THE QUEST FOR PARSIMONY IN BEHAVIORAL ECONOMICS:
NEW METHODS AND EVIDENCE ON THREE FRONTS

Victor Stango
Joanne Yoong
Jonathan Zinman

The Quest for Parsimony in Behavioral Economics: New Methods and Evidence on Three
Fronts
Victor Stango, Joanne Yoong, and Jonathan Zinman
NBER Working Paper No. 23057
January 2017
JEL No. D03,D14,D6,D8,D9,E03,G02

ABSTRACT

Behavioral economics identifies myriad deviations from classical economic assumptions about consumer decision-making, but lacks evidence on how its diverse phenomena fit together and whether they are amenable to modeling as low-dimensional constructs. We pursue such parsimony on three fronts, with success on two and instructive failure on the third. Elicitation parsimony reduces impediments to data collection by streamlining standard methods for directly measuring a person's behavioral tendencies. We do so for 17 potentially behavioral factors per individual in a large, nationally representative sample, and several sets of results indicate that our streamlined elicitations yield low-cost, high-quality data. Behavioral sufficient statistic parsimony aggregates information across behavioral factors, within-person, to create two new lower-dimensional, consumer-level measures of behavioral tendencies. These statistics usefully capture cross-sectional variation in behavioral tendencies, strongly and negatively correlating with a rich index of financial condition even after (over-)controlling for demographics, classical risk attitudes and patience, cognitive skills including financial literacy, and survey effort. Our quest for common factor parsimony largely fails: within-consumer correlations between behavioral factors tend to be low, and the common factor contributing to all 17 behavioral factors within-individual is weakly identified and does not help explain outcomes conditional on the other covariates. Altogether our results provide many new insights into behavioral factors: their distributions, inter-relationships, distinctions from classical factors, and links to outcomes. Our findings also support the two leading approaches to modeling behavioral factors—considering them in relative isolation, and summarizing them with reduced-form sufficient statistics—and provide data and methods for honing both approaches.

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

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

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

The Quest for Parsimony in Behavioral Economics:
New Methods and Evidence on Three Fronts

Victor Stango, Joanne Yoong, and Jonathan Zinman[*]

January 2017

Behavioral economics identifies myriad deviations from classical economic assumptions about consumer decision-making, but lacks evidence on how its diverse phenomena fit together and whether they are amenable to modeling as low-dimensional constructs. We pursue such parsimony on three fronts, with success on two and instructive failure on the third. Elicitation parsimony reduces impediments to data collection by streamlining standard methods for directly measuring a person's behavioral tendencies. We do so for 17 potentially behavioral factors per individual in a large, nationally representative sample, and several sets of results indicate that our streamlined elicitations yield low-cost, high-quality data. Behavioral sufficient statistic parsimony aggregates information across behavioral factors, within-person, to create two new lower-dimensional, consumer-level measures of behavioral tendencies. These statistics usefully capture cross-sectional variation in behavioral tendencies, strongly and negatively correlating with a rich index of financial condition even after (over-)controlling for demographics, classical risk attitudes and patience, cognitive skills including financial literacy, and survey effort. Our quest for common factor parsimony largely fails: within-consumer correlations between behavioral factors tend to be low, and the common factor contributing to all 17 behavioral factors within-individual is weakly identified and does not help explain outcomes conditional on the other covariates. Altogether our results provide many new insights into behavioral factors: their distributions, inter-relationships, distinctions from classical factors, and links to outcomes. Our findings also support the two leading approaches to modeling behavioral factors—considering them in relative isolation, and summarizing them with reduced-form sufficient statistics—and provide data and methods for honing both approaches.

"Everything should be as simple as it can be, but no simpler."
(Attributed to Albert Einstein)

One hundred years of research on human decision making has, in some fields, distilled what first appeared to be countless factors into relatively parsimonious but still-informative constructs like fluid and crystalized intelligence, or the Big Five personality traits. Such constructs can facilitate more tractable and portable measurement and modeling, opening new avenues for research and applications.

Researchers are just beginning to pursue such parsimony in behavioral economics (BE), a younger field that studies deviations from classical economic specifications of preferences, beliefs, and problem-solving approaches. BE is arguably still in the "countless factors" phase of its development (e.g., Fudenberg 2006; Chetty 2015),[1] and thus far behavioral factors ("B-factors") largely have been studied in isolation from each other.[2] Such studies suggest manifold and plentiful behavioral deviations (e.g., DellaVigna 2009), leading micro and macro models alike to employ a variety of assumptions about the magnitude, heterogeneity, and impacts of B-factors (e.g., Akerlof 2002; Campbell 2016; Driscoll and Holden 2014; Kőszegi 2014).

Greater parsimony in BE would clarify and improve applications across many domains. Policymakers invoke B-factors as a basis for specific rulemakings (as done recently by, e.g., the Department of Energy). Many agencies formulate high-level strategy based on assumptions about how B-factors *as a whole* are prevalent and impact decision-making (e.g., the Consumer Financial Protection Bureau, SEC, Financial Conduct Authority, World Bank). "Nudge units" and other centers of applied behavioral social sciences operate on similar premises.

A key obstacle to parsimony has been the lack of empirical evidence on whether and how BE's diverse phenomena fit together and have distinct links to outcomes. We fill this void using a novel research design that generates consumer-level data on outcomes, 17 potentially

---

[1] See also a recent RFP from the National Institutes for Health calling for the "identification and measurement of appropriate economic phenotypes in population-based studies, based on approaches honed in behavioral and experimental studies." (http://grants.nih.gov/grants/guide/rfa-files/RFA-AG-16-010.html).

[2] For other work and discussions re: interactions among behavioral factors and other challenges in behavioral modeling, see, e.g., Benjamin et al. (2016), Dean and Ortoleva (2016), Ericson (forthcoming), Heidhues et al. (2016), and O'Donoghue and Rabin (1999).

behavioral factors and other decision inputs, for a large representative sample. This design permits the pursuit of greater parsimony in BE along three complementary fronts.

"Elicitation parsimony" seeks to reduce research costs, in two ways. One is by reducing the cost of directly eliciting a single B-factor (e.g., present-biased discounting of money) using stylized tasks/questions, thereby expanding possibilities for measuring multiple B-factors (e.g., also measuring loss aversion, overconfidence, limited memory, etc.) in large representative samples.[3] A related question is whether one must necessarily elicit multiple factors for sound research design, or whether one can usefully examine B-factors in relative isolation. If B-factors are strongly correlated with each other within-person, research designs that examine one or a few B-factors might suffer from omitted variable bias. So, we estimate whether those B-factor correlations are strong, and directly estimate omitted variable bias by comparing specifications estimating conditional correlations between outcomes and single B-factors to those controlling for other B-factors.

Our second front pursues "behavioral sufficient statistic parsimony" by aggregating myriad B-factors into summary consumer-level indexes that usefully capture heterogeneity in behavioral tendencies across consumers. This approach furthers the development of reduced-form behavioral sufficient statistic models by providing new data and methods for refining assumptions, testing predictions, and developing applications.[4]

Third, we pursue "common factor parsimony" by looking for a lower-dimensional set of latent or observed variables that drive myriad B-factors and their relationship to outcomes. Such parsimony would reduce elicitation costs, simplify theory and modeling, and perhaps provide an

---

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

[4] For examples of reduced form behavioral sufficient statistic models see Chetty et al. (2009), Mullainathan et al. (2012), Allcott and Taubinsky (2015), Baicker et al. (2015), Farhi and Gabaix (2015), and Gabaix (2016); for overviews see Chetty (2009; 2015).

overarching intuitive explanation for what "being behavioral" means and whence it comes—as fluid and crystallized intelligence have done for work on cognitive skills, for example.[5]

We assay these three fronts by directly eliciting data on *seventeen* behavioral factors within consumers, in a large, representative sample. We streamline standard elicitation methods from recent high-profile papers by shortening, simplifying and combining tasks. Only one of our many elicitations is incentivized on the margin. Our survey modules can be administered quickly and inexpensively, compared to standard methods that use in-person instruction or moderation, many task repetitions, and/or financial incentives.

One set of factors relates to preferences: present-biased discounting (Read and van Leeuwen 1998; Andreoni and Sprenger 2012), loss aversion (Fehr and Goette 2007), preference for certainty (Callen et al. 2014), ambiguity aversion (Dimmock et al. 2016), and choice inconsistency (Choi et al. 2014).[6] Other B-factors capture biased beliefs, biased perceptions, and behavioral decision rules: three varieties of overconfidence (Moore and Healy 2008), narrow bracketing (Rabin and Weizsäcker 2009), exponential growth bias (Stango and Zinman 2009; Levy and Tasoff 2016), statistical fallacies (Dohmen et al. 2009; Benjamin, Moore, and Rabin 2013; Benjamin, Rabin, and Raymond 2016), and limited attention/memory (Ericson 2011). Whether these 17 B-factors are distinct/different from each other is a question we explore in depth, with particular focus on B-factors that are linked theoretically/conceptually (e.g., discounting or preference biases; math biases; overconfidence varieties). Each B-factor potentially links to financial decisions and outcomes, as we detail in the Data Appendix.

Our methods also leave time for collecting rich data on outcomes and control variables, allowing us to estimate conditional correlations between outcomes and behavioral tendencies in the cross-section of consumers. In this paper, we focus on financial outcomes, constructing a rich

---

[5] The only other paper we know of pursuing a common factor approach to BE is Dean and Ortoleva (2016), which examines a student sample and does not explore links to outcomes. See also footnote 10.

[6] This paragraph cites the papers that had the greatest influence on our methods—which rely on direct elicitation—for measuring behavioral factors. We emphasize that neither these cites nor our elicitations are meant to be exhaustive: they do not cover all of the important work on each behavioral factor, nor do they cover the complete set of potential behavioral factors (e.g., we do not attempt to measure projection bias, or anything about social preferences) or the complete set of methods for identifying behavioral biases and their effects. Below we discuss how our direct elicitation methods can complement other methods.

multi-dimensional metric of objective and self-assessed financial condition.[7] Our controls match or go beyond prior exercises linking B-factors to outcomes in the cross-section by including cognitive skills (e.g., fluid intelligence and executive function), knowledge (e.g., financial literacy), classical preferences/attitudes regarding time and risk, and detailed demographics. Our data also track survey response times at the respondent-question level, which we use to assess and correct for any relationships among survey effort, B-factor measures, and outcomes. We lack measures of non-cognitive skills like personality traits, but prior evidence finds weak if any correlations between those skills and B-factors,[8] and hence *a priori* there is little concern that omitting non-cognitive skills will affect inferences about B-factors. But overall, and relative to previous work, we err on the side of over-controlling in estimating links between B-factors and outcomes, although we also estimate more parsimonious specifications of links between outcomes, B-factors, cognitive skills, and classical preferences.

RAND Corporation administered our elicitations in two modules through 2014 and 2015, as part of its nationally representative American Life Panel (ALP). Together the two modules span about 60 minutes of survey time and have been taken by over 1,400 panelists.

We confirm elicitation parsimony in several ways. Some of our elicitations contain standard built-in data quality checks, and those checks indicate quality comparable to methods using more (in-person) instruction, task repetition, and financial incentives. Our estimates of B-factor prevalence, structural parameters, and heterogeneity are in line with those from comparable prior studies. If anything, our elicitations classify fewer people as behavioral, suggesting that streamlined elicitations need not bias toward finding greater prevalence of B-factors.

Our measured B-factors are empirically useful in that they correlate with outcomes. Like in prior work, we find that B-factors are conditionally correlated with financial outcomes in the cross-section, with the overall sign pattern across the 17 B-factors pointing to economically meaningful negative correlations with financial condition, conditional on our rich controls. "Standard" biases (such as present-biased discounting, or underestimating the effects of compounding, or overconfidence) drive those negative relationships with outcomes. "Non-

---

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

[8] See e.g., Becker et al.'s (2012) review article and subsequent papers citing it.

standard" biases (such as future-biased discounting, or overestimating the effects of compounding, or underconfidence) do not exhibit a clear correlational pattern.

We also find that B-factors are distinct correlates of decisions and outcomes: they do not simply capture unmeasured aspects of classical preferences, or cognitive skills, or other covariates. B-factors are weakly correlated with, and poorly fit by, other individual characteristics.[9] Indeed, in most cases our rich set of covariates explains less than 10% of cross-sectional variation in B-factors. Nor does the strength of empirical links between B-factors and outcomes depend on whether one controls for cognitive ability, or survey effort, or classical preferences, arguing against omitted components of those covariates as explaining our B-factors.

B-factors are also largely distinct from *each other*. Although some B-factors are positively correlated within-person, those correlations are often small in magnitude.[10] More critically, we show that point estimates on links between single B-factors and outcomes are essentially invariant to whether one controls for (even the full set of) other B-factors or not. These results are not an artifact of measurement error; e.g., they hold even when controlling for B-factors that have a strong stand-alone conditional correlation with outcomes. And a variety of other findings suggest that, individually and collectively, our B-factor measures are empirically informative.

Success in behavioral sufficient statistic parsimony comes from aggregating across B-factors to create two consumer-level indices for behavioral tendencies. The "B-count" simply sums how

---

[9] Among correlations between measures and behavioral factors and other decision inputs, cognitive skills have been a particular focus in prior work. Our findings are in line with this prior work in the sense that we tend to find negative correlations between behavioral tendencies and fluid intelligence (Benjamin, Brown, and Shapiro 2013; Burks et al. 2009; Frederick 2005), although the correlations are modestly-sized and do not hold across all B-factors (Cesarini et al. 2012; Li et al. 2013). Please see Section 2-C for details. See also Dohmen et al. (2010) on correlations between measures of classical preferences/attitudes and intelligence.

[10] Dean and Ortoleva (2016) and Gillen et al. (2015) find positive correlations within-individual in student samples. Several papers have reported correlations among a smaller number of behavioral factors; see, e.g., Burks et al. (2009) and Li et al. (2013). We focus here on the behavioral economics literature while noting that there are related literatures in other disciplines; e.g., on "deviations from rational thinking." Those literatures are also only beginning to grapple with correlations among factors that economists might consider behavioral (e.g., Stanovich 2016).

many B-factors each individual exhibits, on a 0/1 basis for each. The "B-tile" captures the degree of bias across all B-factors, in normalized percentile units.[11]

Both sufficient statistics show that consumer-level behavioral tendencies are both common and heterogeneous. Nearly everyone exhibits multiple deviations from classical norms, even when one excludes deviations likely stemming from "trembles" and other presumably harmless ways in which humans differ from economic robots. But there is significant variation across consumers in both the breadth and type of behavioral tendencies. That heterogeneity, as with the heterogeneity in single B-factors, is not well-explained by demographics, cognitive skills, preferences, survey effort, and so on. As an example, B-count variation *within* high-income consumers dwarfs the variation between high- and low-income consumers.

Beyond serving as descriptors of consumer-level behavioral tendencies, our behavioral sufficient statistics negatively and robustly correlate with financial outcomes. They also explain a substantial proportion of cross-sectional variation in financial condition. As with the results on single B-factors, the sufficient statistic results hold conditional on our rich set of controls including cognitive skills and knowledge, classical preferences, demographics and survey effort.

Increasing the B-count by one standard deviation is associated with a conditional reduction in financial condition of about 15% on its mean. This is comparable to the magnitude on gender, education, and financial literacy, and it exceeds the magnitude on other aspects of cognitive skills and many other variables commonly thought to be important correlates of financial decisions and outcomes. On a stand-alone basis our sufficient statistics by themselves explain 11% of cross-sectional variation in financial condition, similar to the 12% for cognitive skills when they are the only explanatory variables in the model.[12]

---

[11] Note that the B-tile is a relative, not absolute, measure of the size of behavioral deviations. One could take a similar approach for estimating absolute deviations, e.g., by using parameter estimates for each B-factor. We do not take this approach to constructing a behavioral sufficient statistic because about half of our elicitations are too coarse to yield parameter estimates. But, where possible, we do estimate structural parameters for single B-factors, and discuss their applications, below.

[12] Our findings on behavioral summary statistics and B-factors add to the extensive literature on the cross-sectional correlates and contributions to fit of wealth accumulation and other measures of household financial condition (Poterba 2014; Campbell 2016), by showing that behavioral tendencies are widespread and have economically important links to financial condition. The only other paper we know of that estimates relationships between field outcomes and multiple behavioral factors, in a nationally representative sample, is Goda et al. (2015), which does so for present-biased money discounting and

Additional results provide further guidance on how to construct and interpret behavioral sufficient statistics. Our "B-stats" conditionally correlate with both "hard" outcomes like wealth, savings and stock market participation, and with "soft" self-assessed outcomes like financial satisfaction, financial stress and savings adequacy. Bi-directional biases exhibit the pattern predicted by most existing work: standard biases negatively correlate with financial condition while non-standard biases do not. "Math biases" (where the classical benchmark is demonstrably correct numerically) and all other biases have identical conditional correlations with financial condition. Further extensions and robustness checks yield little evidence that our summary statistics simply reflect survey effort, noise, cognitive skills, classical preferences or anything other than what we set out to measure: behavioral tendencies per se.

The third front in our quest explores prospects for common factor parsimony. We find that standard exploratory factor analysis does suggest the presence of a single behavioral common factor (or at most three). But the few highly-weighted B-factors in that common factor are not linked by any existing theory or construct we can think of.[13] Moreover—and critically for economic applications—we find no evidence of a link between that common factor and financial condition once we condition on observable cognitive skills. Any latent behavioral common factor seems to capture cognitive skills rather than behavioral tendencies per se, and not in any informative way, as our cognitive skill covariates subsume the common factor. We further confirm the distinction between the common factor and behavioral sufficient statistics by decomposing our B-counts and B-tiles, into the piece explained by the common factor and the "residual" component uncorrelated with the common factor. The former does not explain outcomes, while the latter does.

Altogether our results provide new, nationally representative evidence on behavioral factors: their distributions, inter-relationships, distinctions from classical factors, and links to other

exponential growth bias. Bruine de Bruin, Parker, and Fischoff (2007) and Li et al. (2015) also consider a relatively small set of behavioral factors, in convenience samples, as do von Gaudecker et al. (2011) in a representative Dutch sample without exploring links to field behavior. Tanaka et al. (2010) do lab-style elicitations for estimating loss aversion, present-bias, and probability-weighting for 181 Vietnamese villagers, and link those elicitations to survey data (on income, etc.), but they consider each behavioral factor independently.

[13] If one has a strong prior about which B-factors should be inter-correlated and/or load on a common factor, one could use our data to test that prior with confirmatory factor analysis.

covariates and outcomes. As such they have several immediate implications for modeling and applying BE.

They cast doubt on prospects for reducing BE to a low-dimensional set of latent (or observed) characteristics, *a la* fluid and crystalized intelligence for cognitive skills or the Big Five taxonomy for personality. To the contrary: we find that the empirical relationship between separate B-factors and field behavior is roughly additive, meaning that measuring many B-factors explains financial outcomes much better than measuring just a few, or better than measuring many and extracting their common factor. Of course, financial outcomes are but one domain, and links between factor structure and outcomes could differ across domains.

On the other hand, our findings are good news for the current leading approaches to dealing with multiple potentially behavioral factors.

They validate and guide research designs that examine one or a few B-factors at a time. Our results suggest that B-factors are largely distinct from each other, and show directly that omitted B-factor bias is not much of concern when estimating conditional correlations between single B-factors and financial condition. Researchers focused on a particular B-factor, and wishing to err on the side of caution, could use our B-factor correlation matrix to identify which other B-factors are plausible confounding influences and collect data on them, perhaps using our streamlined elicitations. In the absence of auxiliary data on other B-factors, one could also use our results to correct for (or bound) omitted variable bias.

More broadly, our data provides much-needed nationally representative evidence on B-factor statistics and their conditional correlations with each other and with outcomes. One can even use our elicitations to estimate structural parameters, as we detail in the Data Appendix. Going forward, our methods will make it easier to obtain additional data that helps refine, calibrate, estimate and/or test behavioral models, low-dimensional or otherwise.[14]

Our data and methods also help ground and refine the reduced-form behavioral sufficient statistic approach to modeling summarized in Chetty (2009; 2015). Those models rest on

---

[14] We summarize the available nationally representative evidence on B-factors in Section 2-B and provide details for both reduced-form descriptions and structural parameters in the Data Appendix.

heretofore-untested assumptions, some of which we validate[15] and some of which we cast doubt upon.[16] But if the general question is whether it makes sense to use and test models containing summary parameters measuring the prevalence of behavioral agents, the number of behavioral agents on a given margin, and the extent of biases, we provide support for that approach. And while applications of the approach to-date have focused on particular decisions and product markets,[17] we show that behavioral sufficient statistics powerfully correlate with outcomes at a level that interests many policymakers and practitioners—the overall domain of household finance.

The rest of the paper details our methods and their implementation, develops various results informing the quest for parsimony along the three fronts, and then discusses how the full picture of our work informs behavioral modeling and survey design. We conclude with some ideas for extensions and future applications.

1. Research Design

In this section we describe our sample, survey design and elicitation methods, and empirical strategies.

A. *The American Life Panel*

Our data come from the RAND American Life Panel (ALP). The ALP is an online survey panel that was established, in collaboration between RAND and the University of Michigan, to study methodological issues of Internet interviewing. Since its inception in 2003, the ALP has expanded to approximately 6,000 members aged 18 and older.

The ALP takes great pains to obtain a nationally representative sample, combining standard sampling techniques with offers of hardware and a broadband connection to potential

---

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

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

[17] E.g., sales taxes on food and alcohol (Chetty, Looney, and Kroft 2009), health insurance (Baicker, Mullainathan, and Schwartzstein 2015), and lightbulbs (Allcott and Taubinsky 2015). A narrow focus is understandable, given the presumed importance of context in mediating the effect of behavioral tendencies. But as discussed above many policy and business decisions are based on broader, domain-level assumptions about the relationship between behavioral decision making and outcomes.

participants who lack adequate Internet access. ALP sampling weights match the distribution of age, sex, ethnicity, and income to the Current Population Survey.

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

B. Our Research Design and Sample

Speaking broadly, our goal is to design elicitation methods that robustly yield data on the widest possible range of behavioral factors at a reasonable cost. We chose a goal of keeping total elicitation time to an hour. We also sought to use elicitation methods that could be employed online rather than in-person (given that in-person elicitation typically comes at higher cost).

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

Per standard ALP practice, we paid panelists $10 per completed module. Beyond that, all but one of our elicitations are unincentivized on the margin. There is prior evidence that unpaid tasks do not necessarily change inferences about behavioral factors in large representative samples (Von Gaudecker, Van Soest, and Wengström 2011; Gneezy, Imas, and List 2015). Unpaid tasks (with hypothetical rewards) may even offer some conceptual advantages (e.g., Montiel Olea and Strzalecki 2014). In any case, paying marginal incentives is feasible in principle for our elicitations but was impractical for this round of data collection given our budget constraints and strategy of eliciting a rich set of B-factors and other data from a large sample.

After extensive piloting, the ALP fielded the first part of our instrument as ALP module 315, sending standard invitations to panel participants aged 18-60 in November 2014. Given our target of 1,500 respondents, the ALP sent 2,103 initial invitations. The invitation remained open until March 2015, but most respondents completed surveys during the first few weeks after the initial invitation, as is typical in the ALP. 1,515 individuals responded to at least one of our

questions in module 315, and those 1,515 comprise the sample for our study and the sample frame for part two of our instrument.

The ALP fielded the second part of our instrument as ALP module 352, sending invitations to everyone who responded to module 315, starting in January 2015 (to avoid the holidays), with a minimum of two weeks in between surveys. We kept that invitation open until July 2015. 1,427 individuals responded in part or whole to that second module.

Taken together, the two modules yielded a high retention rate (1427/1515 = 94%), low item non-response rate, and high response quality (see below, and Data Appendix)—all features that suggest promise for applying our methods in other contexts. We end up with usable data on a large number of behavioral factors for nearly all 1,515 participants: the respondent-level mean count of measurable behavioral factors is 14 out of a maximum of 16 (we measure two of our B-factors using the same elicitation, and so the max here is 16 instead of 17), with a median of 15 and a standard deviation of 2.9. Below we explore the possibility that missingness in behavioral factors is itself informative in explaining outcomes.

Module 352 also included an invitation to complete a short follow-up survey (module 354) the next day. We use responses to the invitation and actual next-day behavior to measure limited memory as described in the Data Appendix (Section L).

C. Measuring behavioral factors: elicitation methods and key antecedents

Given our goals of directly eliciting useful measures of behavioral factors without breaking the bank, we prioritize elicitation methods that have been featured recently in top journals, did not seek to capture social preferences (we had to draw lines somewhere given our constraints), and were short and simple enough (or could be so modified) to fit into modules that would also allocate substantial survey time to measuring control variables (Section 1-D) and outcome variables (Section 1-E).

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

Deviations from classical norms may be uni-directional, as in the case of choice inconsistency: someone either chooses consistently with the General Axiom of Reveal Preference, or does not. For other factors, deviations from classical norms are bi-directional. For example, in the case of discounting one can be either present-biased or future-biased relative to being time-consistent (unbiased). For bi-directional B-factors we define in each case a "standard" direction based on what has been more commonly observed or cited in prior work. For example, work on Exponential Growth Bias (EGB) more commonly finds that people under-estimate than over-estimate the effects of compounding on future values, and so we count under-estimation as the standard bias and over-estimation as non-standard. Our empirics allow for the possibility that standard vs. non-standard biases have different links to behavior.

The Data Appendix provides further details, for each B-factor, on: i) motivation for trying to measure it; ii) our elicitation method and its key antecedents; iii) data quality indicators, including item non-response; iv) sample size (as it compares to that for other factors); v) definitions and prevalence estimates of behavioral *indicators*, with background on the distinctions between standard vs. non-standard directional biases where applicable, at different cutoffs for classifying a deviation from the classical norm as behavioral; vi) descriptions of the *magnitude* and *heterogeneity* of behavioral deviations, including descriptions of the distribution and—where the data permit—estimates of key parameters used in behavioral models; and vii) estimates of conditional correlations with financial outcomes, including particular components of our financial condition index that have particularly strong links to a given B-factor per theory. Wherever possible we also provide comparisons to prior work.

Section 3-A describes how we aggregate across B-factors, within-person, to create summary statistics (B-stats) that are meant to capture cross-sectional variation in behavioral tendencies.

D. Control variables: Standard covariates, preferences, cognitive skills and survey effort

Our modules also elicit rich measures of cognitive skills, risk attitudes, and patience—measures of human capital and preference parameters that plausibly affect decisions and outcomes in classical models. These serve—among other purposes—as control variables in our outcome regressions linking behavioral indicators to financial outcomes. Table 2 summarizes these variables and how we specify them in the empirics.

