NBER WORKING PAPER SERIES

BEHAVIORAL BIASES ARE TEMPORALLY STABLE

Victor Stango
Jonathan Zinman

Working Paper 27860
http://www.nber.org/papers/w27860

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2020

Thanks to Hannah Trachtman and Sucitro Dwijayana Sidharta for outstanding research assistance; the Sloan/Sage Working Group on Behavioral Economics and Consumer Finance, the Roybal Center (grant # 3P30AG024962), the National University of Singapore, the Michigan Retirement Research Center, and the Pension Research Council/TIAA for funding and patience; Shachar Kariv and Dan Silverman for helping us implement their (with Choi and Muller) interface for measuring choice consistency; Charlie Sprenger for help with choosing the certainty premium elicitation tasks and with adapting the convex time budget tasks; Georg Weizsacker for help in adapting one of the questions we use to measure narrow bracketing; Julian Jamison for advice on measuring ambiguity aversion; Doug Staiger and Josh Schwartzstein for many conversations; and Erik Snowberg for helpful discussions re: measurement error. Special thanks to Joanne Yoong for collaboration on the Round 1 survey design and implementation. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Behavioral Biases are Temporally Stable
Victor Stango and Jonathan Zinman
NBER Working Paper No. 27860
September 2020
JEL No. C36,C81,D90,E70

ABSTRACT

Social scientists often consider temporal stability when assessing the usefulness of a construct and its measures, but whether behavioral biases display such stability is relatively unknown. We estimate stability for 25 biases, in a nationally representative sample, using repeated elicitations three years apart. Bias level indicators are largely stable in the aggregate and within-person. Within-person intertemporal rank correlations imply moderate stability and increase dramatically when using other biases as instrumental variables. Additional results reinforce three key inferences: biases are stable, accounting for classical measurement error in bias elicitation data is important, and eliciting multiple measures of multiple biases is valuable.

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

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

Temporal stability is one of the key criteria for organizing the study of individual characteristics such as preferences, cognitive skills, and personality.[1] Stability is often considered a necessary condition for a characteristic to be a meaningful construct, as Stigler and Becker (1977) famously argue in the case of preferences. A lack of stability makes a characteristic more difficult to model, measure and link to behavior.

Behavioral biases are relatively recent additions to the taxonomy of individual characteristics and differences, so relatively little is known about their temporal stability. How stable are biases at the population level? Answers to this question have implications for macroeconomic models, such as those specifying a mix of biased and unbiased agents. How stable are biases at the person level? Answers to this question have implications for modeling consumer choice. What is the nature and extent of measurement error in data from behavioral bias elicitations, and what is the best way to deal with the impact of measurement error on stability estimates, or when linking biases to outcomes and other characteristics? Answering these questions informs research design for almost any project seeking to employ a behavioral bias measure.

We address each of these questions with empirical evidence on level and rank stability for 25 behavioral biases from 17 potential sources of bias. The data come from lab-style elicitations administered twice to a nationally representative sample of 845 U.S. adults, in surveys three years apart (in 2014 and 2017). Biases are deviations from classical assumptions about consumer preferences, beliefs, or problem-solving approaches. We select biases with clear classical benchmarks, that feature prominently in behavioral economics, and that are amenable to compact elicitation in online surveys.[2]

Our research design produces seven sets of empirical findings that point to behavioral biases being largely stable and illuminate portable strategies for dealing with measurement error.

[1] See, e.g., Mata et al. (2018).

[2] An example: Money discounting is a potential source of bias. It has two potential biases, present-bias or future-bias, vs. the classical benchmark of time-consistent discounting. We measure these using Andreoni and Sprenger’s (2012) Convex Time Budget elicitation but shorten it from 45 choices to 24.

First, sample-wide average bias prevalence, measured discretely for each bias as biased or unbiased relative to the classical benchmark, is remarkably stable across our three year gap.[3] Across each our 25 biases, the average absolute value change in the share of biased individuals is two percentage points, with the biggest change being seven percentage points (on a base of 82 percent for that bias). The sample share of biased observations remains almost constant across the three years, starting at 43 percent and ending at 42 percent. These results support macro models that are microfounded on deep behavioral tendencies and/or assume fixed shares of behavioral and non-behavioral agents over time (e.g., Crucini et al. (2020); Gabaix (forthcoming)).

Second, bias indicators are also quite stable within-person. The sample share of panelists with an unchanged bias/unbiased indicator across rounds ranges from 0.58 to 0.93 across the 25 biases, with a mean of 0.73. This further supports the inference that biases are largely stable in levels, at least on the level margin we measure: the extensive margin. And together with the first set of findings—near-perfect stability of bias indicators at the sample level—it suggests that any measured within-person instability is the artifact of noise around a basically constant mean.

Third, and incorporating information on the intensive margins of deviations from classical benchmarks, panelist rank in the cross-section of each bias is somewhat stable within-person even before adjusting for measurement error.[4] Across our 25 biases, within-person round-to-round Pearson or tetrachoric rank correlations range from 0.02 to 0.59 (with standard errors mostly in the 0.02 to 0.04 range), averaging 0.33. Estimated correlations for the most-comparable standard, classical measures of patience and risk aversion are around 0.30 and 0.40, both in our data and in prior work (Chuang and Schechter 2015; Mata et al. 2018). Estimated stability is substantially higher for cognitive skills and personality traits, which is unsurprising as researchers have spent many more decades conceptualizing and refining elicitations of these constructs (e.g., Jensen 1998; Anusic and Schimmack 2016).

[3] A complementary approach to estimating level stability would be to do so for absolute measures of bias intensity; e.g., by structurally estimating a model-specific parameter for each bias. We do not take this approach because many of our elicitations are too coarse to produce informative measures of level intensity.
[4] Psychometricians tend to prefer rank stability over level stability as a criterion for evaluating constructs. E.g., Schildberg-Hörisch's (2018) classically-focused review of risk preference elicitation explains that rank stability, in contrast to level stability, "accommodates evidence on systematic changes in risk preferences over the life cycle, due to exogenous shocks such as economic crises or natural catastrophes, and due to temporary changes in self-control resources, emotions, or stress" (p. 136).

Fourth, bias stability does not emerge mechanically from a process where classical characteristics like cognitive skills and personality traits are stable, and biases are just functions of those stable characteristics and/or demographics like gender, race and education. To the contrary, biases appear only weakly fit by those other stable characteristics, if at all: adjusted R-squareds from regressions of bias rank on flexibly parameterized measures of cognitive skills, personality traits, patience and risk aversion, and demographics range from -0.01 to 0.22 across our 25 biases, with an average of 0.06.[5] Temporal stability in biases manifests from something that is not just stable, but behaviorally distinct.

Fifth, a standard instrumental variables approach to measurement error (Fuller 2009) dramatically increases estimates of intertemporal stability. Exploiting the fact that biases are correlated with each other, instrumenting for one bias with many others increases the estimated within-person rank correlation, from the aforementioned cross-bias average of 0.33, to around 0.80. Several additional empirical results suggest that the most plausible possible confounds, such as errors correlated across elicitations, would likely push estimated IV correlations downward, not upward towards spurious stability.

Sixth, while the first five sets of results point to classical measurement error attenuating unadjusted (non-instrumented) estimates of stability, we find little evidence of true instability. Various literatures emphasize that changes in opportunities, constraints, or hedonic states can affect decision inputs like behavioral biases, but we find little evidence for such mechanisms in our data.[6] Specifically, we find little evidence that individuals with more stable financial condition, household composition, or subjective well-being exhibit more stable biases across our two snapshots.

Seventh, some plausible alternative approaches to measurement error yield less promising results than measurement error IV ("ME-IV"). We explore whether filtering the sample on measures of survey effort, response quality, cognitive skills, or conscientiousness affects unadjusted bias stability estimates, and find little evidence that it does. We also explore whether

[5] Our results on fit add to a literature that has focused on correlations between biases and classical characteristics (e.g., Dohmen et al. 2010; Benjamin, Brown, and Shapiro 2013).
[6] See, e.g., work on Scarcity (e.g., Mullainathan and Shafir 2013; Carvalho, Meier, and Wang 2016; Lichand and Mani 2019), emotions (e.g., A. Meier 2019) and trauma (e.g., Beine et al. 2020), with the latter two literatures focused primarily on measures of risk and time preferences that impose classical priors.

elicitations that include more task repetition or produce more granular bias measures produce more stable bias measures, and find no evidence that they do.[7] Moreover, biases that should be more stable per theory are not actually more stable in unadjusted terms; in particular, unadjusted stability estimates for behavioral preferences are substantially lower than for non-preference biases. In contrast, the IV estimates exhibit more theoretically appealing patterns.

We contribute most directly to three literatures on the measurement and application of behavioral biases. One, temporal stability was heretofore unstudied for most biases: we could find comparable within-person estimates for only 4 of the 17 potential sources of bias and 4 of the 25 biases we measure, and an aggregate stability estimate for only one. None of those prior studies use as ME-IV, as we do.[8] Two, by distinguishing beliefs from preferences in most of our elicitations, and using elicitations that allow for behavioral deviations from classical benchmarks but do not impose either, we also contribute to the broader literature on measuring the temporal stability of risk and time preferences.[9] Three, we provide concrete guidance for researchers wishing to mitigate measurement error and its consequences in empirical studies of behavioral biases: eliciting multiple biases and/or multiple measures of the same bias is key.[10]

---

[7] A key caveat is that this analysis is merely exploratory; we do not directly test how financial incentives or task repetition affect stability estimates or data quality (see Section 1-E for details).

[8] See Meier and Sprenger (2015), Tasoff and Zhang (2020), Chapman et. al. (2019b), and Chapman, Snowberg, Wang, and Camerer (2019). We discuss that work in Section 2. Callen et al. (2014) examine how exposure to violence changes preference for certainty in a cross-section but does not have within-person estimates of temporal stability. Our work also relates to literatures on the stability of inferences across studies, including replication and external validity issues; see, e.g., DellaVigna and Pope (2019).

[9] E.g., for risk aversion, the most temporally-stable measures come from "stated" or "self-report" elicitations (Mata et al. 2018) like the Dohmen et al. measure we use: "… Are you generally a person who is fully prepared to take financial risks or do you try to avoid taking financial risks? Please indicate your answer on a scale from 0-100…." As has been well-documented (e.g., Falk et al. 2018), this sort of elicitation provides a very useful reduced-form summary of what is more accurately described as risk attitudes instead of risk preferences. But it does not help researchers unpack the extent to which the reported risk-taking, and stability thereof, is driven by: 1) beliefs/perceptions about risk vs. by preferences re: uncertainty *per se*; or 2) any of the many behavioral biases hypothesized to affect beliefs and preferences over risk. Several of our elicitations allow for such distinctions. We allow for the possibility of biased misperceptions of probability like non-belief in the law of large numbers or gambler's fallacies, and for the possibility of biased perceptions of intertemporal prices like exponential growth biases that could affect perceived returns to risk-taking. We also allow for the possibility of behavioral preferences per se like ambiguity aversion, loss aversion or small-stakes risk aversion, preference for certainty, etc.

[10] See also, e.g., Gillen, Snowberg, and Yariv (2019), Giglio et al. (2020), and our two companion papers (Stango and Zinman 2020a; 2020b). In our companion papers, we find that a similar IV approach to the one used here substantially strengthens the estimated links between biases, financial decisions, and subjective well-being. Specifically, we instrument for a *group* of person i's biases at time t+1 with the same group of

Altogether, our findings produce three key inferences. Behavioral biases, or at least the many tendencies toward bias we measure, are temporally stable. Accounting for classical measurement error in bias elicitation data is important. Eliciting multiple measures of multiple biases is an effective method for dealing with that measurement error.

1. Research design and data collection

A. Variables overview

We focus on two sets of consumer characteristics in this paper. Set one contains behavioral biases (detailed in Section 1-E below and Data Appendix Section 1). Set two contains classical characteristics: demographics, measures of presumed-classical time and risk preferences/attitudes, cognitive skills, and personality traits (Section 1-D). We also use data on financial condition, subjective well-being, response times, and response quality, as described in Sections 3-D and 3-E.

B. The American Life Panel

We administered our surveys through the RAND American Life Panel (ALP). The ALP is an online survey panel established in 2003. RAND regularly offers panel members opportunities to participate in surveys designed by researchers for purposes spanning the range of social sciences. Over 400 survey modules have been administered in the ALP, and RAND makes data publicly available after a period of initial embargo.

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

C. Research design and sample

Three principles guide our research design. First, measure the richest set of individual characteristics possible, to minimize potential confounds from omitted variables and to allow exploration of relationships among and between behavioral biases and classical characteristics.

---

person i's biases at time t and vice versa to help examine relationships among the stable components of biases that are linked theoretically ("convergent validity," in psychometrics parlance), and between the stable components of biases and field behavior ("predictive validity"). Here, we instrument for *single* biases with other biases to obtain better estimates of temporal stability.

Second, use standard elicitations and survey questions wherever possible, although in many cases we shorten lab-style elicitations to free up time and budget for measuring a broader set of characteristics. Third, take repeated measurements at different points in time, to describe the temporal stability of behavioral biases and to help account for measurement error.

To those ends, we designed a "round" of about an hour's worth of elicitations and survey questions. We then split this round into two modules designed to be administered roughly two weeks apart, to minimize survey fatigue and per the ALP's advice re: module length and invitation sequencing. After extensive piloting and refinements, we had the ALP field our two Round 1 modules starting in November 2014. We targeted 1,500 working-age respondents, sending 2,103 initial invitations, and ultimately received 1,515 responses to Round 1 Module 1 (ALP #315), with 1,427 subsequently completing Round 1 Module 2 (ALP #352). 95% of respondents completing both modules did so by the end of February 2015.

We undertook Round 2 in October 2017, by inviting the universe of 1,308 panelists who completed both Round 1 modules and remained empaneled to take ALP #474, which is a replica of Round 1 Module 1 (but was not advertised as such). We received 967 responses and then invited those panelists to take the second module. ALP #472 is a replica of Round 1 Module 2 with some additional questions added at the end (but was not advertised as such). We received 845 responses to this second module, creating our sample of 845 panelists who responded to both modules in both rounds.

In refining our elicitations for each potential bias source and other consumer characteristics, we were mindful of the reality that research budgets force tradeoffs between the depth and breadth of measurements, incentives, and sample size. Per standard ALP practice, we paid panelists $10 per completed module. Beyond that, all but one of our elicitations are unincentivized on the margin (limited prospective memory being the exception; see Table 1 for details). Scrutiny of usual motivations for paying marginal incentives casts doubt on their value, given our research objectives, relative to spending research funds on measuring a broader set of consumer characteristics, on a broader sample, at multiple points in time. Researchers often hypothesize that subjects find stylized tasks unpleasant and hence need marginal incentives to engage with the tasks, but the ALP measures panelist engagement and finds evidence to the contrary.[11] Researchers often hypothesize that unincentivized elicitations change inferences, but that hypothesis is not robustly supported empirically (e.g., Von Gaudecker, Van Soest, and Wengström 2011; Gneezy, Imas, and List 2015) and there is a long tradition of using unincentivized lab-style elicitations in surveys (e.g., Barsky et al. 1997; Falk et al. 2018; Bauer, Chytilová, and Miguel 2020). Researchers often assume that marginal incentive mechanisms are the best way to mimic real-world stakes, but this is not generally true for behavioral consumers (Azrieli, Chambers, and Healy 2018), and tasks with hypothetical rewards like ours can offer some conceptual advantages (e.g., Montiel Olea and Strzalecki 2014). In any case, our repeated elicitations and measurement error models should suffice to address concerns about noise.

D. Classical characteristic elicitations and measures

The ALP collects demographics such as income, gender, age, education, and family structure when a panelist first registers, refreshes these measures quarterly, and links them to each module. Our modules measure presumed-classical risk attitudes/preferences with the adaptive lifetime income gamble task developed by Barsky et al. (1997), the financial risk-taking scale from Dohmen et al. (2010; 2011),[12] and patience using the average savings rate across the 24 choices in our version of the Convex Time Budget task (Andreoni and Sprenger 2012). We measure cognitive skills using 4 standard tests for general/fluid intelligence (McArdle, Fisher, and Kadlec 2007), numeracy (Banks and Oldfield 2007), financial literacy (crystalized intelligence for financial decision making) per Lusardi and Mitchell (2014), and executive function/working memory (MacLeod 1991).[13] Pairwise correlations between these four test scores range from 0.16 to 0.42. In our second round of surveying we add elicitations of the Big Five personality traits to the end of our second module.[14]

---

[11] For example, each ALP survey ends with "Could you tell us how interesting or uninteresting you found the questions in this interview?" and roughly 90% of our sample replies that our modules are "Very interesting" or "Interesting", with only 3% replying "Uninteresting" or "Very uninteresting," and 7% "Neither interesting nor uninteresting".

[12] These Barsky and Dohmen et al. measures are correlated 0.14 in our sample. We also elicit Dohmen et al.'s general risk taking scale, which is correlated 0.68 with the financial scale.

[13] The Data Appendix Section 2 provides details on each of these elicitations and measures.

[14] Specifically, we use the validated 10-item version of the Big Five personality trait inventory (Rammstedt and John 2007). We initially decided against eliciting personality measures, given our resource constraints

E. Behavioral bias elicitations and measures

We measure 17 potential sources of biases. These produce 25 total bias measures, because 8 of the 17 sources allow for deviations in two directions from the classical benchmark. For example, discounting can be either present-biased or future-biased relative to time consistency; the gambler's fallacy can be either a "hot hand" or "cold hand" misperception relative to the correct belief; one can be either over- or under-confident; and so on.

Table 1 summarizes our measures and elicitations, with the Data Appendix Section 1 providing details, including granular data descriptions and comparisons of data quality indicators and descriptive statistics to prior work. Some of our elicitations are relatively rich and produce granular information on bias intensity. Others are relatively coarse, with two producing just a 0/1 indicator of whether someone is biased or not: consumption discounting and limited prospective memory. The median number of distinct values produced by our elicitations, across our 25 biases, is ten (Table 1, Column 4).

