Results Appendix. Additional results and robustness checks

Appendix Table 4 confirms that the data warrant pooling. Each column presents an estimate of a B-count coefficient for either Round 1 ORIV (odd-numbered columns) or Round 2 ORIV (even-numbered columns), varying outcomes (objective or subjective financial index) and B-counts (Full B-count and the two Sparsity B-counts). The B-count coefficients are qualitatively similar across rounds, and do not reject equality at conventional p-value cutoffs.

Appendix Table 5 is one of several ways we address the possibility of spurious correlation. Here we focus on a reverse causality interpretation of our main results, through which having lower financial resources produces behavioral biases.[51] We consider this hypothesis by varying our main specification in two ways. One way is using Round 2 data only and instrumenting for the Round 2 B-count with the Round 1 B-count. That "Standard IV" approach uses only the 3-year-earlier measurements of behavioral biases to identify the correlation between our three main B-counts and experienced utility (Columns 1, 4, and 7). The second way is conditioning on objective financial condition while using the subjective financial index as our experienced utility measure (as we do in Table 4); that may err on the side of over-controlling, but allows us to address the possibility that (objectively) low financial resources produce behavioral tendencies by controlling for the former (Columns 2, 5, and 8).[52] Granting that possibility, we then instrument for Round 2 objective financial condition with Round 1 objective financial condition in Columns 3, 6, and 9. The B-count conditional correlation with experienced utility remains strongly negative in each of these nine specifications, suggesting that our main results are not driven by reverse causality.

Re: other spurious correlation hypotheses, we refer the reader back to Table 5 and Section 4-D. The former addresses the standard omitted-variable, unobserved heterogeneity concern by varying control variable specifications. The latter details how our survey design and controls for survey effort minimize the likelihood of spurious correlations between outcome measures and behavioral bias measures.

[51] We say "interpretation" instead of "concern" here, because if reverse causality were to drive the results, that would be important to discover in the sense that it would motivate a revamp of most behavioral models.
[52] We use "produce" instead of "exacerbate" here intentionally, to highlight another benefit of relying on discrete measures of behavioral biases: in our setup it would need to be the case the worse financial condition increases the likelihood that people indicate *any* deviation from classical benchmarks.

Appendix Table 6 decomposes the subjective and objective indexes into their components, and shows that links between B-counts and different outcomes are robustly negative for the component outcomes: all 27 coefficients are negative, 17 of the 27 B-count coefficients have p-values <0.01, and each implies an economically large marginal change in the outcome variable per one standard deviation change in the B-count. There is evidence of some quantitative heterogeneity, however, including within-index. E.g., the Full B-count coefficients on the subjective financial condition index components (each of which have p-values<0.01) range from -0.04 to -0.14 (Panel A, Columns 6-9).

Appendix Table 7 confirms robustness to other functional forms for the B-count: the natural logarithm of the B-count, the ratio of the panelist's B-count to their count of non-missing sources of potential behavioral biases, B-count quartiles (the results on which do not reject a linear relationship between outcomes and the B-count), and the "B-tile," a consumer-level measure of the *magnitude* of behavioral deviations from classical benchmarks.[53] The marginal effects in these alternative specifications hardly differ from those in Table 4 at all—note how similar are the d(Outcome)/d(1 SD B-count) levels across specifications.

Untabulated results, where we estimate the specifications in Table 4 separately for different sub-groups based on demographics, etc., do not reject equality of the B-count coefficient across sub-groups. Subject to the caveat that these tests are under-powered, these results support the assumption of a separable behavioral wedge (Section 4-C). They also fail to support a knife-edge interpretation of our results in which a narrow subset of panelists drives the results. And they cast doubt on the efficacy of targeting behavioral consumers based on more readily observable characteristics (see also Section 7).

---

[53] Some of our bias measures are continuous, permitting percentiles to take on the full range of values from 1 to 100. For discrete-response and uni-directional outcomes like loss aversion, the B–tiles take on fewer values but still measure the degree of deviation from classical benchmarks in useful ways. For example, loss aversion takes on four values: unbiased, and then three ordered responses (whether the individual respondent rejects the compound but not the single lottery, rejects the single but not the compound lottery, or rejects both) coded as 1/2/3. Any respondent accepting both lotteries receives a 0 (meets the classical benchmark), and 37% of individuals share that response. Anyone with the smallest deviation from the benchmark therefore is in the 37th percentile, and 13% of responses fall into that category. Summing, anyone in the next category is in the 50th(=37th+13th), and so on. The B-tile calculates each person's percentile ranking for each of the 17 potential sources of behavioral bias, relative to others in the sample, and sums them. If a person were to be the most biased person in the sample on all 17, that person would have a B-tile of (close to) 17.

Appendix Table 8 shows the full set of coefficients on the covariates in specifications (1)-(3) in Table 4. The table sheds light on the conditional correlations of other variables with outcomes, which are intuitive for the most part. Income is positively correlated with objective financial condition and, consistent with research on happiness, more weakly so with subjective financial well-being, and (weakly) negatively once we control for objective financial condition. Other coefficients in the first and second columns reverse once we control for objective financial condition in column 3, showing its power as a control and highlighting the robustness of the correlation between the B-count and financial condition in relative as well as absolute terms. These columns also highlight the relative magnitude of the coefficient on the B-count (subject to caveats re: over-controlling). E.g., for the objective financial condition outcome, a one-SD increase in the B-count has roughly the same implied correlation as moving down two or three income deciles. For subjective financial condition, the most noteworthy pattern is that the B-count, and missingness thereon, have correlations that are more robust to the inclusion of objective financial condition as an additional covariate than any other variable or group of variables, with the possible exception of survey response times.