We measure aspects of cognitive skills using 4 standard tests. We assess general/fluid intelligence with a standard, 15 question "number series" test (McArdle, Fisher, and Kadlec 2007) that is non-adaptive (i.e., everyone gets the same questions). The mean and median number of correct responses in our sample is 11, with a standard deviation of 3. A second test is comprised of 2 "numeracy" questions,[18] labeled as such and popularized in economics since their deployment in the 2002 English Longitudinal Study of Ageing.[19] Our mean number correct is 1.7, with a standard deviation of 0.6. A third test is a 3-question "financial literacy" quiz developed and popularized by Lusardi and Mitchell (2014).[20] The median respondent gets all 3 correct, with a mean of 2 and a SD of 0.93. We also measure executive function—including working memory and the regulation of attention—using a two-minute Stroop task (MacLeod 1991).[21] Each time the subject chooses an answer that action completes what we refer to as a "round."[22] 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 (Appendix Table 1). In some

---

[18] "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.

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

[20] "Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow?"; "Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, how much would you be able to buy with the money in this account?"; "Please tell me whether this statement is true or false: "Buying a single company's stock usually provides a safer return than a stock mutual fund." Response options are categorical for each of the three questions.

[21] 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).

instances we use the principal component of these four measures when we require a uni-dimensional measure of cognitive ability (e.g., in Figure 3).

We also elicit four standard measures of classical risk attitudes/preferences, although we end up using only two of them in the empirics. The first comes from the adaptive lifetime income gamble task developed by Barsky et al. (1997) and adopted by the Health and Retirement Study and other surveys.[23] 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).[24] Those first two measures of risk aversion are what we use as controls (separately) in the empirics. 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 (Appendix Table 2), 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,[25] but find that the Dohmen et al. and Barsky et al. measures included separately have more explanatory power.

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

The ALP tracks and record survey response time, screen-to-screen, and we use this to construct measures of survey effort. Specifically, for each respondent we measure time spent on

---

[22] 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.

[23] This task starts with: "…. Suppose that you are the only income earner in the family. Your doctor recommends that you move because of allergies, and you have to choose between two possible jobs. The first would guarantee your current total family income for life. The second is possibly better paying, but the income is also less certain. There is a 50% chance the second job would double your current total family income for life and a 50% chance that it would cut it by a third. Which job would you take—the first job or the second job?" Those taking the risky job are then faced with a 50% probability that it cuts it by one-half (and, if they still choose the risky job, by 75%). Those taking the safe job are then faced with lower expected downsides to the risky job (50% chance of 20% decrease, and then, if they still choose the safe job, a 50% chance of a 10% decrease).

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

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

each B-factor, and on the survey as a whole. In much of the empirics we employ decile indicators of survey time spent, either overall, or on specific questions (or groups of questions).

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. Those data also include state of residence, which we include in our empirics as state fixed effects.

E. Measuring financial outcomes

Finally, we designed our instrument to elicit rich data on financial outcomes for use as a dependent variable or variables in our empirics below linking B-factors to financial condition. We chose nine indicators of financial condition that we construct from 15 survey questions, 14 of which are in module 315 (the question on non-retirement savings adequacy is in module 352). We drew the content and wording for these questions from other American Life Panel modules and other surveys (including the National Longitudinal Surveys, the Survey of Consumer Finances, the National Survey of American Families, the Survey of Forces, and the World Values Survey). The questions elicit information on net worth, financial assets, recent savings behavior, severe 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.

Table 3 shows the financial indicators, their sample proportions, and correlations between then. In each case "1" indicates plausibly better financial condition (greater wealth, more financial security, better "financial health," etc.). Our indicators include both stocks and flows. They include 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 indicators of severe financial distress in the last 12 months. They also include four "soft" subjective and self-assessed measures of financial well-being: financial satisfaction above median, financial stress below median, and viewing retirement or non-retirement savings as "adequate" or better.

1,508 of our 1,515 respondents provide data we can use to construct one or more of the nine indicators. The median respondent supplies the full nine, with a mean of 8.8 and standard deviation of 0.6. As Table 3 shows, these indicators are positively correlated with each other.

Our main outcome takes the individual-level mean of these nine indicator variables. In our sample the average value of this summary measure is 0.43, meaning that the average respondent exhibits 4 of our 9 indicators of better financial condition.

F. Estimating links between financial condition and B-factors

In addition to describing B-factors (Section 2-B) and B-stats (Section 3-A), and exploring the relationships among B-factors (Section 4), our primary empirical exercise examines links between behavioral tendencies and financial choices/outcomes. We do this first for single B-factors (Section 2-D), and later for B-counts and B-tiles (Section 3-C), by regressing our summary measure of financial condition on measures of behavioral tendencies and our rich set of controls.

Our goal is to estimate conditional correlations and to winnow down the set of likely interpretations. Given that our quests for parsimony are largely descriptive, we view causal relationships between behavioral tendencies and downstream outcomes as but a subset of many potentially informative ones and do not hang our hats on identifying causal effects.

Our main specification at the single B-factor level is:

$$(1) \; FinCond_i = \alpha + \beta_1 Bfactor\_s_i + \beta_2 Bfactor\_ns_i + \beta_3 Bfactor\_miss_i + \gamma X_i + \varepsilon_i,$$

where *i* indexes individuals. *FinCond* is the summary measure of financial condition described in Section 1-E. The next two variables are B-factor indicators for standard bias and non-standard bias (if applicable). The omitted category is unbiased/classical, unless noted otherwise. *X* contains the full set of control variables described in Table 2. In cases where responses for a particular variable are missing we include a "missing" dummy for that observation and variable. In all we have over 100 control variables including categorical variables, derived from up to 33 different underlying control measures.

One issue affecting the interpretation of B-factor (or B-stat) coefficients is measurement error. Classical measurement error in those variables will bias their coefficients toward zero, and

hence push against our main findings of significant correlations. Section 4-B addresses the question of whether measuring multiple B-factors is itself a measurement error correction, but we find that it is not, largely because the B-factors seem to have distinct and additive relationships with financial outcomes.

Non-classical measurement error is more of a concern. There may be cases where someone exhibits a behavioral deviation that is spurious—due to unmeasured low survey effort, a trembling hand, etc. —rather than an indicator of a tendency toward a behavioral bias per se. But for that to affect inferences it must also be the case the spurious behavioral deviations are correlated with spurious misclassification of financial condition as relatively poor, conditional on our observables. This seems unlikely given that our survey user interfaces do not make it easier to respond in a better or worse direction on the financial questions (Appendix Table 3). Nor is there a straightforward (e.g., linear) pattern of correlations between survey response times and financial condition (Appendix Table 4). There is a correlation in the raw data between very brief survey response times (i.e., low effort) and poorer financial condition, but brief response times are also related to income, and once we control for both in a fully specified regression, there is no clear link between survey response time and financial condition (looking ahead, Table 6).

Our ability to parse behavioral factors into standard vs. non-standard biases also helps identify whether non-classical measurement error has an important influence. Theory predicts an asymmetry, if our B-factor measures have sufficient signal: measures of standard biases should be more strongly correlated with outcomes. Our results are consistent with such an asymmetry (Sections 2-D and 3-C). In contrast a spurious measurement error mechanism would likely affect measures of both standard and non-standard biases—and their correlations with outcomes—in the same way.

Two other closely related issues affecting the interpretation of (1) are unobserved heterogeneity and measurement error in control variables. The interpretation of the behavioral coefficients will be muddled if other variables are correlated with our B-factors, also correlated with financial condition, and omitted from the empirical model. Similarly, if *X* is measured with error, and B-factors are measured with error and correlated with those control variables, then our behavioral coefficients could reflect influence of those other variables.

Primary candidates for such omitted/mis-measured variables include other behavioral characteristics, cognitive skills, classical preference parameters, and survey effort. Omitted behavioral characteristics are less of a concern in the behavioral summary statistic models we estimate in Section 3, but a prominent concern when we consider single B-factors in isolation. As such we consider whether inferences change after adding controls for the other 16 B-factors and find little sensitivity (Section 2-D). Our measures of the control variables are detailed but of course measured imperfectly. Accordingly we examine whether inferences are sensitive to the exclusion of potentially confounding control variables, a la Altonji et al. (2005), and find they are not (Sections 2-D and 3-C). We also directly examine the magnitudes of statistical relationships between B-factors and our *measured* control variables, and find that they are too weak to breed cause for serious concern (Section 2-C). Additionally, we stratify by B-factor response times and find similar results across strata (Section 3-D).

We also consider whether the B-factor/B-count 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 factors and decisions. Nevertheless several pieces of evidence push against the reverse causality hypothesis. First, in theory, reverse causality could just as easily push in the opposite direction of our results, with worse financial condition leading to *more* deliberate consideration of elicitation tasks, less measurement error, and hence fewer deviations from classical norms.[26] Second, reverse causality presupposes instability in measured B-factors, and the limited empirical evidence on this supposition suggests that instability is due to measurement error rather than to marginal changes in financial condition or other life circumstances, although discrete (disastrous) events may play a role.[27] Third, reverse causality likely implies relatively strong correlations

---

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

[27] 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. Callen et al. (2014) find that exposure to violent conflict increases preference for certainty. Li et al. (2013) find moderate (in)stability in present-biased money discounting and in loss aversion, over several months. Carvalho et al. (2016) find small changes in present-biased money discounting around payday in a low-income sample, and no changes in choice inconsistency (or in

between B-stats and "softer" outcomes that are probably themselves relatively unstable and subjectively-measured (Section 1-E). Yet those correlations turn out to be no stronger than ones between B-stats and "harder" outcomes that are relatively sticky and objectively-measured (Section 3-D).

2. Results on Single B-Factors and Elicitation Parsimony

This section collects various results that speak to the quality, validity, and utility of the B-factor data produced by our low-cost elicitations. We summarize specific comparisons to prior work in cases where there is comparable prior work using direct elicitation. The Data Appendix contains more details.

A. *Data quality checks*

Many of our elicitations allow for inferences on whether subjects are actively engaged with the task, and for the most part the response patterns suggest that our elicitations produce data of reasonable quality. We focus here on some key direct comparisons to seminal prior work using more-intensive and/or incentivized elicitations, and provide many additional explorations of response patterns in the Data Appendix. For (biased) discounting with respect to money, our abbreviated version of the Convex Time Budget method produces more interior allocations, and more within-subject variance in response to elicitation variables, than in Andreoni and Sprenger (2012). (On the down side, our version produces more instances of non-monotonic demand.) For choice inconsistency, we verify that neither our 11-round version nor the 25-round version in Choi et al. (2014) produces deterioration in consistency as rounds mount. For preference for certainty, our online version of the in-person Callen et al. (2014) elicitation produces comparable proportions of subjects with strange switching patterns, and a substantially greater proportion of subjects who, encouragingly, switch at different points across the 2 multiple price lists.

B. *Descriptive statistics on single B-factors (with comparisons to prior work)*

Our estimates of prevalence, structural parameters, and heterogeneity suggest that our elicitations produce B-factor distributions that are in line with prior studies (Appendix Table 5).

---

cognitive skills, *contra* , e.g., Shah et al. (2012) and Mani et al. (2013)). There is a larger body of evidence on the reliability of non-behavioral measures of time and risk preferences; see Meier and Sprenger (2015) and Chuang and Schechter (2015) for recent reviews.

Comparing prevalence for the 9 B-factors for which we could find prior studies on nationally representative samples, and focusing on the prior elicitations that are closest to ours (e.g., Rabin and Weizsacker for narrow bracketing) our estimate is nearly equal for 2 of the factors (the two choice inconsistency ones), less for 5 (present-biased money discounting, loss aversion, narrow bracketing, and both EG biases), and greater for 2 (present-biased snack discounting and ambiguity aversion).[28] Those comparisons suggest that our elicitations may be less prone than standard ones to classify people as behavioral. Six of our elicitations are rich enough to permit structural estimation, and in those cases (again focusing on the closest available comparisons) we find very similar parameter estimates for present-biased discounting and for the choice inconsistency measures. We find substantial heterogeneity in B-factors across people, which again is in line with findings from comparable prior direct elicitations.

Table 4 focuses on prevalence and missing values for the B-factors in our sample.[29] Estimated prevalence varies substantially across B-factors. The most common B-factors are inconsistency with GARP (and dominance avoidance), non-belief in the law of large numbers, limited memory, and preference for certainty. The least common are discounting biases re: consumption, gambler's fallacies, and overconfidence. Despite the somewhat lower prevalence in our data compared to prior studies, many B-factors are quite prevalent, with some deviation indicated by at least 50% of the sample for 12 of the 16 factors for which we can estimate prevalence. The "standard" directional bias emphasized by prior literature is indeed more prevalent in our data, in 7 of the 8 B-factors for which we can capture bi-directional biases. Finally, missing values (indicating question-skipping) are uncommon: we have usable responses from more than 90% of the sample for 10 of the 17 B-factors and only two B-factors lack usable data for less than 80% of the sample—one of those cases being due to an inherent limitation of the certainty premium elicitation (only subjects who switch at some point on both multiple price lists help identify the CP), rather than true item non-response.

We can also measure the absolute magnitude of deviations from classical norms for 8 of our B-factors (9 if one counts our ambiguity aversion elicitation), and in these cases we see evidence

---

[28] Unless otherwise noted we define prevalence as exhibiting any deviation from the classical benchmark. Appendix Table 5 also reports prevalence estimates that count only "large" deviations as behavioral; we discuss these below.

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

of substantial deviations (see the Data Appendix, and Appendix Table 5 for a summary). All but one of these factors exhibits substantive average deviations, and all of them include a nontrivial proportion of consumers exhibiting "large" deviations. Below, we more neutrally assess the importance of the *degree* of bias by calculating where in the distribution of bias each consumer lies for each factor, and asking whether scaling B-factors that way improves on a simple 0/1 classification; it does not.

C. B-Factors are distinct from demographics, preferences, cognitive skills and survey effort

A third sort of litmus test for achieving elicitation parsimony is whether streamlined elicitations produce B-factor measures that are statistically distinct from measures of other decision inputs. Table 5 implements this test by regressing the standard bias indicator for each B-factor on the entire set of controls described in Table 2:

(2) $Bfactor\_s_i = \alpha + cX_i + \varepsilon_i,$

Each column in Table 5 presents the results from a single regression, with rows showing results on covariates that are particularly important in prior work or for our empirical strategy (we do not show all 100+ covariates due to conserve space).

The overall and most important takeaway is that B-factors are poorly explained by our rich set of covariates. The bottom rows of Table 5 (second page) show that the adjusted R-squared from including *all* covariates exceeds 0.10 for but 3 of 16 B-factors. In most cases the adjusted R-squared is 0.03 or less. Furthermore, when considered as categories, our classical preference and cognitive skills variables rarely explain more than ten percent of variation in B-factors (2 of 16 factors); demographics similarly fail to explain B-factors (1 of 16 instances); and survey effort as proxied by survey response time explains at best 8 percent of variation in a factor, and most often zero or one percent. Our B-factors are distinct measurements, relative to almost any covariate one can imagine. (We address whether that distinctiveness represents signal or noise in Sections 2-D and 3-C, finding support for signal.)

Turning to the conditional correlations in Table 5, we glean four key takeaways. First, the correlations are in line with prior work; e.g., fluid intelligence is more likely to be negatively than positively correlated with behavioral tendencies (Benjamin, Brown, and Shapiro 2013; Burks et al. 2009; Frederick 2005). Second, the overall pattern suggests that B-factors do not

simply proxy for imperfectly measured classical decision inputs. E.g., being more behavioral is not strongly correlated with being less-educated, and the magnitude of correlations between B-factors and cognitive skills is modest is nearly all cases. Third, the pattern on financial literacy is interesting, as it could support disparate interpretations depending on one's prior. One on hand, 12 of the 16 coefficients are negative, which is consistent with the notion that it makes sense to group lack of financial literacy together with B-factors, at least for the purposes of reduced-form modeling (e.g., Campbell 2016). One the other hand, only 4 of the 16 coefficients have p-values <0.1, with, as one would expect, two of these coefficients on "math biases" that are overwhelmingly one-sided (people tend to underestimate the powers of the LLN and compound interest, and not to overestimate them) and hence perhaps closer conceptually to errors than to biases per se. (One useful conceptual distinction between B-factors and financial literacy is that prior work has identified specific directional biases—as opposed to *any* deviation from classical norms/corrections—as particularly important, and we pursue this distinction between "standard" vs. "non-standard" directional biases below.) Fourth, the pattern on survey effort belies the concern about the simple non-classical measurement error confound discussed in Section 1-F, as response times are uncorrelated with behavioral indicators in many cases, and moderate/longer response times are actually *positively* correlated with behavioral indicators in several cases.

Taken together, these patterns suggest that our B-factor measures capture information that is largely distinct from measures of standard factors and survey effort.

D. B-factors and financial condition: results

A litmus test distinguishing signal from noise in our B-factor elicitations is whether they correlate with financial outcomes. Table 6 shows results of equation (1), regressing our rich measure of financial condition on B-factor indicators and our full set of controls from Table 2, with one regression per B-factor per column. We suppress coefficients on many of the 100+ control variables, focusing on variables of interest. Reading within a column, any applicable standard bias vs. non-standard bias distinction is denoted in Table 1 and detailed in the Data Appendix.

Single B-factors are negatively correlated with financial condition, in ways predicted by standard behavioral theories. Five of the sixteen "standard" bias coefficients are negative and statistically significant. A further six are negative with t-statistics >= 1. Overall fourteen

coefficients are negative and two are positive; a sign test rejects the hypothesis that these coefficients collectively center on zero. In contrast, we cannot reject that the nonstandard bias coefficients are centered on zero. Further encouraging evidence comes from the fact that the "standard" coefficient is *more negative* than the non-standard coefficient in every case but one (preference for certainty and NBLLN each have the non-standard bias as the omitted category, since we lack unbiased responses), although for the most part the estimates are too imprecise to reject equality. One should not read too much into the non-significance of any particular B-factor, if for example that B-factor has been found to matter empirically in other work (see the Data Appendix for comparisons and related discussion). In more parsimonious specifications more similar to prior studies, the size and significance of virtually every B-factor coefficient increases (see as an example the discussion of Appendix Table 7 below).

Magnitudes vary but one can easily interpret them since the RHS variables are dummies, the LHS variable is scaled on [0, 1] and the mean of the LHS is 0.43. For example, in the first column, present-bias on money is associated with a reduction in financial condition of 0.038, about a 9% decline from the base. Limited attention and exponential growth bias (on the asset side) have the most-negative coefficients. Collectively, the biases can have quite influential links to financial condition; in the extreme, turning all standard biases "on" and simply adding the coefficients implies a decline in financial condition of 0.40, which would move someone from the 75th percentile to below the 25th in our data. Of course, few consumers are biased on every B-factor, and biases may not combine linearly. We address the latter point below with our B-stats.

The results are robust to the inclusion or exclusion of controls for survey effort, risk aversion, patience, fluid intelligence, numeracy, financial literacy and executive attention (Appendix Table 6). Looking at the control variables themselves in Table 6, being near retirement age, income, financial literacy, and financial risk aversion are all strongly conditionally correlated with financial condition in the expected directions. The lack of statistical significance on other variables, including ones found to strongly correlate with financial outcomes in other studies, is a bit hard to interpret given the extent of our covariates; e.g., it is possible that we are over-controlling. Appendix Table 7 explores this possibility by dropping survey response time and demographics from the regressions, leaving only a B-factor, cognitive skills and classical preferences on the right-hand-side. The magnitudes and significances of the B-factor coefficients all increase in this model, as one would expect. We also now see the expected strong correlations

between financial condition and: fluid intelligence, numeracy, and financial risk aversion. One new finding is that, conditional on financial risk aversion, aversion to income risks *positively* correlates with financial condition.

E. Elicitation parsimony and research design: are B-factors correlated with each other?

Researchers can further pursue elicitation parsimony by eliciting less than the full set of B-factors as part of research design. The merits of that approach hinge, of course, on how strongly B-factors are correlated with each other within-consumer. Table 7 shows tetrachoric correlations among 15 B-factor indicators. We include 15 instead of 17 B-factors in this analysis because: a) the two choice inconsistency factors are strongly positively correlated by construction; b) we lack a cardinal measure of overconfidence in placement (see Data Appendix Section H for details). For B-factors with bi-directional biases we include only the standard directional indicator.

The overall pattern in Table 7 suggests that consumers who are behavioral on one dimension are somewhat more likely to be behavioral on others, but not strongly so. We find more statistically significant positive correlations than one would expect to find by chance (27/105 with p<=0.1), and no more negative correlations (11/105 with p<=0.1), However, the positive correlations are far from universal and fairly small: only four exceed 0.2. There are some intuitive groupings; e.g., the two overconfidence measures are positively correlated, as are three of the four "math biases" (NBLLN, GF, and the two EGBs). Math biases are also positively correlated with overconfidence, narrow bracketing, and limited attention, suggesting the possibility of a low-dimensional construct that links behavioral *beliefs* and *decision rules/problem-solving approaches*. In contrast, we little evidence of correlations amongst B-factors that plausibly capture something about *preferences*: only 3 of the 14 pairwise correlations among the present-bias measures, inconsistency with GARP, preference for certainty, loss aversion, and ambiguity aversion have p-values <0.1, and 2 of the 3 are negatively correlated.[30]

---

[30] Regressions buttress the inference that behavioral factors are positively correlated within-person, but for the most part only weakly so. In untabulated results, we ask how strongly each single B-factor is correlated with the full set of other B-factors, conditional on our controls. We do that by estimating the models of Table 5 with each B-factor on the LHS and the full set of controls on the RHS, adding as an additional control the number of B-factor indicators present, excluding the one on the LHS. It is a quite conservative test, biased toward finding such relationships if they exist, because for example the "B-count

Several results push against the hypothesis that low correlations between B-factors are an artifact of measurement error. Four of the B-factors that have relatively strong conditional correlations with outcomes in Table 6 are weakly correlated with other B-factors in Table 7, suggesting that weak inter-correlations do not reflect lack of signal. The correlations in Table 7 are no lower for consumers with very short response times (results not tabulated), as one would expect if short response times reflected or produced noise. Table 8 is also quite pertinent, as we discuss next.

F. Does omitting other B-factors bias estimates of a single B-factor on financial condition?

A critical question from a research design perspective, given the large number of studies examining one or a few factors in isolation, is whether omitting other B-factors biases estimates of single B-factors on outcomes. We address this in Table 8. The first column shows the B-factor coefficients from the first row in Table 6, in which we estimate single B-factor models correlating B-factors with financial condition. Those models omit other B-factors.

The second column in Table 8 shows that the point estimates on single B-factors are virtually unchanged when one includes the full vector of other B-factor indicators. In fact, one B-factor that was not statistically significant becomes so with those other B-factors included—not what one expects if including new variables creates a severe collinearity problem. The third column shows results of coefficient equality tests across the two models/specifications, confirming that there is but one instance in which the coefficients differ statistically across the two models. The specification used here belies that concern that our B-factors are measured with too much error to be informative in Table 7 or Table 8 Column 3, as the vector of "all other B-factors" in Table 8 Column 2 contains several that have strong stand-alone conditional correlations with outcomes in Table 6. 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

excluding GARP" does *not* exclude the "GARP with FOSD" factor, which is closely related and indeed based on the same survey responses. We treat other similar-in-spirit factors (present-bias measured two ways, over-confidence measured two ways, EGB measured two ways) similarly conservatively. With all that conservatism, we still find that such relationships between each B-factor and all of the others are not terribly strong. Although eleven of the coefficients on "B-count excluding" are significant, four of the eleven are *negative,* meaning that being more behavioral on other dimensions is negatively correlated with that factor. The magnitude of the correlations is modest here as well, as is the contribution of the other B-factors to fitting a given B-factor indicator.

inferences from Table 8 Column 1 to Column 2, *if* there is a true confounding correlation amongst multiple B-factors and outcomes. We find no evidence of such a confound.

Table 8 has important implications for research design. Omitting other B-factors tends not to affect the single B-factor coefficients, supporting the widely-held assumption/convention of modeling B-factors in isolation from each other, at least with respect to B-factor relationships with financial outcomes. Table 8 also provides guidance for researchers when coupled with the results on elicitation parsimony in Tables 5-7. One can elicit B-factors at lower cost and still extract signal. One can elicit fewer than the full set of factors without fearing serious omitted variable concerns (assuming a rich set of other controls). And, for those researchers taking extra care, our correlations and specifications in Tables 7-8 provide guidance, on a factor-by-factor basis, for which set(s) of factors should be elicited together in the interest of comprehensiveness.

3. B-stats and Behavioral Sufficient Statistic Parsimony

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

A. The "B-count" and "B-tile": person-level summary statistics for behavioral tendencies

Our simplest summary measure is the "B-count": the sum of how many (standard and/or nonstandard) B-factor indicators an individual displays from Table 4. It is a count of extensive margins of behavioral deviations from classical benchmarks. Our primary B-count counts *any* deviation as behavioral, although we also consider other definitions.[31] The maximum possible B-count is 16.[32]

Our more nuanced "B-tile" aggregates B-factor-level measures of *how behavioral* an individual is, across B-factors, to the consumer-level. We quantify that intensive margin by calculating each person's percentile ranking for each factor, compared to others in the sample. Some of our factors are continuous, permitting percentiles to take on the full range of values from 1 to 100. For discrete-response and uni-directional outcomes like loss aversion, the –tiles take on fewer values but still measure the degree of deviation from classical benchmarks in

---

[31] A previous working paper version examined the threshold-for-deviation question in detail, with little change in the key inferences.

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

useful ways.[33] Our approach confers comparability across discretely coded factors, based on how many people are in each category relative to the benchmark. For bi-directional deviations, whether discrete or continuous, we calculate percentiles separately in each direction relative to the classical benchmark.

Specifically, the B-tile is the sum of all of the factor-level percentiles, rescaled for comparability to the B-count. We normalize the classical benchmark at 0 (rather than 1) to comport with the B-count: someone who meets every classical benchmark has a B-count and a B-tile both equal to zero. If a person were to be the *most* biased person in the sample on every factor, that person would have a B-tile of (close to) 17. The B-tile has the advantage of being measurable each of our 17 B-factors, and it adds information beyond the factor-level indicator(s) for 15 of our 17 B-factors.

Table 9 shows that the median B-count, considering all possible deviations, is 9 with a standard deviation of 2.5. Nearly every consumer exhibits at least one deviation (100% with rounding), and the 10th percentile has a B-count of 6. Counting only standard-direction deviations among the bi-directional B-factors produces only slightly lower B-counts than counting any deviation (Column 2 vs. Column 1). Counting only economically large deviations as behavioral (as defined in the Data Appendix) reduces the median B-count to 5, but still leaves 90% of consumers with 2 or more deviations (Column 3).