Two sources of biases relate to discounting, over snacks and money (Read and van Leeuwen 1998; Andreoni and Sprenger 2012). Three relate to decision quality, including inconsistency with the General Axiom of Revealed Preference (Choi et al. 2014), and narrow bracketing (Rabin and Weizsäcker 2009). Three relate to preferences under uncertainty: loss aversion/small-stakes risk aversion (Fehr and Goette 2007), preference for certainty (Callen et al. 2014) and ambiguity aversion (Dimmock et al. 2016). Three are varieties of overconfidence: about absolute performance, relative performance and forecast precision (Moore and Healy 2008). Four sources of "math" biases include two statistical fallacies—both gambler's and non-belief in the law of large numbers (Dohmen et al. 2009; Benjamin, Moore, and Rabin 2017; Benjamin, Rabin, and Raymond 2016)—and exponential growth biases over borrowing costs and investment returns (Stango and Zinman 2009; Levy and Tasoff 2016). We also elicit limited attention and limited memory, in ways that are motivated by behavioral models of inattention (Ericson 2011; Bronchetti et al. 2020).[15]

and the lack of prior evidence of correlations between them and behavioral biases (see, e.g., Becker et al.'s (2012) review article).

[15] Following a common delineation in behavioral economics, we do not measure social preferences. See Dean and Ortoleva (2019) and Chapman et al. (2019a) for evidence on relationships between behavioral biases and social preferences.

For estimating level stability we use the extensive margin of each bias: 1 if biased, 0 otherwise. Measuring the extensive margin is feasible in a wide range of settings because it does not require an elaborate elicitation or granular bias measures. It is desirable because bias indicators are unit-free in cross-bias comparisons and map well into models with a mix of biased and unbiased agents.

For estimating rank stability we use the respondent's cross-sectional percentile rank,[16] measured as the percentile into which the respondent falls for that bias and survey round.[17] As with our measure of bias level, rank is model-free and comparable across biases. Rank has the added benefits of capturing information on cross-sectional heterogeneity along both the extensive and intensive margins of bias, and of conforming to the psychometric standard for measuring stability (see footnote 4).

2. Bias stability measures and correlations not allowing for measurement error

A. Biases are stable in the aggregate

Table 2 describes aggregate bias stability across our two survey rounds. Although bias prevalence varies substantially reading down the rows in Panel A, from 0.06 to 0.89, bias stability remains virtually constant comparing Column 1 to Column 2.[18] Across our 25 biases, the average level change in prevalence across all biases is zero (Panel B). The average absolute value change over three years is two percentage points, on a base prevalence of 43pp in Round 1. The maximum change among our 25 biases is seven percentage points (on a base of 82pp for that bias). The only other estimate we know of for temporal stability of bias sample shares can be inferred from Chapman, Snowberg, Wang, and Camerer's (2019) Figure 3, which suggests that the aggregate share of loss averse consumers is almost perfectly stable over their 6-month horizon.

These results provide support for macroeconomic models assuming fixed (sub-)populations of behavioral agents. The mean-zero change across biases is also noteworthy for inferences about stability at the micro level, as it is consistent with classical measurement error in consumer-level bias measures (see Section 3).

---

[16] There are various ways of measuring rank, and they are very highly correlated with each other in our data.

[17] If the panelist is unbiased, we set the percentile is a "0" for that bias/round. For biases with less-granular measures, the percentiles simply take on values corresponding to the cumulative frequencies of each value.

[18] Using the ALPs population weights produces similar results, here and throughout the paper.

B. Bias levels are stable within-person

We now turn from aggregate to within-person bias stability. Table 3 Column 1 reports, for each of our 25 biases, the share of respondents with an unchanged biased indicator across our two rounds. As an example, 72% of our sample exhibits ambiguity aversion in both rounds or neither round. Other biases show similar stability, with a min/max of [0.58, 0.93]—see also Figure 1a, which orders biases by this measure of level stability. The cross-bias mean is 0.73 (Panel B).[19]

These results further suggest that biases tend to be stable in level terms, at least on the extensive margin, even before allowing for measurement error.

C. Bias percentile ranks are somewhat stable even without allowing for measurement error

Table 3 Column 2 reports, for each of our 25 biases, Pearson or tetrachoric correlation estimates of within-person rank stability across our two survey rounds. We refer to these as "unadjusted" for measurement error. The point estimates range from 0.02 to 0.59—see also Figure 1b, which orders biases from greater to lesser unadjusted rank stability. The cross-bias mean correlation is 0.33 (Panel B). We do not report standard errors here to avoid cluttering the table (but do report them in Table 6 Column 1). All but two SEs imply p-values less than 0.01, with the two exceptions being present-biased money discounting (0.04 with a standard error of 0.03) and certainty premium<0 (0.02 with a standard error of 0.04).

These results align reasonably well with prior work, which although limited in scope yields temporal stability estimates for four of our 25 biases. For present-biased money discounting, the most comparable estimates are from Tasoff and Zhang (2020), which also estimates within-person rank correlations using data from Convex Time Budget elicitations in the ALP, in their case for a short-run discounting parameter. Their estimate is 0.03 over about 16 months, as compared to our estimate of 0.04 for present-bias and 0.11 for future-bias over about three years.[20] For choice inconsistency, Chapman, Snowberg, Wang, and Camerer (2019) find a 6-month correlation 0.24 for a structurally estimated parameter, as compared to our three-year rank correlation estimates of

[19] The prevalence of the two possible changes in bias indicators—from biased to unbiased, or vice versa—tends to be symmetric (we do not report these changes in Table 3 because they are evident from Table 2's near-constant average prevalences over time).

[20] Meier and Sprenger (2015) find one-year within-person level stability correlation of 0.36 for a short-run money discounting parameter that is strongly present-biased on average and elicited using a multiple price list.

0.20 and 0.29 for our two measures of inconsistency with GARP. For loss aversion, Chapman, Snowberg, Wang, and Camerer (2019) find a 6-month within-person correlation of 0.40 for a structurally estimated parameter, as compared to 0.47 for our relatively coarse measure. For ambiguity aversion, Chapman, Dean, Ortoleva, Snowberg, and Camerer (2019b) find a 6-month within-person correlation of 0.21, as compared to our 0.24.

D. Bias stability compared to classical characteristic stability

Because we also measure the classical characteristics described in Section 1-D, we can benchmark classical characteristic stability to our estimated bias stability for the same set of individuals. Table 4 reports our estimated intertemporal, within-person rank correlations for the non-demographic classical characteristics in our data. Each of these correlations has a p-value < 0.01 and so we do not report standard errors in the table.

Starting with comparisons to prior work, our estimated correlation of 0.29 for patience is comparable to Meier and Sprenger's (2015) estimated long-run discount factor stability of 0.25 (using multiple price lists elicited one year apart), and to Chapman, Snowberg, Wang, and Camerer's (2019) estimated discount factor stability of 0.47 (using Dynamically Optimized Sequential Experimentation elicited 6-months apart, and finding greater stability with DOSE than with multiple price lists).[21] For risk aversion, we estimate correlations of 0.38 from the Barsky et al. (1997) income gamble elicitation, and 0.59 from the Dohmen et al. (2010; 2011) qualitative financial risk-taking scale. Mata et al.'s (2018) meta-analysis of risk aversion suggests 3-year stability of around 0.2 for measures gleaned from lottery-choice elicitations like Barsky et al.'s, +/- about 0.1.[22] Mata et al.'s meta-analysis of measures gleaned from self-reported risk-taking elicitations like Dohmen et al.'s suggests precisely estimated 3-year stability of about 0.5.

Since our behavioral bias elicitations are nearly all quantitative, the most apt comparisons are to quantitative elicitations of patience and risk aversion. As discussed in the previous paragraph,

[21] See also Chuang and Schechter's (2015) review.

[22] Subsequently, Chapman, Snowberg, Wang, and Camerer find 6-month stability of 0.44 for their DOSE-elicited lottery choice elicitation, with DOSE yielding greater stability than MPLs. And Tasoff and Zhang (2020) find risk aversion parameter stability, estimated from Barsky et al. elicitations in the ALP administered at varying time intervals, ranging from 0.08 to 0.58. They find that other lottery-choice methods in the ALP have within-person and -elicitation stability ranging from 0.27 to 0.53, again over various horizons.

these tend to produce temporal stability estimates centered in the 0.2 to 0.4 range, both in our sample (0.29 for patience and 0.38 for large-stakes risk aversion) and in prior work. For ease of comparison, we reproduce our cross-bias average unadjusted rank stability correlation of 0.33 from Table 3 as the top row of Table 4. (We examine whether stability varies across groups of related biases in Sections 3-B and 3-E.)

Turning to cognitive skills, our four standard measures have precisely estimated rank intertemporal correlations ranging from 0.48 to 0.81. For personality measures, meta-analysis of the literature on finds a precisely estimated 3-year within-person rank correlation of about 0.6 (Anusic and Schimmack 2016).[23] As noted in the Introduction, the fact that bias stability is only about half that of cognitive skills and personality, before accounting for measurement error, is unsurprising. Research on intelligence and personality has been ongoing for over a century and so their measures should be more refined.

E. Stable biases are not just functions of stable classical characteristics

One might wonder whether bias measures simply reflect relatively stable classical characteristics like cognitive skills and personality traits, a question we address in Table 5. Each cell reports the unweighted mean R-squared (Column 1) or adjusted R-squared (Columns 2-6) across two OLS regressions per bias, one per round, of bias rank on flexibly parameterized measures of cognitive skills, personality traits, patience and risk aversion, and/or plausibly exogenous demographics.

Columns 1 and 2 report the R-squared and adjusted R-squared for regressions including all of these classical characteristics on the right-hand-side. Column 1 shows that, before adjusting for overfitting, together they explain from 11 to 30 percent of the variation in each bias, with a cross-bias mean of 18 percent (Panel B). Column 2 shows that adjusted R-squareds are substantially lower, largely because many characteristics are inter-correlated and contribute incrementally less to overall fit as we add more of them. Adjusted R-squareds range from -0.01 to 0.22 across biases, with a cross-bias mean of 0.06.[24]

[23] On personality, recall that we only elicit those measures in Round 2 and therefore lack in-sample stability estimates.
[24] Appendix Table 1 takes a "kitchen sink" approach and finds that adjusted R-squareds increase modestly in absolute terms; e.g., to a cross-bias mean of 0.11 with the entire expanded set of right-hand-side variables.

Columns 3-6 report the estimated contribution to bias fit for each of the four groups of variables: plausibly exogeneous demographics, risk aversion and patience, cognitive skills, and personality traits. These adjusted R-squareds suggest that demographics and cognitive skills tend to contribute the most to bias fit. But "most" is only relatively speaking, as the adjusted R-squared for any of these variable groups never exceeds 0.18 across 100 estimates (25 biases x 4 groups of RHS variables), and only six exceed 0.10.

In brief, standard measures of classical characteristics poorly explain our bias measures. Temporal stability in our bias measures therefore reflects something distinct and stable about the biases themselves. As noted in the Introduction, these results are also of substantive interest because they add results on fit to a literature on "Who is behavioral?" (Benjamin, Brown, and Shapiro 2013) that has focused on correlations between biases and classical characteristics.

3. Allowing for measurement error

Next, we explore how allowing for measurement error in our biases using an instrumental variable approach ("ME-IV") affects estimated intertemporal rank correlations. At the end of the section we compare the ME-IV method to some other ways of handling measurement error.

A. Measurement error and estimated intertemporal correlations

We are interested in estimating the true intertemporal correlation of bias i (b^*_{i1}, b^*_{i2}), where *i* indexes our 25 biases and *j*=1 or 2 indexes our survey round. That correlation for bias *i* is

$$\rho_i = \frac{\sigma_{i12}}{\sigma_{i1}\sigma_{i2}}$$

Where σ_{i12} is the intertemporal covariance and σ_{i1}, σ_{i2} are the standard deviations of the bias in each round. If the bias has constant standard deviation across rounds, $\sigma_{i1} = \sigma_{i2} = \sigma_i$, the relationship is more simply:

$$\rho_i = \frac{\sigma_{i12}}{\sigma_i^2}$$

One can use regression and the sample standard deviations to estimate the sample correlation,

The additional right-hand-side variables include plausibly endogenous demographics (e.g., income and employment status), measures of survey effort and response quality, and measures of other biases.

$$r_i = \hat{\beta}_i \cdot \frac{s_{i2}}{s_{i1}}$$

Where $\hat{\beta}_i$ is the coefficient from a regression of bias *i* in round 1 on bias *i* in round 2:[25]

$$b_{i1}^{*} = \alpha_i + \beta_i\, b_{i2}^{*} + \nu_{i2}$$

Now, suppose that classical measurement error contaminates our elicited biases:

$$b_{ij} = b_{ij}^{*} + \varepsilon_{ij}$$

As is well-known, that measurement error will attenuate $\hat{\beta}_i$ and thereby the estimated intertemporal correlation r_i by an amount proportional to the measurement error variance $\sigma_{\varepsilon i}^2$, which we assume is the same across rounds for a given bias *i*. If the true intertemporal correlation is ρ_i, with measurement error the estimated correlation is:

$$\hat{\rho}_i = \rho_i \frac{\sigma_i^2}{(\sigma_i^2 + \sigma_{\varepsilon i}^2)}$$

Following a significant body of work using repeated elicitations as instrumental variables to reduce attenuation bias from measurement error—the "ME-IV" approach—we treat contemporaneous elicitations of other biases as instruments.[26] As we detail in Stango and Zinman (2020a), many biases are correlated with each other (see also Chapman et al. (2019a) and Dean and Ortoleva (2019)), particularly those in "related" categories such as discounting, or overconfidence. So, in estimating $\hat{\beta}_i$ we regress round 1 bias on round 2 bias, using other round 2 biases $b_{-i,j}$ in round *j* as instruments for bias in round *j*.[27] To use all of the data, we "stack" round 1 and round 2:

[25] One could identically run the opposite regression, , $b_{i2}^{*} = \alpha_i + \beta_i\, b_{i1}^{*} + \nu_{i1}$, and calculate the correlation by inverting the standard errors in the calculation, $r_i = \hat{\beta}_i \cdot \frac{s_{i1}}{s_{i2}}$.

[26] Gillen et al. (2019) provides a nice summary of the issues and some related recent work.

[27] Each bias-round has missing data for some observations—see the "N" columns in Table 2. We can estimate the Pearson/polychoric correlations in Table 3 in pairwise fashion, using only observations where the bias was observed in both rounds. However, using only nonmissing data in the ME-IV regressions would drop all observations where *any* bias was missing in the instrument set, reducing the sample to ~600 observations rather than ~800. So, in the ME-IV regressions we code missing values for each instrument as "0" and also include a vector of "missing indicators" for each bias instrument, equal to one for any instrument that is missing for that round. We keep the set of observations the same as for the pairwise correlations, by in every model dropping all observations for which the bias of interest is missing in either

Round	LHS	RHS	Instruments
1	b_{i1}	b_{i2}	$b_{-i,2}$
2	b_{i2}	b_{i1}	$b_{-i,1}$

The primary assumptions for the instruments to be valid are that measurement error across the two rounds of elicitation is independent, $E[\varepsilon_{i1}\varepsilon_{i2}] = 0$, and that measurement error is independent from one bias to another, $E[\varepsilon_{ij}\varepsilon_{-i,j}] = 0$.[28] We scrutinize these assumptions in Section 3-B and for now note that they are standard (e.g., Gillen et al. (2019); Giglio et al. (2020)). In our data, they could most plausibly be violated where biases are mechanically related to each other because they are measured from the same elicitation. In most cases, this just means that we exclude the other direction in a bi-directional bias—for example, when we instrument for present-bias on snacks, we exclude future-bias on snacks from the instrument set as the two directions come from the same elicitation. Similarly, the two GARP biases use the same elicitation.[29]

B. Allowing for measurement error roughly doubles estimated intertemporal correlations

Table 6 presents ME-IV estimated intertemporal rank correlations for the each of the 25 biases. Using all other non-mechanically-related biases as instruments, the average correlation rises from 0.33 to 0.80 (Panel B), consistent with what one would expect if measurement error were attenuating the unadjusted correlations. Every IV correlation estimate (Column 2) is larger than its unadjusted estimate (reproduced from Table 3 here in Column 1), and many are close to one. Column 3 shows similar point estimates using the set of related biases only, although the related-bias set produces standard errors that are at least twice as large as the all-bias set in most cases.

round. Coding the instruments this way has little effect on the point estimates relative to dropping all missing values (the average ME-IV correlation is 0.77 in one instance and 0.80 in the other), but substantially improves efficiency since we can use more data.

[28] We also impose the restriction that the (true) coefficients are the same across rounds. This is a weak assumption, as the coefficients should be same, in a statistical sense. But is technically necessary because here, in contrast with OLS case detailed in footnote 25, regressing round 1 on round 2 will not generally lead to the same correlation estimate obtained from regressing round 2 on round 1, because the instrument sets in the two regressions differ.

[29] A few other potential mechanical relationships are more subtle, and we consider them in Section 3-C.

Figure 2 displays the point estimates and 95% confidence intervals for the all-other bias IV estimates in Table 6 Column 2, ordering the biases by correlation magnitude. Of the 25 correlations, 22 have a lower bound that exceeds the average unadjusted estimate of 0.33, and 14 have a confidence interval that includes 1.00. The other striking inference from this figure is that some biases appear to be more stable than others. Although we lack the degrees of freedom and power for a formal analysis of why, we note that the magnitude ordering has some theoretically intuitive appeal per the bias groupings in Table 1 Column 5. Discounting biases are less stable, as they "should" be, given their dependence on expectations and opportunity sets as well as on time preference *per se* (see also Appendix Table 2, row 2, first column of results). In contrast, the eight measures of preferences bias are relatively stable, with five point estimates >0.8 and with 1.0 in their confidence interval (see also Appendix Table 2, row (11), first column of results). The "math" biases are also relatively stable, which makes sense if they reflect something of fluid intelligence (which is known to be quite stable) as well as statistical or perceptual biases (see also Appendix Table 2, row (6), first column of results). One cautionary note however is the strong positive correlation between coefficient magnitude and precision in Figure 2; this suggests that our IV strategy works better for some biases than others, in which case the true bias stability ordering could differ from that suggested by the ME:IV estimates.

C. Assessing the ME-IV approach: The first stage and exogeneity

Table 7 shows test statistics assessing the first stages and exogeneity restrictions associated with our IV approach. Columns 1 and 2 show, for each bias and instrument set, the associated p-value for the F-test of joint significance in the first stage of the IV regression. That p-value assesses the strength of our instruments: do other, non-mechanically related biases jointly strongly correlate with the instrumented bias? By and large, they do. For the all-other bias IV sets, 22 of 25 have p-values < 0.01. For the related bias IV sets, 16 of 25 have p-values < 0.01.

