopher Ruhm. 2014. "Time Preferences and Consumer Behavior." National Bureau of Economic Research.
- 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.
- Buonaccorsi, John P. 2010. *Measurement Error: Models, Methods, and Applications*. CRC Press.
- 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.

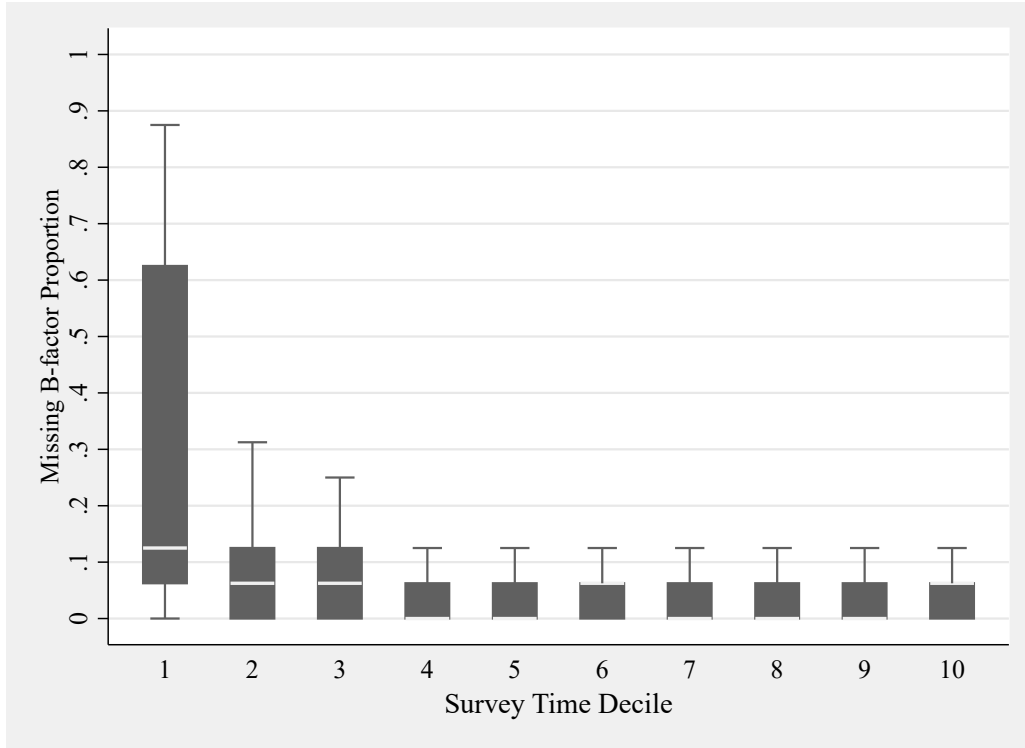


- Camerer, Colin F. 2007. "Neuroeconomics: Using Neuroscience to Make Economic Predictions." *The Economic Journal* 117 (519): C26–42.
- Carvalho, Leandro, Stephan Meier, and Stephanie Wang. 2016. "Poverty and Economic Decision-Making: Evidence from Changes in Financial Resources at Payday." *American Economic Review* 106 (2): 260–84.
- 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.
- 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.
- Cohen, Jonathan D., Keith Ericson, David Laibson, and John Myles White. 2016. "Measuring Time Preferences."
- Dean, Mark, and Pietro Ortoleva. 2016. "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. 2016. "Ambiguity Aversion and Household Portfolio Choice Puzzles: Empirical Evidence." *Journal of Financial Economics* 119 (3): 559–77.
- Dohmen, Thomas, Armin Falk, David Huffman, Felix Marklein, and Uwe Sunde. 2009. "Biased Probability Judgment: Evidence of Incidence and Relationship to Economic Outcomes from a Representative Sample." *Journal of Economic Behavior & Organization* 72 (3): 903–15.
- Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2010. "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review* 100 (3): 1238–60.
- Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association* 9 (3): 522–50.
- Driscoll, John C., and Steinar Holden. 2014. "Behavioral Economics and Macroeconomic Models." *Journal of Macroeconomics* 41 (September): 133–47.
- Ericson, Keith. forthcoming. "On the Interaction of Memory and Procrastination: Implications for Reminders." *Journal of the European Economic Association*.
- . 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."
- Fehr, Ernst, and Lorenz Goette. 2007. "Do Workers Work More If Wages Are High? Evidence from a Randomized Field Experiment." *American Economic Review* 97 (1): 298–317.

- Frederick, Shane. 2005. "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives* 19 (4): 25–42.
- Fudenberg, Drew. 2006. "Advancing Beyond Advances in Behavioral Economics." *Journal of Economic Literature* 44 (3): 694–711.
- Fuller, Wayne A. 2009. *Measurement Error Models*. Vol. 305. John Wiley & Sons.
- Gabaix, Xavier. 2014. "A Sparsity-Based Model of Bounded Rationality." *The Quarterly Journal of Economics* 129 (4): 1661–1710.
- . 2017. "Behavioral Macroeconomics Via Sparse Dynamic Programming."
- Gillen, Ben, Erik Snowberg, and Leeat Yariv. 2017. "Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study."
- Gine, Xavier, Jessica Goldberg, Daniel Silverman, and Dean Yang. forthcoming. "Revising Commitments: Field Evidence on the Adjustment of Prior Choices." *Economic Journal*.
- 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. 2017. "Predicting Retirement Savings Using Survey Measures of Exponential-Growth Bias and Present Bias."
- Gottlieb, Daniel, and Olivia S. Mitchell. 2015. "Narrow Framing and Long-Term Care Insurance."
- Heidhues, Paul, Botond Koszegi, and Philipp Strack. 2017. "Unrealistic Expectations and Misguided Learning." SSRN Scholarly Paper. Rochester, NY.  
<http://papers.ssrn.com/abstract=2703524>.
- Hwang, In Do. 2016. "Prospect Theory and Insurance Demand."
- Kahneman, D. 2003. "Maps of Bounded Rationality: Psychology for Behavioral Economics." *American Economic Review* 93 (5): 1449–75.
- Kőszegi, Botond. 2014. "Behavioral Contract Theory." *Journal of Economic Literature* 52 (4): 1075–1118.
- Levine, David K. 2012. *Is Behavioral Economics Doomed?: The Ordinary versus the Extraordinary*. Open Book Publishers.
- Levy, Matthew, and Joshua Tasoff. 2016. "Exponential-Growth Bias and Lifecycle Consumption." *Journal of the European Economic Association* 14 (3): 545–83.
- Li, Ye, Martine Baldassi, Eric J. Johnson, and Elke U. Weber. 2013. "Complementary Cognitive Capabilities, Economic Decision Making, and Aging." *Psychology and Aging* 28 (3): 595–613.
- 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.

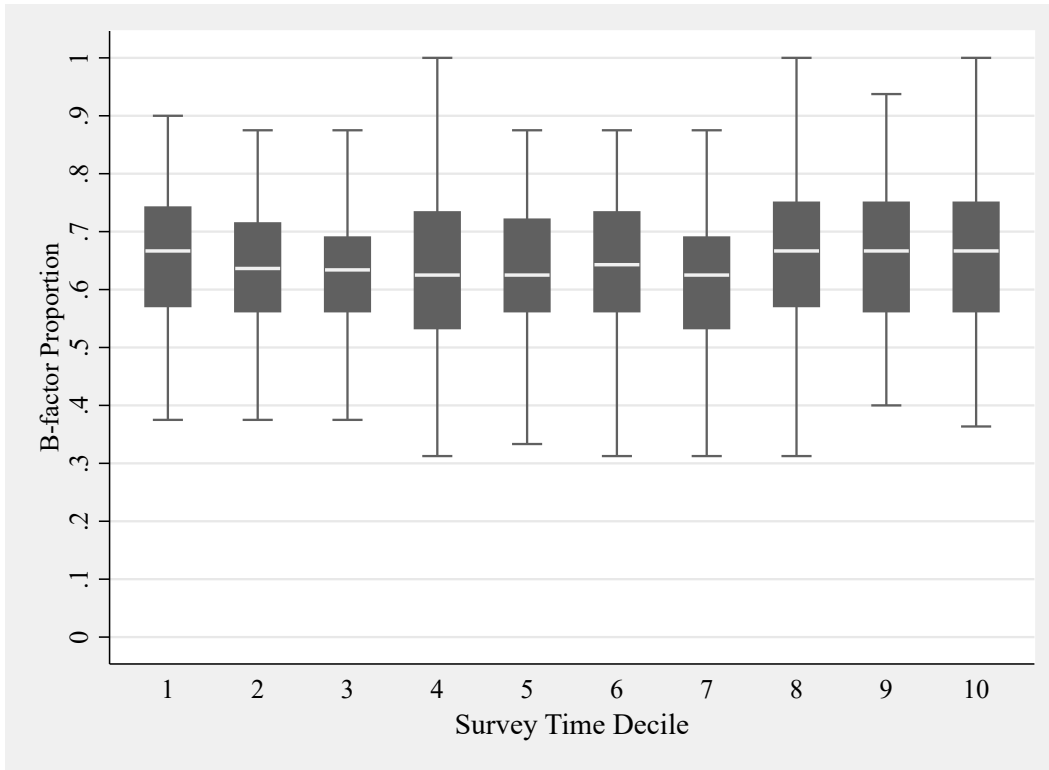
- Meier, Stephan, and Charles D. Sprenger. 2015. "Temporal Stability of Time Preferences." *Review of Economics and Statistics* 97 (2): 273–86.
- Moore, Don A., and Paul J. Healy. 2008. "The Trouble with Overconfidence." *Psychological Review* 115 (2): 502–17.
- O'Donoghue, Ted, and Matthew Rabin. 1999. "Doing It Now or Later." *American Economic Review* 89 (1): 103–24.
- Rabin, M. 1998. "Psychology and Economics." *Journal of Economic Literature* 36 (1): 11–46.
- 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, Joanne Yoong, and Jonathan Zinman. 2017. "The Quest for Parsimony in Behavioral Economics: New Methods and Evidence on Three Fronts."
- Stango, Victor, and Jonathan Zinman. 2009. "Exponential Growth Bias and Household Finance." *The Journal of Finance* 64 (6): 2807–49.
- . 2011. "Fuzzy Math, Disclosure Regulation, and Credit Market Outcomes: Evidence from Truth-in-Lending Reform." *Review of Financial Studies* 24 (2): 506–34.
- 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.

**Figure 1**  
**B-factor non-response as a function of time spent on B-factor questions**



Notes: Box-and-whisker plots show inter-quartile range in box, adjacent values as whiskers, and median as white line. Missing B-factor Proportion = (# of B-factors with missing or unusable data)/17. We calculate this at the individual level and then average it within each Survey Time Decile, which is based on time spent responding across all B-factor elicitations.

**Figure 2**  
**B-factor prevalence (excluding non-response), by time spent on B-factor questions**



Notes: Box-and-whisker plots show inter-quartile range in box, adjacent values as whiskers, and median as white line. B-factor Proportion = (Number of behavioral biases indicated)/(Number of non-missing B-factors). We calculate this at the individual level and then average it within each Survey Time Decile, which is based on time spent responding across all B-factor elicitations.

**Table 1. Research design: Eliciting data on multiple behavioral factors, and defining behavioral bias indicators**

B-factor name: <i>key antecedents</i>	Elicitation method description	Behavioral bias indicator(s), "standard" deviation direction in bold
<b>Time inconsistent discounting of money:</b> <i>Andreoni &amp; Sprenger (2012), Barcellos &amp; Carvalho (2014)</i>	Convex Time Budget. 24 decisions allocating 100 tokens each between smaller-sooner and larger-later amounts; decisions pose varying start dates (today vs. 5 weeks from today), delay lengths (5 or 9 weeks) & savings yields.	<b>Present-biased: discounts more when sooner date is today</b> Future-biased: discounts more when sooner date is 5 weeks from today
<b>Time inconsistent discounting of snacks:</b> <i>Read &amp; van Leeuwen (1998), Barcellos &amp; Carvalho (2014)</i>	Two decisions between two snacks: healthier/less-delicious vs. less healthy/more delicious. Decisions vary only in date snack is delivered: now, or 5 weeks from now.	<b>Present-biased: choose less healthy today, healthy for 5 weeks from now</b> Future-biased: choose healthy for today, less healthy for 5 weeks from now
<b>Violates General Axiom of Revealed Preference: (and/or dominance avoidance)</b> <i>Choi et al (2014)</i>	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.	<b>Violates GARP: potential earnings wasted per CCEI&gt;0</b> <b>Violates GARP and/or dominance avoidance: potential earnings wasted per combined-CCEI&gt;0</b>
<b>Certainty premium:</b> <i>Callen et al (2014)</i>	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 ∈ [0.1, 1.0], p=0.5 on one screen and 1.0 on the other.	<b>Preference for certainty: certainty premium (CP) &gt;0</b> Cumulative prospect theory: certainty premium (CP)<0
<b>Loss aversion/small-stakes risk aversion:</b> <i>Fehr &amp; Goette (2007)</i>	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.	<b>Loss aversion: choosing the certain \$0 payoff in one or more choices.</b>
<b>Narrow bracketing:</b> <i>Rabin &amp; Weizsacker (2009)</i>	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.	<b>Narrow-bracketing: making a choice that is dominated given implications of an earlier decision, on one or both</b>
<b>Ambiguity aversion:</b> <i>Dimmock et al. (2016)</i>	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.	<b>Ambiguity Aversion: prefers bag with 45 green to bag with unknown mix.</b>
<b>(Over)confidence in performance:</b> <i>Larrick et al (2007), Moore &amp; Healy (2008)</i>	"How many of the last 3 questions (the ones on the disease, the lottery and the savings account) do you think you got correct?"	<b>Overconfidence in perform: self-assessment &gt; actual score</b> Underconfidence in perform: self-assessment < actual score
<b>Overconfidence in relative performance:</b> <i>Larrick et al (2007), Moore &amp; Healy (2008)</i>	"... 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?"	<b>Greater diff between self-assessed and actual rank indicates more overconfidence. "Overconfident" = overconfidence above median (no precise cardinal</b>
<b>Overconfidence in precision:</b> <i>Larrick et al (2007), Moore &amp; Healy (2008)</i>	Questions about about likelihoods of different numeracy quiz scores and future income increases.	<b>Overconfidence in precision: responds 100% to one or both questions</b>
<b>Non-belief in the law of large numbers (NBLLN):</b> <i>Benjamin, Moore, and Rabin (2013)</i>	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	Overestimates convergence to 50-50: responds with>78% <b>Underestimates convergence to 50-50: responds with&lt;78%</b>
<b>Gambler's fallacies:</b> <i>Benjamin, Moore, and Rabin (2013)</i>	"Imagine that we had a computer "flip" a fair coin... 10 times. The first 9 are all heads. What are the chances, in % terms, that the 10th flip will be a head?"	<b>Hot-hand fallacy: responds with&gt;50%</b> Cold-hand fallacy: responds with<50%
<b>Exponential growth bias (EGB), debt-side:</b> <i>Stango &amp; Zinman (2009; 2011)</i>	Survey first elicits monthly payment respondent would expect to pay on a \$10,000, 48 month car loan (this response defines the actual APR). Then elicits perceived APR implied by that payment.	<b>Underestimates EG: actual APR&gt;perceived APR</b> Overestimates EG: actual APR<perceived APR
<b>Exponential growth bias (EGB), asset-side:</b> <i>Banks et al (2007)</i>	Elicits perceived future value of \$200, earning 10% annual, after two years.	<b>Underestimates EG: perceived FV&lt;actual FV=\$242</b> Overestimates EG: perceived FV>actual FV=\$242
<b>Limited attention:</b> <i>Author-developed</i>	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 products/services.	<b>Limited attention: Indicates regret about paying too little attention, on one or more of the four questions</b>
<b>Limited prospective memory:</b> <i>Ericson (2011)</i>	"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."	<b>Limited memory: Says will complete task but does not complete</b>

The Data Appendix provides additional details on measuring individual behavioral factors. "Standard" deviation direction, for bi-directional B-factors, is the direction typically theorized/observed in prior work to harm financial condition. "CCEI" = Critical Cost Efficiency Index.