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

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

---

[33] For example, loss aversion takes on four values: unbiased, and then three ordered responses (whether the individual respondent rejects the compound but not the single lottery, rejects the single but not the compound lottery, or rejects both) coded as 1/2/3. Any respondent accepting both lotteries receives a 0 (meets the classical benchmark), and 37% of individuals share that response. Anyone with the smallest deviation from the benchmark therefore is in the 37th percentile, and 13% of responses fall into that category. Summing, anyone in the next category is in the $50^{th}(=37^{th}+13^{th})$, and so on.

Second, they bear directly on key assumptions and inputs to reduced-form behavioral sufficient statistic models (Section 5-G). Third, the substantial heterogeneity in summary behavioral tendencies across individuals suggests that it is worth exploring relationships between B-stats and behavior.

The bottom panel of the table suggests that missing-ness does not overly complicate the construction of B-stats, with the mean (median) respondent supplying data required to measure all but 2 (1) the 17 B-factors. Nevertheless we control for missing-ness in our regressions.

B. Within-group differences in B-stats dwarf cross-group differences

Figures 1-4 show our "standard" B-counts broken out for groups at the opposite ends of the income, education, gender, and cognitive skills distributions. The B-count varies substantially *within* all of the sub-groups we examine. Being behavioral is not confined to those with low cognitive skills, or to males, or to low-income or low-education individuals. In most cases the median level of B-count is similar across splits, and the most striking pattern here is that any cross-group differences are dwarfed by the within-group variation. We see a similar pattern for B-tiles. These patterns further enhance the prospect that heterogeneity in B-stats will correlate with field behavior, conditional on our rich set of controls including the ones considered in Figures 1-4 (recall that Table 2 contains the complete list).

C. Links between financial condition and B-stats

The main litmus test for whether B-stats prove useful is whether they conditionally correlate with field behavior, and so we estimate B-stat versions of equation (1) by regressing our summary measure of financial condition on B-counts, B-tiles, the count of missing B-factors, and the same rich set of controls detailed in the B-factor models. Here we aggregate survey response time across all B-factor questions. We specify B-counts linearly, as more flexible parameterizations do not reject linearity (Appendix Table 8). As with the B-factor regressions we separate both B-counts and B-tiles into "standard" and "non-standard" components. Doing so is informative for a few reasons—not least, to the extent that behavioral theories predict stronger (negative) links to financial condition for "standard" biases, these specifications allow corroboration and inference about whether our sufficient statistics measure behavioral factors versus other omitted variables. We again suppress most of the results on our 100+ control variables to save space in the tables.

Starting with Table 10 columns (1) and (4), we find that B-counts and B-tiles are negatively conditionally correlated with financial condition in an economically meaningful way. The coefficients imply essentially identical magnitudes (recall that the B-tiles are normalized to the same scale as the B-counts). E.g., increasing the B-count or B-tile by four (i.e., adding 4 factors to an individual's "behavioralness") is associated with a reduction in financial condition of roughly 0.10, which is 25% on the mean of 0.43. Put another way, increasing a B-stat by four is equivalent to switching one of the nine indicators of good financial condition from "on" to "off."

Table 10 Column (2) shows a specification in the spirit of Altonji et al. (2005), dropping the cognitive skills and classical preference variables to see if excluding them affects the B-count result. It does not.[34]

Separating B-stats into standard and non-standard biases (Columns 3 and 5), we find the same pattern as with single B-factors: standard deviations from classical benchmarks are associated with larger reductions in financial condition than non-standard deviations. This holds whether we measure deviations using B-counts or B-tiles. One can reject that standard and non-standard coefficients are equal for both B-counts and B-tiles.

When we include both B-counts and B-tiles (Column 6), we find that B-counts are more strongly correlated with financial condition than B-tiles. This is useful to know since, practically speaking, B-counts may more cost-effective to quantify in many settings, based as they are on simple, extensive margin deviations from classical benchmarks.

Harking back to the single B-factor indicator models, recall that there simply summing all of the individual coefficients when "on" suggests a reduction in financial condition of about 0.40. Here, moving the B-count from zero to sixteen factors reduces financial condition by 0.38. This sheds light on whether B-factors have additive/reinforcing links or are simply different metrics of similar underlying tendencies associated with the same outcomes. If the latter were true, higher B-counts might not lead to lower financial condition, past a certain point anyway, because they would be measuring "different parts of the elephant." Similarly, if the B-factor coefficients in Table 6 were heavily biased toward zero due to measurement error, and the B-count simply

---

[34] The coefficient on the missing B-factors variable is affected, although this is driven by the exclusion of controls for missing cognitive skills and classical preferences variables rather than by the exclusion of the cognitive skills and classical preference measures themselves.

provided a correction, one would expect the B-count coefficient in Table 10 to trump the sum of all the B-factor indicators in Table 6 (Dean and Ortoleva 2016; Gillen, Snowberg, and Yariv 2015). That is not the case. Instead our findings suggest that B-factors have reinforcing and roughly additive relationships with financial condition (Section 4-B probes this issue further).

Column 7 asks how much cross-sectional variation in financial condition one can explain using *only* our B-counts and B-tiles, and finds an R-squared of 0.11. By way of comparison, the comparable marginal contribution to fitting financial condition is 0.29 for the 17 income categories, 0.22 for all of the other demographics, 0.12 for our four measures of cognitive skills/knowledge, 0.00 for our five measures of classical risk and time preferences and 0.00 for time spent on the survey. Nor do cognitive skills and our B-counts measure the same thing: if one includes both, the R-squared is 0.18 (result not shown in table).

D. More robustness checks

Recall that the primary empirical concerns are omitted variables, or measurement error, that would spuriously lead to significant links between our B-counts/-tiles and outcomes. Thus far we have addressed these concerns by explicitly controlling for as many confounds as possible, by showing that the B-stat/-factor correlations are robust to omitting key controls a la Altonji et al. (2005), by exploiting the directionality of some biases, and by examining correlations between B-factors and other controls. We undertake some additional robustness checks in Table 11.

Table 11 Column 1 provides additional evidence that B-count results are not driven by a conflation of behavioral characteristics with (math) ability. Here we segment our B-factors into two categories: those that reflect preferences or decision rules, and a set of "math biases" for which the classical benchmark is a clear correct answer. The math bias category includes the EG biases, the gambler's fallacies, and non-belief in the law of large numbers. We then include counts of both math and non-math biases as separate variables, and observe that both are strongly conditionally correlated with outcomes.

Nor is it that case that certain outcomes or B-factors drive the results. Table 11 Columns 2 and 3 subset outcomes into "hard" (quantitative measures: net worth, retirement assets, stock ownership, savings rate, events indicating severe financial problems) and "soft" (qualitative measures: self-assessed savings adequacy, financial satisfaction, and financial stress). The B-

count is strongly conditionally correlated with each subset, and the correlations are not statistically different from each other. Appendix Table 9 further shows that the B-count has a statistically significant negative correlation with each of the 9 component indicators of financial condition. In other specifications we have tested all 16 permutations of removing one B-factor from the B-count, and in all 16 cases still found the B-count to have a statistically and economically significant conditional correlation with financial condition.

In Table 11 columns 4-8 we stratify the sample by education and income, to see whether the B-count result holds across sub-samples. It does, with the exception of the middle-income group. Columns 9-11 stratify by B-factor survey response time, and show that the B-count result is stable across these sub-samples as well. This is yet another piece of evidence that survey effort does not confound the results.

4. Is There a Behavioral Common Factor?

For other inputs to decision making like cognitive and non-cognitive skills, researchers have been able to distill what first appears to be a large set of characteristics/traits into common factors that avoid collinearity and dimensionality problems. We ask whether one can do the same for our set of B-factors. Exploring that possibility also addresses measurement error concerns. If a single factor (or small number of factors) drives the manifestation and effects of different measured B-factors, then eliciting a full set of B-factors is analogous to the repeated elicitations one would employ to correct for measurement error (Gillen, Snowberg, and Yariv 2015), and our B-count and B-tile results above would best be interpreted as measurement error corrections, rather than capturing the cumulative influence of distinct behavioral phenomena. The results in Section 3-C point toward cumulative influence, and here we further explore the two interpretations.

A. *Exploratory Factor Analysis*

Although the findings thus far cast doubt on the prospect that the wide set of B-factors are manifestations of a single (or few) underlying common factors, we conduct a series of exploratory factor analyses (EFAs) to estimate factor structure formally. We use the B-tiles for this analysis, as they contain more variation and avoid some statistical issues associated with indicator variables. Another methodological issue is how to deal with missing values. Although they appear only rarely for any single B-factor, at the consumer level one or more B-factors is

missing for nearly two-thirds of the sample. Facing the choice between estimating common factors on one-third of the sample, or making an assumption that replaces missing values and allows estimation on the full sample, we favor the latter approach. We do so using three different imputations for missing values: that a missing value has no bias, the median level of bias, or the maximum level of bias.[35]

Appendix Table 10 summarizes our initial EFA. Following the approach in Heckman et al. (2013), we use three different methods: Horn's parallel analysis, inspection of scree plots (e.g., Figure 5), and the Onatski (2009) estimator. The estimated number of latent common factors ranges from one to three, with one being the modal count.

Table 12 estimates B-factor loadings in the common factor(s), assuming either one or three common factors, and assuming missing values have either zero/median/maximum bias. The models tend to identify greater loadings under the assumption that a missing value indicates maximum bias, so we focus on those results. In the third column we observe that three B-factors reach what Heckman et al. (2013) calls "dedicated measure" status—a high loading in the common factor, above 0.60. The three B-factors are level overconfidence, EG bias on the asset side, and limited memory. Others close to the threshold (>0.50) include one other overconfidence measure, NBLLN and Gambler's fallacy. Together these three to six B-factors do not seem to capture a clear and intuitive underlying construct. And the findings are even murkier in the three-factor models of the next nine columns. We conclude from this that the best representation of any common factor model underlying the data includes a single common factor, with missing values for bias set equal to the maximum.

B. Structural Equation Model results

A key litmus test for common factor parsimony in economic applications is whether any observed common factor is strongly linked to outcomes. We examine that second stage by estimating structural equation models (SEMs). These models simultaneously estimate the links between the common factor and our observed B-factors, *and* the empirical link between the behavioral common factor(s) and outcomes—essentially replacing the B-count and B-tile of

---

[35] The no-bias imputation strikes us as ill-justified, given that B-factor missingness correlates negatively with financial condition along the same lines as B-factors/B-stats (e.g., Table 10). Nevertheless we present the no-bias imputation in Appendix Table 10 for the sake of completeness.

Table 10 with the behavioral common factor(s). One can also think of SEM as a measurement error model in which the full set of B-factors is treated as a series of repeated measurements of the same latent variable(s).[36]

Table 13 shows results from different SEM specifications of a single common factor model.[37] We impute missing B-factor responses with the maximum level of bias, although the results are similar if we impute them with median bias. All controls are identical to those in Table 10.

The upshot is that the behavioral common factor is not significantly correlated with outcomes (Column 1), although the parameter estimate is noisy. A clue about what creates the noise, and the puzzling positive coefficient, lies below in that same column: the standard errors on the cognitive skill variables are also inflated, suggesting a collinearity problem. Column 2 confirms that by excluding the cognitive skills measures from the SEM estimated in Column 1, at which point the common factor becomes statistically significant with a much lower standard error. The behavioral common factor is seemingly strongly linked to cognitive skills, which is plausible given the correlations between B-factors and cognitive skills found in Table 5 and in prior work (Benjamin, Brown, and Shapiro 2013; Burks et al. 2009; Frederick 2005). However, the common factor appears to add no explanatory power to the model when our cognitive skill measures are included, suggesting that those measures subsume the common factor.

Recall that such a relationship is not true for our B-count and B-tile: dropping cognitive skills from our base specification in Table 10 Column 1 has no effect on the B-count coefficient in Table 10 Column 2. We conduct an even stronger test in Table 13's Columns 3 and 4, by constructing "B-count residuals" from a regression of the B-count on the common factor score, as extracted from the model in column 1.[38] The residuals therefore capture variation in B-counts that is independent of any estimated latent variable common to all B-factors. Column 3 shows that the point estimate on these residuals is even larger than that on the B-count overall (0.34 vs. 0.24). Column 4 further corroborates the distinction between the B-count and cognitive skills:

---

[36] N.B. that the "structural" in SEM is not the typical usage in economics: SEMs do not estimate structural economic parameters.

[37] Results are similar qualitatively when we include more common factors, in that none of the factors relates substantively to outcomes once we condition on cognitive skills.

[38] One can think of the factor score as the fitted value of the common factor, calculated observation by observation using the factor loadings.

now when we drop cognitive skills (from the model in Column 3), the coefficient on the B-count residual remains unchanged.

In sum, the SEM results strongly suggest that any behavioral common factor is *not* linked to outcomes, and rather appears related to cognitive skills. Since we control for cognitive skills in our models any behavioral common factor has no correlation with financial condition above and beyond cognitive skills. In contrast, the variation in our behavioral sufficient statistics that is independent of the common factor, and of cognitive skills, *is* strongly linked to financial condition.

5. Implications and Applications

A. *Survey design*

We start by briefly recapping implications for survey design. Our direct elicitations lower costs for generating new data on potentially behavioral inputs to decision making, across a range of settings and sample sizes. Our existing data and results provide evidence on the relationships among B-factors, and between B-factors and outcomes, that can be used to guide choices of which B-factors to elicit for a given study.

B. *Avoiding and/or controlling for omitted variable bias in single-factor studies*

As discussed in Section 2-F, our findings are good news for the standard approach of focusing on one or a few B-factors and ignoring others. First, our results suggest that, at least for conditional relationships between financial condition and single B-factors, the magnitude of bias from omitting other B-factors is likely to be modest. The same holds for omitting cognitive skills. Second, for researchers still concerned with omitted variable bias, our inter-factor correlations provide two solutions. One is to use the correlations we estimate to bound or correct for omitted variable bias (or the influence of a specific other factor). Another is to use our correlations as guidance on which other B-factors to measure, perhaps using our relatively low-cost elicitations. That approach can be a cost-effective way to address omitted variable bias if measuring a fuller set of B-factors is prohibitively expensive.

C. *B-factors have distinct, multi-dimensional and additive links to financial condition*

Putting together the predictive results on B-factors, B-stats, and common factors, we make two key inferences. First, our behavioral variables capture something distinct from other

covariates: they are not simply reflecting omitted demographics, standard factors (e.g., cognitive skills or classical preferences) or measurement error. Second, our B-factors capture things distinct from *each other:* the B-stats combine distinct B-factors that have separate and cumulative/additive relationships with outcomes, rather than using repeated elicitation of behavioral tendencies to correct for measurement error in a single underlying behavioral common factor. The latter result would be interesting, but it is not borne out by the SEM results, which suggest that the common component of B-factors does not in fact correlate with outcomes, conditional on standard controls for cognitive ability. Nor is it borne out by a comparison of the B-factor results (Table 6) to the B-count results (Table 10). Simply adding up the B-factor coefficients produces a coefficient roughly equal to the B-count coefficient. See also Appendix Table 8, which shows that the data do not reject a linear relationship between the B-stats and financial condition.

D. *Prospects for modeling BE as a low-dimensional set of (latent) traits*

Our methods and results suggest poor prospects for reducing the whole of BE to a manageable number of (latent) traits, as had been done for intelligence and personality. To the contrary, B-factors have distinct and perhaps additive relationships with financial condition. Of course, financial condition is but one of the many outcome and decision domains of interest. If predictive power is to be a key component of litmus tests for parsimony, then linking multiple B-factors to other outcome domains is a critical avenue for future work. The good news is that our methods lower the cost of generating large samples of representative data that include both outcomes and elicitations and hence can be used to validate, test, and refine behavioral models taking various approaches to dimensionality.

E. *Estimating structural parameters*

Our Data Appendix provides details on using our B-factor data to estimate parameters from widely-used, structural, one-factor behavioral models. Specifically, we use our data to estimate parameters for 6 out of our 17 B-factors. (Elicitations for the other 11 B-factors are too coarse to permit structural estimation.) On the whole, our estimates of central tendency and variance are comparable to those found in prior studies (Appendix Table 5). These results suggest that one can use our elicitations—and, more broadly perhaps, our approach of abbreviating elicitations—to obtain useful parameter estimates at relatively low cost.

F. Measuring behavioral prevalence and heterogeneity for practical applications

As noted at the outset, policy and business decisions often build on assumptions about the prevalence or distribution of behavioral tendencies and their links to behavior. We have shown that one can use our methods and data to both describe single B-factors and to describe behavioral tendencies at the consumer-level using our B-stats. The B-stats support the growing interest in applying lessons from behavioral research in the sense that they suggest that behavioral tendencies are nearly ubiquitous in the cross-section of consumers (Table 9), heterogeneous in their magnitude (Table 9), not subsumed by standard covariates such as demographics or cognitive skills (Tables 6 and 10), and strongly conditionally correlated with financial condition (Tables 6, 10, 11, and 13). B-stats may also ultimately provide a tool for targeting programs and/or products, along the lines of how intelligence or personality tests have been used to "type" people.

G. Reduced-form behavioral sufficient statistic modeling

Our work also helps refine and expand empirical implementation of the reduced-form behavioral sufficient statistic approach to modeling summarized in Chetty (2009; 2015).

These models rest on assumptions that are largely untested, and our methods and data open opportunities for examining them. Our estimates of correlations among B-factors support the assumption of positive within-consumer correlation among biases (e.g., Chetty 2015), though not terribly strongly. Our B-stat distributions support the assumption of consumer-level bias that is nonnegative, positive for some, and not mean-zero in the aggregate (Allcott and Taubinsky 2015, p. 2510). At the same time, they cast doubt on the assumption of homogeneity in person-level bias required to use Chetty, Looney, and Kroft's (2009) equivalent price metric to identify the average marginal bias distribution that is a key input to welfare analysis (Allcott and Taubinsky 2015; Mullainathan, Schwartzstein, and Congdon 2012).

Our B-stats (or B-factor data) could be used in estimating key model parameters like the prevalence of behavioral agents, the number of behavioral agents on a given margin, and the extent of their biases (Mullainathan, Schwartzstein, and Congdon 2012). And while applications of the behavioral sufficient statistic approach to-date have focused on particular decisions and product markets—which is desirable in many contexts—we show that B-stats powerfully predict

outcomes at a higher level. This is germane given the growing interest in domain-level applications among policymakers and practitioners.

Going forward we expect that our methods will further complement behavioral sufficient statistic modeling by expanding the set of tools for: testing and refining assumptions; identifying differences between experienced utility and decision utility in the large samples required to accommodate heterogeneity in behavioral biases; and for testing and refining predictions by using behavioral "typing" to, e.g., target/tag or estimate heterogeneous treatment effects.

6. Conclusion

We directly elicit a rich set of potentially behavioral factors in a large representative sample of U.S. consumers, and match those B-factors with rich data on decisions/outcomes and standard/classical inputs. The data produce a rich set of new findings on B-factor distributions, inter-relationships, distinctions from classical factors like cognitive skills, and links to outcomes.

Our pursuit of three types of parsimony in behavioral economics achieves success at two and instructive failure at the third. Our attempts at elicitation parsimony—directly measuring B-factors using low-cost, low-touch, and short adaptations of standard methods—succeed. Behavioral sufficient statistic parsimony—aggregating data across single B-factors to construct summary statistics measuring behavioral tendencies at the consumer level—also works, in that our B-stats are strongly conditionally correlated with financial condition. Our attempts at common factor parsimony fail: it appears that one cannot extract a lower-dimensional latent variable that is linked to both myriad behavioral tendencies and financial outcomes. B-factors are separate constructs, seemingly, with relatively distinct and independent relationships to financial outcomes.

We offer a number of other findings that should illuminate work going forward. Omitting other B-factors from models estimating links between single B-factors and financial condition does not tend to change inferences. The relationship between financial outcomes and B-factors is roughly additive. Nearly everyone in our sample is behavioral on a few dimensions, if not more, even if one takes a conservative view of what "behavioral" means. There is substantial heterogeneity across individuals in the extent of behavioral tendencies, and that heterogeneity is

not well-explained by measures of standard demographics, classical preferences, cognitive skills, or survey effort.

Taken together our findings validate the two leading approaches to dealing with the possibility of supra-dimensionality in behavioral economics. One approach is to consider only a small or singleton set of dimensions at a time. Another approach is to capture myriad dimensions using reduced-form behavioral sufficient statistics. Our methods provide new tools and data for refining both approaches.

Indeed this paper only begins to tap the potential of the new elicitation methods and dataset described herein. On the elicitation side, direct comparisons between ours 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. They should also be helpful in adding to the small but growing body of evidence on the reliability (intertemporal stability) of directly elicited behavioral factors.[39] This in turn is key to unpacking the direction and extent of any causality underlying conditional correlations between behavioral factors and outcomes. In terms of the data used in this paper, there are many possibilities for exploiting the panel, multi-topic architecture of the ALP to explore relationships between our behavioral variables, covariates, and outcomes in yet more domains. That work could include more detailed consideration of standard behavioral models, including structural ones, 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.

---

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

REFERENCES

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

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

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

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

Ashraf, Nava, Dean Karlan, and Wesley Yin. 2006. "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines." *Quarterly Journal of Economics* 121 (2): 673–97.

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

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

Barsky, Robert B, F. Thomas Juster, Miles S Kimball, and Matthew D Shapiro. 1997. "Preference Parameters and Behavioral Heterogeneity; An Experimental Approach in the Health and Retirement Study." *Quarterly Journal of Economics* 112 (2): 537–79.

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.

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

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

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

Camerer, Colin F. 2007. "Neuroeconomics: Using Neuroscience to Make Economic Predictions." *The Economic Journal* 117 (519): C26–42.

Campbell, John. 2016. "Restoring Rational Choice: The Challenge of Consumer Financial Regulation." *American Economic Review* 106 (5): 1–30.

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. 2009. "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods." *Annual Review of Economics* 1 (1): 451–88.

———. 2015. "Behavioral Economics and Public Policy: A Pragmatic Perspective." *American Economic Review* 105 (5): 1–33.

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

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

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

Dean, Mark, and Pietro Ortoleva. 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."

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

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

Frederick, 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.

Gabaix, Xavier. 2016. "Behavioral Macroeconomics Via Sparse Dynamic Programming."

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

Gine, Xavier, Jessica Goldberg, Daniel Silverman, and Dean Yang. 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. 2015. "The Role of Time Preferences and Exponential-Growth Bias in Retirement Savings."

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103 (6): 2052–86.

Heidhues, Paul, Botond Koszegi, and Philipp Strack. 2016. "Unrealistic Expectations and Misguided Learning." SSRN Scholarly Paper. Rochester, NY. http://papers.ssrn.com/abstract=2703524.

Kőszegi, Botond. 2014. "Behavioral Contract Theory." *Journal of Economic Literature* 52 (4): 1075–1118.

Levy, Matthew, and Joshua Tasoff. 2016. "Exponential-Growth Bias and Lifecycle Consumption." *Journal of the European Economic Association* 14 (3): 545–83.

Li, Ye, Martine Baldassi, Eric J. Johnson, and Elke U. Weber. 2013. "Complementary Cognitive Capabilities, Economic Decision Making, and Aging." *Psychology and Aging* 28 (3): 595–613.

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

Lusardi, Annamaria, and Olivia S. Mitchell. 2014. "The Economic Importance of Financial Literacy: Theory and Evidence." *Journal of Economic Literature* 52 (1): 5–44.

MacLeod, Colin M. 1991. "Half a Century of Research on the Stroop Effect: An Integrative Review." *Psychological Bulletin* 109 (2): 163.

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

McArdle, John J., Gwenith G. Fisher, and Kelly M. Kadlec. 2007. "Latent Variable Analyses of Age Trends of Cognition in the Health and Retirement Study, 1992-2004." *Psychology and Aging* 22 (3): 525–45.

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

Montiel Olea, J. L., and T. Strzalecki. 2014. "Axiomatization and Measurement of Quasi-Hyperbolic Discounting." *The Quarterly Journal of Economics* 129 (3): 1449–99.

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

Mullainathan, Sendhil, Joshua Schwartzstein, and William J. Congdon. 2012. "A Reduced-Form Approach to Behavioral Public Finance." *Annual Review of Economics* 4 (1): 511–40.

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

Onatski, Alexei. 2009. "Testing Hypotheses about the Number of Factors in Large Factor Models." *Econometrica* 77 (5): 1447–79.

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

Rabin, Matthew, and Georg Weizsäcker. 2009. "Narrow Bracketing and Dominated Choices." *American Economic Review* 99 (4): 1508–43.

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

Shah, Anuj K, Sendhil Mullainathan, and Eldar Shafir. 2012. "Some Consequences of Having Too Little." *Science* 338 (6107): 682–85.

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

Stanovich, Keith E. 2016. "The Comprehensive Assessment of Rational Thinking." *Educational Psychologist* 51 (1): 23–34.