Columns 3 and 5 present the p-values for the "J-test" of overidentifying restrictions in the IV model. The J-test relies on the assumption that the exclusion restriction holds for at least one of the instruments in the set. This assumption seems especially likely to hold in our all-other bias models, where we have 24 IVs for most instrumented biases after excluding any mechanically-related bias(es) and only need one to be exogenous—without knowing which one—for the test to be empirically informative. If the assumption holds, the J-test identifies whether all of the instruments jointly meet the exclusion restriction, under the null that they do, with a smaller p-value indicating a higher probability the exclusion restriction fails to hold.

The J-test rejects the exclusion restriction for at most a few biases. The IV specifications each have only four p-values < 0.10, and only two < 0.01. Moreover we would expect some of these to be false rejections given that we are testing 50 null hypotheses. But even if we take each rejection as informative, the J-test rejects the null for both sets of IVs for only one or two biases (depending on the p-value cutoff), suggesting that one can usually find a valid IV for a given bias from contemporaneous measures of other biases.

As a final check on the ME-IV identifying assumptions, Table 7 Column 4 tests whether including IVs that are more likely to fail the exclusion restriction leads to a higher likelihood of rejecting the J-test's null hypotheses of exogeneity, relative to our main specification that excludes mechanically-related biases in Column 3, as theory suggests it should. We focus on six cases: including each GARP inconsistency bias in the instrument set for the other, including the level overconfidence biases in the IV set for the future value exponential growth biases and vice versa (they share part of an elicitation, as detailed in the Data Appendix), and including patience in the instrument set for the present- and future-bias in money discounting.[30] As predicted, the p-value falls in each of the six cases, and sharply in five of them. Three (four) of six p-values imply a change from failing to reject in Column 3 to rejecting in Column 4 with a 0.01 (0.10) cutoff. As such, Column 4 provides additional support for the ME-IV identifying assumptions.

Not shown in Table 7, but also informative and reassuring, is that adding the mechanically related IV(s) substantially reduces the estimated stability coefficient in four of six cases.[31] Taken together with the aforementioned pattern of smaller ME-IV coefficients having larger standard errors (Figure 2), this suggests that any endogenous or weak IVs likely push estimates toward OLS and away from our key inference that biases are temporally stable, by-and-large.

[30] As noted in Section 3-A, the two opposing directions of bi-directional biases are also potentially mechanically related by virtue of being measured using the same elicitation, like present- and future-bias on money discounting. But we do not include a bias's opposite direction in the test here because bi-directional bias pairs have a conceptual mechanical relationship as well as a measurement one: if someone is biased in one direction, they cannot be biased in the opposite direction by definition.

[31] Specifically: from 0.97 to 0.45 and from 1.01 to 0.53 for the GARP inconsistency biases, and from 0.95 to 0.68 and from 0.63 to 0.43 for the confidence biases. The EGB future value coefficients are basically unchanged.

D. *Measurement error vs. true instability*

The full picture of evidence thus far strongly suggests that it is important to account for classical measurement error when making inferences from behavioral bias elicitation data, and that ME-IV is a viable strategy for such accounting. Recapping: near perfect sample-level stability in bias levels (Table 2), coupled with substantial but imperfect stability in those same levels within-person (Table 3 Column 1), suggests mean-zero error around the measurement of bias extensive margins. Magnitude estimates for unadjusted bias rank stability (Table 3 Column 2) tend to be similar to comparably-elicited estimates of patience and risk aversion (Table 4). Given that the latter two are commonly presumed by economists to be temporally stable, at least qualitatively speaking, this suggests that unadjusted elicited estimates of preferences, expectations, and decision rules, whether they explicitly allow for behavioral biases or not, are attenuated by classical measurement error. Table 5 and Appendix Table 1 cast serious doubt on the hypothesis that bias stability is due to biases being functions of very stable, non-behavioral characteristics like cognitive skills, personality traits, and demographics. Table 6, Table 7, and Appendix Table 2 suggest that ME-IV is a valid approach that greatly reduces attenuation in bias stability estimates.

This constellation of evidence does not, on its own, rule out meaningful true instability in behavioral biases. One possibility is that ME-IV mistakes true instability for measurement error, although as detailed above this case is unlikely given the plausibility of ME-IV's identifying assumptions on both theoretical and empirical grounds. A more likely possibility is that both measurement error and true instability are important; indeed, the fact that ME-IV increases estimated correlations by a factor of roughly 2.5 implies that measurement error introduces spurious variation that is perhaps only slightly larger than the true variation in biases.[32]

Having said that, we find little empirical evidence consistent with true instability in our data. Motivated in large part by the literatures on Scarcity and emotions discussed in the Introduction, we examine whether households with less temporal variation in opportunities, constraints, and/or hedonic states exhibit less temporal variation in unadjusted bias measures (i.e., whether they exhibit more bias stability). Tables 8a and 8b show that sub-samples of panelists with more-stable household composition and objective measures of finances do not tend to have more stable biases

[32] Recall from Section 3-A that the factor by which measurement error reduces estimated correlations is proportional to the relative variances induced by true variation in the bias.

in level or rank. Nor do sub-samples of panelists with more-stable subjective financial condition, expected financial condition, and life satisfaction have more stable biases in level or rank (Tables 9a and 9b).

E. Other approaches to dealing with measurement error seem less promising than ME-IV

Since ME-IV is costly—it requires rethinking research designs, particularly for projects focused on a single bias, and multiple elicitations—we consider whether less-costly approaches to dealing with measurement error show signs of promise.

First, we consider using response times, a standard input for measuring survey effort, to filter out noisier responses or respondents and thereby reduce attenuation in unadjusted estimates. Appendix Tables 3a (level stability) and 3b (rank stability) show that limiting the sample to respondents exhibiting various standard measures of good survey effort does not produce substantially higher unadjusted stability estimates on average.[33]

Second, we consider using a standard measure of response quality by dropping the 620 panelists with any internally inconsistent response across all bias elicitations (not including the GARP consistency elicitation). The idea is to measure stability on a sub-sample with unusually high presumed response quality. Yet Column 10 of Appendix Tables 3a and 3b shows that this filter does not generally increase bias stability in level or rank, although the cross-bias average is a bit higher in both cases (2pp in each case, on a base of 73pp for level and 33pp for rank).[34]

Third, although we do not randomize task repetition within-elicitation, looking across elicitations one sees that the six biases we measure with data from >10 choices—the two money discounting biases, the two GARP inconsistency biases, and the two certainty premium biases—

[33] Nor does limiting the sample to those with relatively high cognitive skills or conscientiousness (Appendix Tables 4a and 4b), which casts further doubt on the hypothesis that measurement error is produced simply by task difficulty or lack of respondent effort.

[34] We have also extracted "respondent quality" principal components from a broad set of variables that might, in principle, indicate (or be caused by) low survey effort: flexibly parameterized response times, measures of response quality and item non-response, limited attention/memory and decision quality biases. In practice, the data indicate two common factors: one more important (eigenvalue 5.75) that loads on the survey time variables and nothing else, and another of much lesser importance (eigenvalue of 1.70) with high loadings on decision quality, item non-response and response quality. Neither, however, appears meaningfully related to the estimated intertemporal correlations of biases; for example, letting the intertemporal correlations vary by decile of the principal components does not change inferences.

actually have below-average unadjusted stability: 0.68 vs. 0.73 for level, and 0.14 vs. 0.33 for rank. Nor do more granular measures of bias (Table 1, Column 4) yield greater estimates of temporal stability.[35]

Fourth, we consider whether one can use theory to predict which biases will have higher unadjusted stability estimates but find little cause for optimism in the second column of results in Appendix Table 2. There we do not see the expected patterns across bias groupings in unadjusted correlations, in contrast to the more theoretically appealing ordering of the ME-IV magnitudes in the first column (see also Figure 2 and its discussion above in Section 3-B).

4. Conclusion

We increase empirical evidence on the temporal stability of behavioral biases by roughly an order of magnitude with repeated measurements of 25 biases, from 17 potential sources of bias, for a nationally representative sample of 845 U.S. consumers.[36]

Bias levels are remarkably stable across three years in the aggregate, and also quite stable within-person, at least on the margin we measure for levels: the extensive margin. Bias ranks are somewhat stable within-person even before adjusting for measurement error, and comparably so to standard quantitative measures of presumed-classical risk aversion and patience. Those characteristics, and much more stable ones like demographics, cognitive skills, and personality traits, explain very little of bias variation: Bias stability is not due to biases being a function of stable classical characteristics. Instrumenting for a given bias with other biases increases estimated rank correlations dramatically, from 0.33 to around 0.80 on average across the 25 biases. Most of our IV estimates are statistically indistinguishable from one and therefore from complete stability. Several results suggest that the IV identifying assumptions are plausible, and that if they fail to hold they likely produce estimates that understate rather than overstate stability. We find little

[35] Specifically, the bias-level Pearson correlations between the number of unique ranks and the unadjusted stability estimates in Table 3 are -0.12 (SE=0.21) for level and -0.21 (0.20) for rank. For log(unique ranks) the correlations are -0.29 (0.20) and -0.36 (0.19).

[36] The basis for our order of magnitude claim is: (increasing number of biases with temporal stability estimates by a factor of 4 or 6, depending on whether one counts bias sources or bias) x (providing IV estimates as well as unadjusted estimates) x (providing rank as well as level stability estimates). E.g., 4x2x2= 16.

evidence that alternative strategies for addressing measurement error are promising, and little evidence that changes in economic, family, or hedonic circumstances are associated with bias instability.

All told, our results provide multifaceted evidence that behavioral biases are temporally stable, by and large, that it is important to address classical measurement error when eliciting biases, and that measurement error instrumental variables estimators are effective ways of doing so. The latter point builds on recent work showing the value of multiple measures of the same bias (Gillen, Snowberg, and Yariv 2019; Giglio et al. 2020; Stango and Zinman 2020a; 2020b) by showing the added value of instrumenting for one bias with others to help estimate stability.

Much more remains to be done. Although we find little evidence of true instability, we do not rule it out and suspect it would be fruitful to forge tighter connections between our work here and work on Scarcity, emotions, and trauma. Another potential link is to work on meta-awareness: it may be the case that learning about one's biases is limited, and/or that learning leaves underlying biases largely unchanged but affects behavior by increasing sophistication about how to manage them.[37] Similarly, our evidence of stable biases does not rule out dynamics in macroeconomic expectation forecast errors and biases; we do not elicit forecasts but speculate that they emerge from interactions between market forces and the sort of persistent tendencies toward bias that we do measure.[38]

Our results suggest that some biases are harder to measure than others. Examining why, and what to do about it, would likely produce both substantive insights and methodological advances for dealing with measurement error. Related, more systematic testing of elicitation methods and combinations would help researchers optimize research designs for generating informative inferences from behavioral bias elicitations.

[37] Recent work suggests that learning may be limited (e.g., Gagnon-Bartsch, Rabin, and Schwartzstein 2020), and context- and behavior-specific (Allcott et al. 2020; Le Yaouanq and Schwardmann 2020).
[38] This is but one of many classes of theories of belief distortions; for a recent discussion see, e.g., Bianchi, Ludvigson, and Ma (2020).

References

Allcott, Hunt, Joshua Kim, Dmitry Taubinsky, and Jonathan Zinman. 2020. "Are High-Interest Loans Predatory? Theory and Evidence from Payday Lending."

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

Anusic, Ivana, and Ulrich Schimmack. 2016. "Stability and Change of Personality Traits, Self-Esteem, and Well-Being: Introducing the Meta-Analytic Stability and Change Model of Retest Correlations." *Journal of Personality and Social Psychology* 110 (5): 766–81.

Azrieli, Yaron, Christopher P. Chambers, and Paul J. Healy. 2018. "Incentives in Experiments: A Theoretical Analysis." *Journal of Political Economy* 126 (4): 1472–1503.

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.

Bauer, Michal, Julie Chytilová, and Edward Miguel. 2020. "Using Survey Questions to Measure Preferences: Lessons from an Experimental Validation in Kenya."

Becker, Anke, Thomas Deckers, Thomas Dohmen, Armin Falk, and Fabian Kosse. 2012. "The Relationship Between Economic Preferences and Psychological Personality Measures." *Annual Review of Economics* 4 (1): 453–78.

Beine, Michel, Gary Charness, Arnaud Dupuy, and Majlinda Joxhe. 2020. "Shaking Things up: On the Stability of Risk and Time Preferences."

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. 2017. "Biased Beliefs about Random Samples: Evidence from Two Integrated Experiments."

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.

Bianchi, Francesco, Sydney Ludvigson, and Sai Ma. 2020. "Beliefs Distortions and Macroeconomic Fluctuations."

Bronchetti, Erin, Judd Kessler, Ellen Magenheim, Dmitry Taubinsky, and Eric Zwick. 2020. "Is Attention Produced Rationally?"

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.

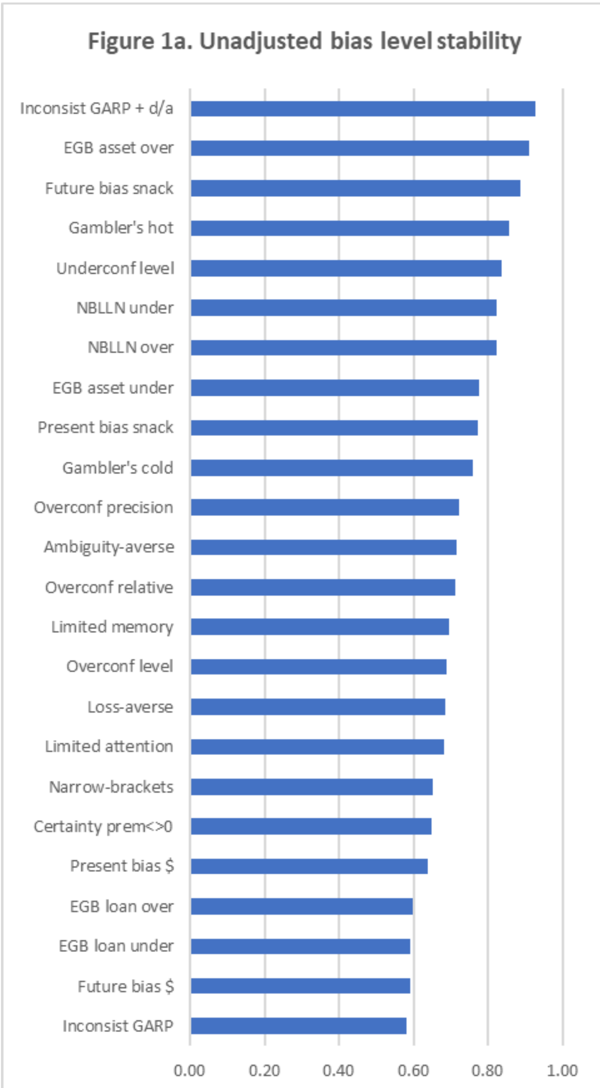
Chapman, Jonathan, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer. 2019a. "Econographics."

———. 2019b. "Willingness to Pay and Willingness to Accept Are Probably Less Correlated than You Think."

Chapman, Jonathan, Erik Snowberg, Stephanie Wang, and Colin Camerer. 2019. "Loss Attitudes in the U.S. Population: Evidence from Dynamically Optimized Sequential Experimentation (DOSE)."

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.
Crucini, Mario, Mototsugu Shintani, and Takayuki Tsuruga. 2020. "A Behavioral Explanation for the Puzzling Persistence of the Aggregate Real Exchange Rate."
Dean, Mark, and Pietro Ortoleva. 2019. "The Empirical Relationship between Non-Standard Economic Behaviors." *Proceedings of the National Academy of Science* 116 (33): 16262–67.
DellaVigna, Stefano, and Devin Pope. 2019. "Stability of Experimental Results: Forecasts and Evidence."
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.
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, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences." *The Quarterly Journal of Economics* 133 (4): 1645–92.
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.
Fuller, Wayne A. 2009. *Measurement Error Models*. Vol. 305. John Wiley & Sons.
Gabaix, Xavier. forthcoming. "A Behavioral New Keynesian Model." *American Economic Review*.
Gagnon-Bartsch, Tristan, Matthew Rabin, and Joshua Schwartzstein. 2020. "Channeled Attention and Stable Errors."
Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus. 2020. "Five Facts about Beliefs and Portfolios."
Gillen, Ben, Erik Snowberg, and Leeat Yariv. 2019. "Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study." *Journal of Political Economy* 127 (4): 1826–63.
Gneezy, Uri, Alex Imas, and John List. 2015. "Estimating Individual Ambiguity Aversion: A Simple Approach."
Jensen, Arthur Robert. 1998. *The g Factor: The Science of Mental Ability*. Praeger.
Le Yaouanq, Yves, and Peter Schwardmann. 2020. "Learning about One's Self."

Levy, Matthew, and Joshua Tasoff. 2016. “Exponential-Growth Bias and Lifecycle Consumption.” *Journal of the European Economic Association* 14 (3): 545–83.
Lichand, Guilherme, and Anandi Mani. 2019. “Cognitive Droughts.”
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.
Mata, Rui, Renato Frey, David Richter, Jürgen Schupp, and Ralph Hertwig. 2018. “Risk Preference: A View from Psychology.” *Journal of Economic Perspectives* 32 (2): 155–72.
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, Armando. 2019. “Emotions, Risk Attitudes, and Patience.”
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, and Eldar Shafir. 2013. *Scarcity: Why Having Too Little Means so Much*. New York: Times Books, Henry Holt and Company.
Rabin, Matthew, and Georg Weizsäcker. 2009. “Narrow Bracketing and Dominated Choices.” *American Economic Review* 99 (4): 1508–43.
Rammstedt, Beatrice, and Oliver P. John. 2007. “Measuring Personality in One Minute or Less: A 10-Item Short Version of the Big Five Inventory in English and German.” *Journal of Research in Personality* 41 (1): 203–12.
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.
Schildberg-Hörisch, Hannah. 2018. “Are Risk Preferences Stable?” *Journal of Economic Perspectives* 32 (2): 135–54.
Stango, Victor, and Jonathan Zinman. 2009. “Exponential Growth Bias and Household Finance.” *The Journal of Finance* 64 (6): 2807–49.
———. 2020a. “Behavioral Common Factors, Financial Decisions and Consumer Well-Being.”
———. 2020b. “We Are All Behavioral, More or Less: Measuring and Using Consumer-Level Behavioral Summary Statistics.”
Stigler, George, and Gary Becker. 1977. “De Gustibus Non Est Disputandum.” *American Economic Review* 67: 76–90.
Tasoff, Joshua, and Wenjie Zhang. 2020. “The Performance of Time-Preference and Risk-Preference Measures in Surveys.”
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.