Appendix Table 9 examines whether using the ALP's sampling weights changes our main empirical results (see Appendix Table 2 for a similar exercise re: B-count descriptive statistics). AT 9 compares weighted estimates to our main unweighted ones from Table 4 and reveals that the weighted coefficients are uniformly more negative (i.e., larger in an economic sense) in point terms, but less precise (e.g., while each of the six unweighted coefficients has a p-value<0.01, two of the weighted coefficients has p-value <0.01 and one has a p>0.10). Mechanically, it must be the case that panelists who are under-sampled by RAND (and therefore over-weighted) have noisier relationships between our outcomes and covariates. Re: external validity, the glass half-empty interpretation of these results and our setup is that ALP sampling weights produce noisier inferences on behavioral summary statistics and, in any case, are based on demographics but not our variables of greatest interest; therefore, the extent to which our inferences our valid for the entire U.S. population is an open question. The glass half-full interpretation is that we have an unusually broad sample compared to most studies in the behavioral social sciences, and that our results on B-count properties and their conditional correlations with outcomes are not unduly sensitive to weighting that is designed to produce valid inferences for the U.S. population.

Appendix Table 1. Other covariates: Measuring classical decision inputs and survey effort

Variable	Definition/specification
Demographics:	
Gender	Indicator, "1" for female.
Age	Four categories: 18-34, 35-45, 46-54, 55+
Education	Four categories: HS or less, some college/associates, BA, graduate
Income	The ALP's 17 categories (collapsed into deciles in some specifications)
Race/ethnicity	Three categories: White, Black, or Other; seperate indicator for Hispanic
Marital status	Three categories: married/co-habitating; separated/divorced/widowed; never married
Household size	Five categories for count of other members: 0, 1, 2, 3, 4+
Employment status	Five categories: working, self-employed, not working, disabled, missing
Immigrated to USA	Indicator, "1" for immigrant
State of residence	Fixed effects
Risk, patience:	
Risk aversion (financial)	100-point scale on financial risk-taking from Dohmen et al., with higher values indicating greater risk aversion
Risk aversion (income)	Adaptive lifetime income scale from Barsky et al., 1-6 with 6 indicating greatest risk aversion
Patience	Average savings rate across the 24 Convex Time Budget decisions, standardized
Cognitive and noncognitive skills	
Fluid intelligence	# correct on standard 15-question, non-adaptive number series quiz
Numeracy	# correct on Banks and Oldfield questions re: division and %
Financial literacy	# correct on Lusardi and Mitchell "Big Three" questions re: interest, inflation, and diversification
Executive attention	# correct on 2-minute Stroop test; respondents instructed to answer as many q's correctly as they can
Big Five Personality Traits	One variable per trait, from Rammstedt and John's validated 10-question test and scorecard (Round 2 only)
Survey effort and attrition	
Time spent on questions	Measured for each B-factor (and other variables), included as decile indicators relative to other respondents
Item non-response	Indicators for variables with non-trivial rates of non-response (although all are <5%): Income, employment status, risk, patience, cognitive skills, non-cognitive skills.

For more details on the cognitive skills measures, please see Data Appendix Section 2.

Appendix Table 2. Key B-count descriptive statistics, without and with population weighting

	(1)	(2)	(3)	(4)	(5)	(6)
weighted?	no	yes	no	yes	no	yes
	Mean (SD), across both rounds		Correlation with full B-count		Correlation (Round1,Round2)	
B-(sub)-count						
Full	9.96	9.97	1.00	1.00	0.44	0.44
	(2.12)	(2.16)				
Sparsity: Narrow	1.20	1.25	0.39	0.40	0.27	0.22
	(0.66)	(0.67)				
Sparsity: Broad	4.12	4.20	0.69	0.69	0.39	0.36
	(1.33)	(1.32)				
Expected biases	8.46	8.43	0.86	0.85	0.44	0.40
	(2.09)	(2.12)				
Non-expected biases	1.50	1.53	0.25	0.26	0.24	0.16
	(1.04)	(1.04)				
Math biases	2.52	2.58	0.57	0.56	0.44	0.38
	(0.90)	(0.89)				
Non-math biases	7.43	7.39	0.90	0.91	0.32	0.33
	(1.74)	(1.78)				
Preference biases	4.28	4.15	0.51	0.54	0.23	0.25
	(1.25)	(1.29)				
Non-preference biases	5.68	5.82	0.82	0.81	0.49	0.44
	(1.74)	(1.71)				
Missing inputs	0.84	0.97	-0.33	-0.38	0.36	0.45
	(1.48)	(1.65)				
N	1690	1690	1690	1690	1690	1690
N panelists	845	845	845	845	845	845

Our data consist of two survey rounds, of two modules each, conducted 3 years apart. We include only those panelists who took all four modules across both rounds (N=845). B-count and B-sub-count definitions are summarized in Table 1 and discussed in Sections 3-A and -B. Round-to-round correlations for B-counts adjust for missing data by conditioning on the count of missing bias measures in each survey round. Column 3 here reproduces Table 2 Column 6. Column 5 here reproduces Table 2 Column 7.

Appendix Table 3. Measuring financial condition and subjective well-being: Definitions, sampling, and descriptive statistics for index components

	Data used				Mean	SD	Pairwise correlation										
	# of questions per module	From our modules?	From other modules?	# panelists with nonmissing	(All rescaled to [0,1])		Net worth>0	Retirement assets>0	Owns stocks	Spent < income	No severe hardship	Financial satisfaction	Retirement saving adequacy	Non-ret saving adequacy	Lack financial stress	Happiness Last 30 days	Happiness in general
Panel A. Objective financial condition index components																	
Net worth>0	2	yes	no	821	0.50	0.50	1.00										
Retirement assets>0	2	yes	no	831	0.60	0.49	0.54	1.00									
Owns stocks	3	yes	no	835	0.54	0.50	0.56	0.96	1.00								
Spent < income in last 12 months	1	yes	no	841	0.41	0.49	0.44	0.35	0.35	1.00							
No severe hardship in last 12 months	4	yes	no	842	0.61	0.49	0.49	0.45	0.49	0.50	1.00						
Panel B. Subjective financial condition index components																	
Financial satisfaction scale	1	yes	no	842	0.59	0.26	0.35	0.32	0.33	0.43	0.49	1.00					
Retirement saving adequacy scale	1	yes	no	842	0.47	0.41	0.44	0.46	0.45	0.46	0.56	0.53	1.00				
Non-retirement saving adequacy scale	1	yes	no	843	0.49	0.37	0.34	0.17	0.18	0.35	0.37	0.31	0.49	1.00			
Lack of financial stress scale	1	yes	no	845	0.47	0.30	0.40	0.31	0.32	0.43	0.53	0.53	0.47	0.35	1.00		
Panel C. Other measures of subjective well-being: Happiness index components																	
Happiness last 30 days	1	no	yes	509	0.62	0.21	0.22	0.22	0.25	0.21	0.33	0.41	0.25	0.12	0.32	1.00	
Happiness in general	1	no	yes	675	0.75	0.26	0.27	0.27	0.25	0.26	0.29	0.39	0.24	0.09	0.34	0.65	1.00