**Table 2. Prevalence and missing values for B-factors and their directional biases**

B-factor and bias (standard direction in bold)	Share biased, conditional on response	Missing detail		
		survey nonresponse	item nonresponse	responded, not usable
<b>Time-inconsistent discounting money: Present-biased</b>	<b>0.26</b>			
Time-inconsistent discounting money: Future-biased	0.36	0.00	0.06	0.00
<b>Time-inconsistent discounting snacks: Present-biased</b>	<b>0.15</b>			
Time-inconsistent discounting snacks: Future-biased	0.07	0.06	0.02	0.00
<b>Violates GARP (based on CCEI)</b>	<b>0.53</b>	0.06	0.10	0.00
<b>Violates GARP (with dominance avoidance)</b>	<b>0.96</b>	0.06	0.10	0.00
<b>Certainty premium: &gt;0=Preference for certainty type</b>	<b>0.77</b>			
Certainty premium: <0=Cumulative prospect theory type	0.23	0.00	0.03	0.28
<b>Loss-averse: prefers certain zero payoff</b>	<b>0.63</b>	0.00	0.00	0.00
<b>Narrow-brackets</b>	<b>0.59</b>	0.00	0.02	0.00
<b>Ambiguity-averse</b>	<b>0.73</b>	0.06	0.03	0.07
<b>Confidence in level performance: Overconfident</b>	<b>0.38</b>			
Confidence in level performance: Underconfident	0.11	0.06	0.03	0.07
<b>Overconfident in precision</b>	<b>0.44</b>	0.06	0.04	0.00
<b>Overconfident in relative performance</b>	<b>0.50</b>	0.06	0.04	0.00
<b>Non-belief in the law of large numbers: Underestimates convergence</b>	<b>0.87</b>			
Non-belief in the law of large numbers: Overestimates convergence	0.13	0.06	0.03	0.00
Gambler's fallacy: hot hand	0.14			
Gambler's fallacy: cold hand	0.26	0.06	0.02	0.00
<b>Exponential growth bias, loan-side: Underestimates APR</b>	<b>0.70</b>			
Exponential growth bias, loan-side: Overestimates APR	0.27	0.00	0.05	0.32
<b>Exponential growth bias, asset-side: Underestimates future value</b>	<b>0.38</b>			
Exponential growth bias, asset-side: Overestimates future value	0.07	0.06	0.03	0.00
<b>Limited attention</b>	<b>0.49</b>	0.00	0.02	0.00
<b>Limited memory</b>	<b>0.86</b>	0.06	0.02	0.02

Unit of observation is the individual respondent, and missing shares are relative to the full sample size of 1,515. Section 1-C provides some details on measuring individual behavioral factors and classifying directional biases as standard vs. non-standard; see the Data Appendix for additional details. "GARP" = General Axiom of Revealed Preference. "CCEI" = Critical Cost Efficiency Index. Proportion exhibiting relative overconfidence is 50% by construction, since our elicitation does not produce a clear cardinal measure (as detailed in Data Appendix Section H). "Share biased" is conditional on non-missing values. "Survey nonresponse" indicates panelists who took our first module but not our second. "Item nonresponse" can occur on either module. The large "unusable" share for the Certainty Premium is partly due to respondents who do not switch on the multiple price lists; this is a limitation of the elicitation rather than an indication of low-quality responses (Data Appendix Section D). The large unusable share for EGB loan-side is due largely to responses that imply a zero APR (Data Appendix Section L).

**Table 3. Research design: Measuring classical factors and respondent survey effort**

Covariate	Definition/specification in empirics
<u>Demographics:</u>	
<b>Gender</b>	Indicator for female
<b>Age</b>	Four categories: 18-34, 35-45, 46-54, 55+
<b>Education</b>	Four categories: HS or less, some college/associates, BA, graduate
<b>Income</b>	The ALP's 17 categories (collapsed into deciles in some specifications)
<b>Race/ethnicity</b>	Three categories: White, Black, or Other; separate indicator for Hispanic
<b>Marital status</b>	Three categories: married/co-habiting; separated/divorced/widowed; never married
<b>Household size</b>	Five categories for count of other members: 0, 1, 2, 3, 4+
<b>Employment status</b>	Five categories: working, self-employed, not working, disabled, missing
<b>Immigrated to USA</b>	Indicator for immigrant
<b>State of residence</b>	Indicator for each state
<u>Risk, patience:</u>	
<b>Risk aversion (financial)</b>	100-point scale on financial risk-taking from Dohmen et al., with higher values indicating greater risk aversion
<b>Risk aversion (income)</b>	6 point scale based on Barsky et al lifetime income gambles, higher values indicate greater aversion
<b>Patience</b>	Average savings rate across the 24 Convex Time Budget decisions, standardized
<u>Cognitive skills:</u>	
<b>Fluid intelligence</b>	# correct on standard 15-question, non-adaptive number series quiz
<b>Numeracy</b>	# correct on Banks and Oldfield questions re: division and %
<b>Financial literacy</b>	# correct on Lusardi and Mitchell "Big Three" questions re: interest, inflation, and diversification
<b>Executive attention</b>	# correct on 2-minute Stroop test; respondents instructed to answer as many q's correctly as they can
<u>Survey effort</u>	
<b>Time spent on questions</b>	Measured for each B-factor (and other variables), included as decile indicators relative to other respondents
<b>Missing variable(s)</b>	Indicators
<b>Did not take our 2nd survey</b>	Indicator for 6% of the sample that took our first module but not the second

Notes: See Section 1-D of the text for details on elicitations and variable construction.



**Table 4. Research design, measuring financial condition: Component indicator prevalence and pairwise correlations**

Variable	Mean of indicator	net worth>0	retirement assets>0	owns stocks	spent < income last 12	financial satisfaction > median	retirement saving adequate	non-ret saving adequate	no severe distress last 12	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
<b>Mean of all indicators</b>	<b>0.43</b>									

Unit of observation is the individual respondent. Unconditional pairwise correlations with p-values: \* 0.10 \*\* 0.05 \*\*\* 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 5. Estimating Equation (1) for each B-factor: Conditional correlations between financial condition and behavioral biases

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
	RHS B-factor shown as column header. LHS variable in all models is summary financial condition (mean = 0.43)																
	Time-inconsistent: money	Time-inconsistent: snack	Violates GARP	Violates GARP FOSD	Certainty Premium	Loss averse	Narrow-brackets	Ambiguity averse	(Over-)confident performance	Over-confident precision	Over-confident relative perf.	Non-belief Law Large Numbers	Gambler's Fallacies	Exponential growth bias: loan	Exponential growth bias: asset	Limited attention	Limited memory
Standard bias indicator	-0.038** (0.016)	-0.020 (0.018)	-0.013 (0.013)	-0.021 (0.034)	-0.006 (0.017)	-0.010 (0.013)	0.002 (0.013)	-0.025* (0.014)	-0.020 (0.015)	-0.018 (0.014)	-0.037** (0.017)	-0.023 (0.020)	-0.028 (0.016)	-0.006 (0.021)	-0.063*** (0.026)	-0.109*** (0.012)	-0.042** (0.018)
Non-standard bias indicator	-0.005 (0.016)	-0.018 (0.025)							0.037* (0.021)				-0.016 (0.016)	0.001 (0.021)	-0.038 (0.026)		
bias missing	-0.009 (0.028)	0.033 (0.051)	-0.012 (0.023)	-0.025 (0.040)	0.016 (0.021)		0.033 (0.048)	0.262 (0.167)	-0.255 (0.275)	0.070* (0.036)	-0.228* (0.122)	0.168** (0.074)	0.060 (0.146)	0.016 (0.030)	-0.023 (0.025)	-0.067 (0.052)	0.016 (0.042)
female	-0.020 (0.013)	-0.020 (0.013)	-0.017 (0.013)	-0.017 (0.013)	-0.019 (0.013)	-0.020 (0.013)	-0.020 (0.013)	-0.019 (0.014)	-0.020 (0.013)	-0.019 (0.013)	-0.019 (0.013)	-0.020 (0.013)	-0.019 (0.013)	-0.019 (0.013)	-0.019 (0.013)	-0.018 (0.013)	-0.018 (0.013)
ed: some college or associates	-0.040** (0.018)	-0.041** (0.018)	-0.039** (0.018)	-0.039** (0.018)	-0.041** (0.018)	-0.040** (0.018)	-0.041** (0.018)	-0.048** (0.019)	-0.043** (0.018)	-0.042** (0.018)	-0.040** (0.018)	-0.042** (0.018)	-0.042** (0.018)	-0.044** (0.018)	-0.044** (0.018)	-0.039** (0.017)	-0.041** (0.018)
highest ed: bachelor's	0.017 (0.021)	0.015 (0.021)	0.016 (0.021)	0.016 (0.021)	0.013 (0.021)	0.015 (0.021)	0.015 (0.021)	0.013 (0.022)	0.012 (0.021)	0.011 (0.021)	0.015 (0.021)	0.012 (0.021)	0.014 (0.021)	0.011 (0.021)	0.013 (0.021)	0.004 (0.020)	0.013 (0.021)
highest ed: graduate	0.031 (0.024)	0.029 (0.024)	0.031 (0.024)	0.030 (0.024)	0.029 (0.024)	0.032 (0.024)	0.030 (0.024)	0.030 (0.025)	0.028 (0.024)	0.027 (0.024)	0.033 (0.024)	0.026 (0.024)	0.028 (0.024)	0.028 (0.024)	0.028 (0.024)	0.011 (0.023)	0.027 (0.024)
age 35-45	-0.016 (0.018)	-0.010 (0.018)	-0.012 (0.018)	-0.012 (0.018)	-0.014 (0.018)	-0.016 (0.018)	-0.016 (0.018)	-0.011 (0.019)	-0.011 (0.018)	-0.014 (0.018)	-0.013 (0.018)	-0.017 (0.018)	-0.012 (0.018)	-0.016 (0.018)	-0.015 (0.018)	-0.016 (0.017)	-0.015 (0.018)
age 46-54	0.003 (0.019)	0.006 (0.019)	0.002 (0.019)	0.003 (0.019)	0.004 (0.019)	0.001 (0.019)	-0.000 (0.019)	0.002 (0.020)	0.007 (0.019)	0.004 (0.019)	0.004 (0.019)	0.003 (0.019)	0.005 (0.019)	-0.002 (0.019)	-0.001 (0.019)	-0.003 (0.018)	0.001 (0.019)
age>=55	0.053** (0.021)	0.057*** (0.021)	0.053** (0.021)	0.053** (0.021)	0.057*** (0.021)	0.050** (0.021)	0.049** (0.021)	0.052** (0.022)	0.056*** (0.021)	0.054** (0.021)	0.053** (0.021)	0.046** (0.021)	0.053** (0.021)	0.047** (0.021)	0.046** (0.021)	0.045** (0.020)	0.049** (0.021)
income decile 2	-0.029 (0.027)	-0.029 (0.027)	-0.028 (0.027)	-0.028 (0.027)	-0.026 (0.027)	-0.030 (0.027)	-0.028 (0.027)	-0.029 (0.028)	-0.020 (0.027)	-0.023 (0.027)	-0.020 (0.027)	-0.026 (0.027)	-0.030 (0.027)	-0.025 (0.027)	-0.025 (0.027)	-0.018 (0.026)	-0.030 (0.027)
income decile 3	0.018 (0.028)	0.021 (0.028)	0.021 (0.028)	0.022 (0.028)	0.026 (0.028)	0.019 (0.028)	0.022 (0.028)	0.027 (0.029)	0.029 (0.028)	0.026 (0.028)	0.028 (0.028)	0.028 (0.028)	0.019 (0.028)	0.025 (0.028)	0.027 (0.028)	0.039 (0.027)	0.022 (0.028)
income decile 4	0.074*** (0.028)	0.076*** (0.028)	0.078*** (0.028)	0.078*** (0.028)	0.078*** (0.028)	0.074*** (0.028)	0.079*** (0.028)	0.073** (0.029)	0.084*** (0.028)	0.081*** (0.028)	0.085*** (0.028)	0.080*** (0.028)	0.073*** (0.028)	0.084*** (0.028)	0.083*** (0.028)	0.097*** (0.027)	0.083*** (0.028)
income decile 5	0.087*** (0.031)	0.089*** (0.031)	0.089*** (0.031)	0.089*** (0.031)	0.092*** (0.031)	0.085*** (0.031)	0.091*** (0.031)	0.091*** (0.032)	0.095*** (0.031)	0.091*** (0.031)	0.096*** (0.031)	0.096*** (0.031)	0.083*** (0.031)	0.095*** (0.031)	0.094*** (0.031)	0.117*** (0.030)	0.094*** (0.031)
income decile 6	0.138*** (0.031)	0.142*** (0.031)	0.139*** (0.031)	0.138*** (0.031)	0.144*** (0.031)	0.137*** (0.031)	0.141*** (0.031)	0.129*** (0.032)	0.143*** (0.031)	0.144*** (0.031)	0.146*** (0.031)	0.147*** (0.031)	0.138*** (0.031)	0.149*** (0.031)	0.146*** (0.031)	0.172*** (0.030)	0.146*** (0.031)
income decile 7	0.174*** (0.031)	0.177*** (0.031)	0.175*** (0.031)	0.175*** (0.031)	0.175*** (0.031)	0.171*** (0.031)	0.178*** (0.031)	0.166*** (0.032)	0.186*** (0.030)	0.179*** (0.031)	0.186*** (0.031)	0.180*** (0.031)	0.168*** (0.031)	0.182*** (0.031)	0.184*** (0.031)	0.201*** (0.030)	0.178*** (0.030)
income decile 8	0.261*** (0.032)	0.263*** (0.032)	0.265*** (0.031)	0.266*** (0.031)	0.269*** (0.031)	0.262*** (0.032)	0.267*** (0.032)	0.274*** (0.033)	0.270*** (0.031)	0.265*** (0.031)	0.269*** (0.031)	0.268*** (0.031)	0.260*** (0.032)	0.273*** (0.032)	0.272*** (0.032)	0.287*** (0.031)	0.272*** (0.031)
income decile 9	0.296*** (0.030)	0.297*** (0.030)	0.299*** (0.030)	0.299*** (0.030)	0.303*** (0.030)	0.296*** (0.030)	0.303*** (0.030)	0.289*** (0.032)	0.304*** (0.030)	0.302*** (0.030)	0.306*** (0.030)	0.304*** (0.030)	0.294*** (0.030)	0.308*** (0.030)	0.302*** (0.030)	0.323*** (0.029)	0.308*** (0.030)
income decile 10	0.374*** (0.044)	0.378*** (0.044)	0.379*** (0.044)	0.381*** (0.044)	0.374*** (0.044)	0.374*** (0.044)	0.381*** (0.044)	0.375*** (0.047)	0.382*** (0.044)	0.380*** (0.044)	0.384*** (0.044)	0.388*** (0.044)	0.368*** (0.044)	0.391*** (0.045)	0.384*** (0.044)	0.391*** (0.043)	0.387*** (0.044)

continued on next page...