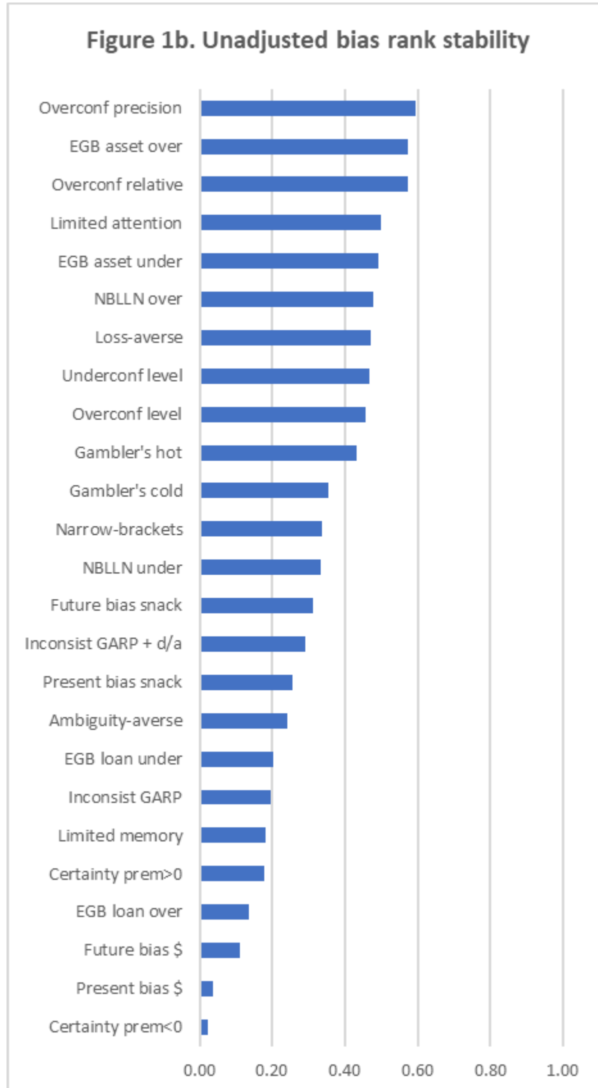
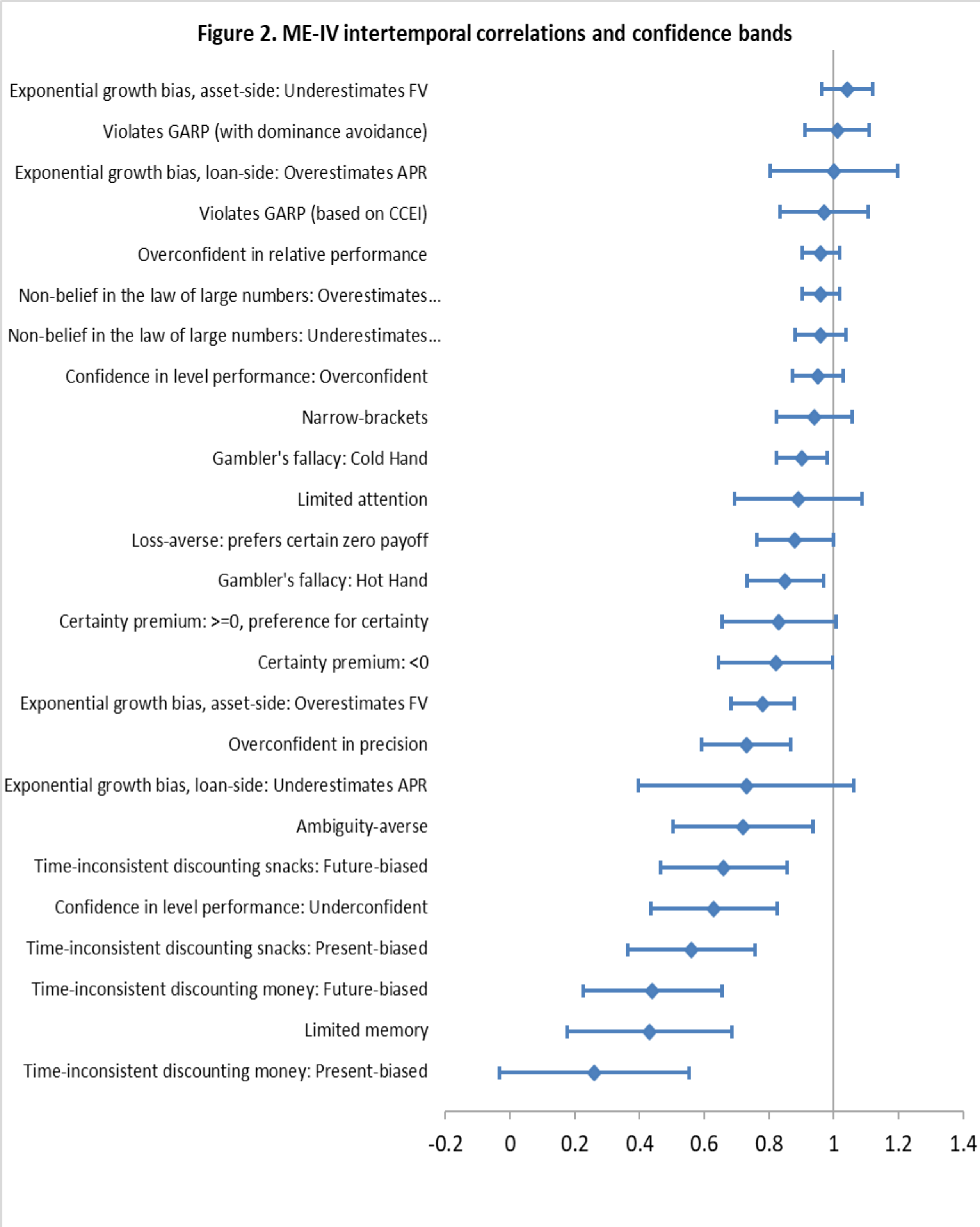


Figure 1a reproduces the proportion of panelists with unchanged bias indicator across our two rounds, measured three years apart, from Table 3 Column 1, ordering the biases from highest proportion unchanged to lowest. Figure 1b does the same for the unadjusted within-person intertemporal rank correlations in Table 3 Column 2.



Correlation point estimates are indicated with diamonds and identical to those in Table 6 Column 2. Bands span the 95% confidence interval.

Table 1. Bias elicitations, definitions, and measurements

Potential source of bias: key antecedent(s) for elicitation design	Elicitation method description	*Bias indicator(s)*	Bias rank(s): unique values	Related bias group
(1)	(2)	(3)	(4)	(5)
Time inconsistent discounting of money: Andreoni & Sprenger (2012), Barcellos & Carvalho (2014)	Convex Time Budget. 24 decisions allocating 100 tokens each between smaller-sooner and larger-later amounts; decisions pose varying start dates (today vs. 5 weeks from today), delay lengths (5 or 9 weeks) & savings yields.	*Present-biased: discounts more when sooner date is today*	28	Discounting
		Future-biased: discounts more when sooner date is 5 weeks from tdy	37	Discounting
Time inconsistent discounting of consumption: Read & van Leeuwen (1998), Barcellos & Carvalho (2014)	Two decisions between two snacks: healthier/less-delicious vs. less healthy/more delicious. Decisions vary only in date snack is delivered: now, or 5 weeks from now.	*Present-biased: choose less healthy now, healthy 5 weeks from now*	2	Discounting
		Future-biased: choose healthy now, less healthy 5 weeks from now	2	Discounting
Violates General Axiom of Revealed Preference (GARP) and dominance avoidance: Choi et al (2014)	Decisions from 11 different linear budget constraints under risk. Subjects choose a point on the line, and then the computer randomly chooses whether to pay the point value of the x-axis or the y-axis.	*Violates GARP: potential earnings wasted per CCEI>0*	54	Decision quality
		Violates GARP and dominance avoidance: potential earnings wasted per combined-CCEI>0	97	Decision quality
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.*	5	Decision quality
Preference for (un)certainty Callen et al (2014)	2 screens of 10 choices each between two lotteries, one a (p, 1-p) gamble over X and Y > X , (p; X, Y), the other a (q, 1-q) gamble over Y and 0, (q; Y, 0). Y=$450, X=$150, q ϵ[0.1, 1.0], p=0.5 on one screen and 1.0 on the other.	*Preference for certainty: certainty premium (CP) >0*	20	Preferences re: uncertainty
		Cumulative prospect theory: certainty premium (CP)<0	14	Preferences re: uncertainty
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 $0 payoff in one or more choices.*	4	Preferences re: uncertainty
Ambiguity aversion: Dimmock et al. (2016)	Two questions re: a game where win $500 if pick green ball. 1. Choose between bag with 45 green-55 yellow and bag with unknown mix. 2. If chose 45-55 bag, how many green balls in 45-55 bag would induce switch.	*Ambiguity Aversion: prefers bag with 45 green to bag with unknown mix.*	17	Preferences re: uncertainty
Over- or under-confidence in performance: Larrick et al (2007), Moore & Healy (2008)	"How many of the last 3 questions (the ones on the disease, the lottery and the savings account) do you think you got correct?"	*Overconfidence in perform: self-assessment > actual score*	4	Confidence
		Underconfidence in perform: self-assessment < actual score	3	Confidence
Overconfidence in precision: Larrick et al (2007), Moore & Healy (2008)	Questions about about likelihoods of different numeracy quiz scores and future income increases.	*Overconfidence in precision: responds 100% to one or both questions*	3	Confidence
Overconfidence in relative performance: Larrick et al (2007), Moore & Healy (2008)	"… what you think about your intelligence as it would be measured by a standard test. How do you think your performance would rank, relative to all of the other ALP members who have taken the test?"	*Greater diff between self-assessed and actual rank indicates more overconfidence. "Overconfident" = overconfidence above median.*	78	Confidence
Non-belief in the law of large numbers (NBLLN): Benjamin, Moore, and Rabin (2013)	Question re: percent chances that, among 1,000 coin flips, the # of heads will fall in ranges [0, 480], [481, 519], and [520, 1000]. NBLLN = distance between response for [481, 519] and 78.	*Overestimates convergence to 50-50: responds with>78%*	21	Math biases
		Underestimates convergence to 50-50: responds with<78%	5	Math biases
Gambler's fallacies: Benjamin, Moore, and Rabin (2013)	"Imagine that we had a computer "flip" a fair coin… 10 times. The first 9 are all heads. What are the chances, in % terms, that the 10th flip will be a head?"	*Gambler's (Cold-hand) fallacy: responds with<50%*	10	Math biases
		Hot-hand fallacy: responds with>50%	8	Math biases
Exponential growth bias (EGB), debt-side: Stango & Zinman (2009; 2011)	Survey first elicits monthly payment respondent would expect to pay on a $10,000, 48 month car loan (this response defines the actual annual percentage rate). Then elicits perceived APR implied by that payment.	*Underestimates EG: actual APR>perceived APR*	50	Math biases
		Overestimates EG: actual APR<perceived APR	34	Math biases
Exponential growth bias (EGB), asset-side: Banks et al (2007)	Elicits perceived future value (FV) of $200, earning 10% annual, after two years.	*Underestimates EG: perceived FV<actual FV=$242*	11	Math biases
		Overestimates EG: perceived FV>actual FV=$242	7	Math biases
Limited attention: Author-developed	Four questions re: whether subject's finances would improve with more attention given the opportunity cost of attention, with questions varying the types of decisions: day-to-day, medium-run, long-run, or choosing financial products/services.	*Limited attention: Indicates regret about paying too little attention given opportunity cost of attention, on one or more of the four questions*	5	Limited attention/memory
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."	*Limited memory: Says will complete task but does not complete.*	2	Limited attention/memory

The Data Appendix Section 1 provides additional details on measuring each behavioral bias. "CCEI" = Critical Cost Efficiency Index. "ALP"=American Life Panel. Unique values is the maximum across our two survey rounds.

Table 2. Behavioral bias stability in the aggregate: Prevalence estimates from the same sample, 3 years apart

Panel A. Within Bias	Round 1	Round 2	N Round 1	N Round 2
Potential source of bias and/or *Bias indicator*	(1)	(2)	(3)	(4)
Time-inconsistent discounting money: *Present-biased*	0.26	0.25	803	819
Time-inconsistent discounting money: *Future-biased*	0.37	0.37		
Time-inconsistent discounting snacks: *Present-biased*	0.14	0.17	835	829
Time-inconsistent discounting snacks: *Future-biased*	0.07	0.06		
Violates GARP (based on CCEI)	0.50	0.50	774	799
Violates GARP (with dominance avoidance)	0.95	0.96		
Narrow-brackets	0.59	0.62	827	837
Preference re certainty: *Certainty premium > 0*	0.77	0.76	620	620
Preference re certainty: *Certainty premium < 0*	0.23	0.24		
Loss-averse	0.64	0.60	843	845
Ambiguity-averse	0.74	0.76	842	821
Mis-confidence in level performance: *Overconfident*	0.36	0.34	829	813
Mis-confidence in level performance: *Underconfident*	0.12	0.13		
Overconfident in precision	0.45	0.45	793	775
Overconfident in relative performance			844	818
Non-belief in the law of large numbers: *Underestimates convergence*	0.85	0.85	833	819
Non-belief in the law of large numbers: *Overestimates convergence*	0.15	0.15		
Gambler's fallacy: *Cold Hand*	0.26	0.24	842	817
Gambler's fallacy: *Hot Hand*	0.12	0.10		
Exponential growth bias, loan-side: *Underestimates APR*	0.51	0.53	778	783
Exponential growth bias, loan-side: *Overestimates APR*	0.34	0.34		
Exponential growth bias, asset-side: *Underestimates future value*	0.44	0.41	761	735
Exponential growth bias, asset-side: *Overestimates future value*	0.08	0.06		
Limited attention	0.45	0.44	832	829
Limited memory	0.82	0.76	825	803

Panel B. Across biases			
Mean prevalence	0.43		0.42
Mean difference across rounds		0.00	
Mean |difference| across rounds		0.02	
Max |difference| across rounds		0.07	

Each results cell in Panel A presents an estimate of the proportion of our sample exhibiting any amount of the bias indicated in the row label. Results are similar using the ALPs sampling weights. Our full sample is 845 respondents, with the lower sample sizes for bias-round measures in Columns 3 and 4 due mostly to item non-response (although the certainty premium can be calculated only for respondents without multiple switch points.) Biases measured using the same elicitation have the same sample size within-round. Relative overconfidence level is not identified in our elicitation. Please see Table 1 for summary of bias measures, and the Data Appendix Section 1 for details. Cross-bias statistics in Panel B are unweighted.

Table 3. How temporally stable are biases within-person, before allowing for measurement error?

Panel A. Within Bias	Across-round: Share with unchanged bias indicator	Across-round: Rank correlation, no measm't error adjustment	
Potential source of bias and/or *Bias*	(1)	(2)	N
Time-inconsistent discounting money: *Present-bias*	0.64	0.04	782
Time-inconsistent discounting money: *Future-bias*	0.59	0.11	
Time-inconsistent discounting snacks: *Present-bias*	0.77	0.26	821
Time-inconsistent discounting snacks: *Future-bias*	0.89	0.31	
GARP violation (based on CCEI)	0.58	0.20	739
GARP violation (with dominance avoidance)	0.93	0.29	
Narrow-bracketing	0.65	0.33	820
Preference re certainty: *Certainty premium > 0*	0.65	0.18	514
Preference re certainty: *Certainty premium < 0*		0.02	
Loss-aversion	0.69	0.47	843
Ambiguity-aversion	0.72	0.24	818, 711
Mis-confidence in level performance: *Overconfidence*	0.69	0.46	799
Mis-confidence in level performance: *Underconfidence*	0.84	0.47	
Overconfidence in precision	0.72	0.59	748
Overconfidence in relative performance	0.71	0.57	817
Non-belief in the law of large #s: *Underestimation of convergence*	0.82	0.33	809
Non-belief in the law of large #s: *Overestimation of convergence*	0.82	0.48	
Gambler's fallacy: *Cold Hand*	0.76	0.35	815
Gambler's fallacy: *Hot Hand*	0.86	0.43	
Exponential growth bias, loan-side: *Underestimation of APR*	0.59	0.20	737
Exponential growth bias, loan-side: *Overestimation of APR*	0.60	0.13	
Exponential growth bias, asset-side: *Underestimation of future value*	0.78	0.49	688
Exponential growth bias, asset-side: *Overestimation of future value*	0.91	0.57	
Limited attention	0.68	0.50	819
Limited memory	0.69	0.18	788
Panel B. Across biases			
Min	0.58	0.02	
Max	0.93	0.59	
Mean	0.73	0.33	

Correlations in Col 2 are Pearson or tetrachoric. Standard errors on correlations are not displayed here, for concision (they are displayed in Table 6 Column 1); all but two SEs imply p-values<0.01, with the exceptions being present-biased money discounting (SE=0.03) and certainty premium <0 (SE=0.04). N is the number of panelists with nonmissing observations in both of our two survey rounds, administered three years apart. Certainty premium indicator stability is not separately identified for the two directional biases, and its N is lower because CP can be calculated only for respondents without multiple switch points. Ambiguity aversion N for rank correlation lower because of internally inconsistent responses. Lacking an absolute measure of overconfidence in performance, we define its bias indicator as confidence>median. Please see Table 1 for summary of bias measures, and the Data Appendix Section 1 for details. Cross-bias statistics in Panel B are unweighted.

Table 4. How stable are biases relative to classical characteristics?

Within-person correlations, no adjustment for measurement error		
	Rank correlation	N
Single behavioral biases	Table 3 Col (2)	Table 3 last column
Mean correlation across biases	0.33	
Patience	0.29	782
Risk-aversion: Lifetime income gamble	0.38	835
Risk-aversion: Financial risk-taking scale	0.59	821
Cognitive skills: Number series	0.75	819
Cognitive skills: Numeracy	0.56	802
Cognitive skills: Financial literacy	0.81	821
Cognitive skills: Stroop	0.48	773

Classical characteristic correlations estimated using Pearson, within-person across our two rounds of data collection 3 years apart. P-values all <0.01. We lack stability estimates for personality traits in our data because we only measured them in Round 2, but Anusic and Schimmack's (2016) meta-analysis of personality measures finds a precisely estimated average 3-year within-person rank correlation of about 0.6.