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.
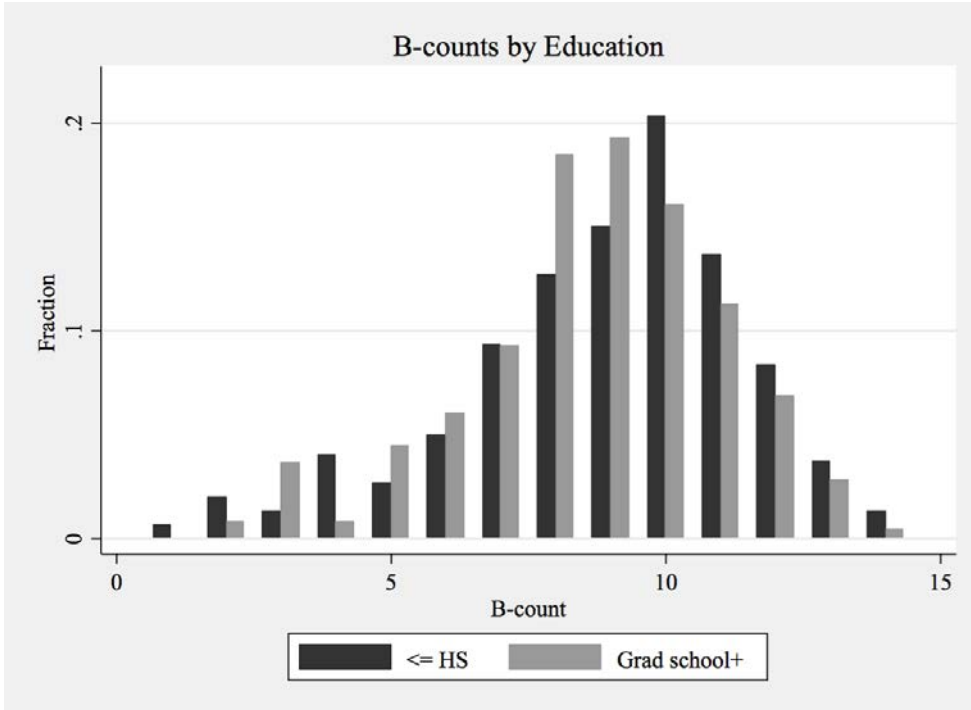
Figure 1
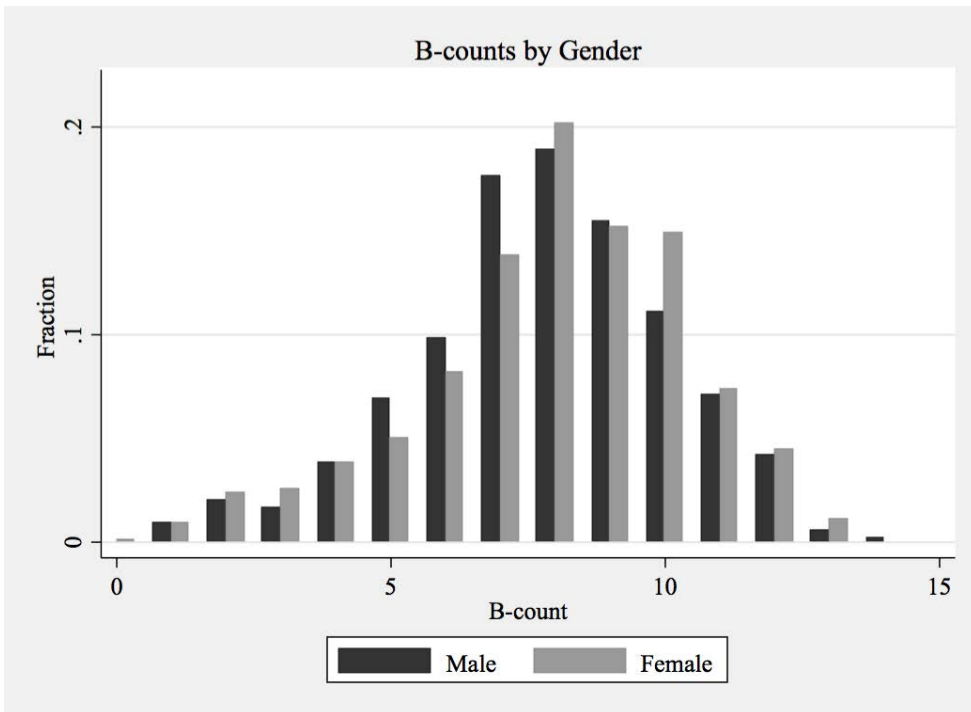
Figure 2
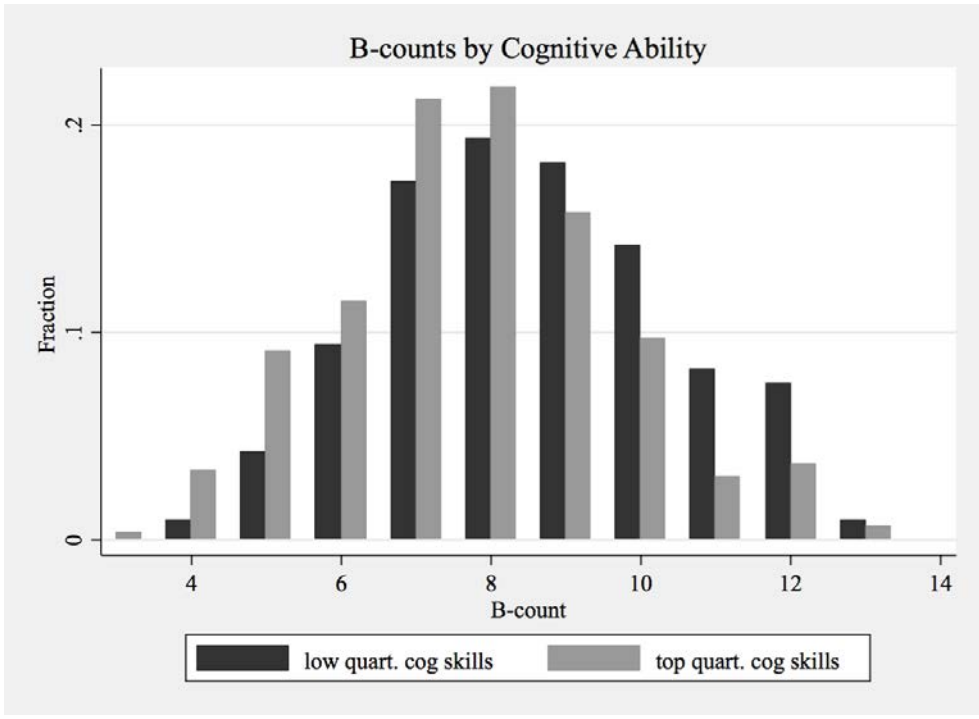
Figure 3. See section I-D for details on defining cognitive skills.
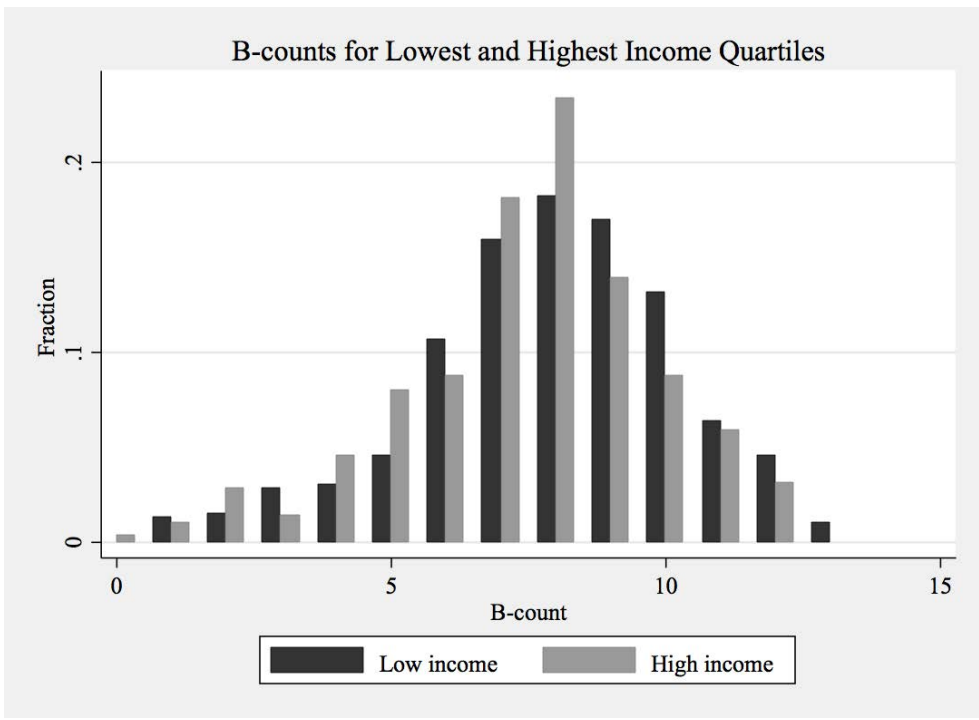
Figure 4
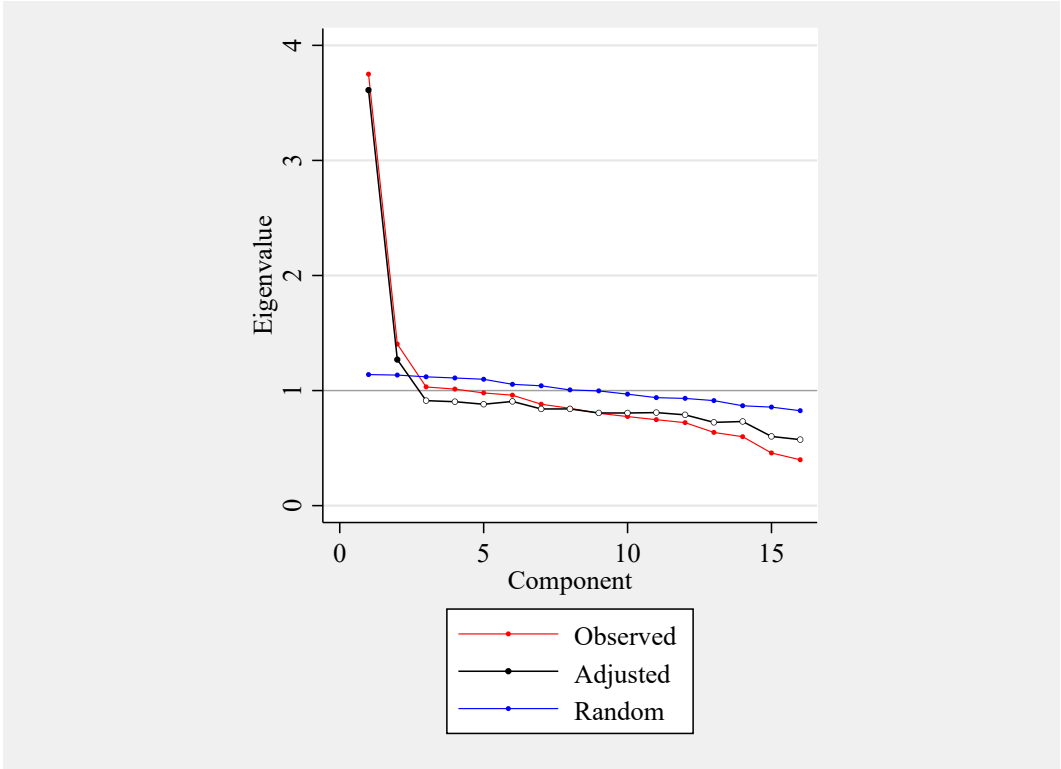
Figure 5. Scree plot illustrating eigenvalues from exploratory factor analysis. "Observed" and "adjusted" are from model; comparing to "random" identifies the number of common factors (those with significantly greater eigenvalues than random).

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

Factor name: *key antecedents*	Elicitation method description	Behavioral indicator(s), "standard" deviation direction in bold
Discounting money: *Andreoni & Sprenger (2012), Barcellos & Carvalho (2014)*	Convex Time Budget. 24 decisions allocating 100 tokens each between smaller-sooner and larger-later amounts; decisions pose varying start dates, delay lengths & savings yields.	**Present-biased: discounts more when sooner date is today** Future-biased: discounts more when sooner date is 5 weeks from tdy
Discounting snacks: *Read & van Leeuwen (1998), Barcellos & Carvalho (2014)*	Two decisions between two snacks: healthier/less-delicious vs. less healthy/more delicious. Decision pose variation only in date snack is delivered: now, or 5 weeks from now.	**Present-biased: choose less healthy tdy, healthy for 5 weeks from now** Future-biased: choose healthy for tdy, less healthy for 5 weeks from now
Choice inconsistency with GARP (with dominance avoidance): *Choi et al (2014)*	Decisions from 11 different linear budget constraints under risk. Subjects choose a point on the line, and then the computer randomly chooses whether to pay the point value of the x-axis or the y-axis.	**GARP only: percentage points of potential earnings wasted (CCEI)** GARP and dominance avoidance: pp of potential earnings wasted (combined-CCEI)
Preference for certainty: *Callen et al (2014)*	2 screens of 10 choices each between two lotteries, one a (p, 1-p) gamble over X and Y > X , (p; X, Y), the other a (q, 1-q) gamble over Y and 0, (q; Y, 0). Y=$450, X=$150, q ϵ[0.1, 1.0], p=0.5 on one screen and 1.0 on the other.	**Preference for cetainty: certainty premium (CP) >0** Cumulative prospect theory: certainty premium (CP)<0
Loss aversion/small-stakes risk aversion: *Fehr & Goette (2007)*	Two choices. Choice 1: between a 50-50 lottery (win $80 or lose $50), and $0. Choice 2: between playing the lottery in Choice 1 six times, and $0.	**Loss aversion: choosing the certain $0 payoff in one or more choices.**
Narrow bracketing: *Rabin & Weizsacker (2009)*	Two tasks of two decisions each. Each decision presents the subject with a choice between a certain payoff and a gamble. Each decision pair appears on the same screen, with an instruction to consider the two decisions jointly.	**Narrow-bracketing: making a choice that is dominated given implications of an earlier decision, on one or both tasks.**
Ambiguity aversion: *Dimmock et al. (forthcoming)*	Two questions re: a game where win $500 if pick green ball. 1. Choose between bag with 45 green-55 yellow and bag with unknown mix. 2. If chose 45-55 bag, how many green balls in 45-55 bag would induce switch to bag with unknown mix.	**Aversion: indicated if prefers 45 green balls to uncertain mix, increases as number of green balls declines**
(Over-)confidence in performance: *Larrick et al (2007), Moore & Healy (2008)*	"How many of the last 3 questions (the ones on the disease, the lottery and the savings account) do you think you got correct?"	**Overconfidence: self-assessment > actual score** Underconfidence: self-assessment < actual score
(Over-)confidence in relative performance: *Larrick et al (2007), Moore & Healy (2008)*	"… what you think about your intelligence as it would be measured by a standard test. How do you think your performance would rank, relative to all of the other ALP members who have taken the test?"	**Greater diff between self-assessed and actual rank indicates more overconfidence. But only ordinal, no cardinal, measures here, because test has < questions than percentiles**
Overconfidence in precision: *Larrick et al (2007), Moore & Healy (2008)*	Questions about about likelihoods of different numeracy quiz scores and future income increases.	**Overconfidence: responds 100% to one or both questions**
Non-belief in the law of large numbers: *Benjamin, Moore, and Rabin (2013)*	Question re: percent chances that, among 1,000 coin flips, the # of heads will fall in ranges [0, 480], [481, 519], and [520, 1000]. NBLLN = distance between response for [481, 519] and 78.	>78 (overestimate convergence to 50-50) **<78 (underestimate convergence to 50-50)**
Gambler's or "hot hand" fallacy: *Benjamin, Moore, and Rabin (2013)*	"Imagine that we had a computer "flip" a fair coin… 10 times. The first 9 are all heads. What are the chances, in percentage terms, that the 10th flip will be a head?"	>50%: "hot hand" fallacy **<50%: gambler's fallacy**
Exponential growth bias, debt-side: *Stango & Zinman (2009; 2011)*	Survey first elicits monthly payment respondent would expect to pay on a $10,000, 48 month car loan. Survey then asks for APR implied by that payment. EGB = difference between actual implied APR and the subject's perceived APR.	**Underestimates** or overestimates APR
Exponential growth bias, asset-side: *Banks et al (2007)*	Elicits future value of $200, earning 10% annual, after two years. EGB = difference between the correct future value ($242), and the subject's perceived future value.	**Underestimates** or overestimates future value
Limited attention: *Author-developed*	Four questions re: whether subject's finances would improve with more attention given the opportunity cost of attention, with questions varying the types of decisions: day-to-day, medium-run, long-run, or choosing financial	**One or more responses indicating regret about amount of attention paid**
Limited prospective memory: *Ericson (2011)*	"The ALP will offer you the opportunity to earn an extra $10.... This special survey has just a few simple questions but will only be open for 24 hours, starting 24 hours from now…. please tell us now whether you expect to do this special survey."	**Says will complete task but does not complete**

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

Table 2. Control variable definitions/specifications.

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

Table 3. Measuring financial condition: Wealth/savings/distress indicators, prevalence (means) and pairwise correlations

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

Unconditional pairwise correlations. * 0.10 ** 0.05 *** 0.01
Pairwise sample sizes range from 1391 to 1508.

Variable definitions: net worth is from two summary questions: "Please think about all of your household assets (including but not limited to investments, other accounts, any house/property you own, cars, etc.) and all of your household debts (including but not limited to mortgages, car loans, student loans, what you currently owe on credit cards, etc.) Are your household assets worth more than your household debts?" and "You stated that your household's [debts/assets] are worth more than your household's [assets/debts]. By how much?" Retirement assets is from questions on IRAs and workplace plans. Stockholding is from questions on stock mutual funds in IRAs, stock mutual funds in 401ks/other retirement accounts, and direct holdings. Spent < income is from a summary question on spending vs. saving over the past year, taken from the Survey of Consumer Finances. Financial satisfaction is based on a 100-point scale responding to "How satisfied are you with your household's overall economic situation?" Savings adequacy questions are placed one each in the two different modules to mitigate mechanical correlations, with response options framed to encourage people to recognize tradeoffs between saving and consumption. Indicators of severe financial distress are taken from the National Survey of American Families: late/missed payment rent, mortgage, heat, or electric; moved in with other people because could not afford housing/utilities; postponed medical care due to financial difficulty; adults in household cut back on food due to lack of money. Financial stress is based on an 100 point scale in response to: "To what extent, if any, are finances a source of stress in your life?"

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

Factor and bias	Share biased	Share "missing"	Missing detail		
			survey nonresponse	item nonresponse	responded, not usable
Discounting money: Present-biased	**0.26**	0.06	0.00	0.06	0.00
Discounting money: Future-biased	0.36				
Discounting snacks: Present-biased	**0.15**	0.07	0.06	0.02	0.00
Discounting snacks: Future-biased	0.07				
Violates GARP (based on CCEI)	**0.53**	0.16	0.06	0.10	0.00
Loses by violating GARP or dominance violations	**0.96**	0.16	0.06	0.10	0.00
Preference for certainty type (positive certainty premium)	**0.77**	0.31	0.00	0.03	0.28
Cumulative prospect theory type (negative certainty premium)	0.23				
Loss-averse: prefers certain zero payoff	**0.63**	0.00	0.00	0.00	0.00
Narrow-brackets	**0.59**	0.02	0.00	0.02	0.00
Ambiguity-averse	**0.73**	0.08	0.06	0.03	0.07
Overconfident in level performance	**0.38**	0.10	0.06	0.03	0.07
Underconfident in level performance	0.11				
Overconfident in precision	**0.44**	0.11	0.06	0.04	0.00
Overconfident in relative performance	n/a		0.06	0.04	0.00
·belief in the law of large numbers: under-estimates convergence	**0.87**	0.09	0.06	0.03	0.00
Non-belief in the law of large numbers: over-estimates convergence	0.13				
Gambler's fallacy	**0.26**	0.08	0.06	0.02	0.00
Hot hand fallacy	0.14				
Exponential growth bias, loan-side: underestimates APR	**0.70**	0.37	0.00	0.05	0.32
Exponential growth bias, loan-side: over-estimates APR	0.27				
Exponential growth bias, asset-side: underestimates future value	**0.47**	0.19	0.06	0.03	0.11
Exponential growth bias, asset-side: over-estimates future value	0.09				
Limited attention with regret	**0.49**	0.02	0.00	0.02	0.00
Limited prospective memory	**0.86**	0.10	0.06	0.02	0.02

Section 1-C provides some details on measuring individual behavioral factors; see the Data Appendix for additional details. "GARP" = General Axiom of Revealed Preference. "CCEI" = Critical Cost Efficiency Index. "Standard" bias classifications are those typically theorized/observed in prior work. Proportion exhibiting overconfidence in relative performance is n/a because we lack an absolute measure of it, as detailed in the Data Appendix Section H. "Share biased" is conditional on non-missing values. Missing shares are relative to the full sample size of 1,515. "Survey nonresponse" generates missing values when the consumer failed to take the second survey module. "Item nonresponse" can occur on either module. "Unusable" reaponses are those that are illogical or internally inconsistent (e.g., negative loan APRs on the EGB questions).

Table 5. B-factor indicators and other covariates.

Variable	(1) Present bias money	(2) Present bias snack	(3) GARP	(4) GARP FOSD	(5) Pref for certainty	(6) Loss averse	(7) Narrow bracket	(8) Ambig averse	(9) OC perf	(10) OC precision	(11) NBLLN	(12) Gambler's fallacy	(13) EGB loan	(14) EGB asset	(15) Limited attention	(16) Limited memory
				Dependent variable shown as column header and is the standard bias indicator for B-factors with potentially bi-directional biases.												
Mean(LHS)	0.26	0.15	0.53	0.96	0.77	0.63	0.59	0.73	0.38	0.44	0.87	0.26	0.70	0.47	0.49	0.86
female	-0.013	-0.005	0.057*	0.025**	0.018	-0.004	0.014	-0.049*	-0.016	-0.068**	0.074***	0.074***	-0.055	0.069**	0.019	0.009
	(0.026)	(0.022)	(0.031)	(0.012)	(0.030)	(0.027)	(0.029)	(0.027)	(0.027)	(0.028)	(0.020)	(0.026)	(0.034)	(0.028)	(0.029)	(0.021)
age 35-45	0.003	0.010	0.006	0.026	0.079*	0.003	-0.000	-0.054	0.045	0.192***	0.008	0.071**	0.121***	0.089**	-0.055	0.013
	(0.035)	(0.030)	(0.043)	(0.017)	(0.040)	(0.037)	(0.039)	(0.036)	(0.038)	(0.039)	(0.027)	(0.035)	(0.046)	(0.039)	(0.039)	(0.029)
age 46-54	0.056	0.033	-0.021	0.026	0.079*	-0.001	-0.020	-0.033	0.074*	0.227***	0.011	0.102***	0.082*	0.044	-0.104**	-0.000
	(0.038)	(0.031)	(0.045)	(0.018)	(0.043)	(0.039)	(0.041)	(0.038)	(0.040)	(0.041)	(0.028)	(0.037)	(0.049)	(0.041)	(0.042)	(0.031)
age>=55	0.046	0.038	0.035	0.030	0.095**	-0.005	-0.035	0.003	0.013	0.257***	-0.019	0.045	0.111**	0.026	-0.107**	-0.024
	(0.042)	(0.034)	(0.051)	(0.020)	(0.047)	(0.043)	(0.045)	(0.042)	(0.044)	(0.045)	(0.031)	(0.041)	(0.053)	(0.045)	(0.046)	(0.034)
ed: some college or associates	0.031	0.043	0.041	0.020	0.039	0.013	0.046	0.098***	0.076**	0.017	0.019	-0.050	-0.072	0.009	-0.014	0.016
	(0.036)	(0.029)	(0.043)	(0.017)	(0.044)	(0.036)	(0.039)	(0.036)	(0.038)	(0.039)	(0.027)	(0.035)	(0.047)	(0.040)	(0.039)	(0.029)
highest ed: bachelor's	0.030	0.004	-0.019	0.010	-0.017	0.042	0.029	0.115***	0.013	0.035	-0.021	-0.103**	-0.044	0.019	-0.098**	-0.009
	(0.041)	(0.034)	(0.050)	(0.020)	(0.047)	(0.042)	(0.045)	(0.042)	(0.044)	(0.045)	(0.031)	(0.041)	(0.053)	(0.045)	(0.045)	(0.034)
highest ed: graduate	0.024	0.021	0.028	0.000	-0.058	-0.030	0.063	0.136***	-0.029	-0.052	-0.053	-0.062	-0.051	-0.026	-0.150***	-0.012
	(0.047)	(0.039)	(0.056)	(0.022)	(0.052)	(0.048)	(0.051)	(0.048)	(0.051)	(0.051)	(0.035)	(0.047)	(0.060)	(0.051)	(0.052)	(0.039)
income decile 2	-0.121**	-0.021	-0.051	-0.013	-0.090	0.036	-0.072	0.030	0.022	0.002	-0.036	-0.003	-0.027	0.053	0.152**	0.068
	(0.056)	(0.045)	(0.066)	(0.026)	(0.071)	(0.056)	(0.060)	(0.056)	(0.059)	(0.061)	(0.042)	(0.054)	(0.076)	(0.063)	(0.061)	(0.045)
income decile 3	-0.065	0.034	-0.028	-0.009	-0.087	0.018	-0.042	-0.006	0.001	0.009	0.007	0.034	0.020	0.003	0.159***	0.052
	(0.057)	(0.045)	(0.066)	(0.026)	(0.072)	(0.057)	(0.060)	(0.056)	(0.059)	(0.061)	(0.041)	(0.054)	(0.075)	(0.062)	(0.061)	(0.045)
income decile 4	-0.054	-0.017	-0.065	-0.033	-0.100	0.019	-0.131**	0.043	-0.035	0.009	-0.023	0.065	-0.054	0.028	0.232***	0.063
	(0.062)	(0.050)	(0.071)	(0.028)	(0.077)	(0.063)	(0.066)	(0.061)	(0.065)	(0.067)	(0.045)	(0.060)	(0.084)	(0.068)	(0.068)	(0.049)
income decile 5	-0.107*	-0.018	0.054	-0.017	-0.057	0.112*	-0.147**	-0.028	-0.017	-0.026	0.037	0.037	0.082	-0.095	0.269***	0.081
	(0.062)	(0.050)	(0.073)	(0.029)	(0.076)	(0.063)	(0.067)	(0.062)	(0.065)	(0.068)	(0.046)	(0.061)	(0.081)	(0.067)	(0.067)	(0.050)
income decile 6	-0.046	-0.020	-0.046	-0.025	-0.074	0.136**	-0.217***	-0.000	0.009	-0.031	0.037	-0.014	-0.003	-0.017	0.219***	0.091*
	(0.061)	(0.050)	(0.072)	(0.029)	(0.075)	(0.063)	(0.066)	(0.062)	(0.065)	(0.067)	(0.046)	(0.060)	(0.083)	(0.068)	(0.067)	(0.050)
income decile 7	-0.123*	-0.040	-0.111	-0.038	-0.048	0.121*	-0.141**	0.069	-0.058	-0.162**	-0.040	0.003	-0.004	-0.075	0.186***	0.076
	(0.063)	(0.052)	(0.075)	(0.030)	(0.077)	(0.064)	(0.068)	(0.063)	(0.067)	(0.070)	(0.047)	(0.062)	(0.083)	(0.069)	(0.069)	(0.051)
income decile 8	-0.092	-0.043	-0.131*	-0.045	-0.076	0.072	-0.205***	0.013	-0.105	-0.109	-0.039	-0.037	-0.007	-0.142**	0.206***	0.154***
	(0.061)	(0.049)	(0.072)	(0.029)	(0.073)	(0.061)	(0.065)	(0.061)	(0.064)	(0.066)	(0.045)	(0.059)	(0.081)	(0.067)	(0.066)	(0.049)
income decile 9	-0.050	0.074	-0.045	0.041	-0.094	0.035	-0.105	0.086	-0.122	-0.014	0.012	-0.015	-0.009	-0.194**	0.182*	0.259***
	(0.088)	(0.072)	(0.105)	(0.042)	(0.100)	(0.091)	(0.095)	(0.089)	(0.093)	(0.095)	(0.066)	(0.086)	(0.115)	(0.095)	(0.097)	(0.073)
income decile 10	-0.039	0.077	-0.033	0.041	-0.098	0.013	-0.084	0.068	-0.124	-0.014	0.020	-0.012	-0.012	-0.194**	0.158	0.259***
	(0.088)	(0.072)	(0.105)	(0.042)	(0.100)	(0.091)	(0.095)	(0.090)	(0.093)	(0.096)	(0.066)	(0.087)	(0.115)	(0.095)	(0.097)	(0.073)

