INSTITUTE FOR SOCIAL RESEARCH • SURVEY RESEARCH CENTER
MICHIGAN RETIREMENT RESEARCH CENTER
UNIVERSITY OF MICHIGAN

Working Paper
WP 2018-378

Quicksand or Bedrock for Behavioral Economics? Assessing Foundational Empirical Questions

Victor Stango, Joanne Yoong, and Jonathan Zinman

MRRC

Project #: UM17-13

Quicksand or Bedrock for Behavioral Economics? Assessing Foundational Empirical Questions

Victor Stango
University of California-Davis

Joanne Yoong
University of Southern California, National University of Singapore, and the London School of Hygiene and Tropical Medicine

Jonathan Zinman
Dartmouth College and NBER

April 2018

Michigan Retirement Research Center
University of Michigan
P.O. Box 1248
Ann Arbor, MI 48104
www.mrrc.isr.umich.edu
(734) 615-0422

Acknowledgements
The research reported herein was performed pursuant to a grant from the U.S. Social Security Administration (SSA) funded as part of the Retirement Research Consortium through the University of Michigan Retirement Research Center Award RRC08098401-09. The opinions and conclusions expressed are solely those of the author(s) and do not represent the opinions or policy of SSA or any agency of the federal government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation or favoring by the United States government or any agency thereof.

Quicksand or Bedrock for Behavioral Economics? Assessing Foundational Empirical Questions

Abstract

Behavioral economics lacks empirical evidence on some foundational 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, behavioral 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.

Citation

Stango, Victor, Joanne Yoong and Jonathan Zinman. 2018. "Quicksand or Bedrock for Behavioral Economics? Assessing Foundational Empirical Questions."Ann Arbor MI: University of Michigan Retirement Research Center (MRRC) Working Paper, WP 2018-378. http://mrrc.isr.umich.edu/wp378/

Authors' acknowledgements

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." 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), the Michigan Retirement Research Center (MRRC) 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.[1]

Yet many economists remain skeptical of the economic importance and usefulness of BE.[2] 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 nonrepresentative samples. Ten 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 show 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

[1] 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.

[2] For critiques of behavioral economics see, e.g., Fudenberg (2006) and Levine (2012).

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.[3] Yet most prior work in BE, both theory and empirics, analyzes one or two B- 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.[4]

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

[3] 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.

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

(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.[5]

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 1,400+ participants in RAND's American Life Panel

[5] 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).

(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 nonclassical 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,[6] and by the success of previous work adapting lab elicitations for administration in surveys.[7] 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. Moreover, we control for respondent survey effort—measured using time spent responding to B-factor elicitations—or down weight low-effort respondents in some specifications. Theory helps as well by generating testable predictions — that certain directional biases ("standard" ones such as over-confidence, and present-bias) should be more prevalent and predictive than others ("nonstandard" biases such

[6] 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).

[7] 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).

as 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 percent 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 five of the seven B-factors with prior comparable studies.[8] 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 six of the seven 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

[8] By our accounting, the seven 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.

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 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.[9] 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 nonstandard 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.[10] The findings are striking:

[9] By our accounting, the three 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.

[10] 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 two 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 versus multiple B-factors in the model. Bruine de Bruin, Parker, and Fischoff (2007) and Li et al. (2015) also consider a relatively small set of behavioral factors, in convenience samples, as do von Gaudecker et al. (2011) in a representative Dutch sample, without exploring links to field behavior. Gottlieb and Mitchell (2015)

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- 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.[11] Our main takeaways here inform research

estimates the relationship between holding long-term care insurance and narrow bracketing in a representative sample of older Americans, and checks robustness by 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.

[11] 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.

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

2013). See also Dohmen et al. (2010) on correlations between measures of classical preferences/attitudes and intelligence. We lack measures of noncognitive 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 noncognitive skills will affect inferences about B-factors.

[12] For some other uses of our existing data see Stango, Yoong, and Zinman (2017).

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 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 18 and older.

The ALP takes great pains to obtain a nationally representative sample, combining standard sampling techniques with offers of hardware and a broadband connection to potential participants who lack adequate internet access. ALP sampling weights match the distribution of age, sex, ethnicity, and income to the Current Population Survey.

Panel members are regularly offered opportunities to participate in surveys, the purposes of which range from basic research to political polling. More than 400 surveys have been administered in the ALP, and 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 30-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 (1,427/1,515 = 94%), low item nonresponse 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 unidirectional, 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 nonstandard and time-consistent as unbiased), over-confidence in performance (with under-confidence as nonstandard 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).[13]

Table 2 shows that missing values from question-skipping/noncompletion ("item nonresponse") or nonsense answers ("responded, not usable") are uncommon: 11 of 17 B-factors we have usable responses from more than 90 percent of the sample. Only two B-factors have usable data for less than 80 percent 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.[14] Standard internal consistency checks for respondent understanding and diligence produce results comparable to prior work, as detailed in the Data Appendix.

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.

---

[13] 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 nonresponse; 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 versus nonstandard 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.

[14] 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.