Unit of observation is the individual respondent, with multiple observations per respondent averaged across survey rounds (for variables in our modules) or across other ALP modules (for variables we merge in from other ALP modules). Other ALP modules used here are all administered *between* our survey rounds; we could not find relevant data collected in modules adminstered after or during our second round. As in most of our main tables, we limit the sample frame here to panelists who completed both of our survey rounds (N=845). Correlations estimated using the two-step "polychoric" procedure in Stata.

Variable definitions: Each variable is scaled so that higher values indicate better financial condition and/or subjective well-being. Each measure here is scaled or rescaled to [0, 1] for comparability.

Net worth is from two summary questions drawn from the National Longitudinal Surveys: "Please think about all of your household assets (including but not limited to investments, other accounts, any house/property you own, cars, etc.) and all of your household debts (including but not limited to mortgages, car loans, student loans, what you currently owe on credit cards, etc.) Are your household assets worth more than your household debts?" and "You stated that your household's [debts/assets] are worth more than your household's [assets/debts]. By how much?"

Retirement assets is from questions asking specifically whether someone has one or more IRA accounts and one or more workplace plans, followed in each case by questions on amounts in such accounts. Questions like these are asked in the Survey of Consumer Finances, the Health and Retirement Study, and many other surveys.

Stockholding is from questions on stock mutual funds in IRAs, stock mutual funds in 401ks/other retirement accounts, and direct holdings. Questions like these are asked in the Survey of Consumer Finances, the Health and Retirement Study, and many other surveys.

Spent < income question is from the Survey of Consumer Finances: "Over the past 12 months, how did your household's spending compare to your household's income? If the total amount of debt you owe decreased, then count yourself as spending less than income. If the total amount of debt you owe increased, then count yourself as spending more than income." Response options are: "Spent more than income", "Spent same as income", and "Spent less than income".

(No) severe hardship questions are taken from the National Survey of American Families: late/missed payment for rent, mortgage, heat, or electric; moved in with other people because could not afford housing/utilities; postponed medical care due to financial difficulty; adults in household cut back on food due to lack of money. Response options for each of the four are Yes or No.

The **Financial satisfaction** question follows standard life and economic satisfaction question wording: "How satisfied are you with your household's overall economic situation?"; responses on a 100-point scale (input using slider or text box).

Retirement and non-retirement savings adequacy questions are placed one each in the two different modules, with different wording, to mitigate mechanical correlations. The questions are: "Using any number from one to five, where one equals not nearly enough, and five equals much more than enough, do you feel that your household is saving and investing enough for retirement? Please consider the income you and any other members of your household expect to receive from Social Security, 401(k) accounts, other job retirement accounts and job pensions, and any additional assets you or other members of your household have or expect to have" and "Now, apart from retirement savings, please think about how your household typically uses the money you have: how much is spent and how much is saved or invested. Now choose which statement best describes your household". These questions are variants on standard ones, but in each case our 5 response options are framed to encourage people to recognize tradeoffs between saving and consumption: any response that includes "saving more" also includes "and borrowing/spending less", and vice versa. In mapping the 5 responses into the variables used here, we code: saved-enough, more-than-enough, and much-more-than-enough as 1 (the latter two responses are rare: 3% of the sample for retirement, and 4% for non-retirement); saved < enough as 0.5; saved << enough as 0.

Financial stress question is taken from The Survey of Forces**:** "To what extent, if any, are finances a source of stress in your life?"; responses on a 100-point scale (respondents can input using slider or text box).

Life satisfaction question is measured using some one of three minor variants on the standard "… how satisfied are you with your life as a whole these days?" asked in many surveys worldwide. For the other-module measure, we take the within-panelist average of non-missing responses to this question across the six ALP modules in which it has appeared subsequent to our round 1 modules, as of this writing. Of the 809/845 panelists with at least one non-missing response, 640 have at

Happiness last 30 days is measured using the standard "During the past 30 days, how much of the time have you been a happy person?" asked in many surveys worldwide. We take the within-panelist average of non-missing responses to this question across the four ALP modules in which it has appeared subsequent to our round 1 modules, as of this writing. Of the 509/845 panelists with at least one non-missing response, 474 have at least two, .

Happiness in general is measured using the standard "Taking all things together, I am generally happy" question asked in many surveys worldwide, incuding ALP module 425.

Appendix Table 4. ORIV estimates are similar across survey rounds
(Compare to Table 4)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Financial outcome index includes:	*Objective*	*Objective*	*Subjective*	*Subjective*	*Objective*	*Objective*	*Subjective*	*Subjective*	*Objective*	*Objective*	*Subjective*	*Subjective*
B-count: Full	-0.073***	-0.053**	-0.094***	-0.080***								
	(0.023)	(0.021)	(0.022)	(0.021)								
B-count: Sparsity Broad					-0.085**	-0.095***	-0.130***	-0.130***				
					(0.034)	(0.033)	(0.034)	(0.031)				
B-count: Sparsity Narrow									-0.193***	-0.300***	-0.292***	-0.397***
									(0.067)	(0.083)	(0.069)	(0.082)
d(LHS)/d(1 SD B-count)	-0.155	-0.112	-0.201	-0.169	-0.113	-0.126	-0.173	-0.172	-0.128	-0.199	-0.194	-0.264
mean(LHS)	0.522	0.539	0.493	0.515	0.522	0.539	0.493	0.515	0.522	0.539	0.493	0.515
Round included?	1 only	2 only	1 only	2 only	1 only	2 only	1 only	2 only	1 only	2 only	1 only	2 only
N panelists	841	844	841	844	841	844	841	844	841	844	841	844
N	1682	1688	1682	1688	1682	1688	1682	1688	1682	1688	1682	1688

* 0.10 ** 0.05 *** 0.01. Standard errors, clustered on panelist, in parentheses. Each column presents results from a single-round Obviously Related Instrumental Variables regression (per Section 4-C) of the LHS variable described in the column label on the variables described in the row labels + the complete set of covariates described in Appendix Table 1. Table 1 provides details on our B-count variable definitions; higher values indicate more behavioral biases. Table 3 provides details on our LHS variable definitions; higher values indicate better financial condition.