Table 5, cont.'d

Variable	(1) Discounting money	(2) Discounting snack	(3) GARP	(4) GARP FOSD	(5) Pref for certainty	(6) Loss averse	(7) Narrow bracket	(8) Ambig averse	(9) OC perf	(10) OC precision	(11) NBLLN	(12) Gambler's fallacy	(13) EGB loan	(14) EGB asset	(15) Limited attention	(16) Limited memory
fluid intell # correct	0.005	-0.003	-0.011*	-0.005**	0.006	0.010*	-0.020***	-0.000	0.004	0.021***	-0.019***	-0.019***	0.009	-0.043***	-0.001	-0.005
	(0.006)	(0.005)	(0.007)	(0.003)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.004)	(0.005)	(0.008)	(0.006)	(0.006)	(0.004)
numeracy # correct	-0.030	0.000	-0.036	0.013	-0.016	-0.052*	0.025	0.048*	-0.285***	0.078***	0.005	0.010	-0.035	0.060**	0.010	-0.022
	(0.027)	(0.021)	(0.030)	(0.012)	(0.035)	(0.027)	(0.029)	(0.026)	(0.027)	(0.028)	(0.019)	(0.025)	(0.036)	(0.028)	(0.029)	(0.021)
financial literacy # correct	-0.007	0.008	-0.025	-0.015*	-0.028	-0.010	-0.017	0.022	0.006	-0.017	-0.029**	-0.021	0.003	-0.096***	-0.026	-0.034**
	(0.018)	(0.014)	(0.021)	(0.008)	(0.021)	(0.018)	(0.019)	(0.018)	(0.018)	(0.019)	(0.013)	(0.017)	(0.024)	(0.020)	(0.020)	(0.014)
exec attention # correct	-0.000	-0.000	-0.001	-0.000	-0.001	0.002***	0.001	-0.001	-0.000	-0.001	0.001	-0.000	-0.000	-0.001**	-0.000	-0.000
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Risk aversion (financial)	0.000	-0.007*	-0.011**	0.001	-0.000	0.024***	-0.002	0.020***	-0.007	-0.006	0.001	-0.002	0.006	-0.001	0.008	-0.002
	(0.005)	(0.004)	(0.005)	(0.002)	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.003)	(0.004)	(0.006)	(0.005)	(0.005)	(0.004)
Risk aversion (income)	-0.010	0.005	0.011	0.005	0.004	0.007	-0.020*	0.020**	0.007	0.015	0.002	0.017*	-0.012	-0.002	-0.017	0.000
	(0.010)	(0.008)	(0.012)	(0.005)	(0.012)	(0.010)	(0.011)	(0.010)	(0.010)	(0.011)	(0.007)	(0.010)	(0.013)	(0.011)	(0.011)	(0.008)
patience (stdized)	-0.026**	0.006	-0.001	-0.004	0.011	-0.029**	0.014	-0.030**	-0.009	0.001	-0.009	0.002	0.001	-0.009	-0.018	0.005
	(0.012)	(0.010)	(0.015)	(0.006)	(0.014)	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)	(0.009)	(0.012)	(0.016)	(0.013)	(0.014)	(0.010)
Survey time spent decile 2	0.103*	0.051	0.015	0.003	-0.175**	-0.046	-0.032	0.005	-0.033	-0.112*	0.019	-0.007	-0.149*	-0.106*	-0.042	0.045
	(0.054)	(0.042)	(0.070)	(0.028)	(0.078)	(0.053)	(0.059)	(0.053)	(0.049)	(0.060)	(0.040)	(0.053)	(0.080)	(0.057)	(0.060)	(0.040)
Time spent decile 3	0.039	0.001	-0.004	0.030	-0.243***	-0.025	-0.044	0.032	0.070	-0.161***	0.081**	-0.005	-0.152*	-0.097*	-0.062	-0.039
	(0.055)	(0.043)	(0.070)	(0.028)	(0.078)	(0.054)	(0.059)	(0.054)	(0.055)	(0.061)	(0.041)	(0.048)	(0.082)	(0.057)	(0.061)	(0.044)
Time spent decile 4	0.016	0.087*	-0.068	0.021	-0.208***	0.067	0.029	0.027	0.027	-0.313***	0.066	0.047	-0.130*	-0.156**	0.005	-0.028
	(0.056)	(0.045)	(0.071)	(0.028)	(0.076)	(0.055)	(0.058)	(0.054)	(0.058)	(0.060)	(0.041)	(0.052)	(0.079)	(0.061)	(0.060)	(0.041)
Time spent decile 5	-0.006	-0.008	-0.079	0.017	-0.151*	0.069	-0.031	0.054	0.048	-0.277***	0.025	-0.012	-0.153*	-0.187***	0.026	-0.041
	(0.055)	(0.042)	(0.071)	(0.028)	(0.077)	(0.055)	(0.058)	(0.053)	(0.060)	(0.061)	(0.041)	(0.052)	(0.082)	(0.057)	(0.060)	(0.045)
Time spent decile 6	0.101*	-0.013	-0.061	0.018	-0.211***	0.015	-0.063	0.053	0.019	-0.380***	0.043	0.056	-0.292***	-0.230***	-0.089	0.030
	(0.055)	(0.043)	(0.071)	(0.028)	(0.077)	(0.055)	(0.058)	(0.052)	(0.063)	(0.061)	(0.043)	(0.050)	(0.079)	(0.059)	(0.061)	(0.046)
Time spent decile 7	0.115**	0.075*	-0.040	0.021	-0.218***	0.037	-0.024	0.010	0.063	-0.425***	0.068	0.072	-0.288***	-0.327***	-0.029	-0.046
	(0.055)	(0.045)	(0.073)	(0.029)	(0.079)	(0.057)	(0.060)	(0.056)	(0.057)	(0.062)	(0.041)	(0.051)	(0.080)	(0.061)	(0.060)	(0.042)
Time spent decile 8	0.143**	0.003	-0.173**	-0.016	-0.201**	0.065	-0.013	-0.019	0.065	-0.491***	0.076*	0.066	-0.281***	-0.268***	0.062	-0.057
	(0.056)	(0.044)	(0.071)	(0.028)	(0.078)	(0.055)	(0.059)	(0.054)	(0.059)	(0.062)	(0.042)	(0.052)	(0.081)	(0.059)	(0.061)	(0.043)
Time spent decile 9	0.094*	0.056	-0.131*	-0.014	-0.233***	0.146***	-0.042	0.083	-0.009	-0.456***	0.145***	0.048	-0.237***	-0.396***	0.018	-0.092**
	(0.056)	(0.044)	(0.071)	(0.028)	(0.078)	(0.055)	(0.059)	(0.053)	(0.058)	(0.062)	(0.042)	(0.052)	(0.080)	(0.058)	(0.061)	(0.043)
Time spent decile 10	0.163***	0.117***	-0.100	-0.004	-0.187**	0.070	-0.083	-0.014	0.053	-0.500***	0.071*	0.118**	-0.242***	-0.355***	-0.005	-0.082*
	(0.056)	(0.044)	(0.071)	(0.028)	(0.079)	(0.055)	(0.059)	(0.054)	(0.060)	(0.063)	(0.042)	(0.050)	(0.079)	(0.060)	(0.061)	(0.043)
Full set of controls from Table 2?	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	1416	1399	1265	1265	1044	1505	1480	1392	1361	1340	1370	1387	952	1217	1477	1353
R-squared	0.08	0.08	0.10	0.11	0.11	0.13	0.08	0.08	0.20	0.20	0.15	0.14	0.11	0.30	0.09	0.10
Adjusted r-squared	0.01	0.02	0.03	0.03	0.02	0.07	0.02	0.01	0.14	0.14	0.09	0.07	0.01	0.24	0.03	0.03
Stand-alone adjusted r-squared:																
risk/patience and cog. skills	0.01	0.00	0.03	0.01	0.01	0.05	0.02	0.01	0.13	0.01	0.07	0.06	0.00	0.17	0.01	0.01
demographics shown in table	0.00	0.00	0.01	0.02	0.01	0.02	0.01	0.01	0.05	0.04	0.06	0.06	0.01	0.09	0.02	0.01
demographics not shown	0.00	0.01	0.00	0.02	0.01	0.03	0.00	0.00	0.03	0.01	0.03	0.02	0.00	0.07	0.01	0.02
time spent	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.08	0.00	0.01	0.02	0.04	0.00	0.01

* 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. Relative to the full set of 17 behavioral factors in Table 1, here we exclude overconfidence in relative performance, since we do not have an absolute indicator of that factor, as detailed in the Data Appendix Section H. All regressions here include the full set of controls described in Table 2: "demographics not shown" are four race/ethnicity categories, state of residence, immigrant indicator, 3 marital status categories, 4 household size categories, and 5 work status categories.. Sample size varies because missing B-factor observations are discarded. Patience is the average savings rate across the 24 CTB decisions; we also include a dummy for missing this variable. See Section 1-C and the Data Appendix for details on B-factor variable construction.
`

Table 6. B-factor indicators and financial condition.

Variable	(1) Discounting money	(2) Discounting snack	(3) GARP	(4) GARP FOSD	(5) Pref for certainty	(6) Loss averse	(7) Narrow bracket	(8) Ambig averse	(9) OC perf	(10) OC precision	(11) NBLLN	(12) Gambler's fallacy	(13) EGB loan	(14) EGB asset	(15) Limited attention	(16) Limited memory
	RHS B-factor shown as column header. LHS variable in all models is summary financial condition (mean = 0.43)															
Standard bias indicator	-0.038**	-0.020	-0.013	-0.021	-0.006	-0.010	0.002	-0.025*	-0.020	-0.018	-0.023	-0.016	0.020	-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)	(0.020)	(0.044)	(0.044)	(0.017)	(0.012)	(0.018)
Non-standard bias indicator	-0.005	-0.018							0.037*			-0.028	0.023	-0.038		
	(0.016)	(0.025)							(0.021)			(0.020)	(0.046)	(0.026)		
bias missing	-0.009	0.033	-0.012	-0.025	0.016	0.000	0.033	0.262	-0.255	0.070*	0.168**	0.058	0.028	-0.023	-0.067	0.016
	(0.028)	(0.051)	(0.023)	(0.040)	(0.021)	(.)	(0.048)	(0.167)	(0.275)	(0.036)	(0.074)	(0.146)	(0.044)	(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	-0.020	-0.020	-0.019	-0.017	-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)
ed: some college or associates	-0.040**	-0.041**	-0.039**	-0.039**	-0.041**	-0.040**	-0.041**	-0.048**	-0.043**	-0.042**	-0.042**	-0.042**	-0.044**	-0.043**	-0.041**	-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)	(0.017)	(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	0.013	0.014	0.010	0.012	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)	(0.020)	(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	0.028	0.028	0.027	0.027	0.014	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)	(0.023)	(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	-0.017	-0.012	-0.016	-0.014	-0.021	-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)	(0.017)	(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	0.002	0.004	-0.002	-0.001	-0.008	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)	(0.018)	(0.019)
age>=55	0.053**	0.057***	0.053**	0.053**	0.057***	0.050**	0.049**	0.052**	0.056***	0.054**	0.047**	0.052**	0.047**	0.045**	0.041**	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)	(0.020)	(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	-0.026	-0.031	-0.025	-0.022	-0.021	-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)	(0.026)	(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	0.026	0.019	0.025	0.030	0.036	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)	(0.027)	(0.028)
income decile 4	0.074***	0.076***	0.078***	0.078***	0.078***	0.074***	0.079***	0.073**	0.084***	0.081***	0.080***	0.074***	0.082***	0.084***	0.089***	0.083***
	(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)	(0.027)	(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***	0.094***	0.084***	0.093***	0.097***	0.112***	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)	(0.030)	(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***	0.145***	0.139***	0.148***	0.147***	0.168***	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)	(0.030)	(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***	0.180***	0.170***	0.181***	0.186***	0.196***	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)	(0.030)	(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***	0.267***	0.262***	0.271***	0.272***	0.281***	0.272***
	(0.032)	(0.032)	(0.031)	(0.031)	(0.031)	(0.032)	(0.032)	(0.033)	(0.031)	(0.031)	(0.032)	(0.032)	(0.031)	(0.031)	(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***	0.303***	0.295***	0.307***	0.300***	0.318***	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)	(0.029)	(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***	0.369***	0.388***	0.380***	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)	(0.045)	(0.044)	(0.043)	(0.044)

Table 6, cont.'d

Variable	(1) Discounting money	(2) Discounting snack	(3) GARP	(4) GARP FOSD	(5) Pref for certainty	(6) Loss averse	(7) Narrow bracket	(8) Ambig averse	(9) OC perf	(10) OC precision	(11) NBLLN	(12) Gambler's fallacy	(13) EGB loan	(14) EGB asset	(15) Limited attention	(16) 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	0.000	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)	(0.003)	(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	0.001	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)
financial literacy # correct	0.028***	0.029***	0.029***	0.029***	0.031***	0.029***	0.027***	0.030***	0.029***	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)
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)
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***	-0.005**	-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)
Risk aversion (income)	0.009*	0.010*	0.009*	0.009*	0.009*	0.009*	0.009*	0.006	0.009*	0.009*	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)
patience (stdized)	0.006	0.007	0.008	0.008	0.007	0.008	0.008	0.008	0.008	0.007	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)
Time spent decile 2	-0.001	-0.043	0.008	0.007	-0.032	0.017	0.003	-0.032	0.010	-0.023	0.034	-0.027	-0.030	-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)	(0.028)	(0.028)	(0.027)	(0.027)	(0.026)	(0.026)
Time spent decile 3	-0.005	-0.040	0.017	0.018	-0.005	0.018	-0.011	-0.062**	0.040	0.019	0.042	-0.007	-0.044	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)	(0.026)	(0.028)	(0.027)	(0.027)	(0.028)
Time spent decile 4	0.018	-0.011	-0.004	-0.003	-0.050*	0.059**	0.010	-0.009	-0.035	0.021	0.023	0.012	-0.043	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)	(0.026)	(0.026)
Time spent decile 5	-0.001	-0.017	0.027	0.028	0.018	0.038	-0.002	-0.002	0.009	-0.005	0.042	0.031	-0.051*	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)	(0.029)	(0.028)	(0.028)	(0.027)	(0.027)	(0.029)
Time spent decile 6	0.017	-0.050*	-0.027	-0.026	-0.049*	0.033	-0.002	-0.039	0.004	0.023	0.046	0.021	-0.027	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)	(0.030)	(0.026)	(0.029)	(0.028)	(0.027)	(0.029)
Time spent decile 7	0.005	-0.033	-0.002	-0.001	-0.023	0.009	0.014	-0.022	-0.030	-0.016	0.002	-0.033	-0.029	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)	(0.029)	(0.027)	(0.029)	(0.029)	(0.026)	(0.027)
Time spent decile 8	0.012	-0.024	0.008	0.009	-0.033	0.014	0.017	-0.026	0.030	-0.007	0.025	-0.024	-0.000	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)	(0.029)	(0.028)	(0.029)	(0.028)	(0.027)	(0.028)
Time spent decile 9	-0.023	-0.018	0.017	0.018	0.004	0.002	0.002	-0.011	-0.029	0.033	0.025	-0.017	-0.030	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)	(0.030)	(0.028)	(0.028)	(0.028)	(0.027)	(0.028)
Time spent decile 10	0.004	-0.014	-0.030	-0.029	-0.039	0.017	0.015	-0.036	0.054*	-0.014	0.052*	-0.006	-0.006	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)	(0.029)	(0.027)	(0.029)	(0.028)	(0.027)	(0.028)
Full set of controls from Table 2?	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R-squared	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.40	0.39	0.39	0.39	0.39	0.40	0.43	0.39
Number of observations	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 the individual, with one regression per column. LHS variable is the summary measure of financial condition: the proportion of indicators with a "1" from Table 3. Relative to the full set of 17 behavioral factors in Table 1, here we exclude overconfidence in relative performance, since we do not have an absolute indicator of that factor, as detailed in the Data Appendix Section H. Models are OLS and also include the full set of controls in Table 2; not shown here to save space are: four race/ethnicity categories, state of residence, immigrant indicator, 3 marital status categories, 4 household size categories, 5 work status categories, fixed state effects, and dummies for missing values associated with each variable and for not taking our 2nd module. Patience is the average savings rate across the 24 CTB decisions; we also include a dummy for missing this variable. Omitted category for Preference for Certainty and Nonbelief in the Law of Large Numbers (NBLLN) is the nonstandard bias, due to lack of unbiased responses; see Section 1-C and the Data Appendix for details on variable construction.

Table 7. Correlations between B-factors.

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

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

Table 8. B-factor indicators and financial condition:
Do B-factor coefficients change when other B-factors are included?

Coefficient on standard bias for B-factor, with one regression per cell:	No other B-factors included (reproduced from Table 6)	All B-factors included	P-value on diff.
Discounting money	-0.038**	-0.039**	0.96
Discounting snack	-0.020	-0.008	0.13
GARP	-0.013	-0.010	0.49
GARP FOSD	-0.021	-0.018	0.78
Pref. for certainty	-0.006	-0.003	0.78
Loss averse	-0.010	-0.008	0.61
Narrow bracket	0.002	-0.000	0.82
Ambig averse	-0.025*	-0.025*	0.74
OC perf	-0.020	0.016	0.00***
OC precision	-0.018	-0.022	0.54
NBLLN	-0.023	-0.014	0.13
Gambler's fallacy	-0.016	-0.016	0.66
EGB loan	0.020	0.027	0.76
EGB asset	-0.063***	-0.058***	0.61
Limited attention	-0.109***	-0.105***	0.34
Limited memory	-0.042**	-0.040**	0.85

* 0.10 ** 0.05 *** 0.01. Unit of observation is the individual, with one regression per cell and N=1,505 in each regression. LHS variable is the financial condition index described in Table 3. Regressions in the first column are identical to those in Table 6. Regressions in the second column add a vector of indicators for all other B-factors (as well as indicators for non-standard bias and bias missing). All models include the full set of controls from Table 2. P-values calculated by stacking the models in a seemingly-unrelated regression (SUR), estimating coefficient covariances and testing for equality across models.

Table 9. B-counts and B-tiles: Summary statistics.

	Indicators included		
		Standard	"Large"
	All	Only	Devs. Only
"B-count" = count of behavioral indicators (N=1511)			
proportion with any bias	1.00	1.00	0.98
mean	9.05	7.90	4.55
SD	2.46	2.40	1.94
10th percentile	6	5	2
25th percentile	8	7	3
50th percentile	9	8	5
75th percentile	11	10	6
90th percentile	12	11	7
"B-tile" = degree of bias for non-missing factors (N=1511)			
mean	6.04	5.03	
SD	1.85	1.77	
10th percentile	3.53	2.69	
25th percentile	4.89	3.87	
50th percentile	6.20	5.06	
75th percentile	7.28	6.25	
90th percentile	8.31	7.33	
Count of missing factors for measuring behavioral indicators (N=1511)			
median	1	1	1
mean	1.75	1.75	1.75
SD	2.60	2.60	2.60

The B-tile is the sum of all of the B-factor percentiles, rescaled for comparability to the B-count (please see Section 3-A for details). Table 1 and the Data Appendix describe and detail behavioral factors and indicator definitions, including large-deviation thresholds.

Table 10. B-counts/-tiles and financial condition.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	LHS variable in all models is summary financial condition (mean = 0.43)						
B-count, all biases	-0.024***	-0.026***					-0.032***
	(0.004)	(0.004)					(0.005)
B-count, standard biases			-0.026***			-0.017**	-0.018**
			(0.004)			(0.007)	(0.008)
B-count, non-standard biases			-0.009				0.055
			(0.007)				(0.068)
B-tile, all biases				-0.027***			
				(0.005)			
B-tile, standard biases					-0.030***	-0.009	-0.041***
					(0.005)	(0.009)	(0.009)
B-tile, non-standard biases					-0.008		-0.098
					(0.008)		(0.076)
b-factor missing count	-0.009	-0.021***	-0.007	-0.004	-0.003	-0.004	-0.036***
	(0.007)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)	(0.004)
female	-0.021	-0.024*	-0.022*	-0.021	-0.022*	-0.023*	
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	
ed: some college or associates	-0.035**	-0.031*	-0.034*	-0.037**	-0.034*	-0.033*	
	(0.018)	(0.017)	(0.018)	(0.018)	(0.018)	(0.018)	
highest ed: bachelor's	0.012	0.018	0.014	0.011	0.013	0.014	
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	
highest ed: graduate	0.030	0.034	0.030	0.031	0.030	0.030	
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	
age 35-45	-0.007	-0.000	-0.006	-0.008	-0.006	-0.005	
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	
age 46-54	0.007	0.017	0.009	0.007	0.011	0.011	
	(0.019)	(0.018)	(0.019)	(0.019)	(0.019)	(0.019)	
age>=55	0.052**	0.063***	0.055***	0.051**	0.055***	0.057***	
	(0.021)	(0.020)	(0.021)	(0.021)	(0.021)	(0.021)	
income decile 2	-0.022	-0.028	-0.023	-0.020	-0.021	-0.023	
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	
income decile 3	0.031	0.028	0.028	0.033	0.030	0.027	
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	
income decile 4	0.089***	0.083***	0.088***	0.091***	0.091***	0.088***	
	(0.028)	(0.027)	(0.028)	(0.028)	(0.028)	(0.028)	
income decile 5	0.099***	0.100***	0.098***	0.101***	0.101***	0.099***	
	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	
income decile 6	0.152***	0.146***	0.151***	0.152***	0.152***	0.151***	
	(0.030)	(0.030)	(0.030)	(0.031)	(0.031)	(0.030)	
income decile 7	0.180***	0.179***	0.181***	0.177***	0.179***	0.180***	
	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	
income decile 8	0.271***	0.274***	0.271***	0.272***	0.272***	0.271***	
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	
income decile 9	0.303***	0.305***	0.302***	0.303***	0.300***	0.301***	
	(0.030)	(0.029)	(0.030)	(0.030)	(0.030)	(0.030)	
income decile 10	0.391***	0.393***	0.394***	0.390***	0.394***	0.395***	
	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
fluid intell # correct	-0.000		-0.000	-0.003	-0.003	-0.001	
	(0.003)		(0.003)	(0.003)	(0.003)	(0.003)	
numeracy # correct	-0.005		-0.005	-0.006	-0.007	-0.005	
	(0.013)		(0.013)	(0.013)	(0.013)	(0.013)	
financial literacy # correct	0.022**		0.021**	0.023**	0.022**	0.022**	
	(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)	
Risk aversion (financial)	-0.005**		-0.005**	-0.006***	-0.006***	-0.005**	
	(0.002)		(0.002)	(0.002)	(0.002)	(0.002)	
Risk aversion (income)	0.010**		0.010**	0.010**	0.011**	0.010**	
	(0.005)		(0.005)	(0.005)	(0.005)	(0.005)	
patience (stdized)	-0.000		0.001	0.000	0.002	0.002	
	(0.006)		(0.006)	(0.006)	(0.006)	(0.006)	
Time spent decile 2	-0.047	-0.051	-0.045	-0.044	-0.039	-0.042	
	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)	
Time spent decile 3	-0.069	-0.073	-0.066	-0.067	-0.062	-0.064	
	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)	
Time spent decile 4	-0.091*	-0.089*	-0.088*	-0.090*	-0.084	-0.085	
	(0.053)	(0.052)	(0.053)	(0.053)	(0.053)	(0.053)	
Time spent decile 5	-0.085	-0.083	-0.079	-0.087	-0.080	-0.077	
	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	
Time spent decile 6	-0.054	-0.056	-0.050	-0.051	-0.044	-0.046	
	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	
Time spent decile 7	-0.055	-0.055	-0.050	-0.055	-0.049	-0.048	
	(0.054)	(0.054)	(0.054)	(0.055)	(0.055)	(0.054)	
Time spent decile 8	-0.080	-0.077	-0.077	-0.079	-0.075	-0.076	
	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)	
Time spent decile 9	-0.091	-0.089	-0.088	-0.094*	-0.089	-0.088	
	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)	
Time spent decile 10	-0.093*	-0.093*	-0.091*	-0.092*	-0.087	-0.089*	
	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	
Other controls listed in Table 2?	yes	yes	yes	yes	yes	yes	no
R-squared	0.42	0.42	0.42	0.42	0.42	0.42	0.11
Observations	1505	1505	1505	1505	1505	1505	1505

* 0.10 ** 0.05 *** 0.01. Each column presents results from an OLS regression where the unit of observation is the individual. LHS variable is the summary measure of financial condition: the proportion of indicators with a "1" from Table 3. B-counts exclude overconfidence in relative performance, since we do not have an absolute indicator of that factor, as detailed in the Data Appendix Section H. B-tiles measure relative performance on all 17 factors. Patience is the average savings rate across the 24 CTB decisions; we also include a dummy for missing this variable. See Section 3-A for details on B-count and B-tile construction, and Table 2 for details on control variables.

.