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 four standard tests. We assess general/fluid intelligence with a standard, 15 question "number series" test (McArdle, Fisher, and Kadlec 2007) that is nonadaptive (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 three. A second test is comprised of two "numeracy" questions,[15] labeled as such and popularized in economics since their deployment in the 2002 English Longitudinal Study of Ageing.[16] Our mean number correct is 1.7, with a standard deviation of 0.6. A third test is a three-question "financial literacy" quiz developed and popularized by Lusardi and Mitchell (2014).[17] The median respondent gets all

[15] "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.

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

[17] "Suppose you had $100 in a savings account and the interest rate was 2 percent 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 percent per year and inflation was 2 percent 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

three correct, with a mean of two and a SD of 0.93. We also measure executive function—including working memory and the regulation of attention—using a two-minute Stroop task (MacLeod 1991).[18] Each time the subject chooses an answer that action completes what we refer to as a "round."[19] 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 by 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

single company's stock usually provides a safer return than a stock mutual fund." Response options are categorical for each of the three questions.

[18] 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).

[19] 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.

and other surveys.[20] We use this to construct an integer scale from 1 (most risk tolerant) to 6 (most risk averse). The second is from Dohmen et al. (2010, 2011): "How do you see yourself: Are you generally a person who is fully prepared to take financial risks" (100 point scale, we transform so that higher values indicate greater risk aversion).[21] 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 records 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 elicitations, and on the entirety of the two modules.

---

[20] 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 percent chance the second job would double your current total family income for life and a 50 percent 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 percent 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 percent chance of 20 percent decrease, and then, if they still choose the safe job, a 50 percent chance of a 10 percent decrease).

[21] We also elicit Dohmen et al.'s general risk-taking scale, which is correlated 0.68 with the financial scale.

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 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, consistent with the hypothesis that they are each signals of some underlying construct such as "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 four of our nine 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:

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

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 nonstandard 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 versus nonstandard 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 nonstandard 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 more than 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 nonclassical and/or correlated measurement error, could affect inferences about each of our foundational questions in various ways.[22] 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 nonresponse, so we control for nonresponse 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

[22] 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).

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.

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 nonstandard 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 nonstandard 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 versus 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 percent of respondents found our modules uninteresting.23[23] This finding pushes against the hypothesis that paying financial incentives is important because subjects find direct elicitation tasks unpleasant.

[23] Each ALP survey ends with "Could you tell us how interesting or uninteresting you found the questions in this interview?" and roughly 90 percent of our sample replies "Very interesting" or "Interesting," with only 3 percent replying "Uninteresting" or "Very uninteresting."

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 presents 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 B-factors are 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.

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 seven 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.[24] Some deviation from classical norms is indicated by at least 50 percent 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 percent of the population).

The "standard" directional bias emphasized by prior literature is, indeed, more prevalent in our data; in six of the seven B-factors where we capture bi-directional biases and there is a clear standard.[25] This pattern is consistent with theory but inconsistent with measurement error stories. Purely noisy B-factor measures should show symmetry between standard and nonstandard 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.

---

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

[25] 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.

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.

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 nonresponse. 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 nonmissing 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 nonmonotonic 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 nonresponse 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 percent 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 three 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 suboptimal 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 versus nonstandard bias distinction is denoted in Table 1 and detailed in the Data Appendix.[26] The omitted B-factor category is unbiased except for Certainty Premium and NBLLN, which each have the nonstandard 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. six (five) of the 17 standard

[26] 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 versus 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).

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 five of six 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.

Sixteen 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 nonstandard bias coefficients are centered on zero. Moreover, the standard coefficient is more negative than the nonstandard 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 versus nonstandard 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 versus nonstandard biases. Likewise, B-factor measures that capture unmeasured aspects of classical factors should produce symmetric (weakly negative) conditional correlations between outcomes and standard versus nonstandard biases.[27]

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

[27] 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.

latter three are the only three among our 17 to have been considered in comparable prior work 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 percent decline from the base. The other standard bias coefficients with p-values below 0.10 imply associated reductions in financial condition ranging from 6 percent to 25 percent.

On balance, we infer that behavioral biases are conditionally correlated with financial condition, to economically and statistically significant extents.[28] 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

[28] We find similar results when replacing the indicators for standard and nonstandard biases with their linear percentiles (Appendix Table 7).

higher in the second to fifth deciles, higher still in the sixth to ninth 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 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 (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.[29]

[29] 29 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

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 a 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, at least in part 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

---

that exposure to violent conflict increases preference for certainty. There is a larger body of evidence on the reliability of nonbehavioral measures of time and risk preferences; see, e.g., Meier and Sprenger (2015) and Chuang and Schechter (2015) for recent reviews.

(column 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 confounding.

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.[30]

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 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 such as 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

[30] 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).

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 needs 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:

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

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 more than 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 regarding 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 such as 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.[31] This is key to unpacking the direction and extent of any causality underlying conditional correlations between behavioral factors and outcomes.

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

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. There are many possibilities for exploiting the panel, multitopic 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.

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.” P*sychology 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.”