Table 5, cont'd

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
	Time- inconsistent: money	Time- inconsistent: snack	Violates GARP	Violates GARP FOSD	Certainty Premium	Loss averse	Narrow- brackets	Ambiguity averse	(Over-) confident performance	Over- confident precision	Over- confident relative perf.	Non-belief Law Large Numbers	Gambler's Fallacies	Exponential growth bias: loan	Exponential growth bias: asset	Limited attention	Limited memory
fluid intell # correct	0.001 (0.003)	0.001 (0.003)	0.000 (0.003)	0.000 (0.003)	0.002 (0.003)	0.001 (0.003)	0.001 (0.003)	0.001 (0.003)	0.001 (0.003)	0.001 (0.003)	-0.004 (0.004)	0.000 (0.003)	0.000 (0.003)	0.001 (0.003)	-0.002 (0.003)	0.001 (0.003)	0.001 (0.003)
numeracy # correct	-0.001 (0.013)	0.000 (0.013)	-0.000 (0.013)	0.000 (0.013)	0.002 (0.013)	-0.001 (0.013)	-0.000 (0.013)	0.003 (0.013)	-0.008 (0.013)	0.005 (0.014)	0.003 (0.013)	0.001 (0.013)	0.001 (0.013)	0.002 (0.013)	0.002 (0.013)	0.001 (0.013)	-0.001 (0.013)
financial literacy # correct	0.028*** (0.009)	0.029*** (0.009)	0.029*** (0.009)	0.029*** (0.009)	0.031*** (0.009)	0.029*** (0.009)	0.027*** (0.009)	0.030*** (0.009)	0.029*** (0.009)	0.031*** (0.009)	0.031*** (0.009)	0.027*** (0.009)	0.027*** (0.009)	0.029*** (0.009)	0.021** (0.009)	0.027*** (0.009)	0.028*** (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)	0.000 (0.000)
risk aversion (financial)	-0.006*** (0.002)	-0.006*** (0.002)	-0.007*** (0.002)	-0.006*** (0.002)	-0.006** (0.002)	-0.006** (0.002)	-0.006*** (0.002)	-0.005** (0.002)	-0.006*** (0.002)	-0.007*** (0.002)	-0.006*** (0.002)	-0.006*** (0.002)	-0.006*** (0.002)	-0.006*** (0.002)	-0.006*** (0.002)	-0.005** (0.002)	-0.006** (0.002)
risk aversion (income)	0.009* (0.005)	0.010* (0.005)	0.009* (0.005)	0.009* (0.005)	0.009* (0.005)	0.009* (0.005)	0.009* (0.005)	0.006 (0.005)	0.009* (0.005)	0.009* (0.005)	0.008* (0.005)	0.009* (0.005)	0.008* (0.005)	0.009* (0.005)	0.009* (0.005)	0.007 (0.005)	0.009* (0.005)
patience (stdized)	0.006 (0.007)	0.007 (0.006)	0.008 (0.006)	0.008 (0.006)	0.007 (0.006)	0.008 (0.006)	0.008 (0.006)	0.008 (0.007)	0.008 (0.006)	0.007 (0.006)	0.006 (0.006)	0.008 (0.006)	0.007 (0.006)	0.007 (0.006)	0.007 (0.006)	0.005 (0.006)	0.009 (0.006)
Response time spent on B- factor in this regression:																	
decile 2	-0.001 (0.027)	-0.043 (0.027)	0.008 (0.028)	0.007 (0.028)	-0.032 (0.028)	0.017 (0.026)	0.003 (0.027)	-0.032 (0.028)	0.010 (0.024)	-0.023 (0.027)	-0.057** (0.027)	0.034 (0.028)	-0.027 (0.028)	-0.027 (0.028)	-0.004 (0.027)	-0.047* (0.026)	0.017 (0.026)
decile 3	-0.005 (0.027)	-0.040 (0.027)	0.017 (0.029)	0.018 (0.029)	-0.005 (0.028)	0.018 (0.026)	-0.011 (0.027)	-0.062** (0.028)	0.040 (0.027)	0.019 (0.027)	-0.047 (0.029)	0.042 (0.029)	-0.007 (0.026)	-0.040 (0.029)	0.041 (0.027)	-0.041 (0.027)	0.008 (0.028)
decile 4	0.018 (0.028)	-0.011 (0.028)	-0.004 (0.029)	-0.003 (0.029)	-0.050* (0.028)	0.059** (0.027)	0.010 (0.027)	-0.009 (0.028)	-0.035 (0.029)	0.021 (0.027)	-0.014 (0.028)	0.023 (0.028)	0.011 (0.028)	-0.039 (0.029)	0.012 (0.029)	-0.018 (0.026)	-0.036 (0.026)
decile 5	-0.001 (0.027)	-0.017 (0.026)	0.027 (0.030)	0.028 (0.030)	0.018 (0.028)	0.038 (0.027)	-0.002 (0.027)	-0.002 (0.028)	0.009 (0.030)	-0.005 (0.027)	-0.038 (0.028)	0.042 (0.029)	0.030 (0.028)	-0.047 (0.029)	0.024 (0.027)	-0.024 (0.027)	-0.020 (0.029)
decile 6	0.017 (0.028)	-0.050* (0.027)	-0.027 (0.029)	-0.026 (0.029)	-0.049* (0.028)	0.033 (0.027)	-0.002 (0.027)	-0.039 (0.027)	0.004 (0.031)	0.023 (0.028)	-0.046* (0.027)	0.046 (0.030)	0.021 (0.026)	-0.023 (0.029)	0.005 (0.028)	0.008 (0.027)	0.034 (0.029)
decile 7	0.005 (0.028)	-0.033 (0.028)	-0.002 (0.030)	-0.001 (0.030)	-0.023 (0.028)	0.009 (0.028)	0.014 (0.028)	-0.022 (0.029)	-0.030 (0.028)	-0.016 (0.028)	-0.051* (0.028)	0.002 (0.029)	-0.033 (0.027)	-0.025 (0.030)	0.051* (0.029)	-0.037 (0.026)	0.031 (0.027)
decile 8	0.012 (0.028)	-0.024 (0.028)	0.008 (0.029)	0.009 (0.029)	-0.033 (0.029)	0.014 (0.027)	0.017 (0.027)	-0.026 (0.028)	0.030 (0.029)	-0.007 (0.028)	-0.055* (0.028)	0.025 (0.029)	-0.024 (0.028)	0.004 (0.030)	0.001 (0.028)	-0.027 (0.027)	-0.028 (0.028)
decile 9	-0.023 (0.028)	-0.018 (0.028)	0.017 (0.030)	0.018 (0.030)	0.004 (0.028)	0.002 (0.027)	0.002 (0.027)	-0.011 (0.028)	-0.029 (0.028)	0.033 (0.028)	-0.034 (0.028)	0.025 (0.030)	-0.018 (0.027)	-0.027 (0.030)	0.008 (0.028)	-0.023 (0.027)	0.018 (0.028)
decile 10	0.004 (0.028)	-0.014 (0.028)	-0.030 (0.029)	-0.029 (0.029)	-0.039 (0.028)	0.017 (0.027)	0.015 (0.027)	-0.036 (0.028)	0.054* (0.029)	-0.014 (0.028)	-0.022 (0.028)	0.052* (0.029)	-0.006 (0.027)	-0.002 (0.030)	0.024 (0.028)	-0.017 (0.027)	0.016 (0.028)
Full controls from Table 3?	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R-squared	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.40	0.39	0.39	0.39	0.39	0.39	0.40	0.43	0.40
Number of observations	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505

\* 0.10 \*\* 0.05 \*\*\* 0.01. Unit of observation is an individual respondent. One regression per column. LHS variable is the summary measure of financial condition: the proportion of indicators with a "1" from Table 3. Models are OLS and also include covariates described in Table 3 but not shown here to save space: four race/ethnicity categories, immigrant indicator, 3 marital status categories, 4 household size categories, 5 work status categories, state of residence dummies, and dummies for missing values associated with each variable and for not taking our 2nd module. Omitted category for B-factors is unbiased (classical), except for Certainty Premium and Nonbelief in the Law of Large Numbers (NBLN), where nonstandard bias is the omitted category due to lack of unbiased responses. See Section 1-C and the Data Appendix for details on B-factor variable construction.

**Table 6. Conditional correlations between financial condition and B-factor (standard bias) indicators, weighted by respondent survey/task effort**

B-factor (standard bias)	(1)	(2)	(3)	(4)
	Survey/task effort weight			
	None: same as Table 5	ln(Total B-Factor time spent)	Total B-Factor time spent decile	Single B-Factor time decile
Time-inconsistent money (present-bias)	-0.038** (0.016)	-0.040** (0.016)	-0.053*** (0.016)	-0.053*** (0.016)
Time-inconsistent snack (present-bias)	-0.020 (0.018)	-0.019 (0.018)	-0.019 (0.017)	-0.009 (0.018)
Violates GARP	-0.013 (0.013)	-0.012 (0.013)	-0.002 (0.013)	-0.006 (0.013)
Violates GARP FOSD	-0.021 (0.034)	-0.024 (0.033)	-0.044 (0.031)	-0.061* (0.032)
Certainty premium (preference for certainty)	-0.006 (0.017)	-0.005 (0.017)	-0.005 (0.017)	0.005 (0.017)
Loss averse	-0.010 (0.013)	-0.011 (0.013)	-0.021* (0.013)	-0.018 (0.013)
Narrow-brackets	0.002 (0.013)	0.002 (0.013)	-0.001 (0.013)	0.001 (0.012)
Ambiguity averse	-0.023 (0.015)	-0.024* (0.014)	-0.024* (0.014)	-0.027* (0.014)
Confidence in level performance (over-)	-0.020 (0.015)	-0.016 (0.015)	-0.009 (0.014)	-0.019 (0.014)
Overconfident precision	-0.018 (0.014)	-0.019 (0.013)	-0.018 (0.013)	-0.025* (0.013)
Overconfident relative perf.	-0.037** (0.017)	-0.037** (0.017)	-0.047*** (0.017)	-0.048*** (0.018)
Non-belief Law Large Numbers (underestimates)	-0.023 (0.020)	-0.025 (0.019)	-0.030 (0.019)	-0.012 (0.020)
Gambler's fallacies (hot hand)	-0.028 (0.020)	-0.027 (0.019)	-0.027 (0.019)	-0.028 (0.019)
Exponential growth bias loan (underestimates)	-0.006 (0.019)	-0.001 (0.019)	0.008 (0.019)	-0.007 (0.019)
Exponential growth bias asset (underestimates)	-0.063*** (0.017)	-0.061*** (0.016)	-0.050*** (0.016)	-0.046*** (0.016)
Limited attention	-0.109*** (0.012)	-0.109*** (0.012)	-0.111*** (0.012)	-0.105*** (0.012)
Limited memory	-0.042** (0.018)	-0.042** (0.018)	-0.048*** (0.017)	-0.044** (0.017)
Average r-squared, all B-factor regressions in column:	0.40	0.40	0.42	0.42
sign test, all coefficients	0.00	0.02	0.00	0.00
sign test, excluding stand-alone sig. coefficients	0.00	0.01	0.02	0.09
Average B-factor effect, all standard B-factors	-0.028***	-0.028***	-0.026***	n/a
Average B-factor effect, excluding stand-alone sig. coeffs.	-0.015***	-0.015***	-0.011**	n/a

\* 0.10 \*\* 0.05 \*\*\* 0.01. Unit of observation is the individual, with one regression per cell and N=1,505 in each regression. Regressions are identical to those in Table 5, except for weighting in columns 2-4 (weighted models exclude survey effort deciles as controls). For B-factors with bi-directional biases we report only the standard bias indicator. The first column is identical to top row of Table 5. Regressions in the second column weight observations by the natural log of total time spent on all B-factor questions. Third column weights by decile of total time spent on all B-factor questions. Last column weights by the time spent on questions used to measure only the B-factor included in a given regression. Sign tests are against the null that the coefficient values are centered on zero. Average effects and standard errors are estimated by stacking the regressions and accounting for covariance across models. "Excluding stand-alone" group is, in each column, the set of coefficients with p>0.10. Average effects cannot be estimated across models in column (4) because each row employs a different weighting scheme.

**Table 7. Testing B-factor separability:****Do conditional correlations between financial condition and a B-factor (standard bias) change when all B-factors are included?**

	(1)	(2)	(3)	(4)
	Unweighted		Weighted by decile of BF survey time	
	Full spec.	Full + all B-factors	Full spec.	Full + all B-factors
Time-inconsistent money (present-bias)	-0.038**	-0.038**	-0.053***	-0.052***
Time-inconsistent snack (present-bias)	-0.020	-0.010	-0.019	-0.008
Violates GARP	-0.013	-0.011	-0.002	0.002
Violates GARP FOSD	-0.021	-0.020	-0.044	-0.045
Certainty premium (preference for certainty)	-0.006	-0.006	-0.005	-0.002
Loss averse	-0.010	-0.007	-0.021*	-0.015
Narrow-brackets	0.002	-0.001	-0.001	-0.006
Ambiguity averse	-0.025*	-0.025*	-0.024*	-0.021
Confidence in level performance (over-)	-0.020	0.017	-0.009	0.023
Overconfident precision	-0.018	-0.023*	-0.018	-0.024*
Overconfident relative perf.	-0.037**	-0.030*	-0.047***	-0.043**
Non-belief Law Large Numbers (underestimates)	-0.023	-0.015	-0.030	-0.019
Gambler's fallacies (hot hand)	-0.016	-0.020	-0.027	-0.026
Exponential growth bias loan (underestimates)	-0.006	-0.009	0.008	-0.004
Exponential growth bias asset (underestimates)	-0.063***	-0.062***	-0.050***	-0.051***
Limited attention	-0.109***	-0.106***	-0.111***	-0.106***
Limited memory	-0.042**	-0.039**	-0.048***	-0.048***

\* 0.10 \*\* 0.05 \*\*\* 0.01. Unit of observation is the individual, with one regression per cell in columns 1 and 3 and N=1,505 in each regression. Regressions are identical to those in Table 5, except for the addition of the full vector of B-factor indicators in columns 2 and 4 and the weighting in columns 3 and 4 (and when we weight we do not also include survey effort deciles as controls). For B-factors with bi-directional biases we report only the standard bias indicator: note that the first column is identical to top row of Table 5. Regression(s) in the second column are identical to those in the first but include all B-factor indicators at once, rather than estimating each in a separate regression. Regressions in the third column weight by total time spent responding to all B-factor elicitations (same as Table 6, column 3). Fourth column is identical to third but includes all B-factors in the model at once. No coefficients are different across the two models with  $p < 0.10$ .

**Table 8. Do B-factor (standard bias) coefficients change when other key covariates are dropped?**

	(1)	(2)	(3)	(4)	(5)
	Full specification	Remove cognitive skills	Remove classical preferences	Remove age, education, gender	Remove survey effort
Time-inconsistent money (present-bias)	-0.038**	-0.041**	-0.042***	-0.039**	-0.037**
Time-inconsistent snack (present-bias)	-0.020	-0.020	-0.017	-0.022	-0.019
Violates GARP	-0.013	-0.017	-0.011	-0.014	-0.013
Violates GARP FOSD	-0.021	-0.030	-0.022	-0.027	-0.022
Certainty premium (preference for certainty)	-0.006	-0.008	-0.005	-0.010	-0.005
Loss averse	-0.010	-0.010	-0.014	-0.011	-0.010
Narrow-brackets	0.002	0.000	0.001	0.000	0.002
Ambiguity averse	-0.025*	-0.021	-0.027*	-0.021	-0.024*
Confidence in level performance (over-)	-0.020	-0.022	-0.019	-0.027*	-0.017
Overconfident precision	-0.018	-0.016	-0.016	-0.015	-0.019
Overconfident relative perf.	-0.037**	-0.032**	-0.034*	-0.036**	-0.034**
Non-belief Law Large Numbers (underestimates)	-0.023	-0.030	-0.024	-0.034*	-0.024
Gambler's fallacies (hot hand)	-0.028	-0.035*	-0.030	-0.034*	-0.026
Exponential growth bias loan (underestimates)	-0.006	-0.006	-0.004	-0.007	-0.002
Exponential growth bias asset (underestimates)	-0.063***	-0.070***	-0.063***	-0.070***	-0.064***
Limited attention	-0.109***	-0.111***	-0.111***	-0.113***	-0.110***
Limited memory	-0.042**	-0.047**	-0.040**	-0.046**	-0.040**

\* 0.10 \*\* 0.05 \*\*\* 0.01. Unit of observation is the individual, with one regression per cell and N=1,505 in each regression. Regressions are identical to those in Table 5, except for the removal of the covariate sets indicated in each column label. For B-factors with bi-directional biases we report only the standard bias indicator: note that the first column is identical to top row of Table 5.

**Table 9. Are B-factors (standard biases) well-explained by all of the other covariates?**

	(1)	(2)	(3)	(4)
	Adjusted R-squared of all RHS variables			
measurement error correction?	no: unweighted	yes: weighted	no: unweighted	yes: weighted
LHS variable = B-factor (standard bias)	indicator		percentile	
Time-inconsistent money (present-bias)	0.01	0.02	0.01	0.03
Time-inconsistent snack (present-bias)	0.02	0.03	0.02	0.04
Violates GARP	0.03	0.04	0.04	0.05
Violates GARP FOSD	0.03	0.05	0.07	0.09
Certainty premium (preference for certainty)	0.01	0.03	0.04	0.07
Loss averse	0.07	0.07	0.07	0.08
Narrow-brackets	0.02	0.04	0.03	0.05
Ambiguity averse	0.02	0.04	0.03	0.03
Confidence in level performance (over-)	0.14	0.09	0.08	0.10
Overconfident precision	0.14	0.07	0.15	0.17
Overconfident relative perf.	0.15	0.15	0.21	0.22
Non-belief Law Large Numbers (underestimates)	0.08	0.08	0.12	0.15
Gambler's fallacies (hot hand)	0.07	0.09	0.04	0.06
Exponential growth bias loan (underestimates)	0.01	0.03	0.00	0.02
Exponential growth bias asset (underestimates)	0.24	0.23	0.22	0.25
Limited attention	0.02	0.05	0.03	0.06
Limited memory	0.03	0.03	0.03	0.03

\* 0.10 \*\* 0.05 \*\*\* 0.01. Unit of observation is the individual, with one OLS regression per column. LHS variable is the 0/1 B-factor indicator (for the standard directional bias where applicable) shown in the column header, or the B-factor percentile. Columns 2 and 4 are weighted using the decile of total time spent responding to B-factor elicitation. RHS variables are the full set of covariates described in Table 3, with two exceptions. The model for "over-confidence in level performance" excludes numeracy, which has a mechanical relationship to the B-factor because it forms the basis for over-confidence in number of answers correct. Also, the model for over-confidence in relative performance excludes fluid intelligence, for the same reason.