Table 11. B-counts and financial condition: Additional robustness checks.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
LHS	Financial condition	"Hard" outcomes	"Soft" outcomes	Financial condition	Financial condition	Financial condition	Financial condition	Financial condition	Financial condition	Financial condition	Financial condition
Mean(LHS)	0.43	0.47	0.38	0.35	0.54	0.25	0.42	0.63	0.41	0.44	0.40
Sub-sample				Low education	High education	Low income	Middle income	High income	Time spent deciles 1-2	Time spent deciles 3-8	Time spent deciles 9-10
B-count, all biases		-0.022***	-0.027***	-0.028***	-0.018***	-0.031***	-0.009	-0.034***	-0.024***	-0.028**	-0.023**
		(0.004)	(0.005)	(0.005)	(0.006)	(0.007)	(0.006)	(0.008)	(0.005)	(0.013)	(0.009)
B-count, math biases	-0.025***										
	(0.004)										
B-count, non-math biases	-0.023***										
	(0.008)										
b-factor missing count	-0.009	-0.016**	-0.001	-0.005	-0.023*	0.012	-0.014	-0.035**	-0.007	-0.061***	-0.009
	(0.007)	(0.008)	(0.009)	(0.008)	(0.014)	(0.011)	(0.012)	(0.017)	(0.009)	(0.022)	(0.017)
female	-0.019	-0.014	-0.025	-0.009	-0.032	-0.039*	-0.016	0.000	-0.010	-0.053	0.005
	(0.013)	(0.015)	(0.017)	(0.017)	(0.022)	(0.023)	(0.023)	(0.027)	(0.016)	(0.044)	(0.033)
highest ed: some college or associates	-0.032*	-0.004	-0.068***			-0.030	-0.037	-0.010	-0.038*	-0.059	0.013
	(0.018)	(0.020)	(0.023)			(0.025)	(0.031)	(0.060)	(0.022)	(0.055)	(0.045)
highest ed: bachelor's	0.014	0.064***	-0.050*			0.007	-0.007	0.061	0.018	-0.098	0.031
	(0.020)	(0.023)	(0.027)			(0.036)	(0.035)	(0.059)	(0.025)	(0.068)	(0.052)
highest ed: graduate	0.031	0.066**	-0.016			0.014	0.019	0.068	0.029	-0.011	0.026
	(0.023)	(0.026)	(0.030)			(0.050)	(0.041)	(0.059)	(0.029)	(0.075)	(0.059)
Risk aversion (financial)	-0.005**	-0.003	-0.007**	-0.000	-0.014***	-0.002	-0.004	-0.015***	-0.003	0.004	-0.013**
	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)	(0.003)	(0.008)	(0.005)
Risk aversion (income)	0.009*	0.004	0.015**	0.005	0.018**	0.008	0.009	0.001	0.011*	0.018	0.003
	(0.005)	(0.005)	(0.006)	(0.006)	(0.009)	(0.007)	(0.009)	(0.013)	(0.006)	(0.016)	(0.011)
patience (stdized)	0.001	0.003	-0.000	-0.005	-0.003	0.003	-0.002	-0.009	0.005	-0.006	-0.020
	(0.006)	(0.007)	(0.008)	(0.008)	(0.010)	(0.011)	(0.011)	(0.013)	(0.008)	(0.021)	(0.016)
fluid intell # correct	-0.000	0.000	-0.000	-0.002	0.002	0.001	-0.005	0.004	-0.001	-0.003	-0.002
	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)	(0.004)	(0.010)	(0.006)
numeracy # correct	-0.005	-0.007	-0.004	-0.004	0.025	-0.024	0.030	0.040	-0.010	-0.053	-0.010
	(0.013)	(0.015)	(0.017)	(0.015)	(0.027)	(0.018)	(0.023)	(0.043)	(0.017)	(0.045)	(0.029)
financial literacy # correct	0.023**	0.034***	0.010	0.026**	0.009	0.016	0.058***	-0.000	0.032***	0.030	-0.001
	(0.009)	(0.010)	(0.012)	(0.010)	(0.019)	(0.013)	(0.015)	(0.025)	(0.011)	(0.031)	(0.021)
exec attention # correct	0.000	0.000	0.000	0.000	0.000	0.001**	0.000	-0.001	0.000	0.000	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)
other controls listed in Table 2?	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R-squared	0.42	0.51	0.50	0.37	0.44	0.32	0.27	0.34	0.67	0.44	0.58
Observations	1505	1505	1505	874	631	477	594	434	981	196	328

* 0.10 ** 0.05 *** 0.01. Each column presents results from an OLS regression where the unit of observation is the individual. LHS variable in cols. 1 and 4-11 is the summary measure of financial condition: the proportion of indicators with a "1" from Table 3. LHS in Column 2 calculates the proportion using "hard" financial outcomes: net worth, savings, stock market participation, lack of financial hardship and retirement savings. LHS in Column 3 uses "soft" outcomes: self-assesed retirement savings adequacy, non-retirement savings adequacy, financial satisfaction and financial stress. "High education" is college degree or higher. "Low/Middle/High" income categories are deciles 1-3, 4-7 and 8-10. "Time spent" is total response time across all B-factor elicitations. RHS variables include the full set of controls in Table 2, including those not shown here to save space. Patience is the average savings rate across the 24 CTB decisions; we also include a dummy for missing this variable. "Math biases" count indicators of B-factors where the classical benchmark is a clear correct answers: the EG biases, the gambler's fallacies, and the NBLLN biases. See Section 3-A for details on B-count construction.

Table 12. B-factor loadings for extracted common factor(s) with greatest eigenvalue(s): Results from 6 different model specifications.

	One-factor models			Three-factor models								
model #:	(1)	(2)	(3)	(4)			(5)			(6)		
Missing values set to:	*unbiased*	*nedian bias*	*full bias*	*unbiased*			*median bias*			*full bias*		
				Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
present bias $	0.03	0.09	0.09	-0.04	0.07	-0.02	0.06	-0.01	0.07	-0.05	0.10	0.14
present bias food	0.15	0.19	0.44	0.13	0.03	0.01	-0.06	0.37	0.03	0.54	-0.07	-0.03
choice inconsist	0.30	0.24	0.41	0.31	-0.01	-0.07	0.06	0.10	0.22	0.41	-0.03	0.17
pref for certainty	-0.05	0.12	0.19	-0.05	0.07	0.22	0.14	-0.15	0.13	-0.09	0.19	0.35
loss aversion	-0.13	-0.15	-0.07	-0.09	0.04	0.29	0.04	-0.02	-0.29	-0.05	0.07	-0.29
narrow bracketing	0.12	0.15	0.09	0.03	0.01	-0.29	-0.04	0.01	0.31	0.05	-0.05	0.32
ambig aver	0.12	-0.04	0.33	0.28	-0.10	0.19	-0.02	0.10	-0.12	0.36	0.02	-0.11
overconf performance	0.51	0.58	**0.67**	0.02	0.57	0.02	**0.62**	0.04	-0.06	0.08	**0.67**	-0.08
oveconf relative	0.54	0.40	0.59	0.46	0.09	-0.13	0.28	-0.02	0.24	0.36	0.20	0.21
overconf precis	0.16	0.12	0.46	0.11	0.13	0.19	0.13	0.11	-0.11	0.39	0.11	-0.05
NBLLN	0.49	0.34	0.55	0.45	0.08	-0.02	0.22	0.02	0.19	0.41	0.14	0.10
Gambler's Fallacy	0.31	0.33	0.55	0.24	0.05	-0.14	0.09	0.21	0.20	0.55	0.00	0.09
EGB debt	-0.03	0.04	0.12	0.01	-0.03	0.03	-0.03	0.02	0.10	0.05	0.01	0.21
EGB future val	0.48	0.58	**0.62**	0.00	0.54	-0.01	0.58	-0.02	0.06	-0.02	**0.67**	0.08
limited attention	0.05	0.10	0.11	-0.05	0.12	0.02	0.04	0.07	0.03	0.06	0.06	0.01
limited memory	0.39	0.31	**0.70**	0.48	-0.03	0.10	0.06	0.42	-0.01	**0.72**	0.04	-0.07

Unit of observation is the individual and N=1515 for each of the six models. We use direct oblique quartimin rotation, following Heckman et al (2013 AER), and Stata's default estimator for factor analysis (principal factor approach). Bolded b-factors satisfy the criteria used in Heckman et al for identifying "dedicated measures" of extracted common factors for a given specification: loading of >=0.6 on one factor, and not loading >0.4 on two or more factors in a multi-factor model.

Table 13. The common B-factor, idiosyncratic variation in B-factors and financial condition

	(1)	(2)	(3)	(4)
Behavioral common factor	0.161	-0.022**		
	(0.115)	(0.010)		
B-count residuals			-0.029***	-0.028***
			(0.004)	(0.003)
female	-0.025	-0.024*	-0.015	-0.020
	(0.018)	(0.013)	(0.013)	(0.013)
age 35-45	-0.061	-0.002	-0.003	0.003
	(0.038)	(0.017)	(0.018)	(0.017)
age 46-54	-0.057	0.021	0.012	0.022
	(0.045)	(0.018)	(0.019)	(0.018)
age>=55	0.011	0.069***	0.055***	0.066***
	(0.038)	(0.020)	(0.021)	(0.020)
highest ed: some college or associates	-0.084**	-0.036**	-0.026	-0.028
	(0.035)	(0.017)	(0.018)	(0.017)
highest ed: bachelor's	0.010	0.010	0.015	0.012
	(0.028)	(0.020)	(0.020)	(0.020)
highest ed: graduate	0.031	0.024	0.028	0.026
	(0.032)	(0.023)	(0.023)	(0.022)
Risk aversion (financial)	0.003		-0.006***	
	(0.006)		(0.002)	
Risk aversion (income)	0.007		0.010**	
	(0.007)		(0.005)	
patience (stdized)	0.011		0.000	
	(0.009)		(0.006)	
fluid intell # correct	0.049		-0.004	
	(0.030)		(0.003)	
numeracy # correct	0.079		-0.013	
	(0.052)		(0.013)	
financial literacy # correct	0.069**		0.019**	
	(0.028)		(0.009)	
exec attention # correct	0.001		0.000	
	(0.000)		(0.000)	
Other controls listed in Table 2?	yes	yes	yes	yes
Observations	1505	1505	1505	1505

* 0.10 ** 0.05 *** 0.01. See Section 1-D or Appendix Table 1 for details on cognitive skills measures.
Each column presents results from a single model; unit of observation is the individual. Models in Columns 1 and 3 contain the full set of controls listed in Table 2; Columns 2 and 4 drop the cognitive skills and classical preference variables. First two columns show results of Structural Equation Modeling (SEM) using the full set of B-factors (measured as -tiles, with missing imputed as fully biased) as endogenous observed variables, stemming from one latent (unobserved) "behavioral common factor" and financial condition as being a function of that latent variable and control variables. The third and fourth columns estimate OLS models similar to those in Table 10, but replace the B-count with "B-count residuals" - the component of the B-count that is independent of the common factor estimated in column 1. We obtain those residuals by estimating the common factor scores from column 1, regressing B-count on the common factor scores, and taking the residuals. Column 4 is the same as column 3 but omits the cognitive skills and classical preference measures.

Appendix Table 1. Pairwise correlations between measures of cognitive skills.

	Fluid intelligence	Numeracy	Financial literacy	Executive attention
Fluid intelligence	1			
mean 10.6, s.d. 2.8,				
min. 0, max. 15	1403			
Numeracy	**0.44**	1		
mean 1.7, s.d. 0.6,	*0.00*			
min. 0, max. 2	1371	1372		
Financial literacy	**0.45**	**0.41**	1	
mean 2.2, s.d. 0.9,	*0.00*	*0.00*		
min. 0, max. 3	1399	1368	1406	
Executive attention	**0.36**	**0.19**	**0.22**	1
mean 65, s.d. 24,	*0.00*	*0.00*	*0.00*	
min. 0, max. 154	1352	1326	1355	1444

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

Appendix Table 2. Pairwise correlations between measures of risk aversion.

	Lifetime income gamble	Financial risk-taking scale	Switch point two-lottery list	Switch point lottery vs. certain
Lifetime income gamble	1			
mean 4.3, s.d. 1.3				
min. 1, max. 6	1503			
Financial risk-taking scale	**0.19**	1		
mean -43, s.d. 23	*0.00*			
min. -100, max. 0	1390	1403		
Switch point two-lottery list	**0.07**	**0.09**	1	
mean 7.6, s.d. 1.5	*0.02*	*0.00*		
min. 2, max. 10	1147	1068	1153	
Switch point lottery vs. certain list	**0.26**	**0.16**	**0.42**	1
mean 6.6, s.d. 1.8	*0.00*	*0.00*	*0.00*	
min. 2, max. 10	1215	1133	1066	1222

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

Appendix Table 3. Does survey formatting bias toward worse financial condition?

Variable	# of questions used	# per q.	orientation	placement of choice(s) indicating worse condition	ordering details
				response options	
net worth>0	1	3	vertical	middle	Assets compared to debts? [Yes/no/about the same]
retirement assets>0	2	2	vertical	n/a*	"Enter total amount: $[fill].00"
owns stocks	3	2	vertical	n/a*	"About what percent of your household's [IRA/KEOGH; 401(k)/other retirement accounts] are invested in stocks or mutual funds (not including money market mutual funds)?"
				n/a*	Aside from anything you have already told us about, do you or another member of your household have any shares of stock or stock mutual funds? If you sold all those and paid off anything you owed on them, about how much would your household have?
spent < income last 12 months	1	3	vertical	top	Spent [more than/same as/less than] income
financial satisfaction > median	1	slider	horizontal	left side of scale	0 to 100 point scale, lower numbers indicate lower satisfaction
retirement saving adequate	1	5	vertical	top	Ordered 1/5 from "not nearly enough" to "much more than enough"
non-ret saving adequate	1	5	vertical	bottom	Ordered 1/5 from "wish my household saved a lot less" to "wish my household saved a lot more"
no severe distress last 12 mos	4	2	vertical	top	Yes/no for each question, with yes on top.
financial stress < median	1	slider	horizontal	right side of scale	0 to 100 point scale, higher numbers indicate higher stress

Please see Table 3 for additional details on variable definitions.

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

Appendix Table 4. Survey response time, income and financial condition components.

| Survey time decile | Income decile | Financial condition component outcomes: Share with indicator | | | | | | | | |
		net worth>0	retirement assets>0	owns stocks	no severe distress last 12 months	spent < income last 12 months	financial satisfaction > median	retirement saving adequate	non-ret saving adequate	fin stress < median
1	4.99	0.33	0.36	0.35	0.60	0.30	0.40	0.27	0.25	0.46
2	5.50	0.37	0.50	0.46	0.60	0.36	0.50	0.26	0.31	0.52
3	5.68	0.47	0.53	0.52	0.58	0.35	0.45	0.24	0.26	0.55
4	5.93	0.52	0.60	0.54	0.60	0.37	0.40	0.28	0.24	0.48
5	5.54	0.47	0.61	0.55	0.56	0.43	0.47	0.26	0.20	0.52
6	5.75	0.49	0.59	0.57	0.59	0.42	0.51	0.25	0.25	0.51
7	5.41	0.50	0.54	0.50	0.47	0.29	0.49	0.21	0.29	0.49
8	4.82	0.46	0.58	0.48	0.51	0.36	0.44	0.30	0.24	0.56
9	4.68	0.41	0.48	0.46	0.58	0.33	0.44	0.24	0.25	0.48
10	4.80	0.39	0.52	0.49	0.50	0.36	0.50	0.35	0.22	0.50
Total	5.31	0.44	0.53	0.49	0.56	0.36	0.46	0.27	0.25	0.51

Notes: Survey time decile is for total survey completion time in minutes. Financial condition components are described in greater detail in Table 3. Results in Table 6 confirm that conditional on income and other covariates, there is no substantive empirical link between survey response time and elicited financial condition, and results in Table 11 confirm that our main empirical results hold even when we stratify by survey time decile. Vertical line separates "hard" from "soft" measures of financial condition.

Appendix Table 5. Behavioral Factor Descriptive Statistics and Comparisons to Prior Work

Factor and potential biases	Prevalence							Parameter mean(sd)		Description of Comps		
	Our sample			Key Comps, Any deviation				Our sample	Key Comp	Source	Method	Sample
	"Lrg" dev threshold	"Large" deviation	Any deviation	AS (2012a)	BC (2014)	*Goda et al*	Bradford et al	Beta				
Money discounting	20pp, where max	0.03	0.26	0.14	0.29*	0.55	0.5	1	1.03	Andreoni and Sprenger	CTB	students
Present-bias	possible is 100pp	0.05	0.36	0.12	0.37*	?	0.26	(varies)**	(0.21)	Barcellos and Carvalho	mod CTB	ALP
Future-bias										Goda et al	MPL	ALP, UAS
										Bradford et al	MPL	Qualtrics
Discounting snacks				BC (2014)	RvL (1998)	Augenblick et al		N/A		Barcellos and Carvalho	hyp snacks	ALP
Present-bias	N/A	N/A	0.15	0.06*	not reported	0.33				Reed and van Leeuwen	real snacks	convenience
Future-bias			0.07	0.09*	not reported	0.21				Augenblick et al	effort tasks	students
				Choi et al (2014), calculated over first 11 rounds				CCEI		Choi et al	same	CentERpanel
Violates GARP		0.07	0.53	0.51				0.05 (0.08)	0.05 (0.10)			
	20pp, where max											
Violates GARP plus dominance avoidance	possible is 100pp	0.31	0.96	0.96				0.16 (0.18)	0.20 (0.20)			
Preference for certainty				Callen et al (2014)				Certainty Premium		Callen et al	same	Afghani voters
Certainty premium > 0	20pp, where max	0.42	0.77	1				0.16	0.37			
Certainty premium < 0	possible is 100pp	0.07	0.23	0				(0.24)	(0.15)			
				FG (2007)	Abeler et al (2011)	Hwang (2016)	vG et al (2011)			Fehr and Goette	same	bike messengers
Loss aversion	Rejects	0.37	0.64	0.67	0.87***	0.70***	0.86****	N/A		Abeler et al	similar	students
	both lotteries									Hwang	Abeler et al	ALP
										von Gaudecker	MPL, structural	CentERpanel
				Rabin-Weiz	Gottlieb-Mitch					Rabin and Weizacker	almost same	KnowledgeNetworks
Narrow Bracketing	Narrow brackets	0.13	0.59		0.3			N/A		Gottlieb and Mitchell	TK epidemic qs	HRS
	on both tasks		Task 2: 0.29	Task 2: 0.53								
			Task 4: 0.50	Task 4: 0.67								
Ambiguity Aversion				DKW	DK et al	Butler et al		N/A		Dimmock, Kouw, Wakk	multiple, adaptive Ellsberg	CentERpanel
	>10pp	0.50	0.73	0.68	0.52	0.52				Dimmock, Kouw, et al	multiple, adaptive Ellsberg	ALP
	>=10pp	0.66								Butler et al	qualitative Ellsberg	Unicredit Survey
Estimation re: performance level				Larrick et al	Moore and Healy			N/A		Larrick et al	similar	students
Overconfidence	deviation >=|2|	0.07	0.39	individual-level results not reported						Moore and Healy	similar	students
Underconfidence	(max 3)	0.01	0.11									
Forecast precision	reports "100%" on			Moore and Healy				N/A		Moore and Healy	subjective probability distributions	students
Overconfidence	both sets of q's re: likelihoods	0.10	0.34	individual-level results not reported								
Estimation re: peformance placement				Larrick et al	Moore and Healy			N/A		Larrick et al	similar	students
Overconfidence	our individual-level measure is ordinal but not cardinal			individual-level results not reported						Moore and Healy	similar	students
Underconfidence	(Data Appendix Section L)											
Nonbelief in the Law of Large Numbers (NBLLN)				Benjamin, Moore, and Rabin				unconditional: |39|		Benjamin et al	similar	food court
Underestimates converg	20pp (max 78)	0.09	0.87	individual-level results not reported				unconditional: -36	not reported			
Overestimates convergence	5pp (max 22)	0.08	0.13					-43pp (17)				
								13pp (9)				
Gambler's Fallacy				Benjamin, Moore, and Rabin		*Dohmen et al*		unconditional: |16|	not reported	Dohmen et al (2009)	almost same	German households
Cold hand	10pp (max 50)	0.26	0.26	individual-level results not reported		0.23		unconditional: -5 (25)		Benjamin et al	almost same	food court
Hot hand		0.13	0.14			0.10		-40 (8)	-29			
								39 (10)	27			
Exponential Growth Bias: Debt-Side				Stango and Zinman (2009, 2011)				N/A	N/A	Stango and Zinman	different time frame for loan	SCF
Underestimates APR	500 basis points	0.47	0.70	0.98								
Overestimates APR		0.10	0.27									
Exponential Growth Bias: Asset-Side				ages 50-60 only								
				our sample	HRS	*Goda et al*	Levy and Tasoff	N/A	N/A	Goda et al	more difficult FV qs	ALP, UAS
Underestimates future value	>20% of growth	0.17	0.47	0.48	0.74	0.69	0.85			Levy and Tasoff	more difficult FV qs	Knowledge Networks
Overestimates FV		0.06	0.09	0.09	0.14	0.09	0.11					
Limited attention with regret	All four indicators	0.06	0.49	N/A				N/A		N/A		
Limited prospective memory		N/A	0.86	Ericson (2011)				N/A		Ericson	incentivized rebate claim	students
				0.30								

Standard or uni-directional biases in bold, see Section 1-C and Data Appendix for details.
* Our calculation using the other paper's data.
** See Data Appendix Section A.
*** We define loss aversion in the Abeler et al task as rejecting one or more of the four small-stakes lotteries with positive expected value.
**** Correspondence with Arthur Van Soest
"pp" = percentage points. "GARP" = General Axiom of Revealed Preference. "CCEI" = Critical Cost Efficiency Index.

Appendix Table 6. Financial condition & behavioral factors: Functional form, and omitting key covariates (compare to Table 6)

functional form of bias:	any			percentile (linear)		
	(1)	(2)	(3)	(4)	(5)	(6)
Present-biased discounting, money	-0.038**	-0.042***	-0.040**	-0.057**	-0.052***	-0.047***
omitted category: unbiased	(0.016)	(0.016)	(0.016)	(0.023)	(0.018)	(0.018)
Present-biased discounting, food	-0.020	-0.017	-0.018	-0.022	-0.020	-0.021
omitted category: unbiased	(0.018)	(0.018)	(0.018)	(0.021)	(0.021)	(0.021)
Inconsistency with GARP	-0.013	-0.014	-0.012	-0.012	-0.016	-0.012
omitted category: consistent	(0.013)	(0.013)	(0.013)	(0.017)	(0.017)	(0.017)
ncons. w/GARP + dominance avoidance	-0.021	-0.030	-0.023	-0.016	-0.021	-0.013
omitted category: consistent	(0.034)	(0.034)	(0.034)	(0.024)	(0.023)	(0.023)
Preference for certainty	-0.006	-0.005	-0.007	-0.016	0.003	-0.018
ted category: preference for uncertainty	(0.017)	(0.017)	(0.017)	(0.038)	(0.036)	(0.037)
Loss aversion	-0.010	-0.015	-0.010	-0.019	-0.030	-0.019
omitted category: not loss averse	(0.013)	(0.013)	(0.013)	(0.023)	(0.022)	(0.023)
Narrow bracketing	0.002	-0.000	-0.001	0.004	0.004	0.005
ted category: does not narrowly bracket	(0.013)	(0.013)	(0.013)	(0.018)	(0.018)	(0.018)
Ambiguity aversion	-0.023	-0.025*	-0.024*	-0.022	-0.022	-0.021
omitted category: not ambiguity averse	(0.015)	(0.014)	(0.014)	(0.019)	(0.019)	(0.019)
Overconfidence in estimation	-0.020	-0.021	-0.016	-0.027	-0.030	-0.025
omitted category: unbiased	(0.015)	(0.014)	(0.015)	(0.022)	(0.020)	(0.022)
Overconfidence in precision	-0.018	-0.014	-0.017	-0.025	-0.023	-0.026
omitted category: not overconfident	(0.014)	(0.014)	(0.013)	(0.021)	(0.021)	(0.020)
Overconfidence in placement	n/a	n/a	n/a	-0.067*	-0.043*	-0.077**
				(0.036)	(0.023)	(0.036)
NBLLN: underestimates	-0.023	-0.030	-0.025	-0.047*	-0.054**	-0.046*
ed category: overestimates convergence	(0.020)	(0.020)	(0.020)	(0.027)	(0.026)	(0.026)
Gambler's Fallacy: cold hand	-0.016	-0.020	-0.019	-0.022	-0.025	-0.023
omitted category: unbiased	(0.016)	(0.015)	(0.016)	(0.019)	(0.018)	(0.019)
EGB debt-side: underestimates	0.020	0.024	0.023	0.001	-0.000	-0.004
;ory mostly overestimates, and unbiased	(0.044)	(0.044)	(0.044)	(0.038)	(0.038)	(0.038)
EGB asset-side: underestimates	-0.063***	-0.068***	-0.064***	-0.057**	-0.066***	-0.060***
omitted category: unbiased	(0.017)	(0.016)	(0.016)	(0.023)	(0.022)	(0.022)
Limited attention with regret	-0.109***	-0.112***	-0.112***	-0.165***	-0.168***	-0.165***
omitted category: no regrets	(0.012)	(0.012)	(0.012)	(0.017)	(0.017)	(0.016)
Limited prospective memory	-0.042**	-0.045**	-0.040**	-0.274**	-0.301**	-0.267**
omitted category:remembers task	(0.018)	(0.018)	(0.018)	(0.123)	(0.123)	(0.123)
Cog. skills and classical preferences?	yes	no	yes	yes	no	yes
Survey effort (time)?	yes	yes	no	yes	yes	no
All other controls from Table 2?	yes	yes	yes	yes	yes	yes

* 0.10 ** 0.05 *** 0.01. Unit of observation is the individual, and N=1,505 in all cases. Each cell in columns 1-6 presents a coefficient and standard error from an OLS regression of our financial condition index on the behavioral factor described in the row label (measured using the functional form described in the column heading), the non-standard bias indicator or percentile where applicable (but not for Preference for Certainty or NBLLN, despite their bi-directional nature, due to lack of unbiased responses), and our usual set of controls (subject to the permutations described in the bottom rows). First column is identical to specifications in Table 6 (reading down rows here, and across columns in Table 6). Additional details on these results, with comparisons to prior work, are in the Data Appendix. See Data Appendix for more details on B-factor definitions and measurement.

Appendix Table 7. Overcontrolling? Financial condition on B-factor indicators, classical preferences, and cognitive skills in a sparser specification.