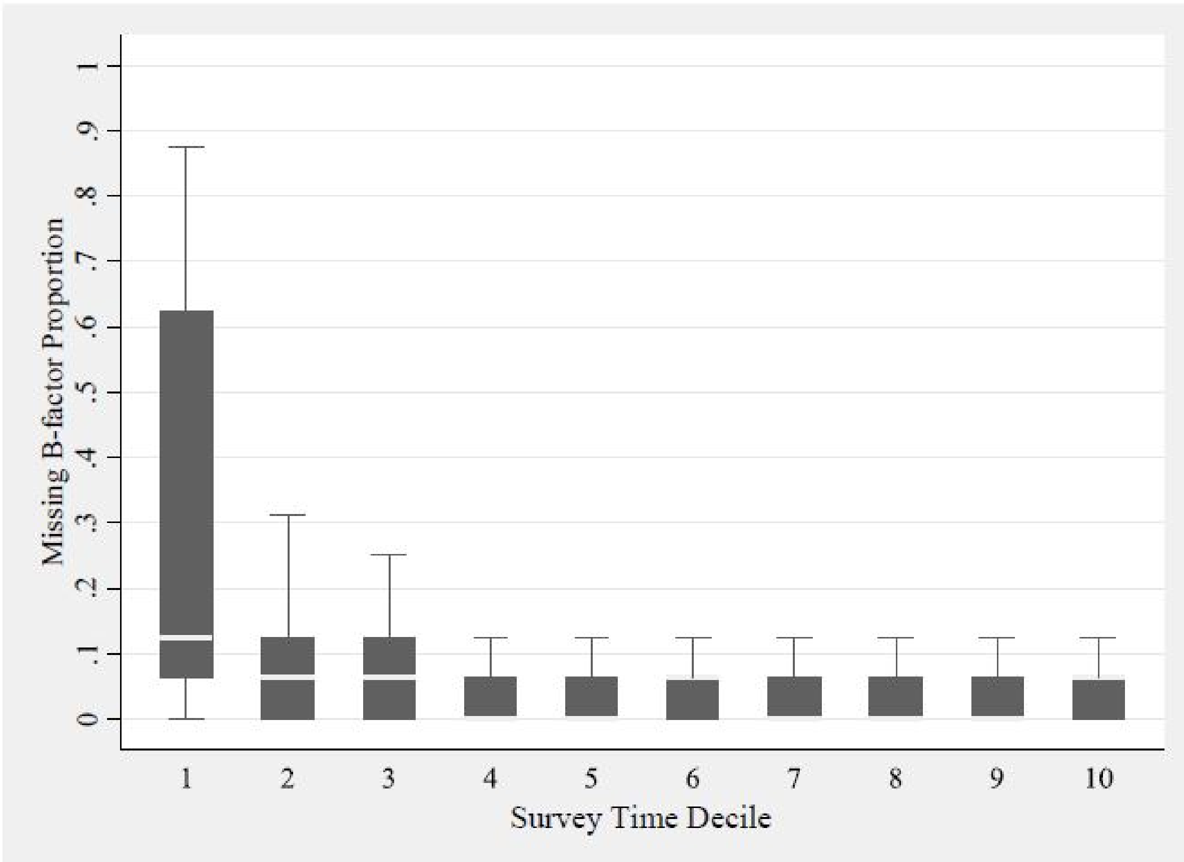
Stango, Victor, and Jonathan Zinman. 2009. “Exponential Growth Bias and Household Finance.” *The Journal of Finance* 64 (6): 2807–49.

———. 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.

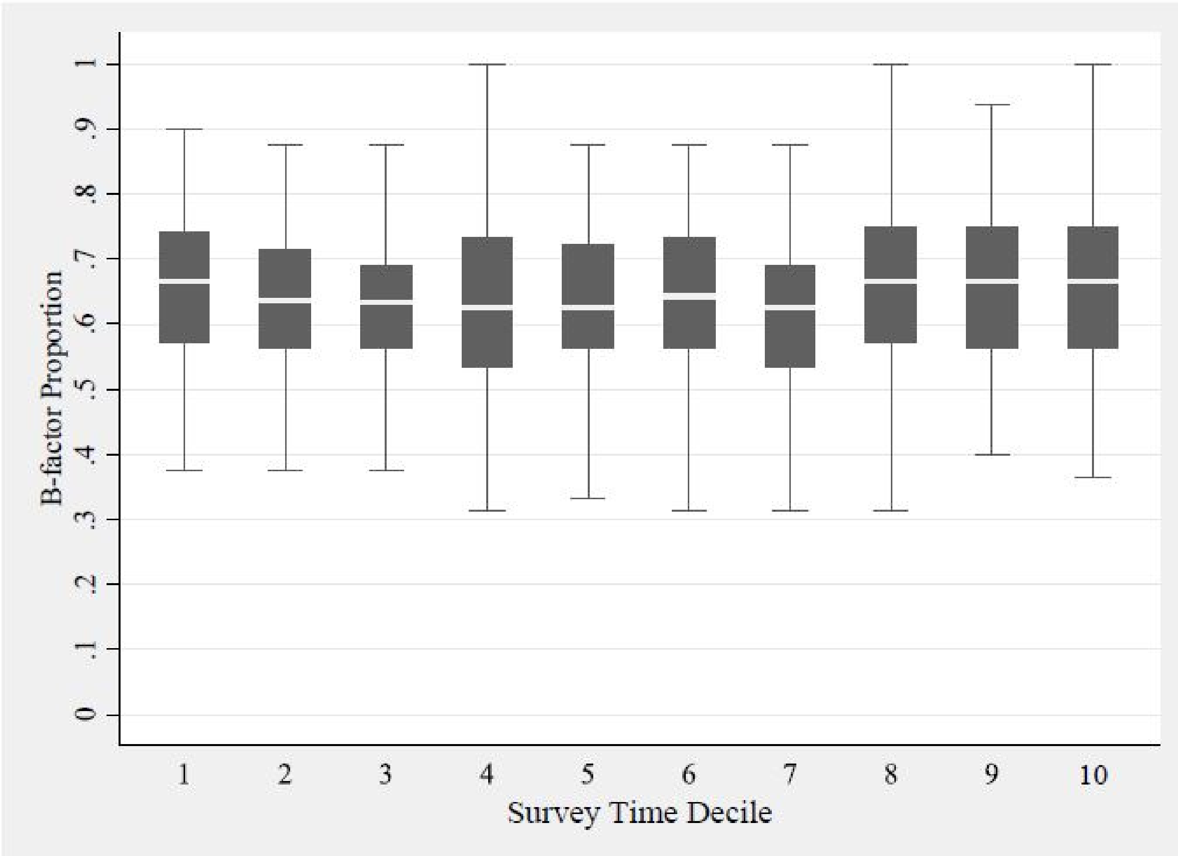
Von Gaudecker, Hans-Martin, Arthur Van Soest, and Erik Wengström. 2011. “Heterogeneity in Risky Choice Behavior in a Broad Population.” *The American Economic Review* 101 (2): 664–94.