Table 5. Behavioral biases are poorly explained by other consumer characteristics

	Cross-round mean of R-squared					
	Unadj. R-sq	Adjusted R-sq				
	All variables		Subsets of variables			
Panel A. Within Bias			Demos	Risk & Time	Cog skills	Personality
Potential source of bias and/or *Bias*	(1)	(2)	(3)	(4)	(5)	(6)
Time-inconsistent discounting money: *Present-bias*	0.13	0.01	0.00	0.00	0.02	0.00
Time-inconsistent discounting money: *Future-bias*	0.13	0.01	0.01	0.00	0.01	-0.01
Time-inconsistent discounting snacks: *Present-bias*	0.13	0.00	0.00	0.00	0.01	0.00
Time-inconsistent discounting snacks: *Future-bias*	0.11	-0.01	0.01	0.00	-0.01	-0.01
GARP violation (based on CCEI)	0.15	0.03	0.03	-0.01	0.03	0.01
GARP violation (with dominance avoidance)	0.21	0.10	0.06	0.00	0.08	0.02
Narrow-bracketing	0.16	0.04	0.02	0.01	0.03	0.01
Preference re certainty: *Certainty premium > 0*	0.16	0.00	0.01	0.00	0.03	-0.02
Preference re certainty: *Certainty premium < 0*	0.15	-0.01	0.01	0.00	0.01	-0.02
Loss-aversion	0.18	0.07	0.03	0.05	0.02	0.00
Ambiguity-aversion	0.15	0.03	0.02	0.00	0.01	0.00
Mis-confidence in level performance: *Overconfidence*	0.22	0.12	0.06	0.03	0.09	0.00
Mis-confidence in level performance: *Underconfidence*	0.15	0.04	0.02	0.00	0.03	0.00
Overconfidence in precision	0.19	0.07	0.04	0.02	0.01	0.02
Overconfidence in relative performance	0.30	0.22	0.11	0.07	0.16	0.03
Non-belief in the law of large #s: *Underestimation of convergence*	0.23	0.12	0.08	0.02	0.10	0.02
Non-belief in the law of large #s: *Overestimation of convergence*	0.19	0.07	0.06	0.00	0.06	0.02
Gambler's fallacy: *Cold Hand*	0.23	0.12	0.08	0.03	0.10	0.00
Gambler's fallacy: *Hot Hand*	0.16	0.04	0.02	0.00	0.03	0.00
Exponential growth bias, loan-side: *Underestimation of APR*	0.12	-0.01	0.00	-0.01	0.00	0.00
Exponential growth bias, loan-side: *Overestimation of APR*	0.16	0.03	0.03	0.00	0.03	0.01
Exponential growth bias, asset-side: *Underestimation of future value*	0.31	0.20	0.10	0.03	0.18	0.02
Exponential growth bias, asset-side: *Overestimation of future value*	0.20	0.08	0.03	0.01	0.05	0.02
Limited attention	0.18	0.07	0.03	0.01	0.01	0.02
Limited memory	0.14	0.02	0.01	0.00	0.01	0.01
Panel B. Across biases						
Mean across biases	0.18	0.06	0.04	0.01	0.04	0.01
Mean across non-binary bias measures	0.18	0.07	0.04	0.01	0.05	0.01

Each cell shows the unweighted mean R-squared or adjusted R-squared across two OLS regressions, one per round, of the bias rank described in the row label on flexibly parameterized measures (one bin per response value or decile) of: plausibly exogenous demographics (education, age, gender, immigration, and race and ethnicity), patience and risk aversion (the risk&time column refers to classical measures of risk aversion and patience), cognitive skills, and personality traits. Money discounting bias regressions drop patience, level overconfidence regressions drop numeracy, and the relative overconfidence regression drops number series (fluid intelligence) so that we are not overfitting by using RHS variables created from same elicitation as the bias measure on the LHS. We only have binary measures of biased discounting of snacks, and of limited memory, and we exclude those measures from the second cross-bias average (reported in the last row). 845 observations per round, but sample sizes are lower than 845 here due to item non-response for biases; see Table 2 for bias-round sample sizes. Cross-bias statistics in Panel B are unweighted.

Table 6. Unadjusted vs. measurement error instrumental variables (ME-IV) estimates of temporal stability

Panel A. Within Bias	Unadjusted: Same as Table 3 Col 2		ME-IV instrument set: "All other" biases		ME-IV instrument set: "Related" biases only	
Potential source of bias and/or *Bias*	(1)		(2)		(3)	
Time-inconsistent discounting money: *Present-bias*	0.04	(0.03)	0.26	(0.15)	-0.02	(0.57)
Time-inconsistent discounting money: *Future-bias*	0.11	(0.03)	0.44	(0.11)	0.04	(0.46)
Time-inconsistent discounting snacks: *Present-bias*	0.26	(0.07)	0.56	(0.10)	-0.22	(0.39)
Time-inconsistent discounting snacks: *Future-bias*	0.31	(0.10)	0.66	(0.10)	0.24	(0.61)
GARP violation (based on CCEI)	0.20	(0.03)	0.97	(0.07)	0.91	(0.36)
GARP violation (with dominance avoidance)	0.29	(0.02)	1.01	(0.05)	1.25	(0.44)
Narrow-bracketing	0.33	(0.03)	0.94	(0.06)	1.10	(0.33)
Preference re certainty: *Certainty premium > 0*	0.18	(0.03)	0.83	(0.09)	0.78	(0.47)
Preference re certainty: *Certainty premium < 0*	0.02	(0.04)	0.82	(0.09)	0.83	(0.25)
Loss-aversion	0.47	(0.04)	0.88	(0.06)	1.20	(0.19)
Ambiguity-aversion	0.24	(0.03)	0.72	(0.11)	0.81	(0.28)
Mis-confidence in level performance: *Overconfidence*	0.46	(0.04)	0.95	(0.04)	0.84	(0.08)
Mis-confidence in level performance: *Underconfidence*	0.47	(0.07)	0.63	(0.10)	0.66	(0.15)
Overconfidence in precision	0.59	(0.03)	0.73	(0.07)	0.61	(0.14)
Overconfidence in relative performance	0.57	(0.02)	0.96	(0.03)	0.95	(0.07)
Non-belief in the law of large #s: *Underestimation of convergence*	0.33	(0.02)	0.96	(0.04)	0.96	(0.06)
Non-belief in the law of large #s: *Overestimation of convergence*	0.48	(0.06)	0.96	(0.03)	1.06	(0.06)
Gambler's fallacy: *Cold Hand*	0.35	(0.02)	0.90	(0.04)	0.87	(0.07)
Gambler's fallacy: *Hot Hand*	0.43	(0.07)	0.85	(0.06)	0.99	(0.12)
Exponential growth bias, loan-side: *Underestimation of APR*	0.20	(0.03)	0.73	(0.17)	0.68	(0.70)
Exponential growth bias, loan-side: *Overestimation of APR*	0.13	(0.03)	1.00	(0.10)	1.14	(0.18)
Exponential growth bias, asset-side: *Underestimation of future value*	0.49	(0.02)	1.04	(0.04)	0.97	(0.08)
Exponential growth bias, asset-side: *Overestimation of future value*	0.57	(0.09)	0.78	(0.05)	0.98	(0.01)
Limited attention	0.50	(0.04)	0.89	(0.10)	1.17	(0.46)
Limited memory	0.18	(0.07)	0.43	(0.13)	0.55	(0.44)

Panel B. Across biases

	(1)	(2)	(3)
Min	0.02	0.26	-0.22
Max	0.59	1.04	1.25
Mean	0.33	0.80	0.77

Cells report correlation and (standard error of correlation). Column 1 is the pairwise Pearson or tetrachoric correlation between round 1 and round 2. Column 2 is the correlation estimated by ME-IV, allowing for measurement error by stacking the data and instrumenting for round 1 (2) bias using round 1 (2) values of "all other" biases except for mechanically-related ones (see Section 3 for details). Column 3 uses only non-mechanically-related "related biases" in the instrument set. See Table 7 for lists of "All other" and "Related" bias IV sets for each bias. Cross-bias statistics in Panel B are unweighted.

Table 7. Tests of IV strength and exogeneity

	P-value of first stage		P-value of J-test			Details on Instrument Sets		
	All biases	Related only	All biases		Related only	Bias number	All	Related
Includes mechanically-related biases?	No	No	No	Yes	No		No	No
Potential source of bias and/or *Bias*	(1)	(2)	(3)	(4)	(5)			
Time-inconsistent discounting money: *Present-bias*	0.18	0.32	0.01		0.19	1	Not 1, 2	3, 4
Time-inconsistent discounting money: *Future-bias*	0.00	0.22	0.01		0.01	2	Not 1, 2	3, 4
Time-inconsistent discounting snacks: *Present-bias*	0.01	0.06	0.25		0.01	3	Not 3, 4	1, 2
Time-inconsistent discounting snacks: *Future-bias*	0.00	0.50	0.30		0.05	4	Not 3, 4	1, 2
GARP violation (based on CCEI)	0.00	0.02	0.54	0.00	0.20	5	Not 5, 6	7
GARP violation (with dominance avoidance)	0.00	0.04	0.22	0.00	0.61	6	Not 5, 6	7
Narrow-bracketing	0.00	0.01	0.57		0.38	7	Not 7	5, 6
Preference re certainty: *Certainty premium > 0*	0.00	0.00	0.71		0.35	8	Not 8, 9	10,11,12
Preference re certainty: *Certainty premium < 0*	0.00	0.00	0.34		0.68	9	Not 8, 9	10,11,12
Loss-aversion	0.00	0.00	0.10		0.40	10	Not 10	8, 9, 11
Ambiguity-aversion	0.02	0.00	0.33		0.48	11	Not 11	8, 9, 10
Mis-confidence in level performance: *Overconfidence*	0.00	0.00	0.13	0.00	0.80	12	Not 12, 13, 22, 23	14, 15
Mis-confidence in level performance: *Underconfidence*	0.00	0.00	0.20	0.07	0.12	13	Not 12, 13, 22, 23	14, 15
Overconfidence in precision	0.00	0.00	0.02		0.15	14	Not 14	12, 13, 15
Overconfidence in relative performance	0.00	0.00	0.78		0.36	15	Not 15	12, 13, 14
Non-belief in the law of large #s: *Underestimation of convergence*	0.00	0.00	0.58		0.41	16	Not 16, 17	18, 19, 20, 21, 22, 23
Non-belief in the law of large #s: *Overestimation of convergence*	0.00	0.00	0.14		0.49	17	Not 16, 17	18, 19, 20, 21, 22, 23
Gambler's fallacy: *Cold Hand*	0.00	0.00	0.91		0.39	18	Not 18, 19	16, 17, 20, 21, 22, 23
Gambler's fallacy: *Hot Hand*	0.00	0.01	0.48		0.53	19	Not 18, 19	16, 17, 20, 21, 22, 23
Exponential growth bias, loan-side: *Underestimation of APR*	0.17	0.99	0.14		0.29	20	Not 20, 21	16, 17, 18, 19, 22, 23
Exponential growth bias, loan-side: *Overestimation of APR*	0.00	0.01	0.39		0.66	21	Not 20, 21	16, 17, 18, 19, 22, 23
Exponential growth bias, asset-side: *Underestimation of future value*	0.00	0.00	0.47	0.26	0.32	22	Not 12, 13, 22, 23	16, 17, 18, 19, 20, 21
Exponential growth bias, asset-side: *Overestimation of future value*	0.00	0.04	0.26	0.25	0.20	23	Not 12, 13, 22, 23	16, 17, 18, 19, 20, 21
Limited attention	0.00	0.08	0.29		0.79	24	Not 24	25
Limited memory	0.00	0.00	0.06		0.08	25	Not 25	24

P-values for the first stage are for an F-test of the joint significance of the instrument set. P-values for the J-test test the null hypothesis that in the set of N instruments, N-1 are also exogenous given that one of the instruments is exogenous. Column 4 is a falsification test of sorts, in the sense that one would be concerned about the J-test's identifying assumption if including mechanically-related biases other than the mirror image of any bi-directional bias did NOT increase the likelihood of rejection. Specifically, in Column 4 we include: the other GARP bias as an additional IV for each GARP bias, the EGB future value biases as additional IVs for each level confidence bias, and the level confidence biases as additional IVs for each EGB future value bias.

Table 8a. Limiting the sample to panelists with more stable household structure or economic status does not generally increase estimates of level stability

		Households that are more stable across rounds						
				Same			Smaller changes in financial condition	
	Full sample	Martial status	# hh members	Work status	Income category	Fin distress indicators	<median	<90th p
Panel A. Within Bias	N=845	N=769	N=613	N=781	N=409	N=658	N=520	N=794
Potential source of bias and/or *Bias*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Time-inconsistent discounting money: *Present-bias*	0.64	0.64	0.64	0.64	0.64	0.63	0.65	0.65
Time-inconsistent discounting money: *Future-bias*	0.59	0.59	0.59	0.59	0.59	0.60	0.60	0.60
Time-inconsistent discounting snacks: *Present-bias*	0.77	0.77	0.76	0.77	0.76	0.78	0.77	0.77
Time-inconsistent discounting snacks: *Future-bias*	0.89	0.89	0.89	0.89	0.88	0.89	0.89	0.89
GARP violation (based on CCEI)	0.58	0.58	0.58	0.58	0.57	0.58	0.60	0.58
GARP violation (with dominance avoidance)	0.93	0.93	0.92	0.93	0.92	0.92	0.92	0.93
Narrow-bracketing	0.65	0.66	0.65	0.65	0.67	0.65	0.64	0.66
Preference re certainty: *Certainty premium > 0* Preference re certainty: *Certainty premium < 0*	0.65	0.65	0.66	0.65	0.63	0.65	0.65	0.64
Loss-aversion	0.69	0.69	0.68	0.69	0.68	0.69	0.69	0.69
Ambiguity-aversion	0.72	0.71	0.71	0.72	0.70	0.70	0.73	0.72
Mis-confidence in level performance: *Overconfidence*	0.69	0.68	0.69	0.69	0.71	0.69	0.72	0.69
Mis-confidence in level performance: *Underconfidence*	0.84	0.84	0.83	0.83	0.83	0.83	0.85	0.84
Overconfidence in precision	0.72	0.72	0.72	0.72	0.74	0.73	0.74	0.73
Overconfidence in relative performance	0.71	0.71	0.71	0.71	0.69	0.72	0.72	0.72
Non-belief in the law of large #s: *Underestimation of convergence*	0.82	0.83	0.82	0.82	0.79	0.82	0.82	0.82
Non-belief in the law of large #s: *Overestimation of convergence*	0.82	0.82	0.81	0.82	0.79	0.82	0.82	0.82
Gambler's fallacy: *Cold Hand*	0.76	0.76	0.76	0.76	0.79	0.77	0.76	0.76
Gambler's fallacy: *Hot Hand*	0.86	0.86	0.85	0.86	0.87	0.86	0.87	0.86
Exponential growth bias, loan-side: *Underestimation of APR*	0.59	0.58	0.59	0.58	0.59	0.59	0.59	0.59
Exponential growth bias, loan-side: *Overestimation of APR*	0.60	0.59	0.60	0.59	0.60	0.60	0.61	0.60
Exponential growth bias, asset-side: *Underestimation of future value*	0.78	0.77	0.79	0.78	0.78	0.78	0.80	0.78
Exponential growth bias, asset-side: *Overestimation of future value*	0.91	0.91	0.91	0.91	0.92	0.91	0.92	0.91
Limited attention	0.68	0.68	0.68	0.69	0.70	0.68	0.68	0.69
Limited memory	0.69	0.69	0.70	0.69	0.72	0.67	0.66	0.69
Panel B. Across Biases								
Min	0.58	0.58	0.58	0.58	0.57	0.58	0.59	0.58
Max	0.93	0.93	0.92	0.93	0.92	0.92	0.92	0.93
Mean	0.73	0.73	0.73	0.73	0.73	0.73	0.74	0.73

Each cell presents the proportion with bias indicator unchanged across the our two survey rounds three years apart. Column 1 is reproduced from Table 3 Column 1. The ALP's measure of income is categorical: 17 bins. Sample in Column 6 is those with same value of our severe financial distress indicator-- indicating >=1 of 4 measures of severe measures of financial distress-- in both rounds. Financial condition is an index of indicators of postive net worth, positive retirement assets, holding equities, having a positive savings rate over the prior 12 months, and not having severe financial hardship during the prior 12 months. It is correlated 0.79 within person across rounds. Certainty premium indicator stability is not separately identified for the two directional biases. Lacking an absolute measure of overconfidence in performance, we define its bias indicator as confidence>median. Please see Data Appendix Section 1 for additional details on bias measures. Cross-bias statistics in Panel B are unweighted.