(compare to Table 6)

Variable	(1) Discounting money	(2) Discounting snack	(3) GARP	(4) GARP FOSD	(5) Pref for certainty	(6) Loss averse	(7) Narrow bracket	(8) Ambig averse	(9) OC perf	(10) OC precision	(11) NBLLN	(12) Gambler's fallacy	(13) EGB loan	(14) EGB asset	(15) Limited attention	(16) Limited memory
	RHS B-factor shown as column header. LHS variable in all models is summary financial condition (mean = 0.43)															
Standard bias indicator	-0.062***	-0.024	-0.016	-0.033	-0.015	0.016	-0.018	-0.016	-0.053***	-0.023	-0.057***	-0.048***	0.049	-0.122***	-0.106***	-0.010
	(0.018)	(0.020)	(0.015)	(0.037)	(0.019)	(0.014)	(0.014)	(0.016)	(0.016)	(0.015)	(0.022)	(0.017)	(0.049)	(0.018)	(0.013)	(0.020)
Non-standard bias indicator	-0.023	-0.033							0.019			-0.053**	0.032	-0.096***	-0.021	0.080*
	(0.017)	(0.028)							(0.024)			(0.022)	(0.051)	(0.029)	(0.052)	(0.047)
bias missing	-0.058*	0.007	0.019	-0.004	0.001		0.016	0.062	-0.059	0.075*	0.070	-0.060	0.033	-0.109***		
	(0.031)	(0.056)	(0.026)	(0.044)	(0.023)		(0.053)	(0.088)	(0.120)	(0.040)	(0.060)	(0.112)	(0.050)	(0.027)		
fluid intell # correct	0.007**	0.007**	0.007**	0.007**	0.008**	0.007**	0.007**	0.007**	0.007**	0.008**	0.006**	0.005	0.007**	0.001	0.007**	0.007**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
numeracy # correct	0.029**	0.032**	0.031**	0.032**	0.034**	0.033**	0.033**	0.034**	0.016	0.039***	0.033**	0.033**	0.033**	0.030**	0.035**	0.035**
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.014)	(0.015)	(0.015)	(0.014)	(0.014)	(0.015)
financial literacy # correct	0.086***	0.087***	0.086***	0.086***	0.087***	0.086***	0.086***	0.087***	0.086***	0.088***	0.085***	0.083***	0.086***	0.069***	0.081***	0.087***
	(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)
Risk aversion (financial)	-0.013***	-0.013***	-0.013***	-0.013***	-0.013***	-0.013***	-0.013***	-0.013***	-0.013***	-0.013***	-0.012***	-0.013***	-0.013***	-0.012***	-0.012***	-0.013***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Risk aversion (income)	0.014***	0.016***	0.015***	0.015***	0.016***	0.016***	0.015***	0.016***	0.016***	0.016***	0.016***	0.016***	0.016***	0.017***	0.014***	0.017***
	(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)	0.004	0.008	0.008	0.008	0.008	0.009	0.009	0.008	0.008	0.008	0.008	0.008	0.009	0.007	0.007	0.009
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
other controls listed in Table 2?	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
N	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505	1505
R-squared	0.15	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.15	0.15	0.15	0.15	0.14	0.17	0.18	0.14

* 0.10 ** 0.05 *** 0.01. Unit of observation is the individual, with one regression per column and N=1,505 across all regressions. LHS variable is the summary measure of financial condition: the proportion of indicators with a "1" from Table 3. Relative to the full set of 17 behavioral factors in Table 1, here we exclude overconfidence in relative performance, since we do not have an absolute indicator of that factor, as detailed in the Data Appendix Section H. Models are OLS and include only the controls shown: B-factors, classical preferences and cognitive skills. Patience is the average savings rate across the 24 CTB decisions; we also include a dummy for missing this variable. Omitted category for Preference for Certainty and Nonbelief in the Law of Large Numbers (NBLLN) is the nonstandard bias, due to lack of unbiased responses; see Section 1-C and the Data Appendix for details on variable construction.

Appendix Table 8. B-counts and financial condition, alternative functional forms.
(compare to Table 10)

	(1)	(2)	(3)
	LHS variable: financial condition.		
B-count, all biases	-0.024***		
	(0.004)		
ln(B-count)		-0.017***	
		(0.003)	
B-count=8 [n=224]			-0.064***
(omitted category is B-count<=7)			(0.023)
B-count = 9 [n=256]			-0.051***
			(0.023)
B-count=10 [n=274]			-0.099***
			(0.023)
B-count = 11 [n=221]			-0.101***
			(0.024)
B-count >=12 [n=215]			-0.166***
			(0.026)
Other controls listed in Table 10?	yes	yes	yes
R-squared	0.42	0.42	0.42

* 0.10 ** 0.05 *** 0.01. Each column presents results from an OLS regression identical to those in Table 10, but for variations in how the B-count is specified.

.

Appendix Table 9. B-counts and individual components of financial condition.
(compare to Table 10)

LHS	(1) net worth>0	(2) retirement assets>0	(3) owns stocks	(4) spent < inc last 12	(5) fin satis > median	(6) adequate ret. Savings	(7) adequate non-ret. Savings	(8) no fin distress recent	(9) no current fin stress
B-count, all biases	-0.025***	-0.020***	-0.015**	-0.024***	-0.023***	-0.040***	-0.020***	-0.026***	-0.025***
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)
B-factor missing count	-0.005	-0.025**	-0.027**	-0.015	0.020	-0.018	0.018	-0.005	-0.023
	(0.014)	(0.012)	(0.013)	(0.014)	(0.014)	(0.013)	(0.014)	(0.014)	(0.015)
female	0.016	0.011	-0.024	-0.058**	-0.027	-0.045*	-0.040	-0.030	0.004
	(0.026)	(0.023)	(0.024)	(0.027)	(0.027)	(0.025)	(0.026)	(0.026)	(0.029)
ed: some college or associates	0.031	-0.014	0.009	-0.029	-0.089**	-0.077**	-0.070**	-0.021	-0.054
	(0.035)	(0.031)	(0.032)	(0.036)	(0.037)	(0.033)	(0.035)	(0.036)	(0.039)
highest ed: bachelor's	0.039	0.122***	0.110***	0.020	-0.054	-0.020	-0.021	0.037	-0.114**
	(0.041)	(0.036)	(0.037)	(0.042)	(0.042)	(0.039)	(0.041)	(0.041)	(0.045)
highest ed: graduate	0.024	0.119***	0.108**	-0.058	-0.073	0.069	0.045	0.119**	-0.084*
	(0.046)	(0.041)	(0.042)	(0.048)	(0.048)	(0.044)	(0.047)	(0.047)	(0.051)
age 35-45	0.082**	0.050	0.045	-0.012	-0.110***	-0.056*	-0.029	-0.008	-0.021
	(0.035)	(0.031)	(0.032)	(0.037)	(0.037)	(0.034)	(0.036)	(0.036)	(0.039)
age 46-54	0.199***	0.070**	0.059*	0.018	-0.159***	-0.011	-0.019	-0.013	-0.069*
	(0.037)	(0.034)	(0.035)	(0.039)	(0.039)	(0.036)	(0.037)	(0.038)	(0.042)
age>=55	0.282***	0.079**	0.096**	0.018	-0.119***	0.050	0.059	0.069*	-0.053
	(0.041)	(0.037)	(0.038)	(0.043)	(0.043)	(0.039)	(0.041)	(0.042)	(0.046)
income decile 2	0.065	-0.056	-0.084*	-0.081	-0.031	0.037	-0.079	0.023	0.004
	(0.052)	(0.047)	(0.048)	(0.054)	(0.055)	(0.050)	(0.052)	(0.053)	(0.058)
income decile 3	0.164***	0.053	0.044	0.036	-0.011	0.005	-0.072	0.074	-0.028
	(0.054)	(0.048)	(0.050)	(0.057)	(0.057)	(0.052)	(0.055)	(0.055)	(0.060)
income decile 4	0.177***	0.184***	0.155***	0.056	0.118**	0.031	-0.123**	0.093*	0.087
	(0.055)	(0.049)	(0.050)	(0.057)	(0.058)	(0.052)	(0.055)	(0.056)	(0.061)
income decile 5	0.183***	0.297***	0.246***	0.003	0.107*	0.018	-0.138**	0.108*	0.041
	(0.060)	(0.054)	(0.055)	(0.063)	(0.063)	(0.058)	(0.060)	(0.061)	(0.067)
income decile 6	0.175***	0.338***	0.290***	0.133**	0.270***	-0.006	-0.199***	0.157**	0.183***
	(0.061)	(0.054)	(0.056)	(0.063)	(0.063)	(0.058)	(0.061)	(0.061)	(0.067)
income decile 7	0.234***	0.356***	0.272***	0.203***	0.199***	0.061	-0.157**	0.253***	0.179***
	(0.060)	(0.054)	(0.056)	(0.063)	(0.063)	(0.058)	(0.061)	(0.061)	(0.067)
income decile 8	0.300***	0.466***	0.421***	0.195***	0.363***	0.148**	-0.064	0.379***	0.202***
	(0.062)	(0.055)	(0.057)	(0.064)	(0.065)	(0.059)	(0.062)	(0.063)	(0.069)
income decile 9	0.450***	0.469***	0.447***	0.247***	0.431***	0.138**	-0.104*	0.358***	0.245***
	(0.059)	(0.053)	(0.054)	(0.061)	(0.062)	(0.057)	(0.060)	(0.060)	(0.066)
income decile 10	0.329***	0.511***	0.511***	0.409***	0.492***	0.317***	0.161*	0.391***	0.357***
	(0.086)	(0.078)	(0.080)	(0.091)	(0.091)	(0.083)	(0.087)	(0.089)	(0.097)
fluid intell # correct	0.002	-0.006	-0.002	0.007	-0.000	-0.007	-0.004	0.004	0.008
	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)
numeracy # correct	0.010	0.007	-0.015	-0.017	-0.040	-0.013	0.010	-0.033	0.030
	(0.026)	(0.023)	(0.024)	(0.027)	(0.027)	(0.025)	(0.025)	(0.026)	(0.029)
financial literacy # correct	0.024	0.049***	0.063***	0.008	0.014	0.008	0.001	0.022	0.014
	(0.018)	(0.016)	(0.016)	(0.018)	(0.019)	(0.017)	(0.017)	(0.018)	(0.019)
exec attention # correct	0.001	0.001*	0.000	-0.000	0.000	0.000	0.001	0.000	-0.001
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Risk aversion (financial)	-0.009*	-0.002	-0.010**	0.004	-0.015***	-0.009**	-0.005	-0.003	0.001
	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)
Risk aversion (income)	-0.004	0.008	0.003	0.000	0.031***	0.012	0.002	0.020**	0.013
	(0.010)	(0.009)	(0.009)	(0.010)	(0.010)	(0.009)	(0.010)	(0.010)	(0.011)
patience (stdized)	0.004	-0.004	0.001	0.024*	-0.032**	-0.002	0.010	-0.019	0.019
	(0.012)	(0.011)	(0.011)	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)	(0.014)
Time spent decile 2	-0.001	-0.107	-0.042	0.106	0.022	-0.097	-0.105	-0.059	-0.204*
	(0.096)	(0.086)	(0.087)	(0.098)	(0.099)	(0.090)	(0.115)	(0.096)	(0.105)
Time spent decile 3	-0.041	-0.055	-0.012	0.087	-0.112	-0.154	-0.122	-0.100	-0.166
	(0.102)	(0.092)	(0.093)	(0.105)	(0.106)	(0.097)	(0.120)	(0.103)	(0.112)
Time spent decile 4	-0.028	-0.112	-0.068	0.099	-0.128	-0.188*	-0.096	-0.115	-0.218*
	(0.105)	(0.095)	(0.096)	(0.109)	(0.109)	(0.100)	(0.121)	(0.106)	(0.116)
Time spent decile 5	-0.054	-0.064	-0.022	0.073	-0.088	-0.173*	-0.168	-0.122	-0.210*
	(0.107)	(0.096)	(0.098)	(0.111)	(0.112)	(0.102)	(0.123)	(0.108)	(0.118)
Time spent decile 6	0.062	-0.107	-0.016	0.116	-0.070	-0.125	-0.104	-0.109	-0.182
	(0.108)	(0.097)	(0.098)	(0.111)	(0.112)	(0.103)	(0.123)	(0.109)	(0.119)
Time spent decile 7	0.105	-0.041	-0.036	0.100	-0.069	-0.156	-0.167	-0.096	-0.189
	(0.109)	(0.097)	(0.099)	(0.112)	(0.113)	(0.104)	(0.124)	(0.110)	(0.120)
Time spent decile 8	0.083	-0.068	-0.020	0.128	-0.150	-0.196*	-0.166	-0.126	-0.245**
	(0.110)	(0.099)	(0.101)	(0.114)	(0.115)	(0.105)	(0.124)	(0.111)	(0.122)
Time spent decile 9	0.027	-0.114	-0.051	0.082	-0.046	-0.161	-0.133	-0.178	-0.263**
	(0.111)	(0.099)	(0.101)	(0.114)	(0.115)	(0.105)	(0.125)	(0.112)	(0.122)
Time spent decile 10	-0.023	-0.065	-0.060	0.016	-0.051	-0.170*	-0.166	-0.148	-0.233*
	(0.109)	(0.097)	(0.099)	(0.112)	(0.113)	(0.103)	(0.124)	(0.109)	(0.119)
Other controls listed in Table 2?	yes	yes	yes	yes	yes	yes	yes	yes	yes
R-squared	0.29	0.43	0.40	0.15	0.21	0.16	0.12	0.25	0.12
Observations	1505	1505	1505	1505	1505	1505	1505	1505	1505

* 0.10 ** 0.05 *** 0.01. Each column presents results from an OLS regression where the unit of observation is the individual. LHS variable is the component of financial condition listed in the column; see Table 3 for details. Models are otherwise identical to those in Table 10.

Appendix Table 10. Exploratory factor analysis on B-factors

Method	Assume missing value is:		
	Unbiased	Median bias	Fully biased
Horn	3	2	2
Scree	1	1	1
Onatski	1	1	2
Number of B-factors	16	16	16
Number of obs	1515	1515	1515

We use the percentile version of each B-factor (for the standard directional bias only for B-factors with the possibility of bi-directional biases). Each model includes 16 instead of 17 B-factors because the two GARP-inconsistency measures are strongly correlated with each other by construction. Results of Horn's parallel analysis using Stata's "paran" command and its default options. Number of common factors is the number of factors with eigenvalues > 1. Scree obtained using Stata's "factor" command with iterated principal factor method, followed visual inspection of "screeplot." We implement Onatski (2009 ECMA) using the Matlab code for extracting static factors provided on Onatski's homepage.

Data Appendix

This Appendix details, for each of the 17 individual behavioral factors:

i)   The motive for eliciting that factor and the mechanism through which that factor might affect financial condition;

ii)   our elicitation method and its key antecedents;

iii)   data quality indicators, including item non-response;

iv)   sample size (as it compares to that for other factors);

v)   definitions and prevalence estimates of behavioral *indicators*, with background on the distinctions between standard vs. non-standard directional biases where applicable;

vi)   descriptions of the *magnitude* and *heterogeneity* of behavioral deviations, including descriptions of the distribution and—where the data permit—estimates of key parameters used in behavioral models;

vii)   estimates of conditional correlations between measures of the behavioral factor and financial outcomes, including particular components of our financial condition index that have particularly strong links to the B-factor per theory. Wherever possible we also provide comparisons to prior work.

Table 1 summarizes our elicitation methods and their key antecedents. Table 4 and Appendix Table 5 present descriptive statistics from our sample and comparisons to prior work. Table 6 and Appendix Table 6 summarize conditional correlations between each behavioral factor and financial outcomes, with Data Appendix Table 1 summarizing comparisons with prior work. Recall that our main outcome measure is an index of 9 correlated measures of financial condition including both objective and self-assessed metrics of financial well-being (Section 1-E of the paper provides details). Table 5 correlates B-factors with our control variables.

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

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

We measure discounting biases with respect to money using the Convex Time Budgets (CTB) method created by Andreoni and Sprenger (2012a). In our version, fielded in ALP module 315 (the first of our two surveys), subjects make 24 decisions, allocating 100 hypothetical tokens

each between (weakly) smaller-sooner and larger-later amounts. See Data Appendix Figure 1 for an example. The 24 decisions are spread across 4 different screens with 6 decisions each. Each screen varies start date (today or 5 weeks from today) x delay length (5 weeks or 9 weeks); each decision within a screen offers a different yield on saving. Among the 1,515 individuals who take our first survey, 1,502 subjects make at least one CTB choice, and the 1,422 who complete at least the first and last decisions on each of the 4 screens comprise our CTB sample.

The CTB already has been implemented successfully in field contexts in the U.S. (Barcellos and Carvalho 2014; Carvalho, Meier, and Wang 2016) and elsewhere (Gine et al. forthcoming). In exploring data quality and prevalence below we focus on comparisons to Andreoni and Sprenger (2012a), and Barcellos and Carvalho (2014).[1] AS draw their sample from university students. BC's sample is drawn from the ALP, like ours (module 212 in their case), but they use a different adaptation of the CTB.

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

We calculate biased discounting, for each individual, by subtracting the consumption rate when the sooner payment date is five weeks from today from the consumption rate when the sooner payment date is today, for each of the two delay lengths. We then average the two differences to get a continuous measure of biased discounting. In keeping with AS, BC and several other recent papers (including Carvalho, Meier, and Wang (2016) and Goda et al.

---

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

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

(2015)), we find little if any present-bias on average, with a median discount bias of zero, and a 1pp mean tilt toward future bias.[3]

Indicators of behavioral deviations here are bi-directional: we label someone as present-biased (future-biased) if the average difference is >0 (<0). We deem present-bias the "standard" direction, since future-bias is relatively poorly understood[4] and could actually lead to more wealth accumulation. Counting any deviation from time-consistent discounting as biased, 26% of our sample is present-biased and 36% is future-biased. As Appendix Table 5 shows, these prevalence estimates fall substantially if we set a higher threshold for classifying someone as behavioral; e.g., if we count only deviations > |20|pp, then only 3% of the sample is present-biased and 5% future-biased. Compared to prior prevalence estimates, our zero-threshold ones are in the middle of the range. E.g., BC's CTB elicitation in the ALP shows 29% with any present-bias, and 37% with any future-bias. Goda et al. use a different elicitation method—a "time-staircase" multiple price list (Falk et al. 2015)—and classify 55% of their nationally representative sample (from the ALP and another online panel) as present-biased. In the AS sample 14% exhibit any present-bias and 12% any future-bias.

Interestingly, if we follow AS and use the CTB data to structurally estimate discounting-bias parameter values for each individual, we find that 90% of our subjects with no monotonicity violations lie within the interval [0.93, 1.07] (Data Appendix Table 2, Columns 11-13).[5] This is noteworthy because behavioral macro papers sometimes assume representative agents with present bias that lies strictly below our 5th percentile (see, e.g., (İmrohoroğlu, İmrohoroğlu, and Joines 2003; Graham and Snower 2013; Pérez Kakabadse and Palacios Huerta 2013). As Harris and Laibson (2013) state: "the short-run discount factor… is typically thought to lie between ½ and 1." Our estimates should give researchers pause before choosing a value much below 1.

Our estimates of conditional correlations between financial condition and discounting biases (Table 6 and Appendix Table 6) differ in several respects from previous studies that estimate relationships between directly elicited discounting biases and outcomes in broad samples

---

[3] Bradford et al. (2014) do find present-bias on average in their Qualtrics sample, classifying >50% as present-biased and 26% as future-biased.
[4] Although see Koszegi and Szeidl (2013) for a theory of future-biased discounting.
[5] The 5th to 95th percentile interval AS' sample is [0.91, 1.11], as reported in their Table 3.

(Bradford et al. 2014; Eisenhauer and Ventura 2006; Goda et al. 2015).[6] We use CTBs rather than Multiple Price Lists, test more flexible functional forms, and control for a much richer set of (behavioral) factors that could be correlated with both discounting and outcomes.[7]

Our findings suggest that future bias is uncorrelated with financial condition, that present bias is, and that the present bias correlation is robust to different functional forms and to excluding various key control variables—including other B-factors. The magnitude of the correlation is substantial, implying e.g., that present-biased individuals have financial condition that is 10% worse than unbiased (i.e., time-consistent) individuals, conditional on our rich set of controls.

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

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

Of the 1427 persons taking our second survey, 1423 answer one of the two snack questions, and 1404 respond to both. 61% choose the healthy snack for today, while 68% choose it for five weeks in the future, with 15% exhibiting present bias (consume treat today, plan to eat healthy in the future) and 7% future bias (consume healthy today, plan to eat treat in the future).[8] Barcellos and Carvalho's ALP subjects answered similar questions in their baseline survey, albeit with only a one-week instead of a five-week delay, with 6% exhibiting present-bias and 9% future-bias. Read and van Leeuwen (1998) offer actual snacks to a convenience sample of employees in

---

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

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

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

Amsterdam but do not calculate individual-level measures of bias. They do find substantial present-bias on average.

Table 6 and Appendix Table 6 present the first estimates we know of correlating measures of consumption discounting biases with field outcomes. We do not find any evidence of statistically significant coefficients, and the implied magnitudes are small. Having said that the present-biased coefficient has the expected negative sign across all specifications (Appendix Table 6). Among the components of our financial condition index, present-bias has been particularly strongly linked to a low savings rate, and indeed in unreported results we find the same relationship between present-bias and savings (one component of our financial condition index).

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

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

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

Following Choi et al., we average across these 11 decisions, within-consumer, to benchmark choices against two different standards of rationality. One benchmark is a complete and transitive preference ordering adhering to the General Axiom of Revealed Preference (GARP),

---

[9] As Section 1-B of the paper details, we were quite constrained on survey time and hence conducted a pilot in which we tested the feasibility of capturing roughly equivalent information with fewer rounds. 58 pilot-testers completed 25 rounds, and we estimated the correlation between measures of choice inconsistency calculated using the full 25 rounds, and just the first 11 rounds. These correlations are 0.62 and 0.88 for the two key measures.

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

as captured by the Afriat (1972) Critical Cost Efficiency Index. 1-CCEI can be interpreted as the subject's degree of choice inconsistency: the percentage points of potential earnings "wasted" per the GARP standard. But as Choi et al. discuss, consistency with GARP is not necessarily the most appealing measure of decision quality because it allows for violations of monotonicity with respect to first-order stochastic dominance (FOSD).[11] Hence, again following Choi et al., our second measure captures inconsistency with both GARP and FOSD.[12] Note that these measures of inconsistency are unidirectional: there is no such thing as being *overly* consistent.

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

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

Choi et al. find that choice inconsistency with GARP is conditionally correlated with lower net worth, but that choice inconsistency with GARP+dominance avoidance is not. Our data show little evidence of a significant conditional correlation between either measure of choice

---

[11] E.g., someone who always allocates all tokens to account X is consistent with GARP if they are maximizing the utility function U(X, Y)=X. Someone with a more normatively appealing utility function—that generates utility over tokens or consumption per se—would be better off with the decision rule of always allocating all tokens to the cheaper account.

[12] The second measure calculates 1-CCEI across the subject's 11 actual decisions and "the mirror image of these data obtained by reversing the prices and the associated allocation for each observation" (Choi et al. p. 1528), for 22 data points per respondent in total.

inconsistency and our broader measure of financial condition, although the coefficients have the expected negative sign (Table 6).

D. Risk attitude re: certainty

Behavioral researchers have long noted a seemingly disproportionate preference for certainty (PFC) among some consumers and posited various theories to explain it: Cumulative Prospect Theory (Daniel Kahneman and Tversky 1979; Amos Tversky and Kahneman 1992), Disappointment Aversion (Bell 1985; Loomes and Sugden 1986; Gul 1991), and u-v preferences (Neilson 1992; Schmidt 1998; Diecidue, Schmidt, and Wakker 2004). PFC may help to explain seemingly extreme risk averse behavior, which could in turn lead to lower wealth in the cross-section.

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

1,463 of 1,505 (97%) of our subjects who started the tasks completed all 20 choices (compared to 977/1127 = 87% in Callen et al.). As is typical with Holt-Laury tasks, we exclude some subjects whose choices indicate miscomprehension of or inattention to the task. 11% of our subjects multiple-switch on our two-lottery task (compared to 10% in Callen et al.), and 9% of our subjects multiple-switch on the lottery vs. certain option tasks (compared to 13% in Callen et al.). 14% of our subjects switch too soon for monotonic utility in the two-lottery—in rows [2, 4]

---

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

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

in the two-lottery task—compared to 13% in Callen et al. All told, 19% of our subjects exhibit a puzzling switch (17% in Callen et al.), leaving us with 1,188 usable observations. Of these subjects, 1,049 switch on both tasks, as is required to estimate CP. Of these 1,049, only 30% switch at the same point on both tasks, in contrast to 63% in Callen et al.

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

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

Callen et al. find significant correlations between the CP and financial outcomes, in particular with avoiding late loan repayments,[15] but their data lack controls for cognitive skills and other B-factors. We do not find evidence of a statistically significant correlation between CP and our index of financial outcomes, and the implied magnitude of our point estimates is small, although our index lacks a direct measure of loan delinquency. Looking particularly at the component of our financial condition index most directly linked to risk-taking—holding public

[15] The marginal relationship between late loan repayments and wealth is a bit ambiguous under limited liability, and the average relationship (not conditioning on borrowing) more ambiguous, since borrowing could lead to (weakly) greater or lesser wealth if consumers are behavioral (Zinman 2014).

equities—we do find stronger correlations, with point estimates of -0.03 for our CP indicator and -0.05 for our CP percentile and p-values of about 1 in each case.

E. Loss aversion/small-stakes risk aversion

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

We measure loss aversion using the two choices developed by Fehr and Goette (2007) in their study of the labor supply of bike messengers (see Abeler et al. (2011) for a similar elicitation method). Choice 1 is between a lottery with a 50% chance of winning $80 and a 50% chance of losing $50, and zero dollars. Choice two is between playing the lottery in Choice 1 six times, and zero dollars. As Fehr and Goette (FG) show, if subjects have reference-dependent preferences, then subjects who reject lottery 1 have a higher level of loss aversion than subjects who accept lottery 1, and subjects who reject both lotteries have a higher level of loss aversion than subjects who reject only lottery 1. In addition, if subjects' loss aversion is consistent across the two lotteries, then any individual who rejects lottery 2 should also reject lottery 1 because a rejection of lottery 2 implies a higher level of loss aversion than a rejection of only lottery 1. Other researchers have noted that, even in the absence of loss aversion, choosing Option B is compatible with small-stakes risk aversion.[16] We acknowledge this but use "loss aversion" instead of "loss aversion and/or small-stakes risk aversion" as shorthand. Small-stakes risk aversion is also often classified as behavioral because it is incompatible with expected utility theory (Rabin 2000).

Response rates suggest a high level of comfort with these questions; only two of our 1,515 subjects skip, and only two more who answer the first question do not answer the second. 37% of our 1,511 respondents reject both lotteries, consistent with relatively extreme loss aversion, compared to 45% of FG's 42 subjects. Another 36% of our subjects accept both lotteries,

---

[16] A related point is that there is no known "model-free" method of eliciting loss aversion (Dean and Ortoleva 2016).

consistent with classical behavior, compared to 33% in FG. The remaining 27% of our subjects (and 21% of FG's) exhibit moderate loss aversion, playing one lottery but not the other, with our main difference from FG being that 14% of our subjects (vs. only 2% of theirs) exhibit the puzzling behavior of playing lottery 1 but not lottery 2. Although one wonders whether these 14% misunderstood the questions, we find only a bit of evidence in support of that interpretation: those playing the single but not compound lottery have slightly lower cognitive skills than other loss averters, conditional on our rich set of covariates, but actually have higher cognitive skills than the most-classical group. And playing the single but not the compound lottery is uncorrelated with our measure of ambiguity aversion, pushing against the interpretation that the compound lottery is sufficiently complicated as to appear effectively ambiguous (Dean and Ortoleva 2016).