Figure 1: B-factor nonresponse 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. 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.

Figure 2: B-factor prevalence (excluding nonresponse), by time spent on B-factor questions



Notes: Box-and-whisker plots show interquartile range in box, adjacent values as whiskers, and median as white line. B-factor Proportion = (Number of behavioral biases indicated)/(Number of nonmissing 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: key antecedents	**Elicitation method description**	**Behavioral bias indicator(s), "standard" deviation direction in bold**
Time inconsistent discounting of money: Andreoni & Sprenger (2012), Barcellos & Carvalho (2014)	Convex Time Budget. 24 decisions allocating 100 tokens each between smaller-sooner and larger-later amounts; decisions pose varying start dates (today vs. 5 weeks from today), delay lengths (5 or 9 weeks) & savings yields.	Present-biased: discounts more when sooner date is today; future-biased: discounts more when sooner date is 5 weeks from today
Time inconsistent discounting of snacks: Read & van Leeuwen (1998), Barcellos & Carvalho (2014)	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.	Present-biased: choose less healthy today, healthy for 5 weeks from now; future-biased: choose healthy for today, less healthy for 5 weeks from now
Violates General Axiom of Revealed Preference: (and/or dominance avoidance), Choi et al (2014)	Decisions from 11 different linear budget constraints under risk. Subjects choose a point on the line, and then the computer randomly chooses whether to pay the point value of the x-axis or the y-axis.	Violates GARP: potential earnings wasted per CCEI>0; violates GARP and/or dominance avoidance: potential earnings wasted per combined-CCEI>0
Certainty premium: Callen et al (2014)	2 screens of 10 choices each between two lotteries, one a (p, 1-p) gamble over X and Y > X , (p; X, Y), the other a (q, 1-q) gamble over Y and 0, (q; Y, 0). Y=$450, X=$150, q ϵ[0.1, 1.0], p=0.5 on one screen and 1.0 on the other.	Preference for cetainty: certainty premium (CP) >0; Cumulative prospect theory: certainty premium (CP)<0
Loss aversion/small-stakes risk aversion: Fehr & Goette (2007)	Two choices. Choice 1: between a 50-50 lottery (win $80 or lose $50), and $0. Choice 2: between playing the lotter in Choice 1 six times, and $0.	Loss aversion: choosing the certain $0 payoff in one or more choices.
Narrow bracketing: Rabin & Weizsacker (2009)	Two tasks of two decisions each. Each decision presents the subject with a choice between a certain payoff and a gamble. Each decision pair appears on the same screen, with an instruction to consider the two decisions jointly.	Narrow-bracketing: making a choice that is dominated given implications of an earlier decision, on one or both
Ambiguity aversion:Dimmock et al. (2016)	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 bag would induce switch?	Ambiguity Aversion: prefers bag with 45 green to bag with unknown mix.
(Over)confidence in performance: Larrick et al (2007), Moore & Healy (2008)	"How many of the last 3 questions (the ones on the disease, the lottery and the savings account) do you think you got correct?"	Overconfidence in perform: self-assessment > actual score; Underconfidence in perform: self-assessment < actual score
Overconfidence in relative performance: Larrick et al (2007), Moore & Healy (2008)	"… what you think about your intelligence as it would be measured by a standard test. How do you think your performance would rank, relative to all of the other ALP members who have taken the test?"	Greater diff between self-assessed and actual rank indicates more overconfidence. "Overconfident" = overconfidence above median (no precise cardinal