Table 8b. Limiting the sample to panelists with more stable household structure or economic status does not generally increase estimates of unadjusted rank stability

		Households that are more stable across rounds							
		Same						Smaller changes in financial condition	
	Full sample	Martial status	# hh members	Work status	Income category	Fin distress indicators		<median	<90th p
Panel A. Within Bias	N=845	N=769	N=613	N=781	N=409	N=658		N=520	N=794
Potential source of bias and/or *Bias*	(1)	(2)	(3)	(4)	(5)	(6)		(7)	(8)
Time-inconsistent discounting money: *Present-bias*	0.04	0.05	0.04	0.04	0.09	0.02		0.03	0.07
Time-inconsistent discounting money: *Future-bias*	0.11	0.11	0.10	0.12	0.09	0.11		0.09	0.13
Time-inconsistent discounting snacks: *Present-bias*	0.26	0.25	0.16	0.25	0.11	0.30		0.25	0.27
Time-inconsistent discounting snacks: *Future-bias*	0.31	0.29	0.31	0.32	0.29	0.32		0.31	0.30
GARP violation (based on CCEI)	0.20	0.21	0.19	0.20	0.21	0.19		0.23	0.19
GARP violation (with dominance avoidance)	0.29	0.29	0.26	0.28	0.26	0.27		0.36	0.29
Narrow-bracketing	0.33	0.35	0.33	0.34	0.38	0.35		0.35	0.35
Preference re certainty: *Certainty premium > 0*	0.18	0.17	0.19	0.18	0.19	0.21		0.18	0.19
Preference re certainty: *Certainty premium < 0*	0.02	0.01	0.02	0.02	-0.01	0.03		0.00	0.01
Loss-aversion	0.47	0.47	0.46	0.48	0.45	0.46		0.45	0.47
Ambiguity-aversion	0.24	0.23	0.18	0.24	0.19	0.24		0.30	0.25
Mis-confidence in level performance: *Overconfidence*	0.46	0.44	0.44	0.47	0.51	0.45		0.48	0.45
Mis-confidence in level performance: *Underconfidence*	0.47	0.47	0.44	0.45	0.44	0.44		0.51	0.48
Overconfidence in precision	0.59	0.59	0.60	0.60	0.62	0.62		0.61	0.60
Overconfidence in relative performance	0.57	0.58	0.58	0.57	0.56	0.56		0.57	0.58
Non-belief in the law of large #s: *Underestimation of convergence*	0.33	0.34	0.31	0.34	0.35	0.35		0.36	0.32
Non-belief in the law of large #s: *Overestimation of convergence*	0.48	0.49	0.45	0.47	0.38	0.48		0.52	0.47
Gambler's fallacy: *Cold Hand*	0.35	0.37	0.36	0.36	0.40	0.35		0.35	0.36
Gambler's fallacy: *Hot Hand*	0.43	0.44	0.39	0.42	0.54	0.47		0.50	0.41
Exponential growth bias, loan-side: *Underestimation of APR*	0.20	0.18	0.19	0.19	0.22	0.19		0.20	0.20
Exponential growth bias, loan-side: *Overestimation of APR*	0.13	0.12	0.14	0.12	0.13	0.13		0.15	0.13
Exponential growth bias, asset-side: *Underestimation of future value*	0.49	0.48	0.51	0.50	0.49	0.49		0.52	0.50
Exponential growth bias, asset-side: *Overestimation of future value*	0.57	0.59	0.61	0.61	0.62	0.57		0.54	0.56
Limited attention	0.50	0.49	0.50	0.51	0.51	0.48		0.51	0.52
Limited memory	0.18	0.18	0.21	0.18	0.30	0.13		0.11	0.18
Panel B. Across Biases									
Min	0.02	0.01	0.02	0.02	-0.01	0.02		0.00	0.01
Max	0.59	0.59	0.61	0.61	0.62	0.62		0.61	0.60
Mean	0.33	0.33	0.32	0.33	0.33	0.33		0.34	0.33

Each cell presents a Pearson or tetrachoric correlation, estimated within-person across our two survey rounds administered three years apart. Column 1 is reproduced from Table 3 Column 2. The ALP's measure of income is categorical: 17 bins. Sample in Column 6 is those with same value of our severe financial distress indicator-- indicating >=1 of 4 measures of severe measures of financial distress-- in both rounds. Financial condition is an index of indicators of postive net worth, positive retirement assets, holding equities, having a positive savings rate over the prior 12 months, and not having severe financial hardship during the prior 12 months. It is correlated 0.79 within person across rounds. Please see Data Appendix Section 1 for additional details on bias measures. Cross-bias statistics in Panel B are unweighted.

Table 9a. Limiting the sample to panelists with more stable subjective well-being does not generally increase estimates of level stability

		Panelists who are more stable across rounds				
	Full sample	Change in subjective financial condition		Change in life satisfaction		No change in expected financial condition
		None	<Largest	<median	<90th p	
Panel A. Within Bias	N=845	N=357	N=711	N=464	N=752	N=515
Potential source of bias and/or *Bias*	(1)	(2)	(3)	(4)	(5)	(6)
Time-inconsistent discounting money: *Present-bias*	0.64	0.65	0.65	0.64	0.63	0.63
Time-inconsistent discounting money: *Future-bias*	0.59	0.58	0.60	0.60	0.59	0.60
Time-inconsistent discounting snacks: *Present-bias*	0.77	0.77	0.77	0.77	0.77	0.78
Time-inconsistent discounting snacks: *Future-bias*	0.89	0.88	0.89	0.90	0.89	0.90
GARP violation (based on CCEI)	0.58	0.59	0.59	0.59	0.59	0.58
GARP violation (with dominance avoidance)	0.93	0.94	0.94	0.91	0.92	0.93
Narrow-bracketing	0.65	0.64	0.65	0.63	0.65	0.67
Preference re certainty: *Certainty premium > 0* Preference re certainty: *Certainty premium < 0*	0.65	0.63	0.64	0.63	0.64	0.62
Loss-aversion	0.69	0.70	0.69	0.68	0.70	0.66
Ambiguity-aversion	0.72	0.74	0.72	0.73	0.72	0.70
Mis-confidence in level performance: *Overconfidence*	0.69	0.69	0.68	0.69	0.68	0.70
Mis-confidence in level performance: *Underconfidence*	0.84	0.83	0.84	0.85	0.83	0.84
Overconfidence in precision	0.72	0.73	0.73	0.73	0.72	0.74
Overconfidence in relative performance	0.71	0.73	0.72	0.71	0.71	0.72
Non-belief in the law of large #s: *Underestimation of convergence*	0.82	0.84	0.82	0.80	0.81	0.85
Non-belief in the law of large #s: *Overestimation of convergence*	0.82	0.83	0.82	0.80	0.81	0.85
Gambler's fallacy: *Cold Hand*	0.76	0.76	0.76	0.77	0.77	0.75
Gambler's fallacy: *Hot Hand*	0.86	0.85	0.86	0.86	0.85	0.88
Exponential growth bias, loan-side: *Underestimation of APR*	0.59	0.56	0.60	0.59	0.59	0.58
Exponential growth bias, loan-side: *Overestimation of APR*	0.60	0.54	0.59	0.60	0.59	0.61
Exponential growth bias, asset-side: *Underestimation of future value*	0.78	0.79	0.78	0.79	0.78	0.78
Exponential growth bias, asset-side: *Overestimation of future value*	0.91	0.91	0.91	0.91	0.91	0.91
Limited attention	0.68	0.69	0.68	0.71	0.70	0.69
Limited memory	0.69	0.66	0.69	0.68	0.70	0.68
Panel B. Across Biases						
Min	0.58	0.54	0.59	0.59	0.59	0.58
Max	0.93	0.94	0.94	0.91	0.92	0.93
Mean	0.73	0.73	0.73	0.73	0.73	0.74

Each cell presents the proportion with unchanged biased indicator across our two survey rounds three years apart. Column 1 is reproduced from Table 3 Column 1. Subjective financial condition is an index of measures of financial satisfaction, retirement savings adequacy, non-retirement savings adequacy, and lack of financial stress. It is correlated 0.65 within-person across rounds. Life satisfaction is measured using a standard question-- '... how satisfied are you with your life as a whole these days?'-- asked in many surveys worldwide. It is correlated 0.58 within-person across rounds. Expected financial condition is measured using 'Do you do you think that a year from now your household will be better off financially, worse off, or about the same?' Certainty premium indicator stability is not separately identified for the two directional biases. Lacking an absolute measure of overconfidence in performance, we define its bias indicator as confidence>median. Please see Data Appendix Section 1 for additional details on bias measures. Cross-bias statistics in Panel B are unweighted.

Table 9b. Limiting the sample to panelists with more stable subjective well-being does not generally increase estimates of unadjusted rank stability

		Panelists who are more stable across rounds				
		Change in subjective financial condition		Change in life satisfaction		No change in expected financial condition
	Full sample	None	<Largest	<median	<90th p	
Panel A. Within Bias	N=845	N=357	N=711	N=464	N=752	N=515
Potential source of bias and/or *Bias*	(1)	(2)	(3)	(4)	(5)	(6)
Time-inconsistent discounting money: *Present-bias*	0.04	0.07	0.06	0.06	0.02	0.02
Time-inconsistent discounting money: *Future-bias*	0.11	0.06	0.13	0.11	0.11	0.15
Time-inconsistent discounting snacks: *Present-bias*	0.26	0.29	0.27	0.18	0.25	0.33
Time-inconsistent discounting snacks: *Future-bias*	0.31	0.17	0.29	0.34	0.26	0.41
GARP violation (based on CCEI)	0.20	0.22	0.22	0.21	0.21	0.19
GARP violation (with dominance avoidance)	0.29	0.34	0.31	0.31	0.33	0.31
Narrow-bracketing	0.33	0.30	0.32	0.26	0.32	0.40
Preference re certainty: *Certainty premium > 0*	0.18	0.12	0.18	0.15	0.19	0.18
Preference re certainty: *Certainty premium < 0*	0.02	-0.04	0.02	-0.02	0.02	-0.03
Loss-aversion	0.47	0.46	0.47	0.43	0.47	0.42
Ambiguity-aversion	0.24	0.28	0.26	0.28	0.26	0.18
Mis-confidence in level performance: *Overconfidence*	0.46	0.42	0.43	0.45	0.43	0.49
Mis-confidence in level performance: *Underconfidence*	0.47	0.42	0.49	0.52	0.45	0.48
Overconfidence in precision	0.59	0.62	0.62	0.59	0.60	0.61
Overconfidence in relative performance	0.57	0.55	0.58	0.54	0.57	0.57
Non-belief in the law of large #s: *Underestimation of convergence*	0.33	0.37	0.34	0.34	0.33	0.36
Non-belief in the law of large #s: *Overestimation of convergence*	0.48	0.52	0.45	0.47	0.46	0.55
Gambler's fallacy: *Cold Hand*	0.35	0.34	0.36	0.34	0.37	0.34
Gambler's fallacy: *Hot Hand*	0.43	0.28	0.43	0.48	0.43	0.51
Exponential growth bias, loan-side: *Underestimation of APR*	0.20	0.13	0.22	0.23	0.21	0.22
Exponential growth bias, loan-side: *Overestimation of APR*	0.13	-0.02	0.13	0.09	0.09	0.14
Exponential growth bias, asset-side: *Underestimation of future value*	0.49	0.50	0.49	0.50	0.49	0.50
Exponential growth bias, asset-side: *Overestimation of future value*	0.57	0.55	0.56	0.31	0.56	0.60
Limited attention	0.50	0.54	0.48	0.54	0.53	0.50
Limited memory	0.18	0.08	0.15	0.12	0.19	0.05
Panel B. Across Biases						
Min	0.02	-0.04	0.02	-0.02	0.02	-0.03
Max	0.59	0.62	0.62	0.59	0.60	0.61
Mean	0.33	0.30	0.33	0.31	0.32	0.34

Each cell presents a Pearson or tetrachoric correlation, estimated within-person across our two survey rounds administered three years apart. Column 1 is reproduced from Table 3 Column 2. Subjective financial condition is an index of measures of financial satisfaction, retirement savings adequacy, non-retirement savings adequacy, and lack of financial stress. It is correlated 0.65 within-person across rounds. Life satisfaction is measured using a standard question-- '... how satisfied are you with your life as a whole these days?'-- asked in many surveys worldwide. It is correlated 0.58 within-person across rounds. Expected financial condition is measured using 'Do you do you think that a year from now your household will be better off financially, worse off, or about the same?' Please see Data Appendix Section 1 for additional details on bias measures. Cross-bias statistics in Panel B are unweighted.

For Online Publication: Appendices

Appendix Table 1. Behavioral biases are poorly explained by even a kitchen sink of other consumer characteristics

	Cross-round mean of R-squared									
	Unadj. R-sq	Adjusted R-squared								
	All variables		Subsets of variables							
			Demographics		Risk &	Cog skills	Personality	Own-bias	Other response	Other
Panel A. Within Bias			Plausibly	Plausibly	Time			response time	quality controls	biases
			exogenous	endogenous				decile		
Potential source of bias and/or *Bias*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Time-inconsistent discounting money: *Present-bias*	0.26	0.04	0.00	0.01	0.00	0.02	0.00	0.02	0.00	0.02
Time-inconsistent discounting money: *Future-bias*	0.28	0.06	0.01	0.01	0.00	0.01	-0.01	0.02	0.00	0.03
Time-inconsistent discounting snacks: *Present-bias*	0.29	0.07	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.06
Time-inconsistent discounting snacks: *Future-bias*	0.24	0.00	0.01	-0.01	0.00	-0.01	-0.01	0.02	0.00	0.01
GARP violation (based on CCEI)	0.36	0.15	0.03	0.01	-0.01	0.03	0.01	0.02	0.03	0.10
GARP violation (with dominance avoidance)	0.38	0.18	0.06	0.02	0.00	0.08	0.02	0.03	0.04	0.07
Narrow-bracketing	0.32	0.01	0.01	0.01	0.00	0.03	-0.02	0.01	0.01	0.01
	0.30	-0.02	0.01	0.01	0.00	0.01	-0.02	0.00	0.00	0.00
Preference re certainty: *Certainty premium > 0*										
Preference re certainty: *Certainty premium < 0*	0.31	0.10	0.03	0.02	0.05	0.02	0.00	0.01	0.01	0.01
Loss-aversion	0.33	0.12	0.02	0.03	0.01	0.03	0.01	0.00	0.03	0.05
Ambiguity-aversion	0.31	0.07	0.02	-0.01	0.00	0.01	0.00	0.00	0.01	0.02
Mis-confidence in level performance: *Overconfidence*	0.41	0.23	0.06	0.05	0.03	0.09	0.00	0.02	0.06	0.12
Mis-confidence in level performance: *Underconfidence*	0.26	0.03	0.02	-0.01	0.00	0.03	0.00	0.00	0.00	0.01
Overconfidence in precision	0.46	0.28	0.04	0.00	0.02	0.01	0.02	0.12	0.01	0.03
Overconfidence in relative performance	0.42	0.24	0.11	0.06	0.07	0.16	0.03	0.01	0.13	0.01
Non-belief in the law of large #s: *Underestimation of convergence*	0.37	0.17	0.08	0.06	0.02	0.10	0.02	0.01	0.05	0.03
Non-belief in the law of large #s: *Overestimation of convergence*	0.31	0.10	0.06	0.04	0.00	0.06	0.02	0.01	0.02	0.03
Gambler's fallacy: *Cold Hand*	0.36	0.16	0.08	0.04	0.03	0.10	0.00	0.02	0.03	0.07
Gambler's fallacy: *Hot Hand*	0.29	0.07	0.02	0.01	0.00	0.03	0.00	0.01	0.01	0.03
Exponential growth bias, loan-side: *Underestimation of APR*	0.24	-0.01	0.00	-0.01	-0.01	0.00	0.00	0.01	0.00	0.00
Exponential growth bias, loan-side: *Overestimation of APR*	0.30	0.06	0.03	0.02	0.00	0.03	0.01	0.02	0.03	0.01
Exponential growth bias, asset-side: *Underestimation of future value*	0.50	0.32	0.10	0.12	0.03	0.18	0.02	0.04	0.06	0.13
Exponential growth bias, asset-side: *Overestimation of future value*	0.33	0.09	0.03	0.03	0.01	0.05	0.02	0.01	0.06	0.01
Limited attention	0.37	0.18	0.03	0.03	0.01	0.01	0.02	0.00	0.00	0.09
Limited memory	0.30	0.07	0.01	0.00	0.00	0.01	0.01	0.01	0.00	0.05
Panel B. Across biases										
Mean across biases	0.33	0.11	0.04	0.02	0.01	0.04	0.01	0.02	0.02	0.04
Mean across non-binary bias measures	0.34	0.12	0.04	0.03	0.01	0.05	0.01	0.02	0.03	0.04

Each cell shows the unweighted mean R-squared or adjusted R-squared across two OLS regressions, one per round, of the bias rank described in the row label on categorical measures of: plausibly exogenous demographics (education, age, gender, immigration, and race and ethnicity); plausibly endogenous demographics (income, employment status, state of residence, marital status, number of household members); patience and risk aversion (the risk&time column refers to classical measures of risk aversion and patience); cognitive skills; personality traits; response time on the elicitation used to measure the LHS variable; other response quality controls: the count of missing bias measures, total time spent on the survey round, and the count of iternally inconsistent or nonsensical responses across all bias elicitations (except for consistency with GARP; there inconsistency is the object of interest); the count of bias indicators exhibited across all biases. Money discounting bias regressions drop patience, level overconfidence regressions drop numeracy, and the relative overconfidence regression drops number series (fluid intelligence) so that we are not overfitting by using RHS variables created from same elicitation as the bias measure on the LHS. We only have binary measures of biased discounting of snacks, and of limited memory, and we exclude those measures from the second cross-bias mean (reported in the last row). 845 observations per round, but sample sizes are lower than 845 here due to item non-response for biases; see Table 2 for bias-round sample sizes. Cross-bias statistics in Panel B are unweighted.

Appendix Table 2. ME-IV and unadjusted stability for groups of related biases

	Cross-bias mean rank		N of biases	Biases
IV set	All	None	in group	included
Component rank correlations from	Table 6, Col 2	Table 3, Col 2		
Bias group				
All (1)	0.80	0.33	25	All
Discounting (2)	0.48	0.18	4	Money and snack
Decision quality (3)	0.97	0.27	3	Violates GARP, narrow bracketing
Preferences re: uncertainty (4)	0.81	0.23	4	Certainty premium, loss aversion, ambiguity aversion
Confidence (5)	0.82	0.52	4	All three over/under-confidence varieties
Math (6)	0.90	0.37	8	Statistical fallacies, exponential growth biases
Limited attention/memory (7)	0.66	0.34	2	Limited attention and memory
Expected direction (8)	0.81	0.33	17	Present-biases, cert prem>0, over-confidence, NBLLN underest., GF cold, EGBs underest., uni-directional biases
Non-expected direction (9)	0.77	0.32	8	Future-biases, cert prem<0, under-confidence, NBLLN overest., GF hot, EGBs overest.
Preferences, including $ discounting (10)	0.72	0.21	10	Discounting, violates GARP, cert prem, loss aversion, ambiguity aversion
Preferences, excluding $ discounting (11)	0.81	0.26	8	See directly above, without money discounting
Non-preferences (12)	0.85	0.41	15	Narrow bracketing, confidence, statistical fallacies, GF, EGBs, limited attention and memory
Non-preferences: Beliefs only (13)	0.87	0.42	12	Confidence, statistical fallacies, GF, EGBs

Cross-bias means are unweighted. Bias classifications in rows (2) through (7) are mutually exclusive, as are (8) and (9), and (10) and (12).

Please see Data Appendix Section 1 for details on bias elicitations and measures.