All told 64% of our subjects indicate some loss aversion, defined as rejecting one or both of the small-stakes lotteries, as do 67% in FG. In Abeler et al.'s (2011) student sample, 87% reject one or more of the four small-stakes lotteries with positive expected value. The Abeler et al. questions were also fielded in an ALP module from early 2013 used by Hwang (2016); 70% of that sample exhibits some loss aversion. In von Gaudecker et al.'s nationally representative Dutch sample, 86% exhibit some loss aversion, as inferred from structural estimation based on data from multiple price lists. We also order sets of deviations to indicate greater degrees of loss aversion, based on whether the individual respondent rejects the compound but not the single lottery, rejects the single but not the compound lottery, or rejects both.

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

important covariates and the only other behavioral factor considered is an interaction between loss aversion and a measure of the Gambler's Fallacy (labeled "Heuristics" in the Hwang paper).[17]

We do not find evidence of a strong conditional correlation between loss aversion and financial condition, although the coefficients are uniformly negative across specifications, which is consistent with the hypothesis that loss aversion is associated with lower wealth. Looking particularly at the component of our financial condition index most directly linked to risk-taking—holding public equities—we do find stronger correlations, with loss averse respondents 2.2pp (SE 2.3pp) less likely to hold stocks than classical respondents (those who play both small-stakes lotteries). Specifying our loss aversion measure categorically, we find that the most loss averse respondents (those who reject both lotteries) are 4.1pp less likely to hold stocks than classical respondents (SE 2.7pp), again conditional on our rich set of controls.

F. Narrow Bracketing and Dominated Choice

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

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

Our first task follows RW's Example 2, with Decision 1 between winning $100 vs. a 50-50 chance of losing $300 or winning $700, and Decision 2 between losing $400 vs. a 50-50 chance

---

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

of losing $900 or winning $100.[18] As RW show, someone who is loss averse and risk-seeking in losses will, in isolation (narrow bracketing) tend to choose A over B, and D over C. But the combination AD is dominated with an expected loss of $50 relative to BC. Hence a broad-bracketer will never choose AD. 29% of our subjects choose AD, compared to 53% in the most similar presentation in RW.

Our second task reproduces RW's Example 4, with Decision 1 between winning $850 vs. a 50-50 chance of winning $100 or winning $1,600, and Decision 2 between losing $650 vs. a 50-50 chance of losing $1,550 or winning $100. As in task one, a decision maker who rejects the risk in the first decision but accepts it in the second decision (A and D) violates dominance, here with an expected loss of $75 relative to BC. 23% of our subjects choose AD, compared to 36% in the most similar presentation in RW. A new feature of task two is that AD sacrifices expected value in the second decision, not in the first. This implies that for all broad-bracketing risk averters AC is optimal: it generates the highest available expected value at no variance. 50% of our subjects choose AC, compared to only 33% in the most similar presentation in RW. I.e., 50% of our subjects do NOT broad-bracket in this task, compared to 67% in RW.

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

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

Research linking directly-elicited measures of NBDC to field outcomes is just beginning. The only paper we know of in this vein, Gottlieb and Mitchell (2015), uses a different method for measuring narrow bracketing—one that does not allow for dominated choice—the Tversky

---

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

and Kahneman (1981) "sensitivity to framing" questions regarding the policy response to an epidemic. 30% of subjects in the Health and Retirement Study choose different policy options under the two different frames, an indicator of framing sensitivity, and this indicator is negatively correlated with the holding of long-term care insurance, conditional on a rich set of covariates include a measure loss aversion.

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

G. Ambiguity Aversion

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

We elicit a coarse measure of ambiguity aversion using just one or two questions about a game that pays $500 if you select a green ball. The first question offers the choice between a Bag One with 45 green and 55 yellow balls vs. a Bag Two of unknown composition. 1,397 subjects respond to this question (out of 1,427 who answer at least one of our questions on ALP module 352). 73% choose the 45-55 bag, and we label them ambiguity averse. The survey then asks these subjects how many green balls would need to be in Bag One to induce them to switch. We subtract this amount from 50, dropping the 99 subjects whose response to the second question is >45 (and the 10 subjects who do not respond), to obtain a continuous measure of ambiguity aversion that ranges from 0 (not averse in the first question) to 50 (most averse=== the three subjects who respond "zero" to the second question). The continuous measure (N=1,288) has a mean of 14 (median=10), and a SD of 13. If we impose a large-deviation threshold of 10 (20% of the max) for labeling someone as ambiguity averse, 50% of our sample exceeds this threshold and another 16% are at the threshold. Our elicitation does not distinguish between ambiguity-neutral and ambiguity-seeking choices (for more comprehensive but still tractable methods see, e.g., Dimmock, Kouwenberg et al. (2016), Dimmock, Kouwenberg, and Wakker (forthcoming), Gneezy et al. (2015)), and so our measure of deviation from ambiguity-neutrality is one-sided.

Despite the coarseness of our elicitation, comparisons to other work suggest that it produces reliable data. Our ambiguity aversion indicator correlates strongly with one constructed from Dimmock et al.'s elicitation in the ALP (0.14, p-value 0.0001, N=789), despite the elicitations taking place roughly 3 years apart. Prevalence at our 10pp large-deviation cutoff nearly matches that from Dimmock, Kouwenberg et al.'s (2016) ALP sample and Butler et al.'s (2014) Unicredit Clients' Survey sample from Italy, and the prevalence of any ambiguity aversion (0% cutoff) nearly matches Dimmock, Kouwenberg, and Wakker's (forthcoming) from the Dutch version of the ALP.

Our predictive analysis builds on the papers by Dimmock and co-authors cited above, which estimate conditional correlations between ambiguity aversion and financial market behavior, by broadening the set of both outcomes and control variables (especially other b-factors).[19] We do find evidence of statistically significant negative correlations between ambiguity aversion and our financial condition index.

H. Overconfidence

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

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

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

---

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

(they instead focus on task-level data, sample-level summary statistics, and/or correlates of cross-sectional heterogeneity in estimation patterns).

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

The third measures confidence in placement (relative performance), using a self-ranking elicited before taking our number series test (Section 2-D): *"We would like to know what you think about your intelligence as it would be measured by a standard test. How do you think your performance would rank, relative to all of the other ALP members who have taken the test?"* We find a better-than-average effect in the sample as a whole (70% report a percentile>median) that disappears when we ask the same question immediately post-test, still not having revealed any scores (50% report a percentile>median). We also construct an individual-level measure of confidence in placement by subtracting the subject's actual ranking from his pre-test self-ranking (N=1,395). This measure is useful for capturing individual-level heterogeneity ordinally, but not for measuring prevalence because the actual ranking is based on a 15-question test and hence its percentiles are much coarser than the self-ranking.

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

We are not aware of any prior work exploring conditional correlations between the sorts of overconfidence measures described above and field outcomes. Our results find the expected negative conditional correlation with financial condition, and a significant positive effect of under-confidence.

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

Under-weighting the importance of the Law of Large Numbers (LLN) can affect how individuals treat risk (as in the stock market), or how much data they demand before making

decisions. In this sense non-belief in LLN (a.k.a. NBLLN) can act as an "enabling bias" for other biases like loss aversion (Benjamin, Rabin, and Raymond 2016).

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

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

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

1,375 subjects respond (out of the 1,427 who answer at least one of our questions in Module 352),[20] with mean (SD) responses of 27 (18), 42 (24), and 31 (20). We measure NBLLN using the distance between the subject's answer for the [481, 519] range and 78. Only one subject gets it exactly right. 87% underestimate; coupled with prior work, this results leads us to designate underestimation as the "standard" directional bias. The modal underestimator responds with 50 (18% of the sample). The other most-frequent responses are 25 (10%), 30 (9%), 33 (8%), and 40 (7%). Few underestimators—only 4% of the sample—are within 10pp of 78, and their mean distance is 43, with an SD of 17. 9% of the sample underestimates by 20pp or less. 13% overestimate relative to 78, with 5% of the sample quite close to correct at 80, and another 5% at 100. Benjamin, Moore, and Rabin (2013) do not calculate individual-level measures of underestimation or overestimation in their convenience sample, but do report that the sample means are 35%, 36%, and 29% for the three bins. The comparable figures in our data are 27%, 42%, and 31%.

We are not aware of any prior work exploring conditional correlations between directly-elicited NBLLN and field outcomes. Our results suggest those correlations do exist in point terms, with those underestimating convergence exhibiting lesser financial condition in the cross-section (Appendix Table 6).

---

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

J.   The Gambler's Fallacy

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

We take a slice of Benjamin, Moore, and Rabin's (2013) elicitation for the gambler's fallacy (GF)/hot hand fallacy (HHF):

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

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

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

---

[21] 34% of the sample in Benjamin, Moore, and Raymond respond "10%"on one or more of their ten questions.

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

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

Dohmen et al. also explore correlations between unemployment or bank overdrafts and directly-elicited GF/HHF, conditional on age, gender, education, income, and wealth. They find evidence of positive correlations between HHF and unemployment and between GF and overdrafting. We find a general pattern of evidence that HHF and GF are negatively conditionally correlated with our financial condition index in point terms (Table 6, Appendix Table 6), but the statistical significance of the results is on the weak side.

K. Exponential Growth Bias

Exponential Growth Bias (EGB) produces a tendency to underestimate the effects of compounding on costs of debt and benefits of saving. It has been linked to a broad set of financial outcomes (Levy and Tasoff 2016; Stango and Zinman 2009).

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

We calculate an individual-level measure of "debt-side EGB" by comparing the difference between the APR *implied* by the monthly payment supplied by that individual, and the *perceived* APR as supplied directly by the same individual. We start by binning individuals into under-

estimators (the standard bias), over-estimators, unbiased, and unknown (37% of the sample).[22] The median level difference between the correct and stated value is 500bp, with a mean of 1,042bp and SD of 1,879bp. Among those with known bias, we count as biased 70%, 64%, 47%, and 34% under error tolerance of zero, 100bp, 500bp, and 1000bp. Under these tolerances we count 3%, 13%, 41%, and 61% as unbiased, and 27%, 22%, 10%, and 3% as negatively biased. This is substantially less EGB than Stango and Zinman (2009; 2011) see from questions in the 1983 Survey of Consumer Finances, where 98% of the sample underestimates, and the mean bias is 1,800bp or 3,800bp depending on the benchmark. The time frames of the questions differ, which may account for the difference.

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

Our second measure of EGB comes from a question popularized by Banks and Oldfield (2007) as part of a series designed to measure basic numeracy: "Let's say you have $200 in a savings account. The account earns 10 percent interest per year. You don't withdraw any money for two years. How much would you have in the account at the end of two years?" 1,389 subjects answer this question (out of the 1,427 respondents to Module 352), and we infer an individual-level measure of "asset-side EGB" by comparing the difference between the correct future value ($242), and the future value supplied by the same individual.[23] We again bin individuals into underestimators (the standard bias), overestimators, unbiased, and unknown (14% of the

---

[22] Non-response is relatively small, as only 4% of the sample does not respond to both questions. Most of those we label as unknown-bias give responses that imply or state a 0% APR. 7% state payment amounts that imply a negative APR, even after being prompted to reconsider their answer. We also classify the 4% of respondents with implied APRs >=100% as having unknown bias.

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

sample).[24] Among those with known bias (N=1,222), the median bias is $0, with a mean of $2 and SD of $14.[25] 44% of our sample provides the correct FV. 47% of our sample underestimates by some amount, with most underestimators (29% of the sample) providing the linearized (uncompounded) answer of $240. Nearly all other underestimates provide an answer that fails to account for even simple interest; the most common reply in this range is "$220." Only 9% of our sample overestimates the FV, with small mass points at 244, 250, 400, and 440.

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

Goda et al. (2015) and Levy and Tasoff (2016) measure asset-side EGB using more difficult questions in their representative samples. They find that 9% and 11% overestimate FVs, while 69% and 85% underestimate.

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

---

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

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

L. Limited Attention/Memory

Prior empirical work has found that limited attention affects a range of financial decisions (e.g., Barber and Odean 2008; DellaVigna and Pollet 2009; Karlan et al. 2016; Stango and Zinman 2014). Behavioral inattention is a very active line of theory inquiry as well (e.g., Bordalo, Gennaioli, and Shleifer 2015; Kőszegi and Szeidl 2013; Schwartzstein 2014).

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

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

Responses are strongly but not perfectly correlated (ranging 0.56 to 0.69 among pairwise expressions of regret). A fourth measure of limited attention is also strongly correlated with the others, based on the question: "Do you believe that you could improve the prices/terms your household typically receives on financial products/services by shopping more?"[26] 18% respond "Yes, and I/we often regret not shopping more," and the likelihood of this response is correlated 0.25 with each of the regret measures above. 1,483 subjects answer all four questions, out of the

---

[26] This question is motivated by evidence that shopping behavior strongly predicts borrowing costs (Stango and Zinman 2016).

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

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

We also seek to measure limited prospective memory, following previous work suggesting that limited memory entails real costs like forgetting to redeem rebates (e.g., Ericson 2011). We offer an incentivized task to subjects taking module 352: "The ALP will offer you the opportunity to earn an extra $10 for one minute of your time. This special survey has just a few simple questions but will only be open for 24 hours, starting 24 hours from now. During this specified time window, you can access the special survey from your ALP account. So we can get a sense of what our response rate might be, please tell us now whether you expect to do this special survey." 97% say they intend to complete the short survey, leaving us with a sample of 1,358. Only 14% actually complete the short survey. Failure to complete is weakly positively correlated with our indicators of attentional regret described above, with correlations ranging from 0.02 to 0.04, and p-values from 0.20 to 0.50.

Our indicator of behavioral limited memory—(do not complete conditional on intending to complete)—is a bit coarse. We suspect that some noise is introduced because our elicitation makes it costless to express an intention to complete (in future research we plan to explore charging a small "sign up" fee), thereby including in the indicator's sample frame some subjects who rationally do not complete the task. Relatedly, although we set the payoff for task completion to be sufficiently high to dominate any attention/memory/time costs in *marginal* terms for most subjects (the effective hourly wage is in the hundreds of dollars), it may well be the case that the *fixed* cost exceeds $10 for some respondents.

Despite the noise, our coarse indicator of limited prospective memory turns out to have the expected negative and significant correlation with financial condition. This is the first work we know of estimating the conditional relationship between limited memory and field outcomes in a broad sample.

References for Data Appendix

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

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

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

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

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

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

Banks, James, Cormac O'Dea, and Zoë Oldfield. 2010. "Cognitive Function, Numeracy and Retirement Saving Trajectories." *The Economic Journal* 120 (548): F381–410.

Barber, Brad, and Terrence Odean. 2008. "All That Glitters: The Effect of Attention on the Buying Behavior of Individual and Institutional Investors." *Review of Financial Studies* 21 (2): 785–818.

Barberis, Nicholas C. 2013. "Thirty Years of Prospect Theory in Economics: A Review and Assessment." *Journal of Economic Perspectives* 27 (1): 173–96.

Barcellos, Silvia, and Leandro Carvalho. 2014. "Information about Self-Control and Intertemporal Choices."

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

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

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

Benjamin, Daniel, Matthew Rabin, and Collin Raymond. 2016. "A Model of Nonbelief in the Law of Large Numbers." *Journal of the European Economic Association* 14 (2): 515–44.

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

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

Burks, Stephen, Jeffrey Carpenter, Lorenz Götte, and Aldo Rustichini. 2012. "Which Measures of Time Preference Best Predict Outcomes: Evidence from a Large-Scale Field Experiment." *Journal of Economic Behavior & Organization* 84 (1): 308–20.

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

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

Carvalho, Leandro, Stephan Meier, and Stephanie Wang. 2016. "Poverty and Economic Decision-Making: Evidence from Changes in Financial Resources at Payday." *American Economic Review* 106 (2): 260–84.

Chabris, Christopher F., David Laibson, Carrie L. Morris, Jonathon P. Schuldt, and Dmitry Taubinsky. 2008. "Individual Laboratory-Measured Discount Rates Predict Field Behavior." *Journal of Risk and Uncertainty* 37 (2–3): 237–69.

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

Daniel, Kent, and David Hirshleifer. 2015. "Overconfident Investors, Predictable Returns, and Excessive Trading." *Journal of Economic Perspectives* 29 (4): 61–88.

Dean, Mark, and Pietro Ortoleva. 2016. "Is It All Connected? A Testing Ground for Unified Theories of Behavioral Economics Phenomena."

DellaVigna, Stefano, and Joshua M Pollet. 2009. "Investor Inattention and Friday Earnings Announcements." *The Journal of Finance* 64 (2): 709–49.

Diecidue, Enrico, Ulrich Schmidt, and Peter P Wakker. 2004. "The Utility of Gambling Reconsidered." *Journal of Risk and Uncertainty* 29 (3): 241–59.

Dimmock, Stephen G., and Roy Kouwenberg. 2010. "Loss-Aversion and Household Portfolio Choice." *Journal of Empirical Finance* 17 (3): 441–59.

Dimmock, Stephen, Roy Kouwenberg, Olivia S. Mitchell, and Kim Peijnenburg. 2016. "Ambiguity Aversion and Household Portfolio Choice Puzzles: Empirical Evidence." *Journal of Financial Economics* 119 (3): 559–77.

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

Dohmen, Thomas, Armin Falk, David Huffman, Felix Marklein, and Uwe Sunde. 2009. "Biased Probability Judgment: Evidence of Incidence and Relationship to Economic Outcomes from a Representative Sample." *Journal of Economic Behavior & Organization* 72 (3): 903–15.

Eisenhauer, Joseph G, and Luigi Ventura. 2006. "The Prevalence of Hyperbolic Discounting: Some European Evidence." *Applied Economics* 38 (11): 1223–34.

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

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

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

Gine, Xavier, Jessica Goldberg, Daniel Silverman, and Dean Yang. forthcoming. "Revising Commitments: Field Evidence on the Adjustment of Prior Choices." *Economic Journal*.

Gneezy, Uri, Alex Imas, and John List. 2015. "Estimating Individual Ambiguity Aversion: A Simple Approach."

Goda, Gopi Shah, Matthew R Levy, Colleen Flaherty Manchester, Aaron Sojourner, and Joshua Tasoff. 2015. "The Role of Time Preferences and Exponential-Growth Bias in Retirement Savings."

Gottlieb, Daniel, and Olivia S. Mitchell. 2015. "Narrow Framing and Long-Term Care Insurance."

Graham, Liam, and Dennis J Snower. 2013. "Hyperbolic Discounting and Positive Optimal Inflation." *Macroeconomic Dynamics* 17 (03): 591–620.

Grubb, Michael D. 2015. "Overconfident Consumers in the Marketplace." *Journal of Economic Perspectives* 29 (4): 9–36.

Gul, Faruk. 1991. "A Theory of Disappointment Aversion." *Econometrica: Journal of the Econometric Society* 59 (3): 667–86.

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

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

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

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

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

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

Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman. 2016. "Getting to the Top of Mind: How Reminders Increase Saving." *Management Science* 62 (12): 3393–3411.

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

Kőszegi, Botond, and Matthew Rabin. 2009. "Reference-Dependent Consumption Plans." *American Economic Review* 99 (3): 909–36.

Kőszegi, Botond, and Adam Szeidl. 2013. "A Model of Focusing in Economic Choice." *The Quarterly Journal of Economics* 128 (1): 53–104.

Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112 (2): 443–77.

Larrick, Richard P, Katherine A Burson, and Jack B Soll. 2007. "Social Comparison and Confidence: When Thinking You're Better than Average Predicts Overconfidence (and When It Does Not)." *Organizational Behavior and Human Decision Processes* 102 (1): 76–94.

Levy, Matthew, and Joshua Tasoff. 2016. "Exponential-Growth Bias and Lifecycle Consumption." *Journal of the European Economic Association* 14 (3): 545–83.

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

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

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

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

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

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

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

Rabin, Matthew. 2000. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica* 68 (5): 1281–92.

Rabin, Matthew, and Dimitri Vayanos. 2010. "The Gambler's and Hot-Hand Fallacies: Theory and Applications." *Review of Economic Studies* 77 (2): 730–78.

Rabin, Matthew, and Georg Weizsäcker. 2009. "Narrow Bracketing and Dominated Choices." *American Economic Review* 99 (4): 1508–43.

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

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

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

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

———. 2011. "Fuzzy Math, Disclosure Regulation, and Credit Market Outcomes: Evidence from Truth-in-Lending Reform." *Review of Financial Studies* 24 (2): 506–34.

———. 2014. "Limited and Varying Consumer Attention: Evidence from Shocks to the Salience of Bank Overdraft Fees." *Review of Financial Studies* 27 (4): 990–1030.

———. 2016. "Borrowing High vs. Borrowing Higher: Price Dispersion and Shopping Behavior in the U.S. Credit Card Market." *Review of Financial Studies* 29 (4): 979–1006.

Sutter, Matthias, Martin G Kocher, Daniela Glätzle-Rützler, and Stefan T Trautmann. 2013. "Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior." *American Economic Review* 103 (1): 510–31.

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

Tversky, A., and D. Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211 (4481): 453–58.

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

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

Zinman, Jonathan. 2014. "Consumer Credit: Too Much or Too Little (or Just Right)?" *Journal of Legal Studies* 43 (S2 Special Issue on Benefit-Cost Analysis of Financial Regulation): S209–37.

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

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

Data Appendix Figure 1. Discounting choices, screenshot
(1 of 4 screens, 6 choices per screen)
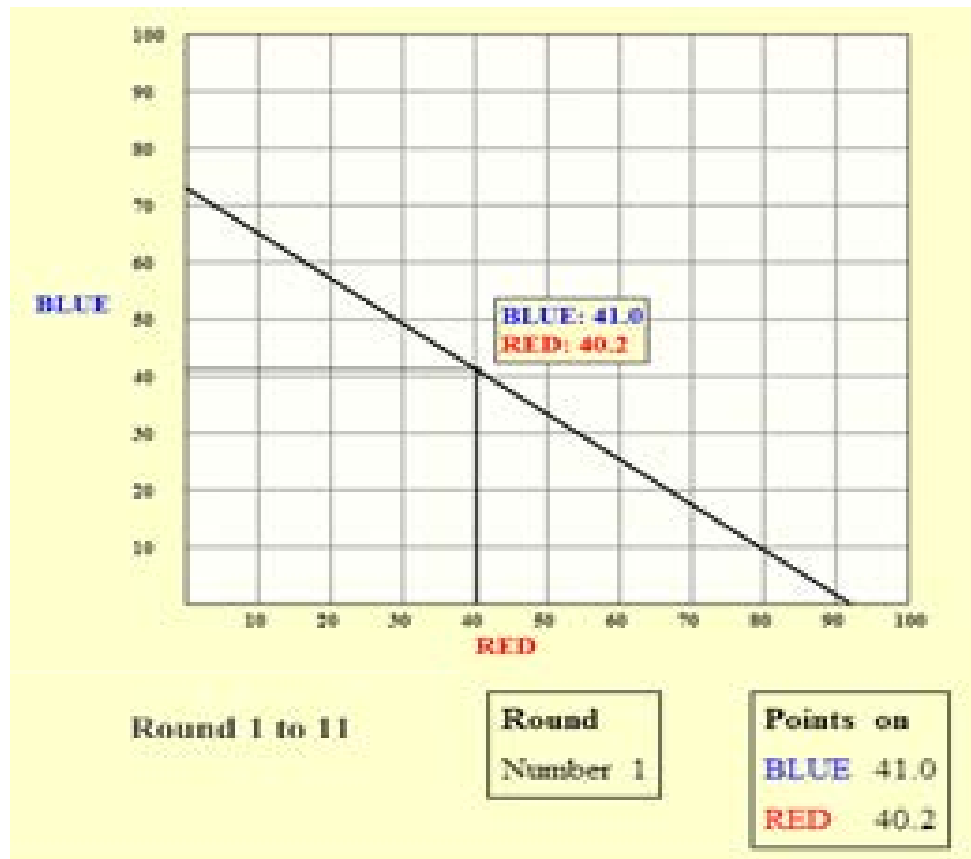
Data Appendix Figure 2. Consistency with GARP choices, screenshot
(1 of 11 rounds, 1 choice per round).

<u>**Data Appendix Table 1. The conditional statistical significance of individual behavioral factors: comparisons to prior work.**</u>

B-factor	Source	Representative data?	With other b-factors?	control for cog skills?	significant corr?
Present-biased discounting, money	Goda et al.	yes	EGB-asset	yes	yes
	Bradford et al.	yes	no	no	yes
	Our paper	yes	yes*	yes**	yes
Present-biased discounting, food	No prior work				
	Our paper	yes	yes*	yes**	no
Inconsistency with GARP	Choi et al.	yes (Dutch)	no	yes	yes
	Our paper	yes	yes*	yes**	no
Inconsistency with GARP + dominance avoidance	Choi et al.	yes (Dutch)	no	yes	no
	Our paper	yes	yes*	yes**	no
Preference for certainty	Callen et al.	no	no	no	yes
	Our paper	yes	yes*	yes**	no
Loss aversion	Fehr and Goette	no	no	individual FE	moderates effect
	Dimmock and Kouw	yes	no	no	yes
	Abeler et al.	no	no	no	yes, with lab effort choices
	Hwang	yes	Gamb Fallacy	no	yes
	Our paper	yes	yes*	yes**	no
Narrow bracketing	Gottlieb and Mitchell	yes (age >49)	loss aversion	yes	yes
	Our paper	yes	yes*	yes**	no
Ambiguity aversion	Dimmock, Kouw, et al.	yes	no	yes	yes
	Dimmock, Kouw, Wakker	yes	no	yes	no
	Sutter et al.	no	no	yes	no
	Our paper	yes	yes*	yes**	yes
Overconfidence in estimation	No prior work				
	Our paper	yes	yes*	yes**	yes***
Overconfidence in precision	No prior work				
	Our paper	yes	yes*	yes**	no
Overconfidence in placement	No prior work				
	Our paper	yes	yes*	yes**	yes
NBLLN: underestimates	No prior work				
	Our paper	yes	yes*	yes**	yes
Gambler's Fallacy: cold hand	Dohmen et al. (2009)	yes	no	no	yes
	Our paper	yes	yes*	yes**	yes***
EGB debt-side: underestimates	Stango and Zinman	yes	no	no	yes
	Our paper	yes	yes*	yes**	no
EGB asset-side: underestimates	Goda et al.	yes	no	no	yes
	Our paper	yes	yes*	yes**	yes
Limited attention with regret	No prior work				
	Our paper	yes	yes*	yes**	yes
Limited prospective memory	No prior work				
	Our paper	yes	yes*	yes**	yes

See Data Appendix for more details on B-factor definitions and measurement. "Our paper" refers to specifications reported in Table 6 and Appendix Table 6, except where noted by ***. Measure of cognitive skills are detailed in Section 1-D of the paper and Appendix Table 1.
* Appendix Table 6.
** Table 6.
*** In Appendix Table 7.

Data Appendix Table 2. Estimated Distributions of Individual-Level Present Bias Parameter from our data

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

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

Background consumption:

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

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

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

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

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

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