Overconfidence in precision: Larrick et al (2007), Moore & Healy (2008)	Questions about likelihoods of different numeracy scores and future income increases.	Overconfidence in precision: responds 100 percent to one or both questions
Non-belief in the law of large numbers (NBLLN): Benjamin, Moore, and Rabin (2013)	Question re: percent chances that, among 1,000 coin flips, the # of heads will fall in ranges [0, 480], [481, 519], and [520, 1000]. NBLLN = distance between response for	Overconfidence in precision: responds 100 percent to one or both questions
Gambler's fallacies: Benjamin, Moore, and Rabin (2013)	"Imagine that we had a computer "flip" a fair coin… 10 times. The first 9 are all heads. What are the chances, in % terms, that the 10th flip will be a head?"	Hot-hand fallacy: responds with>50%; Cold-hand fallacy: responds with<50%
Exponential growth bias (EGB), debt-side: Stango & Zinman (2009; 2011)	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.	Underestimates EG: actual APR>perceived APR; Overestimates EG: actual APR<perceived APR
Exponential growth bias (EGB), asset-side: Banks et al (2007)	Elicits perceived future value of $200, earning 10% annual, after two years.	Underestimates EG: perceived FV<actual FV=$242; Overestimates EG: perceived FV>actual FV=$242
Limited attention: Author-developed	Four questions re: whether subject's finances would improve with more attention given the opportunity cost of attention, with questions varying the types of decisions: day-to-day, medium-run, long-run, or choosing financial products/services.	Limited attention: Indicates regret about paying too little attention, on one or more of the four questions
Limited prospective memory: Ericson (2011)	"The ALP will offer you the opportunity to earn an extra $10.... This special survey has just a few simple questions but will only be open for 24 hours, starting 24 hours from Limited memory: Says will complete task but does not complete now…. please tell us now whether you expect to do this special survey."	Limited memory: Says will complete task but does not complete

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 EfficiencyIndex.

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

B-factor and bias (standard direction marked with *)	Share biased, conditional on response	Survey nonresponse	Missing detail: Item nonresponse	Missing detail: Responded, not usable
Time-inconsistent discounting money: Present-biased *	0.26 *	0.00	0.06	0.00
Time-inconsistent discounting money: Future-biased	0.36			
Time-inconsistent discounting snacks: Present-biased *	0.15 *	0.06	0.02	0.00
Time-inconsistent discounting snacks: Future-biased	0.07			
Violates GARP (based on CCEI) *	0.53 *	0.06	0.10	0.00
Violates GARP (with dominance avoidance) *	0.96 *	0.06	0.10	0.00
Certainty premium: >0=Preference for certainty type *	0.77 *	0.00	0.03	0.28
Certainty premium: <0=Cumulative prospect theory type	0.23			
Loss-averse: prefers certain zero payoff *	0.63 *	0.00	0.00	0.00
Narrow-brackets *	0.59 *	0.00	0.02	0.00
Ambiguity-averse *	0.73 *	0.06	0.03	0.07
Confidence in level performance: Overconfident *	0.38 *	0.06	0.03	0.07
Confidence in level performance: Underconfident	0.11			
Overconfident in precision *	0.44 *	0.06	0.04	0.00
Overconfident in relative performance *	0.50 *	0.06	0.04	0.00
Nonbelief in the law of large numbers: Underestimates convergence *	0.87 *	0.06	0.03	0.00
Nonbelief in the law of large numbers: Overestimates convergence	0.13			
Gambler's fallacy: hot hand	0.14	0.06	0.02	0.00
Gambler's fallacy: cold hand	0.26			
Exponential growth bias, loan-side: Underestimates APR*	0.70 *	0.00	0.05	0.32
Exponential growth bias, loan-side: Overestimates APR	0.27			
Exponential growth bias, asset-side: Underestimates future value *	0.38 *	0.06	0.03	0.00
Exponential growth bias, asset-side: Overestimates future value	0.07			
Limited attention *	0.49 *	0.00	0.02	0.00
Limited memory *	0.86 *	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. nonstandard; 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 nonmissing 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 indicaiton 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
Demographics	Gender	Indicator for female
	Age	Four categories: 18-34, 35-45, 46-54, 55+
	Education	Four categories: HS or less, some college/associates, BA, graduate
	Income	The ALP's 17 categories (collapsed into deciles in some specifications)
	Race/ethnicity	Three categories: White, Black, or Other; separate indicator for Hispanic
	Marital status	Three categories: married/co-habitating; separated/divorced/widowed; never married
	Household size	Five categories for count of other members: 0, 1, 2, 3, 4+
	Employment status	Five categories: working, self-employed, not working, disabled, missing
	Immigrated to USA	Indicator for immigrant
	State of residence	Indicator for each state
Risk, patience	Risk aversion (financial)	100-point scale on financial risk-taking from Dohmen et al., with higher values indicating greater risk aversion
	Risk aversion (income)	6 point scale based on Barsky et al lifetime income gambles, higher values indicate greater aversion
	Patience	Average savings rate across the 24 Convex Time Budget decisions, standardized
Cognitive skills	Fluid intelligence	# correct on standard 15-question, nonadaptive number series quiz
	Numeracy	# correct on Banks and Oldfield questions re: division and %
	Financial literacy	# correct on Lusardi and Mitchell "Big Three" questions re: interest, inflation, and diversification
	Executive attention	# correct on 2-minute Stroop test; respondents instructed to answer as many q's correctly as they can
Survey effort	Time spent on questions	Measured for each B-factor (and other variables), included as decile indicators relative to other responde
	Missing variable(s)	Indicators
	Did not take our 2nd survey	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

		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	.033***	1							
	Owns stocks	0.49	0.34***	0.82***	1						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				
Variable	**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

Unit of observation is the individual respondent. Unconditional pairwise correlations with p-values: * 0.10 ** 0.05 *** 0.01. Pairwise sample sizes range from 1,391 to 1,508.