Appendix Table 3a. Dropping respondents with indicators of low survey effort or response quality does not generally increase estimates of level stability

	Full sample	Response quality adjustments								
		Drop fastest decile response times			Drop slowest decile response times			Drop largest diffs in resp. times		Drop if any
Panel A. Within Bias		Elicitation level	All bias qs	All other qs	Elicitation level	All bias qs	All other qs	>median	>90th p	internal inconsist
Potential source of bias and/or *Bias*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Time-inconsistent discounting money: *Present-bias*	0.64	0.63	0.63	0.64	0.64	0.63	0.63	0.65	0.64	0.68
Time-inconsistent discounting money: *Future-bias*	0.59	0.59	0.59	0.60	0.60	0.58	0.59	0.60	0.59	0.62
Time-inconsistent discounting snacks: *Present-bias*	0.77	0.77	0.77	0.77	0.78	0.77	0.78	0.79	0.78	0.74
Time-inconsistent discounting snacks: *Future-bias*	0.89	0.88	0.89	0.89	0.90	0.89	0.89	0.89	0.89	0.90
GARP violation (based on CCEI)	0.58	0.59	0.60	0.59	0.57	0.57	0.58	0.60	0.58	0.60
GARP violation (with dominance avoidance)	0.93	0.92	0.93	0.93	0.93	0.93	0.92	0.93	0.93	0.90
Narrow-bracketing	0.65	0.66	0.66	0.66	0.65	0.65	0.65	0.64	0.66	0.63
Preference re certainty: *Certainty premium > 0* Preference re certainty: *Certainty premium < 0*	0.65	0.64	0.64	0.64	0.67	0.64	0.62	0.64	0.66	0.60
Loss-aversion	0.69	0.68	0.69	0.68	0.69	0.70	0.69	0.67	0.70	0.75
Ambiguity-aversion	0.72	0.72	0.72	0.73	0.73	0.72	0.72	0.71	0.72	0.78
Mis-confidence in level performance: *Overconfidence*	0.69	0.69	0.70	0.70	0.70	0.68	0.69	0.68	0.70	0.73
Mis-confidence in level performance: *Underconfidence*	0.84	0.84	0.85	0.85	0.84	0.83	0.85	0.81	0.84	0.88
Overconfidence in precision	0.72	0.72	0.71	0.71	0.73	0.72	0.72	0.73	0.74	0.74
Overconfidence in relative performance	0.71	0.70	0.70	0.71	0.70	0.70	0.72	0.70	0.71	0.72
Non-belief in the law of large #s: *Underestimation of convergence*	0.82	0.82	0.82	0.82	0.83	0.83	0.81	0.81	0.83	0.74
Non-belief in the law of large #s: *Overestimation of convergence*	0.82	0.82	0.82	0.82	0.83	0.83	0.81	0.81	0.82	0.73
Gambler's fallacy: *Cold Hand*	0.76	0.75	0.76	0.75	0.78	0.77	0.77	0.77	0.77	0.87
Gambler's fallacy: *Hot Hand*	0.86	0.85	0.86	0.86	0.87	0.85	0.86	0.85	0.86	0.95
Exponential growth bias, loan-side: *Underestimation of APR*	0.59	0.59	0.58	0.58	0.59	0.58	0.59	0.59	0.59	0.62
Exponential growth bias, loan-side: *Overestimation of APR*	0.60	0.60	0.60	0.60	0.60	0.60	0.61	0.60	0.60	0.63
Exponential growth bias, asset-side: *Underestimation of future value*	0.78	0.76	0.77	0.77	0.79	0.77	0.79	0.81	0.79	0.79
Exponential growth bias, asset-side: *Overestimation of future value*	0.91	0.90	0.91	0.91	0.93	0.90	0.92	0.90	0.91	0.97
Limited attention	0.68	0.68	0.70	0.69	0.68	0.69	0.69	0.68	0.69	0.68
Limited memory	0.69	0.69	0.69	0.70	0.70	0.70	0.66	0.67	0.69	0.69
Panel B. Across Biases										
Min	0.58	0.59	0.58	0.58	0.57	0.57	0.58	0.59	0.58	0.60
Max	0.93	0.92	0.93	0.93	0.93	0.93	0.92	0.93	0.93	0.97
Mean	0.73	0.73	0.73	0.73	0.74	0.73	0.73	0.73	0.74	0.75

Each cell presents the proportion with bias indicator unchanged across our two survey rounds three years apart. Column 1 reproduces Table 3 Column 1. In columns 2-4 (5-7) we drop if the respondent is in the fastest (slowest) decile in -either- round. In columns 8-9 we use within-elicitation differences in response times; e.g., in the first two rows of column 8 we drop respondents with above-median difference in response times for the money discounting task across our two rounds. Internal inconsistency in Column 10 refers to behavioral bias elicitations only. Excluding the violates-GARP elicitation, where inconsistency is the object of interest, 620 of our panelists exhibit internal inconsistency in at least one round on at least one of the other five elicitations where we can identify it, leaving us with a sample of 225 for Column 10. Certainty premium indicator stability is not separately identified for the two directional biases. Lacking an absolute measure of overconfidence in performance, we define its bias indicator as confidence>median. Please see Data Appendix Section 1 for additional details on bias measures. Cross-bias statistics in Panel B are unweighted.

Appendix Table 3b. Dropping respondents with indicators of low survey effort or response quality does not generally increase estimates of unadjusted rank stability

	Full sample	Response quality adjustments								
		Drop fastest decile response times			Drop slowest decile response times			Drop largest diffs in resp. times		Drop if any
Panel A. Within Bias		Elicitation level	All bias qs	All other qs	Elicitation level	All bias qs	All other qs	>median	>90th p	internal inconsist
Potential source of bias and/or *Bias*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Time-inconsistent discounting money: *Present-bias*	0.04	0.04	0.04	0.04	0.01	0.00	0.01	-0.04	0.02	-0.06
Time-inconsistent discounting money: *Future-bias*	0.11	0.13	0.11	0.12	0.11	0.09	0.10	0.13	0.11	0.18
Time-inconsistent discounting snacks: *Present-bias*	0.26	0.24	0.28	0.30	0.29	0.26	0.23	0.27	0.28	0.04
Time-inconsistent discounting snacks: *Future-bias*	0.31	0.28	0.29	0.34	0.26	0.29	0.35	0.37	0.29	0.50
GARP violation (based on CCEI)	0.20	0.22	0.23	0.22	0.19	0.18	0.18	0.16	0.19	0.22
GARP violation (with dominance avoidance)	0.29	0.30	0.32	0.32	0.31	0.30	0.29	0.29	0.30	0.39
Narrow-bracketing	0.33	0.34	0.35	0.35	0.31	0.32	0.35	0.34	0.34	0.28
Preference re certainty: *Certainty premium > 0*	0.18	0.16	0.14	0.16	0.19	0.14	0.15	0.28	0.18	0.05
Preference re certainty: *Certainty premium < 0*	0.02	0.01	-0.02	-0.01	0.05	-0.01	-0.01	0.11	0.04	-0.02
Loss-aversion	0.47	0.45	0.47	0.46	0.49	0.48	0.48	0.53	0.48	0.55
Ambiguity-aversion	0.24	0.24	0.24	0.26	0.29	0.26	0.23	0.34	0.26	0.31
Mis-confidence in level performance: *Overconfidence*	0.46	0.48	0.47	0.47	0.47	0.44	0.47	0.47	0.47	0.38
Mis-confidence in level performance: *Underconfidence*	0.47	0.50	0.54	0.52	0.45	0.43	0.48	0.45	0.42	0.35
Overconfidence in precision	0.59	0.60	0.57	0.58	0.61	0.59	0.57	0.70	0.62	0.63
Overconfidence in relative performance	0.57	0.56	0.57	0.58	0.57	0.57	0.57	0.59	0.57	0.52
Non-belief in the law of large #s: *Underestimation of convergence*	0.33	0.34	0.34	0.32	0.34	0.32	0.34	0.40	0.34	0.46
Non-belief in the law of large #s: *Overestimation of convergence*	0.48	0.51	0.47	0.49	0.52	0.49	0.46	0.60	0.52	0.40
Gambler's fallacy: *Cold Hand*	0.35	0.34	0.37	0.33	0.37	0.36	0.36	0.45	0.35	0.45
Gambler's fallacy: *Hot Hand*	0.43	0.44	0.44	0.46	0.45	0.35	0.48	0.50	0.45	0.65
Exponential growth bias, loan-side: *Underestimation of APR*	0.20	0.20	0.18	0.18	0.22	0.19	0.19	0.22	0.22	0.26
Exponential growth bias, loan-side: *Overestimation of APR*	0.13	0.11	0.13	0.12	0.14	0.13	0.15	0.13	0.14	0.06
Exponential growth bias, asset-side: *Underestimation of future value*	0.49	0.44	0.47	0.48	0.53	0.48	0.50	0.59	0.53	0.62
Exponential growth bias, asset-side: *Overestimation of future value*	0.57	0.37	0.55	0.46	0.68	0.59	0.61	0.81	0.63	0.81
Limited attention	0.50	0.50	0.52	0.52	0.51	0.52	0.51	0.56	0.50	0.53
Limited memory	0.18	0.19	0.19	0.19	0.16	0.21	0.14	0.17	0.16	0.21
Panel B. Across Biases										
Min	0.02	0.01	-0.02	-0.01	0.01	-0.01	-0.01	-0.04	0.02	-0.06
Max	0.59	0.60	0.57	0.58	0.68	0.59	0.61	0.81	0.63	0.81
Mean	0.33	0.32	0.33	0.33	0.34	0.32	0.33	0.38	0.34	0.35

Each cell presents a Pearson or tetrachoric correlation of bias rank across our two survey rounds three years apart. Column 1 reproduces Table 3 Column 2. In columns 2-4 (5-7) we drop if the respondent is in the fastest (slowest) decile in -either- round. In columns 8-9 we use within-elicitation differences in response times; e.g., in the first two rows of column 8 we drop respondents with above-median difference in response times for the money discounting task across our two rounds. Internal inconsistency in Column 10 refers to behavioral bias elicitations only. Excluding the violates-GARP elicitation, where inconsistency is the object of interest, 620 of our panelists exhibit internal inconsistency in at least one round on at least one of the other five elicitations where we can identify it, leaving us with a sample of 225 for Column 10. Cross-bias statistics in Panel B are unweighted.

Appendix Table 4a. Limiting the sample to those with high cognitive skills or conscientiousness does not generally increase estimates of level stability

	Full sample	Education	Cog skills>median		Conscientiousness	
Panel A. Within Bias		>=B.A.	Round 1	Both Rounds	>median	>=median
Potential source of bias and/or *Bias*	(1)	(2)	(3)	(4)	(5)	(6)
Time-inconsistent discounting money: *Present-bias*	0.64	0.64	0.67	0.65	0.61	0.63
Time-inconsistent discounting money: *Future-bias*	0.59	0.58	0.60	0.59	0.57	0.58
Time-inconsistent discounting snacks: *Present-bias*	0.77	0.76	0.79	0.80	0.81	0.80
Time-inconsistent discounting snacks: *Future-bias*	0.89	0.91	0.90	0.88	0.90	0.88
GARP violation (based on CCEI)	0.58	0.55	0.57	0.60	0.60	0.60
GARP violation (with dominance avoidance)	0.93	0.91	0.91	0.90	0.93	0.92
Narrow-bracketing	0.65	0.65	0.64	0.65	0.63	0.63
Preference re certainty: *Certainty premium > 0* Preference re certainty: *Certainty premium < 0*	0.65	0.61	0.62	0.60	0.65	0.64
Loss-aversion	0.69	0.72	0.75	0.76	0.69	0.69
Ambiguity-aversion	0.72	0.75	0.77	0.78	0.72	0.74
Mis-confidence in level performance: *Overconfidence*	0.69	0.70	0.73	0.76	0.69	0.70
Mis-confidence in level performance: *Underconfidence*	0.84	0.84	0.87	0.88	0.85	0.85
Overconfidence in precision	0.72	0.73	0.75	0.75	0.75	0.74
Overconfidence in relative performance	0.71	0.74	0.74	0.79	0.69	0.70
Non-belief in the law of large #s: *Underestimation of convergence*	0.82	0.76	0.76	0.75	0.80	0.81
Non-belief in the law of large #s: *Overestimation of convergence*	0.82	0.76	0.76	0.75	0.80	0.81
Gambler's fallacy: *Cold Hand*	0.76	0.82	0.86	0.90	0.73	0.73
Gambler's fallacy: *Hot Hand*	0.86	0.91	0.92	0.94	0.86	0.86
Exponential growth bias, loan-side: *Underestimation of APR*	0.59	0.57	0.60	0.58	0.60	0.60
Exponential growth bias, loan-side: *Overestimation of APR*	0.60	0.60	0.64	0.63	0.62	0.62
Exponential growth bias, asset-side: *Underestimation of future value*	0.78	0.78	0.79	0.80	0.77	0.77
Exponential growth bias, asset-side: *Overestimation of future value*	0.91	0.93	0.93	0.95	0.92	0.92
Limited attention	0.68	0.72	0.72	0.71	0.67	0.67
Limited memory	0.69	0.70	0.67	0.67	0.69	0.71
Panel B. Across Biases						
Min	0.58	0.55	0.57	0.58	0.57	0.58
Max	0.93	0.93	0.93	0.95	0.93	0.92
Mean	0.73	0.73	0.75	0.75	0.73	0.73

Each cell presents the proportion with bias indicator unchanged across our two survey rounds three years apart. Column 1 reproduces Table 3 Column 1. For column 2 here, we have 394 panelists with education >=B.A. in round 1. Cognitive skills measured here using the first principal component of our four test scores (please see Data Appendix Section 2 for details). Certainty premium indicator stability is not separately identified for the two directional biases. Lacking an absolute measure of overconfidence in performance, we define its bias indicator as confidence>median. Please see Data Appendix Section 1 for additional details on bias measures. Cross-bias statistics in Panel B are unweighted.

Appendix Table 4b. Limiting the sample to those with high cognitive skills or conscientiousness does not generally increase estimated unadjusted rank stability

	Full sample	Education	Cog skills>median		Conscientiousness	
Panel A. Within Bias		>=B.A.	Round 1	Both Rounds	>median	>=median
Potential source of bias and/or *Bias*	(1)	(2)	(3)	(4)	(5)	(6)
Time-inconsistent discounting money: *Present-bias*	0.04	0.05	0.02	0.01	0.02	0.05
Time-inconsistent discounting money: *Future-bias*	0.11	0.07	0.11	0.09	0.06	0.07
Time-inconsistent discounting snacks: *Present-bias*	0.26	0.16	0.27	0.10	0.17	0.25
Time-inconsistent discounting snacks: *Future-bias*	0.31	0.30	0.36	0.26	0.17	0.17
GARP violation (based on CCEI)	0.20	0.14	0.15	0.17	0.21	0.22
GARP violation (with dominance avoidance)	0.29	0.29	0.32	0.33	0.24	0.25
Narrow-bracketing	0.33	0.35	0.36	0.38	0.30	0.32
Preference re certainty: *Certainty premium > 0*	0.18	0.14	0.10	0.08	0.14	0.16
Preference re certainty: *Certainty premium < 0*	0.02	-0.01	-0.03	-0.03	0.03	0.03
Loss-aversion	0.47	0.50	0.53	0.53	0.45	0.48
Ambiguity-aversion	0.24	0.24	0.34	0.39	0.18	0.27
Mis-confidence in level performance: *Overconfidence*	0.46	0.37	0.39	0.37	0.45	0.47
Mis-confidence in level performance: *Underconfidence*	0.47	0.40	0.54	0.53	0.52	0.51
Overconfidence in precision	0.59	0.62	0.61	0.65	0.63	0.63
Overconfidence in relative performance	0.57	0.61	0.53	0.52	0.54	0.55
Non-belief in the law of large #s: *Underestimation of convergence*	0.33	0.33	0.32	0.32	0.23	0.28
Non-belief in the law of large #s: *Overestimation of convergence*	0.48	0.41	0.41	0.40	0.42	0.45
Gambler's fallacy: *Cold Hand*	0.35	0.38	0.38	0.40	0.33	0.32
Gambler's fallacy: *Hot Hand*	0.43	0.48	0.58	0.42	0.52	0.48
Exponential growth bias, loan-side: *Underestimation of APR*	0.20	0.18	0.23	0.20	0.23	0.24
Exponential growth bias, loan-side: *Overestimation of APR*	0.13	0.10	0.17	0.16	0.18	0.18
Exponential growth bias, asset-side: *Underestimation of future value*	0.49	0.46	0.37	0.52	0.52	0.51
Exponential growth bias, asset-side: *Overestimation of future value*	0.57		0.36	0.57	0.59	0.61
Limited attention	0.50	0.56	0.58	0.55	0.49	0.47
Limited memory	0.18	0.16	0.15	0.15	0.24	0.27
Panel B. Across Biases						
Min	0.02	-0.01	-0.03	-0.03	0.02	0.03
Max	0.59	0.62	0.61	0.65	0.63	0.63
Mean	0.33	0.30	0.33	0.32	0.31	0.33

Each cell presents a Pearson or tetrachoric correlation of bias rank across our two survey rounds three years apart. Blank cell has undefined estimates due to low but nonzero prevalence of consumers with that bias, in that sub-sample. Column 1 reproduces Table 3 Column 2. For column 2 here, we have 394 panelists with >=B.A. in round 1. Cognitive skills measured here using the first principal component of our four test scores (please see Data Appendix Section 2 for details). Cross-bias statistics in Panel B are unweighted.

Data Appendix

1. Measuring Behavioral Biases

This section details, for each of the 17 potential sources of behavioral bias we measure:

i) The motive for eliciting that potential source of bias (B-factor) and the mechanism through which that factor might affect financial condition;
ii) our elicitation method and its key antecedents;
iii) data quality indicators, including item non-response;
iv) sample size (as it compares to that for other B-factors);
v) definitions and prevalence estimates of behavioral *indicators*, with background on the distinctions between expected direction (standard) vs. less-expected (non-standard) direction 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;

Since our empirical work here is purely descriptive, we focus on our Round 1 data (ALP modules 315 and 352) to get the largest possible sample of panelists. We provide comparisons to prior work wherever possible.

A. Present- or future-biased discounting (money)

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

We measure discounting biases with respect to money using the Convex Time Budgets (CTB) method created by Andreoni and Sprenger (2012). In our version, fielded in ALP module 315 (the first of our two surveys), subjects make 24 decisions, allocating 100 hypothetical tokens each between (weakly) smaller-sooner and larger-later amounts. See Data Appendix Figure 1 for an example. The 24 decisions are spread across 4 different screens with 6 decisions each. Each screen varies start date (today or 5 weeks from today) x delay length (5 weeks or 9 weeks); each decision within a screen offers a different yield on saving. Among the 1,515 individuals who take our first module in Round 1, 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 (Giné et al. 2018). In exploring data quality and prevalence below we focus on comparisons to Andreoni and Sprenger (2012), 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 overall—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. (2019)), 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]

---

[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.