**For Online Publication:**  
**Appendix Tables and Data Appendix**



**Appendix Table 1. Does survey formatting bias toward worse financial condition?**

Variable	# of questions used	# per q.	orientation	response options		ordering details
				choice(s) indicating worse condition	placement of	
net worth>0	1	3	vertical	middle		Assets compared to debts? [Yes/no/about the same]
retirement assets>0	2	2	vertical	n/a*		"Enter total amount: \$[fill].00"
owns stocks	3	2	vertical	n/a*		"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)?"
				n/a*		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?
spent < income last 12 months	1	3	vertical	top		Spent [more than/same as/less than] income
financial satisfaction > median	1	slider	horizontal	left side of scale		0 to 100 point scale, lower numbers indicate lower satisfaction
retirement saving adequate	1	5	vertical	top		Ordered 1/5 from "not nearly enough" to "much more than enough"
non-retirement saving adequate	1	5	vertical	bottom		Ordered 1/5 from "wish my household saved a lot less" to "wish my household saved a lot more"
no severe distress last 12 mos	4	2	vertical	top		Yes/no for each question, with yes on top.
financial stress < median	1	slider	horizontal	right side of scale		0 to 100 point scale, higher numbers indicate higher stress

See Table 4 for additional details on variable definitions.

\* - these responses provided check-boxes indicating "zero" as answers, generally below the section for the continuous response.

**Appendix Table 2. Survey response time and financial condition components**

Survey time decile	Financial condition component outcomes: Share with indicator of better condition									Overall
	"Hard" outcomes: Balance sheet positions, flows, and events					"Soft" outcomes: Subjective perceptions				
	net worth>0	retirement assets>0	owns stocks	no severe distress last 12 months	spent < income last 12 months	financial satisfaction > median	retirement saving adequate	non-ret saving adequate	fin stress < median	
1	0.33	0.36	0.35	0.60	0.30	0.40	0.27	0.25	0.46	0.37
2	0.37	0.50	0.46	0.60	0.36	0.50	0.26	0.31	0.52	0.43
3	0.47	0.53	0.52	0.58	0.35	0.45	0.24	0.26	0.55	0.44
4	0.52	0.60	0.54	0.60	0.37	0.40	0.28	0.24	0.48	0.45
5	0.47	0.61	0.55	0.56	0.43	0.47	0.26	0.20	0.52	0.45
6	0.49	0.59	0.57	0.59	0.42	0.51	0.25	0.25	0.51	0.46
7	0.50	0.54	0.50	0.47	0.29	0.49	0.21	0.29	0.49	0.42
8	0.46	0.58	0.48	0.51	0.36	0.44	0.30	0.24	0.56	0.44
9	0.41	0.48	0.46	0.58	0.33	0.44	0.24	0.25	0.48	0.41
10	0.39	0.52	0.49	0.50	0.36	0.50	0.35	0.22	0.50	0.42

Notes: Survey time decile is for total survey completion time in minutes. Financial condition components are described in greater detail in Table 4.

Appendix Table 3. Univariate correlations between B-factors and other variables, by decile of survey effort

	Decile(s) of survey effort, measured as total spent responding across all B-factor elicitations			
	1	2-5	6-9	10
Financial condition	-0.09	-0.28	-0.34	-0.21
Cognitive skills (1st principal component)	-0.13	-0.32	-0.38	-0.40
Income (decile, ordinal)	-0.02	-0.21	-0.25	-0.15
Education (category, ordinal)	-0.12	-0.16	-0.26	-0.13

Notes: B-factors are summarized as the fraction of 17 B-factor indicators displayed at the respondent level. So each cell presents an estimate of a univariate correlation between that B-factor summary measure and the linearized variable described in the row label. Cognitive skill is a continuous principal component score derived from the four cognitive skill variables described in Table 3. Education categories are defined in Table 3 and coded on a 1-4 scale.

Appendix Table 4. Behavioral bias prevalence: Comparisons to prior work using representative samples

	(U.S. samples in bold)		
	Our sample	Prior work	
		Comp 1	Comp 2
Time-inconsistent money discounting: Present-biased	<b>0.26</b>	<b>0.29</b> <sup>1</sup>	<b>0.55</b> <sup>2</sup>
Time-inconsistent money discounting: Future-biased	<b>0.36</b>	<b>0.37</b>	
Time-inconsistent snack discounting: Present-biased	<b>0.15</b>	<b>0.06</b> <sup>1</sup>	
Time-inconsistent snack discounting: Future-biased	<b>0.07</b>	<b>0.09</b>	
Violates GARP	<b>0.53</b>	0.51 <sup>3</sup>	
Violates GARP plus dominance avoidance	<b>0.96</b>	0.96	
Loss-averse	<b>0.64</b>	<b>0.70</b> <sup>4</sup>	0.86 <sup>5</sup>
Narrow-brackets	<b>0.59</b>		<b>0.30</b> <sup>7</sup>
	<b>Task 2: 0.29</b>	<b>Task 2: 0.53</b> <sup>6</sup>	
	<b>Task 4: 0.50</b>	<b>Task 4: 0.67</b>	
Ambiguity-averse	<b>0.73</b>	<b>0.52</b> <sup>8</sup>	0.68 <sup>9</sup>
Gambler's Fallacy: Hot hand	<b>0.14</b>	0.10	
Gambler's fallacy: Cold hand	<b>0.26</b>	0.23 <sup>10</sup>	
Exponential growth bias, loan-side: Underestimates APR	<b>0.7</b>	<b>0.98</b> <sup>11</sup>	
Exponential growth bias, loan-side: Overestimates APR	<b>0.27</b>	<b>0.00</b>	
Exponential growth bias, asset-side: Underestimates FV	<b>0.47</b>	<b>0.69</b> <sup>2</sup>	<b>0.85</b> <sup>12</sup>
Exponential growth bias, asset-side: Overestimates FV	<b>0.09</b>	<b>0.09</b>	<b>0.11</b>

Notes: The B-factors not listed here but included in other tables are those for which we could not find a prevalence estimate from a representative sample. See Data Appendix for details on elicitations, prevalence and distributions. In some cases we take comparisons directly from prior work, and in others we use data from other papers to perform our own calculations. "GARP" = General Axiom of Revealed Preference. "APR" = Annual Percentage Rate. "FV" = Future Value.

Footnotes:

<sup>1</sup> - Barcellos and Carvahlo (2014), source data are from ALP.

<sup>2</sup> - Goda et al. (2017), sources are ALP and Understanding America Survey.

<sup>3</sup> - Choi et al. (2011), source is CentER panel (Netherlands).

<sup>4</sup> - Hwang (2016), source is ALP. We define loss aversion as rejecting one or more of the four small-stakes lotteries with positive expected value.

<sup>5</sup> - von Gaudeker et al. (2011), source is CentER panel (Netherlands).

<sup>6</sup> - Rabin and Weizacker (2009), source is KnowledgeNetworks

<sup>7</sup> - Gottlieb and Mitchell (2015), source is Health and Retirement Study (older Americans).

<sup>8</sup> - Dimmock et al. (2016), source is ALP.

<sup>9</sup> - Dimmock, Kouwenberg and Wakker (forthcoming), source is CentER panel (Netherlands).

<sup>10</sup> - Dohmen et al. (2009), source is German SocioEconomic Panel.

<sup>11</sup> - Stango and Zinman (2009, 2011), source is Survey of Consumer Finances.

<sup>12</sup> - Levy and Tasoff (2016), source is KnowledgeNetworks

Appendix Table 5. B-factor non-response by survey time spent

	Decile of survey time spent on B-factors										Total
	1	2	3	4	5	6	7	8	9	10	
Time-inconsistent money discounting	0.11	0.04	0.06	0.06	0.05	0.03	0.03	0.05	0.08	0.11	0.06
Time-inconsistent snack discounting	0.25	0.13	0.07	0.09	0.05	0.02	0.04	0.02	0.01	0.07	0.07
Violates GARP	0.38	0.19	0.16	0.15	0.15	0.15	0.12	0.07	0.10	0.14	0.16
Certainty premium	0.55	0.33	0.28	0.25	0.22	0.32	0.27	0.30	0.29	0.27	0.31
Loss aversion	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00
Narrow bracketing	0.03	0.01	0.02	0.03	0.03	0.01	0.00	0.01	0.01	0.04	0.02
Ambiguity aversion	0.29	0.15	0.06	0.08	0.04	0.05	0.05	0.01	0.01	0.04	0.08
(Over)confidence in performance	0.32	0.19	0.09	0.07	0.07	0.05	0.05	0.03	0.07	0.07	0.10
Overconfidence in precision	0.38	0.19	0.09	0.11	0.08	0.07	0.05	0.03	0.05	0.07	0.11
Overconfidence in relative performance	0.29	0.15	0.06	0.07	0.04	0.05	0.05	0.01	0.01	0.05	0.08
Non-belief Law Large Numbers	0.33	0.17	0.09	0.08	0.05	0.05	0.04	0.03	0.02	0.09	0.09
Gambler's fallacies	0.30	0.15	0.07	0.07	0.04	0.05	0.04	0.01	0.01	0.06	0.08
Exponential growth bias: loan	0.22	0.09	0.16	0.05	0.09	0.08	0.03	0.05	0.05	0.09	0.09
Exponential growth bias: asset	0.43	0.27	0.24	0.15	0.10	0.16	0.12	0.15	0.16	0.17	0.19
Limited attention	0.06	0.01	0.03	0.03	0.01	0.01	0.00	0.00	0.02	0.04	0.02
Limited memory	0.34	0.19	0.11	0.09	0.05	0.06	0.05	0.04	0.02	0.09	0.10
Average across B-factors	0.27	0.14	0.10	0.09	0.07	0.07	0.06	0.05	0.06	0.09	0.10
Mean survey time within decile, in minutes	2	14	19	22	26	29	34	41	53	354	59

Notes: Each cell represents the share of panelists with a missing value for that B-factor, by decile of total survey time spent on all B-factor questions. Results are similar by decile of time spent on questions for a particular B-factor.

**Appendix Table 6. Is behavioral bias prevalence related to survey/task response time (conditional on a response)?**

	Decile of survey time spent on B-factors										Total
	1	2	3	4	5	6	7	8	9	10	
Time-inconsistent money discounting: Present-biased	0.19	0.25	0.22	0.27	0.23	0.27	0.27	0.23	0.31	0.33	0.26
Time-inconsistent money discounting: Future-biased	0.25	0.33	0.34	0.37	0.38	0.30	0.34	0.40	0.38	0.38	0.36
Time-inconsistent snack discounting: Present-biased	0.05	0.15	0.18	0.08	0.17	0.13	0.15	0.13	0.19	0.19	0.15
Time-inconsistent snack discounting: Future-biased	0.16	0.07	0.09	0.04	0.03	0.08	0.07	0.11	0.06	0.08	0.07
Violates GARP (based on CCEI)	0.56	0.63	0.58	0.48	0.53	0.52	0.45	0.50	0.56	0.50	0.53
Violates GARP/FOSD	1.00	0.98	0.98	0.97	0.95	0.97	0.93	0.95	0.95	0.94	0.96
Preference for certainty type (positive certainty premium)	0.82	0.74	0.69	0.82	0.82	0.73	0.77	0.83	0.78	0.74	0.77
Cumulative prospect theory type (negative certainty premium)	0.18	0.26	0.31	0.18	0.18	0.27	0.23	0.17	0.22	0.26	0.23
Loss-averse	0.61	0.56	0.61	0.69	0.66	0.67	0.67	0.63	0.60	0.62	0.64
Narrow-brackets	0.68	0.70	0.58	0.57	0.57	0.55	0.54	0.60	0.59	0.62	0.59
Ambiguity-averse	0.53	0.67	0.71	0.75	0.71	0.75	0.75	0.72	0.79	0.77	0.73
Overconfident in level performance	0.38	0.38	0.29	0.37	0.32	0.34	0.43	0.39	0.46	0.47	0.38
Underconfident in level performance	0.25	0.12	0.11	0.10	0.11	0.16	0.10	0.12	0.09	0.07	0.11
Overconfident in precision	0.27	0.37	0.43	0.43	0.49	0.51	0.45	0.44	0.45	0.39	0.44
Overconfident in relative performance	0.59	0.57	0.55	0.51	0.45	0.45	0.42	0.48	0.51	0.48	0.50
Non-belief in the law of large numbers: Underestimates	0.88	0.89	0.87	0.85	0.88	0.86	0.89	0.88	0.86	0.87	0.87
Non-belief in the law of large numbers: Overestimates	0.13	0.11	0.13	0.15	0.12	0.14	0.11	0.12	0.14	0.13	0.13
Gambler's Fallacy: Cold hand	0.41	0.24	0.30	0.22	0.26	0.21	0.28	0.31	0.28	0.25	0.26
Gambler's Fallacy: Hot hand	0.12	0.16	0.10	0.11	0.13	0.13	0.14	0.16	0.17	0.13	0.14
Exponential growth bias, loan-side: Underestimates APR	0.69	0.71	0.77	0.65	0.78	0.67	0.73	0.75	0.64	0.59	0.70
Exponential growth bias, loan-side: Overestimates APR	0.31	0.24	0.22	0.34	0.18	0.25	0.26	0.24	0.33	0.37	0.27
Exponential growth bias, asset-side: Underestimates FV	0.56	0.59	0.44	0.44	0.41	0.44	0.46	0.45	0.48	0.52	0.47
Exponential growth bias, asset-side: Overestimates FV	0.33	0.10	0.10	0.10	0.09	0.03	0.08	0.13	0.09	0.07	0.09
Limited attention	0.47	0.50	0.48	0.40	0.53	0.47	0.46	0.51	0.53	0.49	0.48
Limited memory	1.00	0.88	0.89	0.87	0.92	0.81	0.80	0.83	0.87	0.82	0.86
Average prevalence across standard B-factors	0.55	0.57	0.55	0.54	0.56	0.54	0.55	0.56	0.57	0.56	0.56
Mean survey time within decile, in minutes	2	14	19	22	26	29	34	41	53	354	59

Notes: Each cell shows the share of respondents measured as exhibiting the bias indicated in the row label, conditional on providing a response. See Appendix Table 5 for data on missing values as a function of survey response time, by B-factor. Survey time spent is measured across all B-factor questions. Results are similar using only time spent on questions for a particular B-factor.