Appendix Table 5. Identifying relationships between outcomes and B-counts: Reverse causality looks unlikely

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Financial outcome index includes:					*Subjective measures*				
B-count: Full	-0.055***	-0.046***	-0.042**						
	(0.018)	(0.017)	(0.017)						
B-count: Sparsity Broad				-0.123***	-0.093***	-0.076***			
				(0.030)	(0.027)	(0.026)			
B-count: Sparsity Narrow							-0.315***	-0.231***	-0.177***
							(0.064)	(0.059)	(0.054)
Objective financial index		0.332***	0.488***		0.333***	0.513***		0.300***	0.490***
		(0.036)	(0.054)		(0.037)	(0.055)		(0.041)	(0.057)
d(LHS)/d(1 SD B-count)	-0.122	-0.102	-0.092	-0.164	-0.124	-0.102	-0.213	-0.156	-0.120
Data used	Round 2	Round 2	Round 2	Round 2	Round 2	Round 2	Round 2	Round 2	Round 2
IV for B-count with Round 1?	yes	yes	yes	yes	yes	yes	yes	yes	yes
IV for objective financial index with Round 1?	no	no	yes	no	no	yes	no	no	yes
mean(LHS)	0.515	0.515	0.515	0.515	0.515	0.515	0.515	0.515	0.515
N = N panelists	844.000	844.000	844.000	844.000	844.000	844.000	844.000	844.000	844.000

* 0.10 ** 0.05 *** 0.01. Standard errors in parentheses. Each column presents results from a single two-stage least square regression of the LHS variable described in the column label on the variables described in the row labels + the complete set of covariates described in Appendix Table 1. Table 1 provides details on our B-count variable definitions; higher values indicate more behavioral biases. Table 3 provides details on the subjective financial condition index construction; higher values indicate better financial condition and higher experienced utility. The difference between this table and our main specifications is that here we only use "replicate 2" and "standard IV": we use Round 2 data for all variables except for instruments.

Appendix Table 6. B-counts are strongly conditionally correlated with financial index components
(Compare to Table 4)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Index:	*Objective financial condition*					*Subjective financial condition*			
Component:	*Net worth>0*	*Retirement assets>0*	*Owns stocks*	*Spent < income*	*No severe hardship*	*Financial satisfaction*	*Retirement saving adequacy*	*Non-ret saving adequacy*	*Lack financial stress*
Panel A.									
B-count: Full	-0.077***	-0.067**	-0.047*	-0.041	-0.080***	-0.042***	-0.067***	-0.137***	-0.088***
	(0.029)	(0.026)	(0.027)	(0.029)	(0.029)	(0.016)	(0.019)	(0.030)	(0.024)
d(LHS)/d(1 SD B-count)	-0.164	-0.142	-0.099	-0.087	-0.170	-0.089	-0.142	-0.291	-0.182
mean(LHS)	0.500	0.598	0.543	0.407	0.610	0.586	0.473	0.468	0.490
N	3300	3322	3338	3360	3362	3362	3368	3362	3310
Panel B.									
B-count: Sparsity Broad	-0.103**	-0.089**	-0.050	-0.091**	-0.121***	-0.061**	-0.105***	-0.205***	-0.140***
	(0.043)	(0.041)	(0.039)	(0.043)	(0.044)	(0.024)	(0.030)	(0.046)	(0.038)
d(LHS)/d(1 SD B-count)	-0.136	-0.118	-0.066	-0.120	-0.160	-0.081	-0.139	-0.272	-0.183
mean(LHS)	0.500	0.598	0.543	0.407	0.610	0.586	0.473	0.468	0.490
N	3300	3322	3338	3360	3362	3362	3368	3362	3310
Panel C.									
B-count: Sparsity Narrow	-0.276***	-0.197**	-0.159*	-0.317***	-0.231***	-0.142***	-0.298***	-0.479***	-0.391***
	(0.095)	(0.089)	(0.087)	(0.095)	(0.089)	(0.048)	(0.067)	(0.099)	(0.086)
d(LHS)/d(1 SD B-count)	-0.183	-0.131	-0.106	-0.211	-0.153	-0.094	-0.197	-0.318	-0.259
mean(LHS)	0.500	0.598	0.543	0.407	0.610	0.586	0.473	0.468	0.490
N	3300	3322	3338	3360	3362	3362	3368	3362	3310

Each panel*column reports results from a single regression, using the same specification as Table 4 Col 1 and 2 (in Panel A here), Table 4 Col 4 and 5 (in Panel B here), or Table 4 Col 7 and 8 (in Panel C here). Sample sizes are slightly smaller here than in Table 4 because of non-response in index components. See Appendix Table 3 for index component variable definitions and statistics.