[3] See also Imai et al's (2020) meta-analysis of average estimates (imposing homogeneity in a given sample) of the quasi-hyperbolic discounting model's present-bias parameter. They find "many studies did *not* find strong evidence to reject the null of PB = 1…" (see, e.g., their Figure 1). Bradford et al. (2017) do find present-bias on average in their Qualtrics sample, classifying >50% as present-biased and 26% as future-biased.

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] Counting any deviation from time-consistent discounting as biased, 26% of our sample is present-biased and 36% is future-biased. These prevalence estimates fall 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.

Our prevalence estimates are similar to those from other studies of broad populations that allow for the possibility of future- or present-bias (Data Appendix Table 1). E.g., BC's CTB elicitation in the ALP shows 29% with any present-bias, and 37% with any future-bias. Carvalho et al (2019) find 28% with any present-bias and 31% with any future-bias in a sample of account aggregation software users in Iceland.[5]

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 our second Round1 module.

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).[6] Barcellos

[4] Although see Koszegi and Szeidl (2013) for a theory of future-biased discounting.
[5] Goda et al. use a different elicitation method—a "time-staircase" multiple price list (Falk et al. 2018)—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.
[6] 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).

and Carvalho's ALP subjects answered similar questions in their baseline survey, albeit with only a one-week instead of a five-week delay, with 6% exhibiting present-bias and 9% future-bias. Read and van Leeuwen (1998) offer actual snacks to a convenience sample of employees in Amsterdam but do not calculate individual-level measures of bias. They do find substantial present-bias on average. We do not know of any prior work estimating correlations between measures of consumption discounting biases and field outcomes.

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.[7] 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 Round 1 module make all 11 decisions, and comprise our sample for measuring choice inconsistency.[8] See Data Appendix Figure 2 for an example.

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

[7] 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.

[8] 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.

respect to first-order stochastic dominance (FOSD).[9] Hence, again following Choi et al., our second measure captures inconsistency with both GARP and FOSD.[10] 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.

Data Appendix Table 1 shows that our prevalence estimates are also nearly identical to those from the Choi et al (2014) data. In our data, 53% of subjects exhibit any inconsistency with GARP, and 96% exhibit any inconsistency with GARP or FOSD. If we set a 20pp threshold for classifying someone as inconsistent, only 7% are inconsistent with GARP, and 31% are inconsistent with GARP or FOSD. Looking more directly at heterogeneity, we see standard deviations of 0.08 and 0.18, and 10^{th}-90^{th} percentile ranges of 0.16 and 0.41.

D. Risk attitude re: certainty (certainty premium)

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

[9] 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.
[10] 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.

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[11] for measuring a subject's *certainty premium* (CP).[12] Similar to Holt and Laury tasks, in one of the Callen et al. tasks subjects make 10 choices between two lotteries, one a (p, 1-p) gamble over X and Y > X , (p; X, Y), the other a (q, 1-q) gamble over Y and 0, (q; Y, 0). Both Callen et al. and we fix Y and X at 450 and 150 (hypothetical dollars in our case, hypothetical Afghanis in theirs), fix p at 0.5, and have q range from 0.1 to 1.0 in increments of 0.1. In the other task, p = 1, so the subject chooses between a lottery and a certain option. Our two tasks are identical to Callen et al.'s except for the currency units. But our settings, implementation, and use of the elicited data are different. Callen et al. administer the tasks in-person, using trained surveyors, at polling centers and homes in Afghanistan. They use the data to examine the effects of violence on risk preferences.

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

We estimate CP for each respondent i by imputing the likelihoods q* at which i expresses indifference as the midpoint of the q interval at which i switches, and then using the two likelihoods to estimate the indirect utility components of the CP formula. As Callen et al. detail, the CP "is defined in probability units of the high outcome, Y, such that one can refer to certainty of X being worth a specific percent chance of Y relative to its uncertain value." We estimate a

[11] Callen et al. describes its task as "a field-ready, two-question modification of the uncertainty equivalent presented in Andreoni and Sprenger (2016)."

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

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. Their 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.

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 (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.[13] We acknowledge this but use "loss aversion" instead of "loss aversion and/or small-stakes risk aversion" as shorthand. Small-stakes risk aversion is also often classified as behavioral because it is incompatible with expected utility theory (Rabin 2000).

Response rates suggest a high level of comfort with these questions; only two of our 1,515 subjects skip, and only two more who answer the first question do not answer the second. 37% of our 1,511 respondents reject both lotteries, consistent with relatively extreme loss aversion, compared to 45% of FG's 42 subjects. Another 36% of our subjects accept both lotteries, consistent with classical behavior, compared to 33% in FG. The remaining 27% of our subjects (and 21% of FG's) exhibit moderate loss aversion, playing one lottery but not the other, with our main difference from FG being that 14% of our subjects (vs. only 2% of theirs) exhibit the puzzling behavior of playing lottery 1 but not lottery 2. Although one wonders whether these 14% misunderstood the questions, we find only a bit of evidence in support of that interpretation: those playing the single but not compound lottery have slightly lower cognitive skills than other loss averters, conditional on our rich set of covariates, but actually have higher cognitive skills than the most-classical group. And playing the single but not the compound lottery is uncorrelated with our measure of ambiguity aversion, pushing against the interpretation that the compound lottery is sufficiently complicated as to appear effectively ambiguous (Dean and Ortoleva 2019).

All told 64% of our subjects indicate some loss aversion, defined as rejecting one or both 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.

[13] A related point is that there is no known "model-free" method of eliciting loss aversion (Dean and Ortoleva 2019).

F. Narrow bracketing and dominated choice

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

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

Our first task follows RW's Example 2, with Decision 1 between winning \$100 vs. a 50-50 chance of losing \$300 or winning \$700, and Decision 2 between losing \$400 vs. a 50-50 chance of losing \$900 or winning \$100.[14] 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. As RW discuss, a new feature of task two is that AD sacrifices expected value in the second decision, not in the first. This implies that for all broad-bracketing risk averters AC is optimal: it generates the highest available expected value at no

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

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. RW do not create summary indicators across tasks, but, as noted above, their subjects exhibit substantially more narrow bracketing at the task level than our subjects do.

G. Ambiguity aversion

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

We elicit a coarse measure of ambiguity aversion using just one or two questions about a game that pays $500 if you select a green ball. The first question offers the choice between a Bag One with 45 green and 55 yellow balls vs. a Bag Two of unknown composition. 1,397 subjects respond to this question (out of 1,427 who answer at least one of our questions on ALP module 352). 73% choose the 45-55 bag, and we label them ambiguity averse. The survey then asks these subjects how many green balls would need to be in Bag One to induce them to switch.[15] We subtract this amount from 50, dropping the 99 subjects whose response to the second

[15] Because not everyone answers the second question, we measure time spent responding to the ambiguity aversion elicitation using only the first question.

question is >45 (and the 10 subjects who do not respond), to obtain a continuous measure of ambiguity aversion that ranges from 0 (not averse in the first question) to 50 (most averse=== the three subjects who respond "zero" to the second question). The continuous measure (N=1,288) has a mean of 14 (median=10), and a SD of 13. If we impose a large-deviation threshold of 10 (20% of the max) for labeling someone as ambiguity averse, 50% of our sample exceeds this threshold and another 16% are at the threshold. Our elicitation does not distinguish between ambiguity-neutral and ambiguity-seeking choices (for more comprehensive but still tractable methods see, e.g., Dimmock, Kouwenberg et al. (2016), Dimmock, Kouwenberg, and Wakker (2016), Gneezy et al. (2015)), and so our measure of deviation from ambiguity-neutrality is one-sided.

Despite the coarseness of our elicitation, comparisons to other work suggest that it produces reliable data. Our ambiguity aversion indicator correlates 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 our prevalence of any ambiguity aversion, 0.73 is similar to Dimmock, Kouwenberg, and Wakker's (2016) 0.68 from the Dutch version of the ALP .

H. Overconfidence: Three varieties

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

The second measures overconfidence in precision, as indicated by responding "100%" on two sets of questions about the likelihoods (of different possible Banks and Oldfield quiz scores or of future income increases). This is a coarse adaptation of the usual approaches of eliciting 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: *"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.

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),[16] with mean (SD) responses of 27 (18), 42 (24), and 31 (20). We measure NBLLN using the distance between the subject's answer for the [481, 519] range and 78. Only one subject gets it exactly right. 87% underestimate; coupled with prior work, this result leads us to designate underestimation as the "standard" directional bias. The modal underestimator responds with 50 (18% of the sample). The other most-frequent responses are 25 (10%), 30 (9%), 33 (8%), and 40 (7%). Few underestimators—only 4% of the sample—are within 10pp of 78, and their mean distance is 43, with an SD of 17. 9% of the sample underestimates by 20pp or less. 13% overestimate relative to 78, with 5% of the sample quite close to correct at 80, and another 5% at 100. Benjamin, Moore, and Rabin (2017) 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%.

J. Gambler's Fallacies

The Gambler's Fallacies involve falsely attributing statistical dependence to statistically independent events, in either expecting one outcome to be less likely because it has happened recently (recent reds on roulette make black more likely in the future) or the reverse, a "hot hand" view that recent events are likely to be repeated. Gambler's fallacies can lead to 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 (2017) elicitation for the fallacies:

---

[16] 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.

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

1,392 subjects respond, out of the 1,427 respondents to module 352. The cold-hand fallacy implies a response < 50%, while the hot-hand fallacy implies a response > 50%. Our mean response is 45% (SD=25), which is consistent with the cold-hand but substantially above the 32% in Benjamin, Moore, and Rabin. Another indication that we find less evidence of the cold-hand fallacy is that, while they infer that "at the individual level, the gambler's fallacy [cold-hand] appears to be the predominant pattern of belief" (2013, p. 16), we find only 26% answering < "50." 14% of our sample responds with >"50" (over half of these responses are at "90" or "100"). So 60% of our sample answers correctly. Nearly everyone who responds with something other than "50" errs by a substantial amount—e.g., only 2 % of the sample is [30, 50) or (50, 70]. Sixteen percent of our sample answers "10,"[17] which Benjamin, Moore, and Rabin speculates is an indicator of miscomprehension; we find that while subjects with this indicator do have significantly lower cognitive skills than the unbiased group, they actually have higher cognitive skills than the rest of subjects exhibiting a gambler's fallacy.

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

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

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

[17] 34% of the sample in Benjamin, Moore, and Raymond respond "10%" on one or more of their ten questions.

K. Exponential growth bias: Two varieties

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 given market rates; 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 (15% of the sample).[18] 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 51% and 34% as negatively biased (overestimating APR) under error tolerance of zero. This is less EGB than Stango and Zinman (2009; 2011) see from questions in the 1983 Survey of Consumer Finances, where 98% of the sample underestimates, and the mean bias is 1,800bp or 3,800bp depending on the benchmark. The time frames of the questions differ, which may account for the difference (and is why we do not estimate an EGB structural model parameter to compare with our prior work or that of Levy and Tasoff).

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

[18] Non-response is relatively small, as only 4% of the sample does not respond to both questions. 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.

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.[19] We again bin individuals into underestimators (the standard bias), overestimators, unbiased, and unknown (14% of the sample).[20] Among those with known bias (N=1,222), the median bias is $0, with a mean of $2 and SD of $14.[21] 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. (2019) and Levy and Tasoff (2016) measure asset-side EGB using more difficult questions in their representative samples. They find that 9% and 11% overestimate FVs, while 69% and 85% underestimate. We do not construct an EGB parameter to compare to theirs, because our questions lack their richness and yield heavy mass points at unbiased and linear-biased responses.

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

[20] 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.

[21] 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 and limited 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 forthcoming; 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?"[22] 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

[22] 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.

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.

Our indicator of behavioral limited memory— (not completing the follow-up task 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.

2. Measuring Cognitive Skills

We measure fluid intelligence using a 15-question, non-adaptive number series (McArdle, Fisher, and Kadlec 2007). Number series scores correlate strongly with those from other fluid intelligence tests like IQ and Raven’s.

We measure numeracy using: “If 5 people split lottery winnings of two million dollars ($2,000,000) into 5 equal shares, how much will each of them get?” and “If the chance of getting a disease is 10 percent, how many people out of 1,000 would be expected to get the disease?” (Banks and Oldfield 2007). Response options are open-ended. These questions have been used in economics as numeracy and/or financial literacy measures since their deployment in the 2002 English Longitudinal Study of Ageing, with subsequent deployment in the Health and Retirement Study and other national surveys.

We measure financial literacy using Lusardi and Mitchell’s (2014) “Big Three”: “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?”; and “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.

We measure executive function using a two-minute Stroop task (MacLeod 1991). 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). Before starting the task, the computer shows demonstrations of two choices (movie-style)—one with a correct response, and one with an incorrect response—and then gives the subject the opportunity to practice two choices on her own. After practice ends, the task lasts for two minutes.

References for Data Appendix

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

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

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

———. 2016. "Prospect Theory Revisited: Unconfounded Experimental Tests of Probability Weighting."

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

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

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

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

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

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

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

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

Benjamin, Daniel, Don Moore, and Matthew Rabin. 2017. "Biased Beliefs about Random Samples: Evidence from Two Integrated Experiments."

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. forthcoming. "Memory, Attention, and Choice." *Quarterly Journal of Economics*.

Bradford, David, Charles Courtemanche, Garth Heutel, Patrick McAlvanah, and Christopher Ruhm. 2017. "Time Preferences and Consumer Behavior." *Journal of Risk and Uncertainty* 55 (2–3): 119–45.

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.

Carvalho, Leandro, Arna Olafsson, and Dan Silverman. 2019. “Misfortune and Mistake: The Financial Conditions and Decision-Making Ability of High-Cost Loan Borrowers.”

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. 2019. “The Empirical Relationship between Non-Standard Economic Behaviors.” *Proceedings of the National Academy of Science* 116 (33): 16262–67.

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

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

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

Dimmock, Stephen, Roy Kouwenberg, and Peter P Wakker. 2016. “Ambiguity Attitudes in a Large Representative Sample.” *Management Science* 62 (5): 1363–80.

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

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, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. “Global Evidence on Economic Preferences.” *The Quarterly Journal of Economics* 133 (4): 1645–92.

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.

Giné, Xavier, Jessica Goldberg, Dan Silverman, and Dean Yang. 2018. “Revising Commitments: Field Evidence on the Adjustment of Prior Choices.” *The Economic Journal* 128 (608): 159–88.

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. 2019. “Predicting Retirement Savings Using Survey Measures of Exponential-Growth Bias and Present Bias.” *Economic Inquiry* 57 (3): 1636–58.

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.

Hwang, In Do. 2016. “Prospect Theory and Insurance Demand.”

Imai, Taisuke, Tom Rutter, and Colin F. Camerer. 2020. “Meta-Analysis of Present-Bias Estimation Using Convex Time Budgets.”
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.
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.
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.

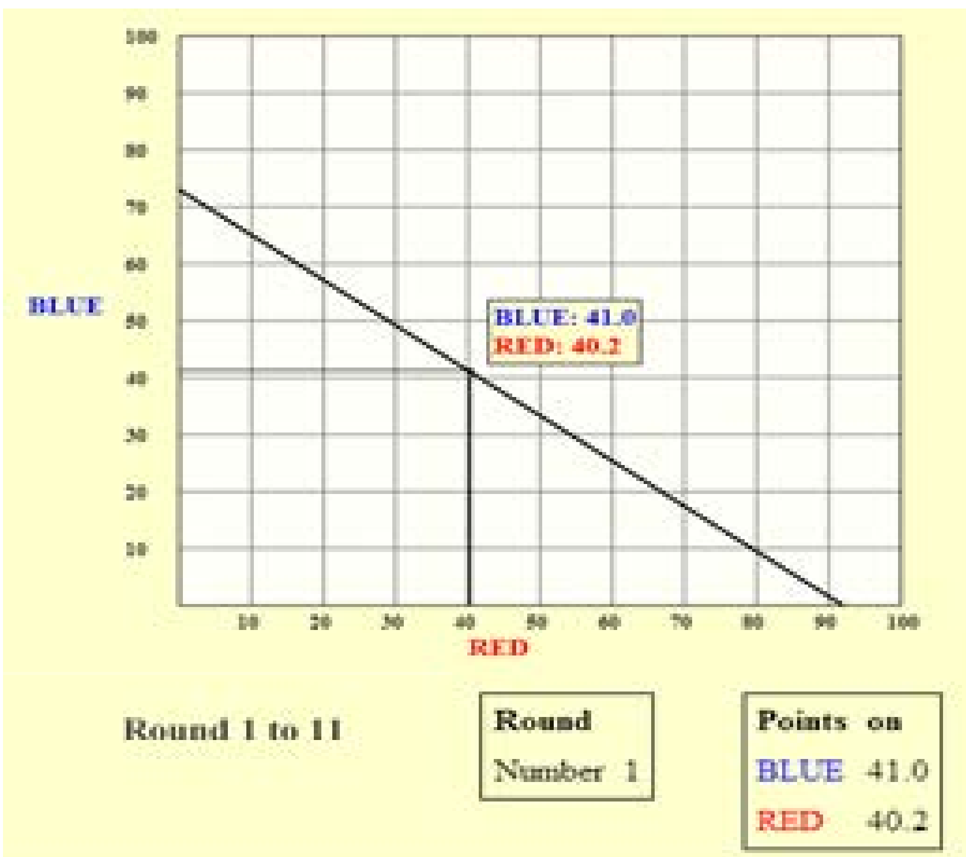
Toubia, Olivier, Eric Johnson, Theodoros Evgeniou, and Philippe Delquié. 2013. “Dynamic Experiments for Estimating Preferences: An Adaptive Method of Eliciting Time and Risk Parameters.” *Management Science* 59 (3): 613–40.

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

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).

Data Appendix Table 1. Behavioral bias prevalence: Comparisons to prior work using representative samples

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

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

Footnotes:

[1] - Barcellos and Carvahlo (2014), source data are from ALP.

[2] - Goda et al. (2017), sources are ALP and Understanding America Survey.

[3] - Choi et al. (2011), source is CentER panel (Netherlands).

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

[5] - von Gaudeker et al. (2011), source is CentER panel (Netherlands).

[6] - Rabin and Weizacker (2009), source is KnowledgeNetworks

[7] - Gottleib and Mitchell (2015), source is Health and Retirement Study (older Americans).

[8] - Dimmock et al. (2016), source is ALP.

[9] - Dimmock, Kouwenberg and Wakker (forthcoming), source is CentER panel (Netherlands).

[10] - Dohmen et al. (2009), source is German SocioEconomic Panel.

[11] - Stango and Zinman (2009, 2011), source is Survey of Consumer Finances.

[12] - Levy and Tasoff (2016), source is KnowledgeNetworks