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

RHS B-factor shown as column header. LHS variable in all models is summary financial condition (mean = 0.43)

	(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**	**Overconfident precision**	**Overconfident relative perf.**	**Nonbelief Law Large Numbers**	**Gambler's fallacies**	**Exponential growth bias: loan**	**Exponential growth bias: asset**	**Limited attention**	**Limited memory**
Standard bias indicator	-0.038**	-0.020	-0.013	-0.021	-0.006	-0.010	0.002	-0.025*	-0.020	-0.018	-0.037**	-0.023	-0.028	-0.006	-0.063***	-0.109***	-0.042**
	(0.016)	(0.018)	(0.013)	(0.034)	(0.017)	(0.013)	(0.013)	(0.014)	(0.015)	(0.014)	(0.017)	(0.020)	(0.020)	(0.019)	(0.017)	(0.012)	(0.018)
Non-standard bias indicator	-0.005	-0.018						0.037*				-0.016	0.001	-0.038			
	(0.016)	(0.025)						(0.021)			(0.016)	(0.021)	(0.026)				
bias missing	-0.009	0.033	-0.012	-0.025	0.016	0.033	0.262	-0.255	0.070*	-0.228*	0.168**	0.060	0.016	-0.023	-0.067	0.016	
	(0.028)	(0.051)	(0.023)	(0.040)	(0.021)	(0.048)	(0.167)	(0.275)	(0.036)	(0.122)	(0.074)	(0.146)	(0.030)	(0.025)	(0.052)	(0.042)	
female	-0.020	-0.020	-0.017	-0.017	-0.019	-0.020	-0.020	-0.019	-0.020	-0.019	-0.019	-0.020	-0.019	-0.019	-0.019	-0.018	-0.018
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Ed: some college/ associates	-0.040**	-0.041**	-0.039**	-0.039**	-0.041** -0.040**	-0.041**	-0.048**	-0.043**	-0.042**	-0.040**	-0.042**	-0.042**	-0.044**	-0.044**	-0.039**	-0.041**	
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.017)	(0.018)
highest ed: bachelor's	0.017	0.015	0.016	0.016	0.013	0.015	0.015	0.013	0.012	0.011	0.015	0.012	0.014	0.011	0.013	0.004	0.013
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.022)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.020)	(0.021)
highest ed: graduate	0.031	0.029	0.031	0.030	0.029	0.032	0.030	0.030	0.028	0.027	0.033	0.026	0.028	0.028	0.028	0.011	0.027
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.025)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.023)	(0.024)
age 35-45	-0.016	-0.010	-0.012	-0.012	-0.014	-0.016	-0.016	-0.011	-0.011	-0.014	-0.013	-0.017	-0.012	-0.016	-0.015	-0.016	-0.015
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.017)	(0.018)
age 46-54	0.003	0.006	0.002	0.003	0.004	0.001	-0.000	0.002	0.007	0.004	0.004	0.003	0.005	-0.002	-0.001	-0.003	0.001
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.020)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.018)	(0.019)
age>=55	0.053**	0.057***	0.053**	0.053**	0.057***	0.050**	0.049**	0.052**	0.056***	0.054**	0.053**	0.046**	0.053**	0.047**	0.046**	0.045**	0.049**
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.022)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.020)	(0.021)
income decile 2	-0.029	-0.029	-0.028	-0.028	-0.026	-0.030	-0.028	-0.029	-0.020	-0.023	-0.020	-0.026	-0.030	-0.025	-0.025	-0.018	-0.030
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.028)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.026)	(0.027)
income decile 3	0.018	0.021	0.021	0.022	0.026	0.019	0.022	0.027	0.029	0.026	0.028	0.028	0.019	0.025	0.027	0.039	0.022
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.029)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.027)	(0.028)
income	0.074**	0.076***	0.078*	0.078*	0.078*	0.074*	0.079*	0.073**	0.084***	0.081*	0.085*	0.080*	0.073*	0.084***	0.083***	0.097*	0.083*