**Appendix Table 7. Estimating equation (1) for each B-factor, with percentiles instead of indicators: Conditional correlations between financial condition and behavioral biases**  
 (compare to Table 5)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
	RHS B-factor shown as column header. LHS variable in all models is summary financial condition (mean = 0.43)																
	Time-inconsistent: money	Time-inconsistent: snack	Violates GARP	Violates GARP FOSD	Certainty premium	Loss averse	Narrow-brackets	Ambiguity averse	(Over-)confident performance	Over-confident precision	Over-confident relative perf.	Non-belief Law Large Numbers	Gambler's Fallacies	Exponential growth bias: loan	Exponential growth bias: asset	Limited attention	Limited memory
Standard bias percentile	-0.045** (0.019)	-0.024 (0.021)	-0.014 (0.017)	-0.017 (0.024)	-0.006 (0.023)	-0.021 (0.023)	0.007 (0.018)	-0.022 (0.019)	-0.028 (0.022)	-0.029 (0.021)	-0.085** (0.036)	-0.048** (0.022)	-0.027* (0.017)	-0.005 (0.024)	-0.055** (0.023)	-0.162*** (0.017)	-0.278** (0.123)
Non-standard bias percentile	-0.005 (0.019)	-0.019 (0.027)							0.043* (0.024)					0.004 (0.024)	-0.023 (0.028)		
bias missing	-0.009 (0.028)	0.033 (0.051)	-0.011 (0.023)	-0.013 (0.025)	0.017 (0.019)	-0.141 (0.139)	0.034 (0.047)	0.272 (0.169)	0.053 (0.109)	0.066* (0.036)		0.108** (0.055)	-0.097 (0.099)	0.020 (0.029)	-0.001 (0.025)	-0.028 (0.046)	0.017 (0.042)
female	-0.020 (0.013)	-0.020 (0.013)	-0.017 (0.013)	-0.017 (0.013)	-0.018 (0.013)	-0.020 (0.013)	-0.020 (0.013)	-0.019 (0.014)	-0.020 (0.013)	-0.019 (0.013)	-0.019 (0.013)	-0.020 (0.013)	-0.019 (0.013)	-0.019 (0.013)	-0.019 (0.013)	-0.018 (0.013)	-0.018 (0.013)
ed: some college or associates	-0.040** (0.018)	-0.041** (0.018)	-0.040** (0.018)	-0.039** (0.018)	-0.040** (0.018)	-0.040** (0.018)	-0.041** (0.018)	-0.049*** (0.019)	-0.043** (0.018)	-0.043** (0.018)	-0.040** (0.018)	-0.042** (0.018)	-0.042** (0.018)	-0.044** (0.018)	-0.044** (0.018)	-0.039** (0.017)	-0.041** (0.018)
highest ed: bachelor's	0.017 (0.021)	0.015 (0.021)	0.016 (0.021)	0.016 (0.021)	0.015 (0.021)	0.015 (0.021)	0.014 (0.021)	0.012 (0.021)	0.012 (0.021)	0.011 (0.021)	0.015 (0.021)	0.012 (0.021)	0.014 (0.021)	0.011 (0.021)	0.013 (0.021)	0.004 (0.020)	0.013 (0.021)
highest ed: graduate	0.032 (0.024)	0.029 (0.024)	0.030 (0.024)	0.030 (0.024)	0.031 (0.024)	0.032 (0.024)	0.030 (0.024)	0.029 (0.025)	0.027 (0.024)	0.026 (0.024)	0.033 (0.024)	0.026 (0.024)	0.028 (0.024)	0.028 (0.024)	0.028 (0.024)	0.011 (0.023)	0.027 (0.024)
age 35-45	-0.016 (0.018)	-0.010 (0.018)	-0.012 (0.018)	-0.012 (0.018)	-0.015 (0.018)	-0.015 (0.018)	-0.016 (0.018)	-0.011 (0.019)	-0.012 (0.018)	-0.013 (0.018)	-0.013 (0.018)	-0.013 (0.018)	-0.012 (0.018)	-0.016 (0.018)	-0.015 (0.018)	-0.016 (0.017)	-0.015 (0.018)
age 46-54	0.004 (0.019)	0.006 (0.019)	0.002 (0.019)	0.003 (0.019)	0.004 (0.019)	0.002 (0.019)	-0.000 (0.019)	0.002 (0.020)	0.007 (0.019)	0.004 (0.019)	0.004 (0.019)	0.003 (0.019)	0.005 (0.019)	-0.002 (0.019)	-0.001 (0.019)	-0.003 (0.018)	0.001 (0.019)
age>=55	0.053** (0.021)	0.057*** (0.021)	0.054** (0.021)	0.054** (0.021)	0.056*** (0.021)	0.050** (0.021)	0.049** (0.021)	0.051** (0.022)	0.055*** (0.021)	0.054** (0.021)	0.053** (0.021)	0.046** (0.021)	0.053** (0.021)	0.047** (0.021)	0.046** (0.021)	0.045** (0.020)	0.049** (0.021)
income decile 2	-0.029 (0.027)	-0.029 (0.027)	-0.028 (0.027)	-0.027 (0.027)	-0.024 (0.027)	-0.030 (0.027)	-0.028 (0.028)	-0.029 (0.027)	-0.020 (0.027)	-0.023 (0.027)	-0.020 (0.027)	-0.026 (0.027)	-0.030 (0.027)	-0.025 (0.027)	-0.025 (0.027)	-0.018 (0.026)	-0.030 (0.027)
income decile 3	0.018 (0.028)	0.021 (0.028)	0.021 (0.028)	0.022 (0.028)	0.027 (0.028)	0.019 (0.028)	0.022 (0.028)	0.027 (0.029)	0.029 (0.028)	0.026 (0.028)	0.028 (0.028)	0.028 (0.028)	0.019 (0.028)	0.025 (0.028)	0.027 (0.028)	0.039 (0.027)	0.022 (0.028)
income decile 4	0.074*** (0.028)	0.076*** (0.028)	0.078*** (0.028)	0.079*** (0.028)	0.080*** (0.028)	0.073*** (0.028)	0.079*** (0.028)	0.074** (0.029)	0.084*** (0.028)	0.084*** (0.028)	0.085*** (0.028)	0.080*** (0.028)	0.073*** (0.028)	0.084*** (0.028)	0.083*** (0.028)	0.097*** (0.027)	0.083*** (0.028)
income decile 5	0.086*** (0.031)	0.089*** (0.031)	0.090*** (0.031)	0.091*** (0.031)	0.094*** (0.031)	0.085*** (0.031)	0.091*** (0.031)	0.091*** (0.032)	0.096*** (0.031)	0.091*** (0.031)	0.096*** (0.031)	0.096*** (0.031)	0.083*** (0.031)	0.095*** (0.031)	0.094*** (0.031)	0.117*** (0.030)	0.094*** (0.031)
income decile 6	0.137*** (0.031)	0.142*** (0.031)	0.139*** (0.031)	0.139*** (0.031)	0.146*** (0.031)	0.136*** (0.031)	0.141*** (0.031)	0.130*** (0.032)	0.144*** (0.031)	0.145*** (0.031)	0.146*** (0.031)	0.147*** (0.031)	0.138*** (0.031)	0.149*** (0.031)	0.146*** (0.031)	0.172*** (0.030)	0.146*** (0.031)
income decile 7	0.174*** (0.031)	0.177*** (0.031)	0.176*** (0.030)	0.176*** (0.030)	0.176*** (0.031)	0.171*** (0.031)	0.178*** (0.031)	0.167*** (0.032)	0.186*** (0.030)	0.179*** (0.031)	0.186*** (0.031)	0.180*** (0.031)	0.168*** (0.031)	0.182*** (0.031)	0.184*** (0.031)	0.201*** (0.030)	0.178*** (0.030)
income decile 8	0.261*** (0.032)	0.263*** (0.032)	0.266*** (0.031)	0.266*** (0.031)	0.271*** (0.031)	0.262*** (0.032)	0.268*** (0.032)	0.274*** (0.033)	0.271*** (0.031)	0.268*** (0.031)	0.269*** (0.031)	0.268*** (0.031)	0.260*** (0.032)	0.273*** (0.032)	0.272*** (0.032)	0.287*** (0.031)	0.272*** (0.031)
income decile 9	0.295*** (0.030)	0.297*** (0.030)	0.299*** (0.030)	0.299*** (0.030)	0.303*** (0.030)	0.295*** (0.030)	0.304*** (0.030)	0.290*** (0.032)	0.305*** (0.030)	0.304*** (0.030)	0.306*** (0.030)	0.304*** (0.030)	0.294*** (0.030)	0.308*** (0.030)	0.302*** (0.030)	0.323*** (0.029)	0.308*** (0.030)
income decile 10	0.374*** (0.044)	0.378*** (0.044)	0.379*** (0.044)	0.380*** (0.044)	0.375*** (0.044)	0.374*** (0.044)	0.382*** (0.045)	0.376*** (0.047)	0.384*** (0.044)	0.380*** (0.044)	0.384*** (0.044)	0.388*** (0.044)	0.368*** (0.044)	0.391*** (0.045)	0.384*** (0.044)	0.391*** (0.043)	0.387*** (0.044)

continued on next page...

Appendix Table 7, cont.'d

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
	Time- inconsistent: money	Time- inconsistent: snack	Violates GARP	Violates GARP FOSD	Certainty premium	Loss averse	Narrow- brackets	Ambiguity averse	(Over-) confident performance	Over- confident precision	Over- confident relative perf.	Non-belief Law Large Numbers	Gambler's Fallacies	Exponential growth bias: loan	Exponential growth bias: asset	Limited attention	Limited memory
fluid intellig # correct	0.001 (0.003)	0.001 (0.003)	0.000 (0.003)	0.000 (0.003)	0.002 (0.003)	0.001 (0.003)	0.001 (0.003)	0.001 (0.003)	0.001 (0.003)	0.001 (0.003)	-0.007 (0.004)	-0.000 (0.003)	-0.000 (0.003)	0.001 (0.003)	-0.000 (0.003)	0.001 (0.003)	0.001 (0.003)
numeracy # correct	-0.001 (0.013)	0.000 (0.013)	-0.000 (0.013)	0.000 (0.013)	0.003 (0.013)	-0.001 (0.013)	-0.000 (0.013)	0.002 (0.013)	-0.011 (0.014)	0.005 (0.013)	0.002 (0.013)	0.000 (0.013)	0.002 (0.013)	0.002 (0.013)	-0.001 (0.013)	0.001 (0.013)	-0.001 (0.013)
financial literacy # correct	0.028*** (0.009)	0.029*** (0.009)	0.029*** (0.009)	0.029*** (0.009)	0.030*** (0.009)	0.028*** (0.009)	0.028*** (0.009)	0.030*** (0.009)	0.030*** (0.009)	0.031*** (0.009)	0.029*** (0.009)	0.025*** (0.009)	0.027*** (0.009)	0.029*** (0.009)	0.024*** (0.009)	0.026*** (0.009)	0.028*** (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)	0.000 (0.000)
Risk aversion (financial)	-0.006*** (0.002)	-0.006*** (0.002)	-0.007*** (0.002)	-0.007*** (0.002)	-0.006*** (0.002)	-0.006** (0.002)	-0.006*** (0.002)	-0.006** (0.002)	-0.006*** (0.002)	-0.007*** (0.002)	-0.007*** (0.002)	-0.006*** (0.002)	-0.006*** (0.002)	-0.006*** (0.002)	-0.006*** (0.002)	-0.005** (0.002)	-0.006** (0.002)
Risk aversion (income)	0.009* (0.005)	0.010* (0.005)	0.009* (0.005)	0.009* (0.005)	0.009* (0.005)	0.009* (0.005)	0.009* (0.005)	0.006 (0.005)	0.009* (0.005)	0.009* (0.005)	0.009* (0.005)	0.009* (0.005)	0.008 (0.005)	0.009* (0.005)	0.009* (0.005)	0.007 (0.005)	0.009* (0.005)
patience (stdized)	0.006 (0.007)	0.007 (0.006)	0.008 (0.006)	0.008 (0.006)	0.007 (0.006)	0.008 (0.006)	0.008 (0.006)	0.009 (0.007)	0.008 (0.006)	0.007 (0.006)	0.007 (0.006)	0.008 (0.006)	0.007 (0.006)	0.007 (0.006)	0.008 (0.006)	0.005 (0.006)	0.009 (0.006)
Response time spent on the B-factor included in this regression:																	
decile 2	-0.000 (0.027)	-0.043 (0.027)	0.007 (0.028)	0.007 (0.028)	-0.029 (0.028)	0.017 (0.026)	0.003 (0.027)	-0.033 (0.028)	0.009 (0.024)	-0.031 (0.026)	0.005 (0.024)	0.036 (0.028)	-0.026 (0.028)	-0.028 (0.028)	-0.004 (0.027)	-0.051* (0.026)	0.017 (0.026)
decile 3	-0.005 (0.027)	-0.040 (0.027)	0.017 (0.029)	0.017 (0.029)	-0.001 (0.028)	0.018 (0.026)	-0.010 (0.027)	-0.062** (0.028)	0.038 (0.027)	0.011 (0.027)	0.030 (0.027)	0.043 (0.029)	-0.007 (0.026)	-0.041 (0.029)	0.039 (0.027)	-0.048* (0.027)	0.008 (0.028)
decile 4	0.017 (0.028)	-0.011 (0.028)	-0.005 (0.029)	-0.005 (0.029)	-0.046* (0.028)	0.059** (0.027)	0.010 (0.027)	-0.010 (0.029)	-0.037 (0.029)	0.013 (0.027)	-0.041 (0.029)	0.024 (0.028)	0.013 (0.028)	-0.040 (0.029)	0.012 (0.029)	-0.029 (0.026)	-0.037 (0.026)
decile 5	-0.002 (0.027)	-0.017 (0.026)	0.027 (0.030)	0.027 (0.030)	0.021 (0.028)	0.037 (0.027)	-0.001 (0.027)	-0.003 (0.028)	0.008 (0.030)	-0.014 (0.027)	0.000 (0.030)	0.044 (0.029)	0.032 (0.028)	-0.048 (0.029)	0.025 (0.027)	-0.031 (0.026)	-0.020 (0.029)
decile 6	0.017 (0.028)	-0.050* (0.027)	-0.028 (0.029)	-0.028 (0.029)	-0.045 (0.028)	0.032 (0.027)	-0.002 (0.027)	-0.039 (0.027)	0.002 (0.031)	0.015 (0.028)	-0.005 (0.031)	0.047 (0.030)	0.022 (0.026)	-0.024 (0.029)	0.007 (0.028)	-0.003 (0.027)	0.034 (0.029)
decile 7	0.005 (0.028)	-0.033 (0.028)	-0.003 (0.030)	-0.003 (0.030)	-0.019 (0.028)	0.009 (0.028)	0.014 (0.028)	-0.022 (0.029)	-0.032 (0.028)	-0.025 (0.028)	-0.037 (0.028)	0.004 (0.029)	-0.032 (0.027)	-0.026 (0.029)	0.054* (0.029)	-0.045* (0.026)	0.031 (0.027)
decile 8	0.012 (0.028)	-0.024 (0.028)	0.007 (0.029)	0.008 (0.029)	-0.030 (0.029)	0.013 (0.027)	0.017 (0.027)	-0.027 (0.028)	0.026 (0.029)	-0.016 (0.028)	0.020 (0.028)	0.027 (0.029)	-0.024 (0.028)	0.003 (0.030)	0.003 (0.028)	-0.035 (0.027)	-0.029 (0.027)
decile 9	-0.023 (0.028)	-0.018 (0.028)	0.016 (0.030)	0.016 (0.030)	0.007 (0.028)	0.001 (0.027)	0.002 (0.027)	-0.012 (0.028)	-0.031 (0.028)	0.024 (0.028)	-0.035 (0.028)	0.029 (0.030)	-0.017 (0.027)	-0.028 (0.030)	0.014 (0.028)	-0.035 (0.027)	0.018 (0.028)
decile 10	0.004 (0.028)	-0.014 (0.028)	-0.031 (0.029)	-0.031 (0.029)	-0.036 (0.028)	0.017 (0.027)	0.016 (0.027)	-0.037 (0.028)	0.051* (0.029)	-0.023 (0.028)	0.043 (0.029)	0.055* (0.029)	-0.006 (0.027)	-0.003 (0.029)	0.029 (0.028)	-0.030 (0.027)	0.016 (0.028)
Full set of controls from Table 3?	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R-squared	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.40	0.39	0.39	0.39	0.39	0.40	0.39	0.43	0.39
Number of observations	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505

\* 0.10 \*\* 0.05 \*\*\* 0.01. These are the same specifications as in Table 5, but with behavioral bias(es) measured as percentiles instead of indicators.