Appendix Table 7. Functional form robustness of the Full B-count's conditional correlation with financial outcomes
(Columns 1 and 6 here are same specifications as Columns 1 and 2 in Table 4)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Financial outcome index includes:	*Objective measures*					*Subjective measures*				
Full B-count	-0.061***					-0.084***				
	(0.018)					(0.018)				
ln(B-count)		-0.647***					-0.873***			
		(0.200)					(0.209)			
B-count proportion			-0.974***					-1.203***		
			(0.277)					(0.271)		
B-tile: Average percentile across all biases				-1.293***					-1.571***	
				(0.357)					(0.329)	
B-count: 2nd quartile					-0.156					-0.124
					(0.096)					(0.103)
B-count: 3rd quartile					-0.303**					-0.399***
					(0.146)					(0.144)
B-count: 4th quartile					-0.446***					-0.580***
					(0.138)					(0.141)
d(LHS)/d(1 SD B-count variable)	-0.130	-0.154	-0.130	-0.127		-0.179	-0.208	-0.160	-0.154	
dy/d(1 SD B-count quartile 2)					-0.068					-0.054
dy/d(1 SD B-count quartile 3)					-0.141					-0.186
dy/d(1 SD B-count quartile 4)					-0.208					-0.271
mean(LHS)	0.531	0.531	0.531	0.531	0.531	0.504	0.504	0.504	0.504	0.504
N	3370	3370	3370	3370	3370	3370	3370	3370	3370	3370

* 0.10 ** 0.05 *** 0.01. Standard errors, clustered on panelist, in parentheses. Each column presents results from a single pooled Obviously Related Instrumental Variables regression (equation 3 in the text) of the LHS variable described in the column label on the variable(s) described in the row labels + the complete set of covariates described in Appendix Table 1. Table 1 provides details on our Full B-count variable definition; higher values indicate more behavioral biases. Table 3 provides details on our LHS variable definitions; higher values indicate better financial condition.

Appendix Table 8. Main specifications for estimating correlation between financial condition and the Full B-count, showing results on all of the other covariates. (Same specifications as Table 4, Columns 1-3.)

	(1)	(2)	(3)
Financial outcome index includes:	*Objective measures*	*Subjective measures*	*Subjective measures*
B-count: Full	-0.061***	-0.084***	-0.064***
	(0.018)	(0.018)	(0.016)
Missing bias count	-0.044***	-0.055***	-0.040***
	(0.013)	(0.013)	(0.012)
Female	0.019	0.019	0.012
	(0.020)	(0.018)	(0.015)
Education: Some college	-0.033	-0.032	-0.020
	(0.025)	(0.023)	(0.020)
Education: B.A.	0.031	-0.016	-0.026
	(0.027)	(0.026)	(0.023)
Education: Grad school	0.050*	0.047	0.030
	(0.030)	(0.030)	(0.026)
Income: 2nd decile	0.061**	-0.011	-0.032
	(0.026)	(0.026)	(0.023)
Income: 3rd decile	0.114***	0.004	-0.034
	(0.033)	(0.032)	(0.029)
Income: 4th decile	0.182***	-0.007	-0.068**
	(0.030)	(0.030)	(0.028)
Income: 5th decile	0.241***	0.017	-0.065**
	(0.036)	(0.035)	(0.031)
Income: 6th decile	0.297***	0.047	-0.054*
	(0.034)	(0.033)	(0.030)
Income: 7th decile	0.304***	0.044	-0.059*
	(0.035)	(0.036)	(0.032)
Income: 8th decile	0.358***	0.088**	-0.034
	(0.036)	(0.036)	(0.032)
Income: 9th decile	0.360***	0.073*	-0.050
	(0.039)	(0.041)	(0.035)
Income: Top decile	0.504***	0.191***	0.020
	(0.039)	(0.048)	(0.045)
Age 35-45	0.026	-0.031	-0.040**
	(0.023)	(0.022)	(0.019)
Age 46-54	0.086***	-0.014	-0.043**
	(0.024)	(0.023)	(0.020)
Age 55+ (Max 60)	0.121***	0.028	-0.013
	(0.025)	(0.024)	(0.021)
Race: Black	-0.037	0.025	0.037
	(0.030)	(0.027)	(0.023)
Race: Other non-white	-0.064**	-0.023	-0.001
	(0.027)	(0.029)	(0.027)
Latino	-0.041	0.013	0.027
	(0.027)	(0.025)	(0.023)
Immigrant	0.050*	0.026	0.009
	(0.027)	(0.027)	(0.025)
Peviously married	0.010	0.019	0.016
	(0.021)	(0.019)	(0.017)
Never married	0.026	-0.015	-0.024
	(0.022)	(0.021)	(0.018)
Other household members: 1	-0.012	-0.033*	-0.029*
	(0.020)	(0.018)	(0.016)
Other household members: 2	-0.015	-0.020	-0.015
	(0.021)	(0.021)	(0.019)
Other household members: 3	-0.035	-0.040	-0.028
	(0.026)	(0.026)	(0.022)
Other household members: 4	-0.037	-0.029	-0.016
	(0.033)	(0.030)	(0.026)
Work status: Self-employed	-0.022	-0.042	-0.035
	(0.032)	(0.030)	(0.026)
Work status: Not working	-0.030	0.007	0.017
	(0.028)	(0.025)	(0.023)
Work status: Disabled	-0.155***	-0.087***	-0.034
	(0.032)	(0.031)	(0.027)
Work status: Unknown	-0.135**	0.015	0.061
	(0.063)	(0.078)	(0.068)
Patience in CTB task on 0 to 1 scale	0.019	0.022	0.015
	(0.029)	(0.027)	(0.023)
Patience missing	0.029	0.028	0.018
	(0.035)	(0.034)	(0.030)