decile 4	*		**	**	**	**	**			**	**	**	**			**	**
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.029)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.027)	(0.028)
income decile 5	0.087***	0.089*** 0.089***	0.089***	0.092*** 0.085***	0.091***	0.091*** 0.095***	0.091***	0.096*** 0.096***	0.083***	0.095***	0.094***	0.117***	0.094***				
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.032)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.030)	(0.031)
income decile 6	0.138***	0.142***	0.139***	0.138***	0.144***	0.137***	0.141***	0.129***	0.143***	0.144***	0.146***	0.147***	0.138***	0.149***	0.146***	0.172***	0.146***
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.032)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.030)	(0.031)
income decile 7	0.174***	0.177*** 0.175***	0.175***	0.175*** 0.171***	0.178***	0.166*** 0.186***	0.179***	0.186*** 0.180***	0.168***	0.182***	0.184***	0.201***	0.178***				
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.032)	(0.030)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.030)	(0.030)
income decile 8	0.261***	0.263***	0.265***	0.266***	0.269***	0.262***	0.267***	0.274***	0.270***	0.265***	0.269***	0.268***	0.260***	0.273***	0.272***	0.287***	0.272***
	(0.032)	(0.032)	(0.031)	(0.031)	(0.031)	(0.032)	(0.032)	(0.033)	(0.031)	(0.031)	(0.031)	(0.031)	(0.032)	(0.032)	(0.032)	(0.031)	(0.031)
income decile 9	0.296***	0.297***	0.299***	0.299***	0.303***	0.296***	0.303***	0.289***	0.304***	0.302***	0.306***	0.304***	0.294***	0.308***	0.302***	0.323***	0.308***
	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	(0.032)	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	(0.029)	(0.030)
income decile 10	0.374***	0.378***	0.379***	0.381***	0.374***	0.374***	0.381***	0.375***	0.382***	0.380***	0.384***	0.388***	0.368***	0.391***	0.384***	0.391***	0.387***
	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	(0.047)	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	(0.045)	(0.044)	(0.043)	(0.044)

Table 5, cont.: Estimating Equation (1) for each B-factor: Conditional correlations between financial condition and behavioral biases

RHS B-factor shown as column header. LHS variable in all models is summary financial condition (mean = 0.43)

	(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 percision**	**Over-confident relative perf.**	**Nonbelief Law of Large Numbers**	**Gambler's fallacies**	**Exponential growth bias: loan**	**Exponential growth bias: asset**	**Limited attention**	**Limited memory**
fluid intell # correct	0.001	0.001	0.000	0.000	0.002	0.001	0.001	0.001	0.001	0.001	-0.004	0.000	0.000	0.001	-0.002	0.001	0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
numeracy # correct	-0.001	0.000	-0.000	0.000	0.002	-0.001	-0.000	0.003	-0.008	0.005	0.003	0.001	0.001	0.002	0.002	0.001	-0.001
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
financial literacy # correct	0.028***	0.029***	0.029***	0.029***	0.031***	0.029***	0.027***	0.030***	0.029***	0.031***	0.031***	0.027***	0.027***	0.029***	0.021**	0.027***	0.028***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(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***	-0.006***	-0.007***	-0.006***	-0.006**	-0.006**	-0.006***	-0.005**	-0.006***	-0.007***	-0.006***	-0.006***	-0.006***	-0.006***	-0.006***	-0.005**	-0.006**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
risk aversion (income)	0.009*	0.010*	0.009*	0.009*	0.009*	0.009*	0.009*	0.006	0.009*	0.009*	0.008*	0.009*	0.008*	0.009*	0.009*	0.007	0.009*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
patience (stdized)	0.006	0.007	0.008	0.008	0.007	0.008	0.008	0.008	0.008	0.007	0.006	0.008	0.007	0.007	0.007	0.005	0.009
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Response time spent on B- factor in this regression:																	
decile 2	-0.001	-0.043	0.008	0.007	-0.032	0.017	0.003	-0.032	0.010	-0.023	-0.057**	0.034	-0.027	-0.027	-0.004	-0.047*	0.017
	(0.027)	(0.027)	(0.028)	(0.028)	(0.028)	(0.026)	(0.027)	(0.028)	(0.024)	(0.027)	(0.027)	(0.028)	(0.028)	(0.028)	(0.027)	(0.026)	(0.026)
decile 3	-0.005	-0.040	0.017	0.018	-0.005	0.018	-0.011	-0.062**	0.040	0.019	-0.047	0.042	-0.007	-0.040	0.041	-0.041	0.008
	(0.027)	(0.027)	(0.029)	(0.029)	(0.028)	(0.026)	(0.027)	(0.028)	(0.027)	(0.027)	(0.029)	(0.029)	(0.026)	(0.029)	(0.027)	(0.027)	(0.028)
decile 4	0.018	-0.011	-0.004	-0.003	-0.050*	0.059**	0.010	-0.009	-0.035	0.021	-0.014	0.023	0.011	-0.039	0.012	-0.018	-0.036
	(0.028)	(0.028)	(0.029)	(0.029)	(0.028)	(0.027)	(0.027)	(0.028)	(0.029)	(0.027)	(0.028)	(0.028)	(0.028)	(0.029)	(0.029)	(0.026)	(0.026)