Appendix Table 8. Are B-factor (standard biases) well-explained by other covariates? Coefficients on some of the RHS variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
Dependent variable shown as column header and is the standard bias indicator for B-factors with potentially bi-directional biases.																	
LHS Variable	Present-biased \$	Present-biased snack	Violates GARP	Violates GARP	Prof. for certainty	Loss averse	Narrow-brackets	Ambiguity averse	Overconf level perf.	Overconf precision	Overconf relative perf.	NBLN underest	Gambler's hot hand	EGB loan underest	EGB asset underest	Limited attention	Limited memory
Mean(LHS), non-missing values	0.26	0.15	0.53	0.96	0.77	0.64	0.59	0.73	0.38	0.44	0.50	0.87	0.14	0.70	0.47	0.48	0.86
female	-0.013 (0.026)	-0.005 (0.022)	0.057* (0.031)	0.025** (0.012)	0.018 (0.030)	-0.004 (0.027)	0.014 (0.029)	-0.049* (0.027)	-0.003 (0.029)	-0.068** (0.028)	0.021 (0.028)	0.074*** (0.020)	0.006 (0.020)	-0.042 (0.030)	0.069** (0.028)	0.019 (0.029)	0.009 (0.021)
age 35-45	0.003 (0.035)	0.010 (0.030)	0.006 (0.043)	0.026 (0.017)	0.079* (0.040)	0.003 (0.037)	-0.000 (0.039)	-0.054 (0.036)	0.038 (0.040)	0.192*** (0.039)	0.053 (0.038)	0.008 (0.027)	0.010 (0.028)	0.055 (0.041)	0.089** (0.039)	-0.055 (0.039)	0.013 (0.029)
age 46-54	0.056 (0.038)	0.033 (0.031)	-0.021 (0.045)	0.026 (0.018)	0.079* (0.043)	-0.001 (0.039)	-0.020 (0.041)	-0.033 (0.038)	0.087** (0.042)	0.227*** (0.041)	0.097** (0.040)	0.011 (0.028)	0.041 (0.029)	0.041 (0.044)	0.044 (0.041)	-0.104** (0.042)	-0.000 (0.031)
age>=55	0.046 (0.042)	0.038 (0.034)	0.035 (0.051)	0.030 (0.020)	0.095** (0.047)	-0.005 (0.043)	-0.035 (0.045)	0.003 (0.042)	0.015 (0.046)	0.257*** (0.045)	0.121*** (0.044)	-0.019 (0.031)	0.006 (0.032)	0.082* (0.048)	0.026 (0.045)	-0.107** (0.046)	-0.024 (0.034)
ed: some college or associates	0.031 (0.036)	0.043 (0.029)	0.041 (0.043)	0.020 (0.017)	0.039 (0.044)	0.013 (0.036)	0.046 (0.039)	0.098*** (0.036)	0.050 (0.040)	0.017 (0.039)	-0.022 (0.038)	0.019 (0.027)	0.015 (0.028)	-0.041 (0.041)	0.009 (0.040)	-0.014 (0.039)	0.016 (0.029)
highest ed: bachelor's	0.030 (0.041)	0.004 (0.034)	-0.019 (0.050)	0.010 (0.020)	-0.017 (0.047)	0.042 (0.042)	0.029 (0.045)	0.115*** (0.042)	-0.007 (0.046)	0.035 (0.045)	-0.089** (0.044)	-0.021 (0.031)	-0.012 (0.032)	-0.063 (0.047)	0.019 (0.045)	-0.098** (0.045)	-0.009 (0.034)
highest ed: graduate	0.024 (0.047)	0.021 (0.039)	0.028 (0.056)	0.000 (0.022)	-0.058 (0.052)	-0.030 (0.048)	0.063 (0.051)	0.136*** (0.048)	-0.045 (0.053)	-0.052 (0.051)	-0.119** (0.050)	-0.053 (0.035)	-0.032 (0.037)	-0.054 (0.054)	-0.026 (0.051)	-0.150*** (0.052)	-0.012 (0.039)
income decile 2	-0.044 (0.054)	0.003 (0.043)	-0.010 (0.062)	0.000 (0.025)	-0.057 (0.072)	-0.061 (0.055)	-0.035 (0.058)	0.057 (0.053)	0.011 (0.059)	0.079 (0.058)	-0.040 (0.056)	-0.022 (0.040)	0.102** (0.041)	-0.029 (0.065)	0.033 (0.061)	0.086 (0.059)	-0.003 (0.043)
income decile 3	-0.121** (0.056)	-0.021 (0.045)	-0.051 (0.066)	-0.013 (0.026)	-0.090 (0.071)	0.036 (0.056)	-0.072 (0.060)	0.030 (0.056)	-0.016 (0.062)	0.002 (0.061)	-0.054 (0.059)	-0.036 (0.042)	0.030 (0.043)	-0.010 (0.066)	0.053 (0.063)	0.152** (0.061)	0.068 (0.045)
income decile 4	-0.065 (0.057)	0.034 (0.045)	-0.028 (0.066)	-0.009 (0.026)	-0.087 (0.072)	0.018 (0.057)	-0.042 (0.060)	-0.006 (0.056)	-0.028 (0.061)	0.009 (0.061)	0.019 (0.059)	0.007 (0.041)	0.030 (0.043)	0.049 (0.067)	0.003 (0.062)	0.159*** (0.061)	0.052 (0.045)
income decile 5	-0.054 (0.062)	-0.017 (0.050)	-0.065 (0.071)	-0.033 (0.028)	-0.100 (0.077)	0.019 (0.063)	-0.131** (0.066)	0.043 (0.061)	-0.084 (0.067)	0.009 (0.067)	0.043 (0.067)	-0.023 (0.065)	-0.023 (0.045)	-0.021 (0.047)	0.028 (0.073)	0.232*** (0.068)	0.063 (0.049)
income decile 6	-0.107* (0.062)	-0.018 (0.050)	0.054 (0.073)	-0.017 (0.029)	-0.057 (0.076)	0.112* (0.063)	-0.147** (0.067)	-0.028 (0.062)	-0.088 (0.068)	-0.026 (0.068)	-0.014 (0.066)	0.037 (0.046)	0.032 (0.048)	0.097 (0.073)	-0.095 (0.067)	0.269*** (0.067)	0.081 (0.050)
income decile 7	-0.046 (0.061)	-0.020 (0.050)	-0.046 (0.072)	-0.025 (0.029)	-0.074 (0.075)	0.136** (0.063)	-0.217*** (0.066)	-0.000 (0.062)	-0.054 (0.068)	-0.031 (0.067)	0.114* (0.065)	0.037 (0.046)	-0.011 (0.048)	0.021 (0.073)	-0.017 (0.068)	0.219*** (0.067)	0.091* (0.050)
income decile 8	-0.123* (0.063)	-0.040 (0.052)	-0.111 (0.075)	-0.038 (0.030)	-0.048 (0.077)	0.121* (0.064)	-0.141** (0.068)	0.069 (0.063)	-0.122* (0.070)	-0.162** (0.070)	-0.010 (0.067)	-0.040 (0.047)	0.016 (0.049)	0.018 (0.075)	-0.075 (0.069)	0.186*** (0.069)	0.076 (0.051)
income decile 9	-0.092 (0.061)	-0.043 (0.049)	-0.131* (0.072)	-0.045 (0.029)	-0.076 (0.073)	0.072 (0.061)	-0.205*** (0.065)	0.013 (0.061)	-0.162** (0.067)	-0.109 (0.066)	-0.073 (0.064)	-0.039 (0.045)	0.051 (0.047)	-0.016 (0.072)	-0.142** (0.067)	0.206*** (0.066)	0.154*** (0.049)
income decile 10	-0.050 (0.088)	0.074 (0.072)	-0.045 (0.105)	0.041 (0.042)	-0.094 (0.100)	0.035 (0.091)	-0.105 (0.095)	0.086 (0.089)	-0.191** (0.096)	-0.014 (0.095)	0.044 (0.094)	0.012 (0.066)	-0.055 (0.069)	0.047 (0.103)	-0.194** (0.095)	0.182* (0.097)	0.259*** (0.073)

continued on next page...

Appendix Table 8, cont.'d

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
Dependent variable shown as column header and is the standard bias indicator for B-factors with potentially bi-directional biases.																	
Variable	Time-inconsistent:	Time-inconsistent:	Violates GARP	Violates GARP	Prof. for certainty	Loss averse	Narrow-brackets	Ambiguity averse	Over-confident	Over-confident	Over-confident	Non-belief Law Large	Gambler's fallacy (hot)	Exponential growth bias:	Exponential growth bias:	Limited attention	Limited memory
fluid intell # correct	0.005 (0.006)	-0.003 (0.005)	-0.011* (0.007)	-0.005** (0.003)	0.006 (0.007)	0.010* (0.006)	-0.020*** (0.006)	-0.000 (0.006)	-0.011* (0.006)	0.021*** (0.006)							
numeracy # correct	-0.030 (0.027)	0.000 (0.021)	-0.036 (0.030)	0.013 (0.012)	-0.016 (0.035)	-0.052* (0.027)	0.025 (0.029)	0.048* (0.026)		0.078*** (0.028)	-0.090*** (0.026)	0.005 (0.019)	-0.005 (0.020)	-0.058* (0.031)	0.060** (0.028)	0.010 (0.029)	-0.022 (0.021)
financial literacy # correct	-0.007 (0.018)	0.008 (0.014)	-0.025 (0.021)	-0.015* (0.008)	-0.028 (0.021)	-0.010 (0.018)	-0.017 (0.019)	0.022 (0.018)	-0.028 (0.019)	-0.017 (0.019)	-0.072*** (0.018)	-0.029** (0.013)	-0.021 (0.013)	0.003 (0.021)	-0.096*** (0.020)	-0.026 (0.020)	-0.034** (0.014)
exec attention # correct	-0.000 (0.001)	-0.000 (0.000)	-0.001 (0.001)	-0.000 (0.000)	-0.001 (0.001)	0.002*** (0.001)	0.001 (0.001)	-0.001 (0.001)	-0.000 (0.001)	-0.001 (0.001)	-0.002*** (0.001)	0.001 (0.000)	0.000 (0.000)	0.000 (0.001)	-0.001** (0.001)	-0.000 (0.001)	-0.000 (0.000)
Risk aversion (financial)	0.000 (0.005)	-0.007* (0.004)	-0.011** (0.005)	0.001 (0.002)	-0.000 (0.005)	0.024*** (0.005)	-0.002 (0.005)	0.020*** (0.004)	-0.005 (0.005)	-0.006 (0.005)	-0.016*** (0.005)	0.001 (0.003)	0.004 (0.003)	0.005 (0.005)	-0.001 (0.005)	0.008 (0.005)	-0.002 (0.004)
Risk aversion (income)	-0.010 (0.010)	0.005 (0.008)	0.011 (0.012)	0.005 (0.005)	0.004 (0.012)	0.007 (0.010)	-0.020* (0.011)	0.020** (0.010)	0.008 (0.011)	0.015 (0.011)	0.015 (0.010)	0.002 (0.007)	-0.014* (0.008)	-0.007 (0.011)	-0.002 (0.011)	-0.017 (0.011)	0.000 (0.008)
patience (stdized)	-0.026** (0.012)	0.006 (0.010)	-0.001 (0.015)	-0.004 (0.006)	0.011 (0.014)	-0.029** (0.013)	0.014 (0.013)	-0.030** (0.012)	-0.003 (0.014)	0.001 (0.013)	-0.014 (0.013)	-0.009 (0.009)	0.005 (0.010)	-0.010 (0.014)	-0.009 (0.013)	-0.018 (0.014)	0.005 (0.010)
Survey time spent decile 2	0.103* (0.054)	0.051 (0.042)	0.015 (0.070)	0.003 (0.028)	-0.175** (0.078)	-0.046 (0.053)	-0.032 (0.059)	0.005 (0.053)	-0.051 (0.051)	-0.112* (0.060)	-0.088 (0.055)	0.019 (0.040)	0.035 (0.042)	0.009 (0.069)	-0.106* (0.057)	-0.042 (0.060)	0.045 (0.040)
Time spent decile 3	0.039 (0.055)	0.001 (0.043)	-0.004 (0.070)	0.030 (0.028)	-0.243*** (0.078)	-0.025 (0.054)	-0.044 (0.059)	0.032 (0.054)	0.064 (0.058)	-0.161*** (0.061)	-0.052 (0.060)	0.081** (0.041)	0.045 (0.038)	-0.030 (0.070)	-0.097* (0.057)	-0.062 (0.061)	-0.039 (0.044)
Time spent decile 4	0.016 (0.056)	0.087* (0.045)	-0.068 (0.071)	0.021 (0.028)	-0.208*** (0.076)	0.067 (0.055)	0.029 (0.058)	0.027 (0.054)	0.000 (0.061)	-0.313*** (0.060)	-0.060 (0.057)	0.066 (0.041)	0.072* (0.041)	0.030 (0.069)	-0.156** (0.061)	0.005 (0.060)	-0.028 (0.045)
Time spent decile 5	-0.006 (0.055)	-0.008 (0.042)	-0.079 (0.071)	0.017 (0.028)	-0.151* (0.077)	0.069 (0.055)	-0.031 (0.058)	0.054 (0.053)	0.037 (0.063)	-0.277*** (0.061)	-0.075 (0.056)	0.025 (0.041)	0.115*** (0.041)	-0.015 (0.071)	-0.187*** (0.057)	0.026 (0.060)	-0.041 (0.045)
Time spent decile 6	0.101* (0.055)	-0.013 (0.043)	-0.061 (0.071)	0.018 (0.028)	-0.211*** (0.077)	0.015 (0.055)	-0.063 (0.058)	0.053 (0.052)	-0.005 (0.065)	-0.380*** (0.061)	-0.237*** (0.056)	0.043 (0.043)	0.043 (0.040)	-0.072 (0.070)	-0.230*** (0.059)	-0.089 (0.061)	0.030 (0.046)
Time spent decile 7	0.115** (0.055)	0.075* (0.045)	-0.040 (0.073)	0.021 (0.029)	-0.218*** (0.079)	0.037 (0.057)	-0.024 (0.060)	0.010 (0.056)	0.071 (0.060)	-0.425*** (0.062)	-0.128** (0.057)	0.068 (0.041)	0.050 (0.040)	-0.068 (0.071)	-0.327*** (0.061)	-0.029 (0.060)	-0.046 (0.042)
Time spent decile 8	0.143** (0.056)	0.003 (0.044)	-0.173** (0.071)	-0.016 (0.028)	-0.201** (0.078)	0.065 (0.055)	-0.013 (0.059)	-0.019 (0.054)	0.015 (0.061)	-0.491*** (0.062)	-0.122** (0.057)	0.076* (0.042)	0.001 (0.042)	-0.093 (0.071)	-0.268*** (0.059)	0.062 (0.061)	-0.057 (0.043)
Time spent decile 9	0.094* (0.056)	0.056 (0.044)	-0.131* (0.071)	-0.014 (0.028)	-0.233*** (0.078)	0.146*** (0.055)	-0.042 (0.059)	0.083 (0.053)	-0.026 (0.060)	-0.456*** (0.062)	-0.172*** (0.057)	0.145*** (0.042)	0.060 (0.041)	-0.025 (0.071)	-0.396*** (0.058)	0.018 (0.061)	-0.092** (0.043)
Time spent decile 10	0.163*** (0.056)	0.117*** (0.044)	-0.100 (0.071)	-0.004 (0.028)	-0.187** (0.079)	0.070 (0.055)	-0.083 (0.059)	-0.014 (0.054)	0.069 (0.063)	-0.500*** (0.063)	-0.080 (0.057)	0.071* (0.042)	0.059 (0.040)	-0.023 (0.070)	-0.355*** (0.060)	-0.005 (0.061)	-0.082* (0.043)
Other covariates from Table 3?	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	1416	1399	1265	1265	1044	1505	1480	1392	1361	1340	1390	1370	1387	1373	1217	1477	1353
R-squared	0.08	0.09	0.10	0.10	0.10	0.13	0.08	0.09	0.13	0.21	0.21	0.14	0.11	0.08	0.31	0.09	0.10
Adjusted r-squared	0.01	0.02	0.03	0.03	0.01	0.07	0.02	0.02	0.07	0.14	0.15	0.08	0.04	0.01	0.24	0.02	0.03

\* 0.10 \*\* 0.05 \*\*\* 0.01. Unit of observation is the individual, with one OLS regression per column. LHS variable is the 0/1 B-factor indicator (for the standard directional bias where applicable) shown in the column header. GARP=General Axiom of Revealed Preference. NBLLN=Non-belief in the Law of Large Numbers. EGB=Exponential Growth Bias. All regressions here include the full set of controls described in Table 3, with two exceptions. The model for "over-confidence in level performance" excludes numeracy, which has a mechanical relationship to the B-factor because it forms the basis for over-confidence in number of answers correct. Also, the model for over-confidence in relative performance excludes fluid intelligence, for the same reason. Not shown are four race/ethnicity categories, state of residence, immigrant indicator, 3 marital status categories, 4 household size categories, and 5 work status categories. 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 1-C and the Data Appendix for details on B-factor variable construction.

## Data Appendix

This Appendix details, for each of the 17 individual behavioral factors (B-factors):

- i) The motive for eliciting that B-factor and the mechanism through which that factor might affect financial condition;
- ii) our elicitation method and its key antecedents;
- iii) data quality indicators, including item non-response;
- iv) sample size (as it compares to that for other B-factors);
- v) definitions and prevalence estimates of behavioral *indicators*, with background on the distinctions between standard vs. non-standard directional biases where applicable;
- vi) descriptions of the *magnitude* and *heterogeneity* of behavioral deviations, including descriptions of the distribution and—where the data permit—estimates of key parameters used in behavioral models;
- vii) estimates of conditional correlations between measures of the behavioral factor and financial outcomes, including particular components of our financial condition index (itself described in Section 1-E of the paper) that have particularly strong links to the B-factor per theory. Wherever possible we also provide comparisons to prior work.

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

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

We measure discounting biases with respect to money using the Convex Time Budgets (CTB) method created by Andreoni and Sprenger (2012). In our version, fielded in ALP module 315 (the first of our two surveys), subjects make 24 decisions, allocating 100 hypothetical tokens each between (weakly) smaller-sooner and larger-later amounts. See Data Appendix Figure 1 for an example. The 24 decisions are spread across 4 different screens with 6 decisions each. Each screen varies start date (today or 5 weeks from today) x delay length (5 weeks or 9 weeks); each decision within a screen offers a different yield on saving. Among the 1,515 individuals who take our first 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 2016) and elsewhere (Gine et al. forthcoming). In exploring data quality and prevalence below we focus on comparisons to Andreoni and Sprenger (2012), and Barcellos and Carvalho (2014).<sup>1</sup> AS draw their sample from university students. BC's sample is drawn from the ALP, like ours (module 212 in their case), but they use a different adaptation of the CTB.

Indicators of response quality are encouraging for the most part. Interior allocations are more common in our sample than in AS, and comparable to BC. More of our subjects exhibit some variance in their allocations than AS or BC. Our subjects are internally consistent 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 elicitation but high compared to the 8% in AS.<sup>2</sup>

We calculate biased discounting, for each individual, by subtracting the consumption rate when the sooner payment date is five weeks from today from the consumption rate when the sooner payment date is today, for each of the two delay lengths. We then average the two differences to get a continuous measure of biased discounting. In keeping with AS, BC and several other recent papers (including Carvalho, Meier, and Wang (2016) and Goda et al. (2017)), we find little if any present-bias on average, with a median discount bias of zero, and a 1pp mean tilt toward future bias.<sup>3</sup>

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<sup>4</sup> 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. These prevalence estimates fall

---

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

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

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

<sup>4</sup> Although see Koszegi and Szeidl (2013) for a theory of future-biased discounting.

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., BC's CTB elicitation in the ALP shows 29% with any 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 the AS 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] (Data Appendix Table 1, Columns 11-13).<sup>5</sup> This is noteworthy because behavioral macro papers sometimes assume representative agents with present bias that lies strictly below our 5<sup>th</sup> 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  $\frac{1}{2}$  and 1.” Our estimates should give researchers pause before choosing a value much below 1.