Risk aversion: Financial on -1 to 0 scale	-0.058*	-0.035	-0.016
	(0.033)	(0.030)	(0.026)
Risk aversion: financial missing	-0.023	-0.001	0.007
	(0.080)	(0.096)	(0.099)
Risk aversion: lifetime income	0.012**	0.009*	0.005
	(0.006)	(0.005)	(0.005)
Risk aversion: income missing	0.036	0.097	0.084
	(0.107)	(0.081)	(0.077)
Fluid intelligence score	-0.006	-0.009*	-0.007*
	(0.005)	(0.005)	(0.004)
Fluid intelligence missing	0.022	-0.046	-0.053
	(0.091)	(0.088)	(0.087)
Numeracy score	0.004	-0.011	-0.013
	(0.015)	(0.012)	(0.011)
Numeracy missing	-0.027	-0.084*	-0.075*
	(0.055)	(0.046)	(0.045)
Financial literacy score	0.026**	-0.013	-0.022**
	(0.011)	(0.011)	(0.009)
Financial literacy missing	0.019	-0.070	-0.076
	(0.091)	(0.121)	(0.131)
Stroop score/100	0.013	-0.011	-0.015
	(0.032)	(0.028)	(0.025)
Stroop missing	0.001	-0.039	-0.039
	(0.037)	(0.039)	(0.035)
Survey effort: 2nd decile	0.033	-0.039	-0.051**
	(0.030)	(0.026)	(0.023)
Survey effort: 3rd decile	0.028	-0.032	-0.041*
	(0.029)	(0.028)	(0.025)
Survey effort: 4th decile	0.008	-0.068**	-0.071***
	(0.031)	(0.027)	(0.024)
Survey effort: 5th decile	0.016	-0.061**	-0.067***
	(0.032)	(0.029)	(0.025)
Survey effort: 6th decile	0.022	-0.070**	-0.078***
	(0.031)	(0.029)	(0.026)
Survey effort: 7th decile	0.038	-0.033	-0.046*
	(0.031)	(0.029)	(0.026)
Survey effort: 8th decile	0.037	-0.057**	-0.070***
	(0.031)	(0.029)	(0.025)
Survey effort: 9th decile	0.007	-0.076***	-0.078***
	(0.031)	(0.029)	(0.026)
Survey effort: 10th decile	0.013	-0.026	-0.031
	(0.030)	(0.028)	(0.026)
Extraversion score	0.002	0.006	0.006
	(0.004)	(0.004)	(0.003)
Agreeableness score	0.002	0.008*	0.008*
	(0.005)	(0.004)	(0.004)
Conscientiousness score	0.015***	0.010**	0.005
	(0.005)	(0.005)	(0.004)
Neuroticism score	-0.005	-0.009**	-0.007**
	(0.004)	(0.004)	(0.004)
Openness score	-0.014***	-0.008*	-0.003
	(0.005)	(0.004)	(0.004)
Personality variables missing	-0.026	-0.013	-0.004
	(0.046)	(0.045)	(0.037)
Objective financial index			0.340***
			(0.026)
State of residence fixed effects		Individual states not shown	
pval demographics=0	0.000	0.000	0.004
pval cognitive skills=0	0.304	0.121	0.023
pval noncognitive skills=0	0.001	0.000	0.005
pval classical preferences=0	0.109	0.223	0.542
pval survey effort=0	0.898	0.104	0.028
pval state FE=0	0.000	0.000	0.000
mean(LHS)	0.531	0.504	0.504
N	3370	3370	3370

* 0.10 ** 0.05 *** 0.01. Standard errors, clustered on panelist, in parentheses. Each column presents results from a single pooled Obviously Related Instrumental Variables regression (equation 4 in the text) of the LHS variable described in the column label on the variables described in the row labels. Appendix Table 1 provides details on the other covariate definitions. Table 1 provides details on our B-count variable definitions; higher values indicate more behavioral biases. Table 3 provides details on our LHS variable definitions; higher values indicate better financial condition.

Appendix Table 9. B-count conditional correlations with financial outcomes: Unweighted vs. unweighted

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Financial outcome index includes:	*Objective*		*Subjective*		*Objective*		*Subjective*		*Objective*		*Subjective*	
B-count: Full	-0.061***	-0.065**	-0.084***	-0.114***								
	(0.018)	(0.032)	(0.018)	(0.035)								
B-count: Sparsity Broad					-0.090***	-0.152*	-0.128***	-0.238***				
					(0.028)	(0.078)	(0.028)	(0.090)				
B-count: Sparsity Narrow									-0.236***	-0.500	-0.328***	-0.753*
									(0.063)	(0.334)	(0.065)	(0.451)
d(LHS)/d(1 SD B-count)	-0.130	-0.138	-0.179	-0.242	-0.119	-0.202	-0.169	-0.315	-0.157	-0.332	-0.218	-0.500
Sampling weights?	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
Same/analogous specification in Table 4	col 1	col 1	col 2	col 2	col 4	col 4	col 5	col 5	col 7	col 7	col 8	col 8
mean(LHS)	0.531	0.484	0.504	0.489	0.531	0.484	0.504	0.489	0.531	0.484	0.504	0.489
N	3370	3370	3370	3370	3370	3370	3370	3370	3370	3370	3370	3370

* 0.10 ** 0.05 *** 0.01. Odd-numbered columns are reproduced from Table 4; even-numbered columns use the same specification as the preceding column but with sampling weights.

Appendix Table 10. OLS coefficients are attenuated and sensitive to dropping other covariates
(Compare to Table 5)

LHS=Subjective financial index	(1)	(2)	(3)	(4)	(5)	(6)
B-count: Full	-0.019***	-0.035***				
	(0.003)	(0.003)				
B-count: Sparsity Broad			-0.026***	-0.051***		
			(0.005)	(0.005)		
B-count: Sparsity Narrow					-0.067***	-0.083***
					(0.009)	(0.010)
Covariates in Appendix Table 2 included?	All	B-miss only	All	B-miss only	All	B-miss only
Comparable ORIV spec in Table 5	Pan A Col 1	Pan A Col 6	Pan B Col 1	Pan B Col 6	Pan C Col 1	Pan C Col 6
d(LHS)/d(1 SD B-count)	-0.040	-0.075	-0.034	-0.068	-0.044	-0.055
mean(LHS)	0.504	0.505	0.504	0.505	0.504	0.505
N	1685	1690	1685	1690	1685	1690

* 0.10 ** 0.05 *** 0.01. OLS, with standard errors clustered on panelist. Each column presents results from a single OLS regression, using both rounds of data (two obs per panelist), of the subjective financial index on the variables described in the row labels. "B-miss" refers to the count of missing behavioral biases.