decile 5	-0.001	-0.017	0.027	0.028	0.018	0.038	-0.002	-0.002	0.009	-0.005	-0.038	0.042	0.030	-0.047	0.024	-0.024	-0.020
	(0.027)	(0.026)	(0.030)	(0.030)	(0.028)	(0.027)	(0.027)	(0.028)	(0.030)	(0.027)	(0.028)	(0.029)	(0.028)	(0.029)	(0.027)	(0.027)	(0.029)
decile 6	0.017	-0.050*	-0.027	-0.026	-0.049*	0.033	-0.002	-0.039	0.004	0.023	-0.046*	0.046	0.021	-0.023	0.005	0.008	0.034
	(0.028)	(0.027)	(0.029)	(0.029)	(0.028)	(0.027)	(0.027)	(0.027)	(0.031)	(0.028)	(0.027)	(0.030)	(0.026)	(0.029)	(0.028)	(0.027)	(0.029)
decile 7	0.005	-0.033	-0.002	-0.001	-0.023	0.009	0.014	-0.022	-0.030	-0.016	-0.051*	0.002	-0.033	-0.025	0.051*	-0.037	0.031
	(0.028)	(0.028)	(0.030)	(0.030)	(0.028)	(0.028)	(0.028)	(0.029)	(0.028)	(0.028)	(0.028)	(0.029)	(0.027)	(0.030)	(0.029)	(0.026)	(0.027)
decile 8	0.012	-0.024	0.008	0.009	-0.033	0.014	0.017	-0.026	0.030	-0.007	-0.055*	0.025	-0.024	0.004	0.001	-0.027	-0.028
	(0.028)	(0.028)	(0.029)	(0.029)	(0.029)	(0.027)	(0.027)	(0.028)	(0.029)	(0.028)	(0.028)	(0.029)	(0.028)	(0.030)	(0.028)	(0.027)	(0.028)
decile 9	-0.023	-0.018	0.017	0.018	0.004	0.002	0.002	-0.011	-0.029	0.033	-0.034	0.025	-0.018	-0.027	0.008	-0.023	0.018
	(0.028)	(0.028)	(0.030)	(0.030)	(0.028)	(0.027)	(0.027)	(0.028)	(0.028)	(0.028)	(0.028)	(0.030)	(0.027)	(0.030)	(0.028)	(0.027)	(0.028)
decile 10	0.004	-0.014	-0.030	-0.029	-0.039	0.017	0.015	-0.036	0.054*	-0.014	-0.022	0.052*	-0.006	-0.002	0.024	-0.017	0.016
	(0.028)	(0.028)	(0.029)	(0.029)	(0.028)	(0.027)	(0.027)	(0.028)	(0.029)	(0.028)	(0.028)	(0.029)	(0.027)	(0.030)	(0.028)	(0.027)	(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 (NBLLN), 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

	(1)	(2)	(3)	(4)
	Survey/task effort weight			
B-factor (standard bias)	**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**	-0.040**	-0.053***	-0.053***
	(0.016)	(0.016)	(0.016)	(0.016)
Time-inconsistent snack (present-bias)	-0.020	-0.019	-0.019	-0.009
	(0.018)	(0.018)	(0.017)	(0.018)
Violates GARP	-0.013	-0.012	-0.002	-0.006
	(0.013)	(0.013)	(0.013)	(0.013)
Violates GARP FOSD	-0.021	-0.024	-0.044	-0.061*
	(0.034)	(0.033)	(0.031)	(0.032)
Certainty premium (preference for certainty)	-0.006	-0.005	-0.005	0.005
	(0.017)	(0.017)	(0.017)	(0.017)
Loss averse	-0.010	-0.011	-0.021*	-0.018
	(0.013)	(0.013)	(0.013)	(0.013)
Narrow-brackets	0.002	0.002	-0.001	0.001
	(0.013)	(0.013)	(0.013)	(0.012)
Ambiguity averse	-0.023	-0.024*	-0.024*	-0.027*
	(0.015)	(0.014)	(0.014)	(0.014)
Confidence in level performance (over-)	-0.020	-0.016	-0.009	-0.019
	(0.015)	(0.015)	(0.014)	(0.014)
Overconfident precision	-0.018	-0.019	-0.018	-0.025*
	(0.014)	(0.013)	(0.013)	(0.013)
Overconfident relative perf.	-0.037**	-0.037**	-0.047***	-0.048***
	(0.017)	(0.017)	(0.017)	(0.018)
Non-belief Law Large Numbers (underestimates)	-0.023	-0.025	-0.030	-0.012
	(0.020)	(0.019)	(0.019)	(0.020)
Gambler's fallacies (hot hand)	-0.028	-0.027	-0.027	-0.028
	(0.020)	(0.019)	(0.019)	(0.019)
Exponential growth bias loan (underestimates)	-0.006	-0.001	0.008	-0.007
	(0.019)	(0.019)	(0.019)	(0.019)
Exponential growth bias asset (underestimates)	-0.063***	-0.061***	-0.050***	-0.046***
	(0.017)	(0.016)	(0.016)	(0.016)
Limited attention	-0.109***	-0.109***	-0.111***	-0.105***
	(0.012)	(0.012)	(0.012)	(0.012)
Limited memory	-0.042**	-0.042**	-0.048***	-0.044**
	(0.018)	(0.018)	(0.017)	(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. Average B-factor effect, all standard B-factors Average B-factor effect, excluding stand-alone sig. coeffs.

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) Full specification	(2) Remove cognitive skills	(3) Remove classical preferences	(4) Remove age, education, gender	(5) 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 elicitations. 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.