Our estimates of conditional correlations between financial condition and discounting biases (Tables 5-8 and Appendix Table 7) differ 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. 2017).<sup>6</sup> 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.<sup>7</sup>

Our findings suggest that future bias is uncorrelated with financial condition, that present bias is, and that the present bias correlation is robust to different functional forms and to excluding various key control variables—including other B-factors. The magnitude of the correlation is substantial, implying e.g., that present-biased individuals have financial condition

---

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

<sup>6</sup> Other papers have explored links between discounting biases and field behavior using direct elicitations on narrower samples, with narrower sets of covariates; see e.g., Chabris et al. (2008), Meier and Sprenger (2010), Burks et al. (2012), and Li et al. (2015).

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

that is 10% worse than unbiased (i.e., time-consistent) individuals, conditional on our rich set of controls.

### *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 persons taking our second survey, 1423 answer one of the two snack questions, and 1404 respond to both. 61% choose the healthy snack for today, while 68% choose it for five weeks in the future, with 15% exhibiting present bias (consume treat today, plan to eat healthy in the future) and 7% future bias (consume healthy today, plan to eat treat in the future).<sup>8</sup> 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.

Tables 5-8 and Appendix Table 7 present the first estimates we know of correlating measures of consumption discounting biases with field outcomes. We do not find any evidence of statistically significant coefficients. Having said that the present-biased coefficient has the expected negative sign across all specifications. Among the components of our financial condition index, present-bias has been particularly strongly linked to a low savings rate, and indeed in unreported results we find the same relationship between present-bias and savings (one component of our financial condition index).

---

<sup>8</sup> 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).

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

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

We use the same task and user interface as in Choi et al. (2014) but abbreviate it from 25 decisions to 11.<sup>9</sup> Each decision confronts respondents with a linear budget constraint under risk: subjects choose a point on the line, and then the computer randomly chooses whether to pay the point value of the x-axis or the y-axis. 1,270 of the 1,427 individuals taking our second survey make all 11 decisions, and comprise our sample for measuring choice inconsistency.<sup>10</sup> See Data Appendix Figure 2 for an example.

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

---

<sup>9</sup> We were quite constrained on survey time and hence conducted a pilot in which we tested the feasibility of capturing roughly equivalent information with fewer rounds. 58 pilot-testers completed 25 rounds, and we estimated the correlation between measures of choice inconsistency calculated using the full 25 rounds, and just the first 11 rounds. These correlations are 0.62 and 0.88 for the two key measures.

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

<sup>11</sup> 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.

second measure captures inconsistency with both GARP and FOSD.<sup>12</sup> Note that these measures of inconsistency are unidirectional: there is no such thing as being *overly* consistent.

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

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

Choi et al. find that choice inconsistency with GARP is conditionally correlated with lower net worth, but that choice inconsistency with GARP+dominance avoidance is not. Our data show little evidence of a significant conditional correlation between either measure of choice inconsistency and our broader measure of financial condition, although the coefficients have the expected negative sign (Tables 5-8 and Appendix Table 7).

#### *D. Risk attitude re: certainty (Certainty Premium)*

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

---

<sup>12</sup> 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.



(Neilson 1992; Schmidt 1998; Diecidue, Schmidt, and Wakker 2004). PFC may help to explain seemingly extreme risk averse behavior, which could in turn lead to lower wealth in the cross-section.

We use Callen et al.'s (2014) two-task method<sup>13</sup> for measuring a subject's *certainty premium* (CP).<sup>14</sup> Similar to Holt and Laury tasks, in one of the Callen et al. tasks subjects make 10 choices between two lotteries, one a  $(p, 1-p)$  gamble over  $X$  and  $Y > X$ ,  $(p; X, Y)$ , the other a  $(q, 1-q)$  gamble over  $Y$  and  $0$ ,  $(q; Y, 0)$ . Both Callen et al. and we fix  $Y$  and  $X$  at 450 and 150 (hypothetical dollars in our case, hypothetical Afghanis in theirs), fix  $p$  at 0.5, and have  $q$  range from 0.1 to 1.0 in increments of 0.1. In the other task,  $p = 1$ , so the subject chooses between a lottery and a certain option. Our two tasks are identical to Callen et al.'s except for the currency units. But our settings, implementation, and use of the elicited data are different. Callen et al. administer the tasks in-person, using trained surveyors, at polling centers and homes in Afghanistan. They use the data to examine the effects of violence on risk preferences.

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

We estimate CP for each respondent  $i$  by imputing the likelihoods  $q^*$  at which  $i$  expresses indifference as the midpoint of the  $q$  interval at which  $i$  switches, and then using the two likelihoods to estimate the indirect utility components of the CP formula. As Callen et al. detail, the CP “is defined in probability units of the high outcome,  $Y$ , such that one can refer to certainty

---

<sup>13</sup> Callen et al. describes its task as “a field-ready, two-question modification of the uncertainty equivalent presented in Andreoni and Sprenger (2016).”

<sup>14</sup> The Callen et al. tasks also elicit non-parametric measures of classical risk aversion: a higher switch point indicates greater risk aversion. We discuss these measures in Section 1-D of the paper.

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 both our data and Callen et al.’s.

Callen et al. find significant correlations between the CP and financial outcomes, in particular with avoiding late loan repayments,<sup>15</sup> but their data lack controls for cognitive skills and other B-factors. We do not find evidence of a statistically significant correlation between CP and our index of financial outcomes, and the implied magnitude of our point estimates is small, although our index lacks a direct measure of loan delinquency. Looking particularly at the component of our financial condition index most directly linked to risk-taking—holding public equities—we do find stronger correlations, with point estimates of -0.03 for our CP indicator and -0.05 for our CP percentile and p-values of about 1 in each case.

#### *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.

---

<sup>15</sup> The theoretical mapping from late loan repayments to our index of financial condition is unclear under limited liability, and the average relationship (not conditioning on borrowing) more ambiguous, since borrowing could lead to (weakly) greater or lesser wealth if consumers are behavioral (Zinman 2014).

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.<sup>16</sup> We acknowledge this but use “loss aversion” instead of “loss aversion and/or small-stakes risk aversion” as shorthand. Small-stakes risk aversion is also often classified as behavioral because it is incompatible with expected utility theory (Rabin 2000).

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

---

<sup>16</sup> A related point is that there is no known “model-free” method of eliciting loss aversion (Dean and Ortleva 2016).

compound lottery is sufficiently complicated as to appear effectively ambiguous (Dean and Ortleva 2016).

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); 70% of that sample exhibits some loss aversion. In von Gaudecker et al.'s nationally representative Dutch sample, 86% exhibit some loss aversion, as inferred from structural estimation based on data from multiple price lists. We also order sets of deviations to indicate greater degrees of loss aversion, based on whether the individual respondent rejects the compound but not the single lottery, rejects the single but not the compound lottery, or rejects both.

Despite the massive amount of work on loss aversion, research exploring links between directly elicited measures of loss aversion and field behavior is only beginning. von Gaudecker et al. (2011) do not explore links between loss aversion and field 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 (labeled "Heuristics" in the Hwang paper).<sup>17</sup>

We do not find evidence of a strong conditional correlation between loss aversion and financial condition, although the coefficients are uniformly negative across specifications (Tables 5-8 and Appendix Table 7), which is consistent with the hypothesis that loss aversion is associated with lower wealth. Looking particularly at the component of our financial condition

---

<sup>17</sup> Hwang (2016) also discusses the potential (mediating) role of narrow framing/bracketing but lacks a directly elicited measure of such.

index most directly linked to risk-taking—holding public equities—we do find stronger correlations, with loss averse respondents 2.2pp (SE 2.3pp) less likely to hold stocks than classical respondents (those who play both small-stakes lotteries). Specifying our loss aversion measure categorically, we find that the most loss averse respondents (those who reject both lotteries) are 4.1pp less likely to hold stocks than classical respondents (SE 2.7pp), again conditional on our rich set of controls.

#### *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 Weizsacker (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.<sup>18</sup> 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

---

<sup>18</sup> Given the puzzling result that RW's Example 2 was relatively impervious to a broad-bracketing treatment, we changed our version slightly to avoid zero-amount payoffs. Thanks to Georg Weizsacker for this suggestion.

risk in the first decision but accepts it in the second decision (A and D) violates dominance, here with an expected loss of \$75 relative to BC. 23% of our subjects choose AD, compared to 36% in the most similar presentation in RW. As RW discuss, a new feature of task two is that AD sacrifices expected value in the second decision, not in the first. This implies that for all broad-bracketing risk averters AC is optimal: it generates the highest available expected value at no 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 unidirectional indicators: we either classify someone as narrow-bracketing, or not. RW do not create summary indicators across tasks, but, as noted above, their subjects exhibit substantially more narrow bracketing at the task level than our subjects do.

Research linking directly-elicited measures of NBDC to field 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.

We do not find any statistically significant conditional correlations between our measures of narrow bracketing and our financial condition index.

### *G. Ambiguity Aversion*

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

We elicit a coarse measure of ambiguity aversion using just one or two questions about a game that pays \$500 if you select a green ball. The first question offers the choice between a Bag One with 45 green and 55 yellow balls vs. a Bag Two of unknown composition. 1,397 subjects respond to this question (out of 1,427 who answer at least one of our questions on ALP module 352). 73% choose the 45-55 bag, and we label them ambiguity averse. The survey then asks these subjects how many green balls would need to be in Bag One to induce them to switch. 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. (2016), Dimmock, Kouwenberg, and Wakker (2016), Gneezy et al. (2015)), and so our measure of deviation from ambiguity-neutrality is one-sided.

Despite the coarseness of our elicitation, comparisons to other work suggest that it produces reliable data. Our ambiguity aversion indicator correlates 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 (2016) 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 (2016) from the Dutch version of the ALP.

Our predictive analysis builds on the papers by Dimmock and co-authors cited above, which estimate conditional correlations between ambiguity aversion and financial market behavior, by broadening the set of both outcomes and control variables (especially other B-factors).<sup>19</sup> We too find evidence of statistically significant negative correlations between ambiguity aversion and financial outcomes (Table 5-8 and Appendix Table 7).

#### *H. Overconfidence*

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

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

The first measures it in level/absolute terms, by following the three Banks and Oldfield questions (Section 1-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 Banks and Oldfield quiz scores or of future income increases). This is a coarse adaptation of the usual approaches of eliciting

---

<sup>19</sup> The other paper we know of examining correlations between ambiguity attitudes and field behavior is Sutter et al.'s (2013) study of adolescents in Austria.



several confidence intervals or subjective probability distributions (Moore and Healy). In our data 34% of 1,345 responding to both sets respond 100% on  $\geq 1$  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 1-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 pre-test self-ranking (N=1,395). This measure is useful for capturing individual-level heterogeneity ordinally, but not for measuring prevalence because the actual ranking is based on a 15-question test and hence its percentiles are much coarser than the self-ranking.

Among the three sets of pairwise correlations between measures of overconfidence 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 described above and field outcomes. Our results find the expected negative conditional correlations with financial condition (especially strong for overconfidence in relative performance), and some evidence for significant positive effect of under-confidence.

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

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

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

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

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

1,375 subjects respond (out of the 1,427 who answer at least one of our questions in Module 352),<sup>20</sup> with mean (SD) responses of 27 (18), 42 (24), and 31 (20). We measure NBLLN using the distance between the subject's answer for the [481, 519] range and 78. Only one subject gets it exactly right. 87% underestimate; coupled with prior work, this result leads us to designate underestimation as the “standard” directional bias. The modal underestimator responds with 50 (18% of the sample). The other most-frequent responses are 25 (10%), 30 (9%), 33 (8%), and 40 (7%). Few underestimators—only 4% of the sample—are within 10pp of 78, and their mean distance is 43, with an SD of 17. 9% of the sample underestimates by 20pp or less. 13% overestimate relative to 78, with 5% of the sample quite close to correct at 80, and another 5% at 100. Benjamin, Moore, and Rabin (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 are 27%, 42%, and 31%.

We are not aware of any prior work exploring conditional correlations between directly-elicited NBLLN and field outcomes. The sign pattern suggests those underestimating convergence exhibit lesser financial condition in the cross-section, with larger underestimation more strongly negatively correlated than the simple indicator (Appendix Table 7 vs. Tables 5-8).

### *J. Gambler's Fallacies*

The Gambler's Fallacies involve falsely attributing statistical dependence to statistically independent events, in either expecting one outcome to be less likely because it has happened recently (recent reds on roulette make black more likely in the future) or the reverse, a “hot hand” view that recent events are likely to be repeated. Gambler's fallacies can lead to

---

<sup>20</sup> 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.

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). Because the hot hand fallacy is more closely linked to harmful financial behaviors such as “return-chasing” or over-valuing the talent of stock-pickers (Rabin and Vayanos 2010), for analyses linking the fallacies to field behavior we denote hot-hand as the “standard” bias and cold-hand as “non-standard.”

We take a slice of Benjamin, Moore, and Rabin’s (2013) elicitation for the fallacies:

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

1,392 subjects respond, out of the 1,427 respondents to module 352. The cold-hand fallacy implies a response  $< 50\%$ , while the hot-hand fallacy implies a response  $> 50\%$ . Our mean response is 45% (SD=25), which is consistent with the cold-hand but substantially above the 32% in Benjamin, Moore, and Rabin. Another indication that we find less evidence of the cold-hand fallacy is that, while they infer that “at the individual level, the gambler’s fallacy [cold-hand] appears to be the predominant pattern of belief” (2013, p. 16), we find only 26% answering  $< “50.”$  14% of our sample responds with  $> “50”$  (over half of these responses are at “90” or “100”). So 60% of our sample answers correctly. Nearly everyone who responds with something other than “50” errs by a substantial amount—e.g., only 2 % of the sample is [30, 50) or (50, 70]. Sixteen percent of our sample answers “10,”<sup>21</sup> which Benjamin, Moore, and Rabin speculates is an indicator of miscomprehension; we find that while subjects with this indicator do have significantly lower cognitive skills than the unbiased group, they actually have higher cognitive skills than the rest of subjects exhibiting a gambler’s fallacy.

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

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

---

<sup>21</sup> 34% of the sample in Benjamin, Moore, and Raymond respond “10%” on one or more of their ten questions.

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

Dohmen et al. also explore correlations between unemployment or bank overdrafts and their directly-elicited fallacy measures, conditional on age, gender, education, income, and wealth. They find evidence of positive correlations between hot-hand and unemployment and between cold-hand and overdrafting. We find a general pattern of evidence that both fallacies are negatively conditionally correlated with our financial condition index in point terms (Tables 5-8 and Appendix Table 7).

### *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 a broad set of financial outcomes (Levy and Tasoff 2016; Stango and Zinman 2009).

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

We calculate an individual-level measure of “debt-side EGB” by comparing the difference between the APR *implied* by the monthly payment supplied by that individual, and the *perceived* APR as supplied directly by the same individual. We start by binning individuals into under-estimators (the standard bias), over-estimators, unbiased, and unknown (37% of the sample).<sup>22</sup>

---

<sup>22</sup> 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  $\geq 100\%$  as having unknown bias.

The median level difference between the correct and stated value is 500bp, with a mean of 1,042bp and SD of 1,879bp. Among those with known bias, we count as biased 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 less EGB than Stango and Zinman (2009, 2011) see from questions in the 1983 Survey of Consumer Finances, where 98% of the sample underestimates, and the mean bias is 1,800bp or 3,800bp depending on the benchmark. The time frames of the questions differ, which may account for the difference (and is why we do not estimate an EGB structural model parameter to compare with our prior work or that of Levy and Tasoff).

Stango and Zinman (2009; 2011) find that more debt-side EGB is strongly conditionally correlated with debt allocation, worse loan terms, and less savings and wealth. But those papers lack direct controls for cognitive skills and other behavioral factors. Here we find no evidence of a statistically significant correlation—whether or not we control for variables omitted in SZ, and whether or not we consider our financial condition index as a whole or saving behavior or wealth in particular.

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.<sup>23</sup> We again bin individuals into underestimators (the standard bias), overestimators, unbiased, and unknown (14% of the sample).<sup>24</sup> Among those with known bias (N=1,222), the median bias is \$0, with a mean of \$2

---

<sup>23</sup> Responses to this question are correlated with responses to two other questions, drawn from Levy and Tasoff (2016), that can also be used to measure asset-side EGB, but our sample sizes are smaller for those two other questions and hence we do not use them here.

<sup>24</sup> 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.

and SD of \$14.<sup>25</sup> 44% of our sample provides the correct FV. 47% of our sample underestimates by some amount, with most underestimators (29% of the sample) providing the linearized (uncompounded) answer of \$240. Nearly all other underestimates provide an answer that fails to account for even simple interest; the most common reply in this range is “\$220.” Only 9% of our sample overestimates the FV, with small mass points at 244, 250, 400, and 440.