Appendix Table 11. Identifying relationships between outcomes and B-counts: Sensitivity to covariate specifications (Same as Table 5, but with objective financial index as dependent variable instead of subjective financial index)

	LHS=Objective financial index								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Full B-Count									
Full B-count	-0.061***	-0.082***	-0.063***	-0.059***	-0.059***	-0.098***	-0.058***	-0.060***	-0.053***
	(0.018)	(0.016)	(0.018)	(0.015)	(0.018)	(0.012)	(0.018)	(0.019)	(0.017)
dY/d(1 SD B-count)	-0.130	-0.174	-0.134	-0.124	-0.125	-0.208	-0.122	-0.127	-0.112
Panel B. Sparsity Broad B-count									
Sparsity biases: attention+	-0.090***	-0.108***	-0.094***	-0.093***	-0.091***	-0.165***	-0.087***	-0.089***	-0.086***
	(0.028)	(0.026)	(0.029)	(0.025)	(0.028)	(0.022)	(0.028)	(0.029)	(0.028)
dY/d(1 SD B-count)	-0.119	-0.144	-0.124	-0.123	-0.120	-0.219	-0.115	-0.117	-0.114
Panel C. Sparsity Narrow B-count									
Sparsity biases: attention only	-0.236***	-0.169***	-0.240***	-0.237***	-0.241***	-0.233***	-0.235***	-0.236***	-0.232***
	(0.063)	(0.052)	(0.063)	(0.063)	(0.061)	(0.055)	(0.065)	(0.063)	(0.064)
dY/d(1 SD B-count)	-0.157	-0.112	-0.159	-0.157	-0.160	-0.155	-0.156	-0.157	-0.154
Missing bias count included?	yes	yes	yes	yes	yes	yes	yes	yes	yes
Demographics included?	yes	no	yes	yes	yes	no	yes	yes	yes
Classical preferences included?	yes	yes	no	yes	yes	no	yes	yes	yes
Cognitive skills included?	yes	yes	yes	no	yes	no	yes	yes	yes
Non-cognitive skills included?	yes	yes	yes	yes	no	no	yes	yes	yes
IV for B-count?	yes	yes	yes	yes	yes	yes	yes	yes	yes
IV for classical preferences	no	no	no	no	no	no	yes	no	yes
IV for cognitive skills	no	no	no	no	no	no	no	yes	yes
mean(LHS)	0.531	0.532	0.531	0.531	0.531	0.532	0.531	0.531	0.531
N	3370	3380	3370	3370	3370	3380	3370	3370	3370

* 0.10 ** 0.05 *** 0.01. Standard errors, clustered on panelist, in parentheses. Each panel-column presents results from a single ORIV regression of our objective financial index on the B-count described in the Panel title and row label and the other covariates described in rows at the bottom of the table. I.e., this table presents results for specifications identical to those in Table 5 except for the LHS variable.

Appendix Table 12. Identifying relationships between outcomes and B-counts: Unpacking the Math B-count

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Financial condition index includes:	*Objective*	*Objective*	*Subjective*	*Subjective*	*Objective*	*Objective*	*Subjective*	*Subjective*
Math biases (1)	-0.011	-0.018	-0.014	0.003				
	(0.044)	(0.040)	(0.041)	(0.038)				
Non-math biases (2)	-0.083***	-0.078***	-0.115***	-0.110***	-0.083***	-0.078***	-0.114***	-0.111***
	(0.029)	(0.027)	(0.030)	(0.028)	(0.029)	(0.027)	(0.032)	(0.031)
Exepcted Math Biases (3)					-0.016	-0.025	-0.040	-0.029
					(0.044)	(0.042)	(0.047)	(0.045)
Unexpected Math Biases (4)					0.012	0.010	0.108	0.128
					(0.121)	(0.107)	(0.138)	(0.128)
Fluid intelligence score	-0.005		-0.008*		-0.005		-0.006	
	(0.005)		(0.005)		(0.006)		(0.006)	
Fluid intelligence missing	0.037		-0.024		0.034		-0.038	
	(0.094)		(0.087)		(0.093)		(0.090)	
Numeracy score	0.004		-0.010		0.006		-0.003	
	(0.015)		(0.013)		(0.017)		(0.016)	
Numeracy missing	-0.022		-0.076		-0.015		-0.042	
	(0.057)		(0.047)		(0.064)		(0.057)	
Financial literacy score	0.031***		-0.006		0.031***		-0.007	
	(0.012)		(0.011)		(0.012)		(0.012)	
Financial literacy missing	0.033		-0.050		0.033		-0.047	
	(0.094)		(0.119)		(0.095)		(0.119)	
Stroop score	0.000		0.000		0.000		0.000	
	(0.000)		(0.000)		(0.000)		(0.000)	
Stroop missing	0.006		-0.031		0.006		-0.031	
	(0.038)		(0.041)		(0.038)		(0.043)	
pval (1)=(2)	0.240	0.304	0.091	0.051				
pval (2)=(3)					0.274	0.373	0.245	0.196
pval (2)=(4)					0.462	0.454	0.143	0.099
reproduced from Table 7?	col 1		col 2					
mean(LHS)	0.531	0.531	0.504	0.504	0.531	0.531	0.504	0.504
N panelists	843	843	843	843	843	843	843	843
N	3370	3370	3370	3370	3370	3370	3370	3370

* 0.10 ** 0.05 *** 0.01. One ORIV regression per column of the LHS variable described in the column label on the RHS variables described in the row label plus all of the additional covariates described in Appendix Table 1, *except* that even-numbered columns here do not include the cognitive skills covriates. Standard errors clustered on panelist.

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. (2017)), we find little if any present-bias on average, with a median discount bias of zero, and a 1pp mean tilt toward future bias.[3]

Indicators of behavioral deviations here are bi-directional: we label someone as present-biased (future-biased) if the average difference is >0 (<0). We deem present-bias the "standard" direction, since future-bias is relatively poorly understood.[4] Counting any deviation from time-

[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] Bradford et al. (2017) do find present-bias on average in their Qualtrics sample, classifying >50% as present-biased and 26% as future-biased.

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

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. Compared to prior prevalence estimates, our zero-threshold ones are in the middle of the range (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. Goda et al. use a different elicitation method—a "time-staircase" multiple price list (Falk et al. 2016)—and classify 55% of their nationally representative sample (from the ALP and another online panel) as present-biased. In the AS sample 14% exhibit any present-bias and 12% any future-bias.