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

Goda et al. (2017) and Levy and Tasoff (2016) measure asset-side EGB using more difficult questions in their representative samples. They find that 9% and 11% overestimate FVs, while 69% and 85% underestimate. We do not construct an EGB parameter to compare to theirs, because our questions lack their richness and yield heavy mass points at unbiased and linear-biased responses.

The only prior paper we know looking directly at links between a measure of asset-side EGB and field outcomes is Goda et al., who use data on fewer behavioral factors. They find significant negative correlations between asset-side EGB and retirement savings. We also find negative correlations, both with our financial condition index (Tables 5-8 and Appendix Table 7) and with self-assessed retirement preparedness in particular.

---

<sup>25</sup> For calculating the mean and SD we truncate bias at -42 for the 4% sample answering with future values  $284 < FV < 485$ , to create symmetric extrema in the bias distribution since our definition caps bias at 42.

### *L. Limited Attention/Memory*

Prior empirical work has found that limited attention affects a range of financial decisions (e.g., Barber and Odean 2008; DellaVigna and Pollet 2009; Karlan et al. 2016; Stango and Zinman 2014). Behavioral inattention is a very active line of theory inquiry as well (e.g., Bordalo, Gennaioli, and Shleifer 2017; Kőszegi and Szeidl 2013; Schwartzstein 2014).

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

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

Responses are strongly but not perfectly correlated (ranging 0.56 to 0.69 among pairwise expressions of regret). A fourth measure of limited attention is also strongly correlated with the others, based on the question: “Do you believe that you could improve the prices/terms your household typically receives on financial products/services by shopping more?”<sup>26</sup> 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

---

<sup>26</sup> This question is motivated by evidence that shopping behavior strongly predicts borrowing costs (Stango and Zinman 2016).

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

Limited attention strongly and negatively correlates with financial condition across all specifications. It is worth considering these correlations are somewhat mechanical, in the sense that people with worse outcomes might be more likely to express regret ex-post (perhaps even misattributing their lack of attention as a contributing factor). We designed the response options to try to dampen this possibility, by building in consideration of tradeoffs, but nevertheless grant that the possibility of a mechanical correlation looms larger here than for our other B-factors.

We also seek to measure limited prospective memory, following previous work suggesting that limited memory entails real costs like forgetting to redeem rebates (e.g., Ericson 2011). We offer an incentivized task to subjects taking module 352: “The ALP will offer you the opportunity to earn an extra \$10 for one minute of your time. This special survey has just a few simple questions but will only be open for 24 hours, starting 24 hours from now. During this specified time window, you can access the special survey from your ALP account. So we can get a sense of what our response rate might be, please tell us now whether you expect to do this special survey.” 97% say they intend to complete the short survey, leaving us with a sample of 1,358. Only 14% actually complete the short survey. Failure to complete is weakly positively correlated with our indicators of attentional regret described above, 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 conditional on intending to complete)—is a bit coarse. We suspect that some noise is introduced because our elicitation makes it costless to express an intention to complete (in future research we plan to explore charging a small “sign up” fee), thereby including in the indicator’s sample frame some subjects who rationally do not complete the task. Relatedly, although we set the payoff for task completion to be sufficiently high to dominate any attention/memory/time costs in *marginal* terms for most subjects (the effective hourly wage is in the hundreds of dollars), it may well be the case that the *fixed* cost exceeds \$10 for some respondents.

Despite the noise, our coarse indicator of limited prospective memory turns out to have the expected negative and significant correlation with financial condition. This is the first work we



know of estimating the conditional relationship between limited memory and field outcomes in a broad sample.

### References for Data Appendix

- Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2011. "Reference Points and Effort Provision." *American Economic Review* 101 (2): 470–92.
- Afriat, S. N. 1972. "Efficiency Estimation of Production Functions." *International Economic Review* 13 (3): 568.
- Andreoni, James, and Charles Sprenger. 2012. "Estimating Time Preferences from Convex Budgets." *The American Economic Review* 102 (7): 3333–56.
- . 2016. "Prospect Theory Revisited: Unconfounded Experimental Tests of Probability Weighting."
- Augenblick, Ned, Muriel Niederle, and Charles Sprenger. 2015. "Working over Time: Dynamic Inconsistency in Real Effort Tasks." *The Quarterly Journal of Economics* 130 (3): 1067–1115.
- Banks, J., and Z. Oldfield. 2007. "Understanding Pensions: Cognitive Function, Numerical Ability, and Retirement Saving." *Fiscal Studies* 28 (2): 143–70.
- Banks, James, Cormac O’Dea, and Zoë Oldfield. 2010. "Cognitive Function, Numeracy and Retirement Saving Trajectories." *The Economic Journal* 120 (548): F381–410.
- Barber, Brad, and Terrence Odean. 2008. "All That Glitters: The Effect of Attention on the Buying Behavior of Individual and Institutional Investors." *Review of Financial Studies* 21 (2): 785–818.
- Barberis, Nicholas C. 2013. "Thirty Years of Prospect Theory in Economics: A Review and Assessment." *Journal of Economic Perspectives* 27 (1): 173–96.
- Barcellos, Silvia, and Leandro Carvalho. 2014. "Information about Self-Control and Intertemporal Choices."
- Bell, David E. 1985. "Disappointment in Decision Making under Uncertainty." *Operations Research* 33 (1): 1–27.
- Benartzi, S., and R. H. Thaler. 1995. "Myopic Loss Aversion and the Equity Premium Puzzle." *The Quarterly Journal of Economics* 110 (1): 73–92.
- Benjamin, Daniel, Don Moore, and Matthew Rabin. 2013. "Misconceptions of Chance: Evidence from an Integrated Experiment."
- Benjamin, Daniel, Matthew Rabin, and Collin Raymond. 2016. "A Model of Nonbelief in the Law of Large Numbers." *Journal of the European Economic Association* 14 (2): 515–44.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2017. "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.
- Burks, Stephen, Jeffrey Carpenter, Lorenz Götte, and Aldo Rustichini. 2012. "Which Measures of Time Preference Best Predict Outcomes: Evidence from a Large-Scale Field Experiment." *Journal of Economic Behavior & Organization* 84 (1): 308–20.
- Butler, Jeffrey V., Luigi Guiso, and Tullio Jappelli. 2014. "The Role of Intuition and Reasoning in Driving Aversion to Risk and Ambiguity." *Theory and Decision* 77 (4): 455–84.
- Callen, Michael, Mohammad Isaqzadeh, James D Long, and Charles Sprenger. 2014. "Violence and Risk Preference: Experimental Evidence from Afghanistan." *The American Economic Review* 104 (1): 123–48.
- Carvalho, Leandro, Stephan Meier, and Stephanie Wang. 2016. "Poverty and Economic Decision-Making: Evidence from Changes in Financial Resources at Payday." *American Economic Review* 106 (2): 260–84.

- 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.
- 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. 2016. "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.
- Dimmock, Stephen, Roy Kouwenberg, Olivia S. Mitchell, and Kim Peijnenburg. 2016. "Ambiguity Aversion and Household Portfolio Choice Puzzles: Empirical Evidence." *Journal of Financial Economics* 119 (3): 559–77.
- Dimmock, Stephen, Roy Kouwenberg, and Peter P Wakker. 2016. "Ambiguity Attitudes in a Large Representative Sample." *Management Science* 62 (5): 1363–80.
- Dohmen, Thomas, Armin Falk, David Huffman, Felix Marklein, and Uwe Sunde. 2009. "Biased Probability Judgment: Evidence of Incidence and Relationship to Economic Outcomes from a Representative Sample." *Journal of Economic Behavior & Organization* 72 (3): 903–15.
- Eisenhauer, Joseph G, and Luigi Ventura. 2006. "The Prevalence of Hyperbolic Discounting: Some European Evidence." *Applied Economics* 38 (11): 1223–34.
- Ericson, Keith. 2011. "Forgetting We Forget: Overconfidence and Memory." *Journal of the European Economic Association* 9 (1): 43–60.
- Falk, Armin, Anke Becker, Thomas Dohmen, 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. forthcoming. "Revising Commitments: Field Evidence on the Adjustment of Prior Choices." *Economic Journal*.
- 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. 2017. "Predicting Retirement Savings Using Survey Measures of Exponential-Growth Bias and Present Bias."
- 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 (03): 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.

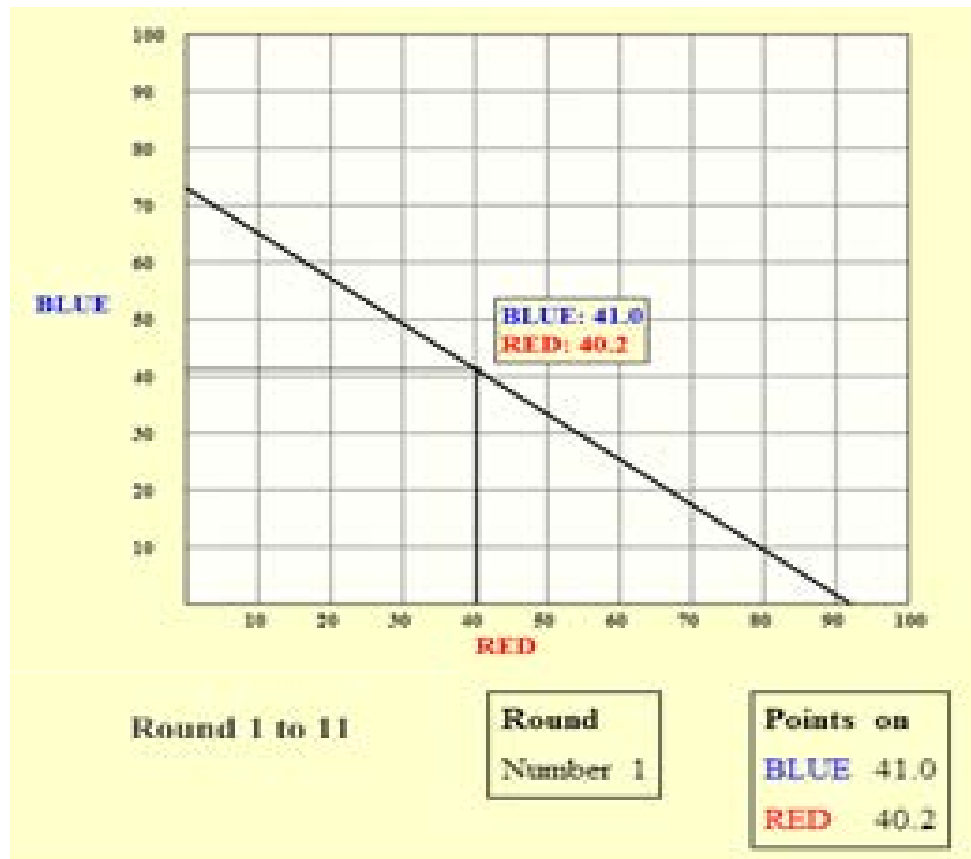
- Harris, Christopher, and David Laibson. 2013. "Instantaneous Gratification." *The Quarterly Journal of Economics* 128 (1): 205–48.
- Hwang, In Do. 2016. "Prospect Theory and Insurance Demand."
- İmrohoroğlu, Ayşe, Selahattin İmrohoroğlu, and Douglas H. Joines. 2003. "Time-Inconsistent Preferences and Social Security." *The Quarterly Journal of Economics* 118 (2): 745–84.
- Kahneman, D, and A Tversky. 1972. "Subjective Probability: A Judgement of Representativeness." *Cognitive Psychology* 3: 430–54.
- Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47 (2): 263–91.
- Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman. 2016. "Getting to the Top of Mind: How Reminders Increase Saving." *Management Science* 62 (12): 3393–3411.
- Kartashova, Katya. 2014. "Private Equity Premium Puzzle Revisited." *American Economic Review* 104 (10): 3297–3334.
- Kőszegi, Botond, and Matthew Rabin. 2009. "Reference-Dependent Consumption Plans." *American Economic Review* 99 (3): 909–36.
- Kőszegi, Botond, and Adam Szeidl. 2013. "A Model of Focusing in Economic Choice." *The Quarterly Journal of Economics* 128 (1): 53–104.
- Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112 (2): 443–77.
- Larrick, Richard P, Katherine A Burson, and Jack B Soll. 2007. "Social Comparison and Confidence: When Thinking You're Better than Average Predicts Overconfidence (and When It Does Not)." *Organizational Behavior and Human Decision Processes* 102 (1): 76–94.
- Levy, Matthew, and Joshua Tasoff. 2016. "Exponential-Growth Bias and Lifecycle Consumption." *Journal of the European Economic Association* 14 (3): 545–83.
- 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. 2009. "Exponential Growth Bias and Household Finance." *The Journal of Finance* 64 (6): 2807–49.
- . 2011. "Fuzzy Math, Disclosure Regulation, and Credit Market Outcomes: Evidence from Truth-in-Lending Reform." *Review of Financial Studies* 24 (2): 506–34.
- . 2014. "Limited and Varying Consumer Attention: Evidence from Shocks to the Salience of Bank Overdraft Fees." *Review of Financial Studies* 27 (4): 990–1030.
- . 2016. "Borrowing High vs. Borrowing Higher: Price Dispersion and Shopping Behavior in the U.S. Credit Card Market." *Review of Financial Studies* 29 (4): 979–1006.
- 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.
- Toubia, Olivier, Eric Johnson, Theodoros Evgeniou, and Philippe Delquié. 2013. "Dynamic Experiments for Estimating Preferences: An Adaptive Method of Eliciting Time and Risk Parameters." *Management Science* 59 (3): 613–40.
- Tversky, A., and D. Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211 (4481): 453–58.
- 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.
- Zinman, Jonathan. 2014. "Consumer Credit: Too Much or Too Little (or Just Right)?" *Journal of Legal Studies* 43 (S2 Special Issue on Benefit-Cost Analysis of Financial Regulation): S209–37.

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

	Token value 5 weeks from today	Token value 14 weeks from today	Decision: How many of the 100 tokens would you like to allocate to the <b>sooner</b> payment 5 <b>weeks from today?</b>	Tokens received 5 weeks from today	Tokens remaining 14 weeks from today	Total payment 5 weeks from today	Total payment 14 weeks from today
1	\$1	\$1	<input type="text" value="0"/> out of 100 tokens	0	100	\$0.00	\$100.00
2	\$1	\$1.02	<input type="text" value="0"/> out of 100 tokens	0	100	\$0.00	\$102.00
3	\$1	\$1.04	<input type="text" value="0"/> out of 100 tokens	0	100	\$0.00	\$104.00
4	\$1	\$1.07	<input type="text" value="0"/> out of 100 tokens	0	100	\$0.00	\$107.00
5	\$1	\$1.11	<input type="text" value="0"/> out of 100 tokens	0	100	\$0.00	\$111.00
6	\$1	\$1.17	<input type="text" value="0"/> out of 100 tokens	0	100	\$0.00	\$117.00

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



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

**Data Appendix Table 1. Estimated Distributions of Individual-Level Present Bias Parameter from our data**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
<u>present-bias parameter</u>													
p50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
p5	0.00	0.06	0.00	0.01	0.00	0.01	0.00	0.00	0.68	0.73	0.93	0.95	0.96
p95	1158	99	539	421	710	219	397	343	1.62	1.6	1.07	1.05	1.06
<u>concavity parameter</u>													
starting value	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.9	0.7	0.9	0.7	0.7	0.9
restricted > 0?	no	no	no	no	no	no	no	no	yes	yes	no	no	yes
<u>background consumption</u>													
assume same across time?	yes	yes	yes	yes	yes	yes	no	yes	yes	yes	yes	yes	yes
assume same across people?	yes	yes	yes	no	no	no	no	yes	yes	yes	yes	yes	yes
assumed value(s)?	0	estimated	see below	see below	see below	see below	see below	0	0	0	0	0	0
<u>response quality</u>													
drop if any non-monotonicity	no	no	no	no	no	no	no	no	no	no	yes	yes	yes
drop if no variance in choice	no	no	no	no	no	no	no	no	no	no	no	yes	yes
N individuals	1259	1244	1258	1258	590	590	524	1250	1236	1237	715	689	689

All models estimated using the nonlinear least squares version of the model in Andreoni and Sprenger (2012).

Background consumption:

In (2), estimated as a model parameter at the individual-level.

In (3), assumed to be the median value of individual-level daily spending as measured in ALP module 417 (\$16.50), calculated over all respondents to that survey.

In (4), assumed to be the median value of individual-level daily spending as measured in ALP module 417, calculated over all respondents to that survey, multiplied by the number of household members reported in our module containing the CTB (ALP 315).

In (5), measured directly using data on individual-level daily spending from ALP module 417.

In (6), measured directly using data on individual-level daily spending from ALP module 417, multiplied by the number of household members reported in our module containing the CTB (ALP 315).

In (7), measured directly using individual-level spend data from two different points in time (modules 400 and 417).