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

Previous studies estimate relationships between directly elicited discounting biases and outcomes in broad samples (Bradford et al. 2017; Eisenhauer and Ventura 2006; Goda et al. 2017).[6] We use CTBs rather than Multiple Price Lists, test more flexible functional forms, and control for a much richer set of (behavioral) factors that could be correlated with both discounting and outcomes.[7]

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

[5] The 5th to 95th percentile interval AS' sample is [0.91, 1.11], as reported in their Table 3.
[6] Other papers have explored links between discounting biases and field behavior using direct elicitations on narrower samples, with narrower sets of covariates; see e.g., Chabris et al. (2008), Meier and Sprenger (2010), Burks et al. (2012), and Li et al. (2015).
[7] Other key differences include Bradford et al. (2017) lacking controls for cognitive skills, and Eisenhauer and Ventura (2006) only controlling for income.

(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).[8] Barcellos and Carvalho’s ALP subjects answered similar questions in their baseline survey, albeit with only a one-week instead of a five-week delay, with 6% exhibiting present-bias and 9% future-bias. Read and van Leeuwen (1998) offer actual snacks to a convenience sample of employees in Amsterdam but do not calculate individual-level measures of bias. They do find substantial present-bias on average. 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.[9] Each decision confronts respondents with a linear budget constraint under risk: subjects choose a point on the line, and then the computer randomly chooses whether to pay the point value of the x-axis or the y-axis. 1,270 of the 1,427 individuals taking our second Round 1

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

[9] We were quite constrained on survey time and hence conducted a pilot in which we tested the feasibility of capturing roughly equivalent information with fewer rounds. 58 pilot-testers completed 25 rounds, and we estimated the correlation between measures of 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.

module make all 11 decisions, and comprise our sample for measuring choice inconsistency.[10] See Data Appendix Figure 2 for an example.

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

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

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

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

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

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

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.

Choi et al. find that choice inconsistency with GARP is conditionally correlated with lower net worth, but that choice inconsistency with GARP+dominance avoidance is not.

D. Risk attitude re: certainty (certainty premium)

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

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

[13] Callen et al. describes its task as "a field-ready, two-question modification of the uncertainty equivalent presented in Andreoni and Sprenger (2016)."
[14] 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.

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 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.

Callen et al. find significant correlations between the CP and financial outcomes, in particular with avoiding late loan repayments,[15] but their data lack controls for cognitive skills and other B-factors.

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

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

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

[16] A related point is that there is no known “model-free” method of eliciting loss aversion (Dean and Ortoleva 2018).

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 2018).

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.

Despite the massive amount of work on loss aversion, research exploring links between directly elicited measures of loss aversion and field behavior is only beginning. von Gaudecker et al. (2011) do not explore links between loss aversion and field behavior. Dimmock and Kouwenberg (2010) do, like von Gaudecker et al. using CentERdata, but lack many important covariates. Fehr and Goette (2007) find that loss aversion moderates the effect of a wage increase, but their sample includes only bike messengers and lacks measures of many other potentially moderating factors. Abeler et al. (2011) find that loss aversion is strongly correlated with effort choices in the lab among their student sample, but again they lack data on many covariates of interest. Hwang (2016) uses the Abeler et al. measures to infer a strong correlation between loss aversion and insurance holdings in an earlier ALP module, but lacks many important covariates and the only other behavioral factor considered is an interaction between loss aversion and a measure of the Gambler's Fallacy (labeled "Heuristics" in the Hwang paper).[17]

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

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

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

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

in the most similar presentation in RW. As RW discuss, a new feature of task two is that AD sacrifices expected value in the second decision, not in the first. This implies that for all broad-bracketing risk averters AC is optimal: it generates the highest available expected value at no variance. 50% of our subjects choose AC, compared to only 33% in the most similar presentation in RW. I.e., 50% of our subjects do NOT broad-bracket in this task, compared to 67% in RW.

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

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

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

G. Ambiguity aversion

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

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

Despite the coarseness of our elicitation, comparisons to other work suggest that it produces reliable data. Our ambiguity aversion indicator correlates 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 .

Our examination of links to field behaviors builds on the papers by Dimmock and co-authors cited above, which estimate conditional correlations between ambiguity aversion and financial market behavior. We broaden the set of both outcomes and control variables (especially other B-factors).[19]

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-

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

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.

We are not aware of any prior work exploring conditional correlations between the sorts of overconfidence measures described above and field outcomes.

I. Non-belief in the Law of Large Numbers

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

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

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

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

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

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

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%.

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

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

We take a slice of Benjamin, Moore, and Rabin's (2017) elicitation for the fallacies:

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

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

Dohmen et al. (2009) measure 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).

Dohmen et al. also explore correlations between unemployment or bank overdrafts and their directly-elicited fallacy measures, conditional on age, gender, education, income, and wealth. They find evidence of positive correlations between hot-hand and unemployment and between cold-hand and overdrafting.

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).

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

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 (37% of the sample).[22] The median level difference between the correct and stated value is 500bp, with a mean of 1,042bp and SD of 1,879bp. Among those with known bias, we count as biased 70%, 64%, 47%, and 34% under error tolerance of zero, 100bp, 500bp, and 1000bp. Under these tolerances we count 3%, 13%, 41%, and 61% as unbiased, and 27%, 22%, 10%, and 3% as negatively biased. This is less EGB than Stango and Zinman (2009, 2011) see from questions in the 1983 Survey of Consumer Finances, where 98% of the sample underestimates, and the mean bias is 1,800bp or 3,800bp depending on the benchmark. The time frames of the questions differ, which may account for the difference (and is why we do not estimate an EGB structural model parameter to compare with our prior work or that of Levy and Tasoff).

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

Our second measure of EGB comes from a question popularized by Banks and Oldfield (2007) as part of a series designed to measure basic numeracy: "Let's say you have $200 in a

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

savings account. The account earns 10 percent interest per year. You don't withdraw any money for two years. How much would you have in the account at the end of two years?" 1,389 subjects answer this question (out of the 1,427 respondents to Module 352), and we infer an individual-level measure of "asset-side EGB" by comparing the difference between the correct future value ($242), and the future value supplied by the same individual.[23] We again bin individuals into underestimators (the standard bias), overestimators, unbiased, and unknown (14% of the sample).[24] Among those with known bias (N=1,222), the median bias is $0, with a mean of $2 and SD of $14.[25] 44% of our sample provides the correct FV. 47% of our sample underestimates by some amount, with most underestimators (29% of the sample) providing the linearized (uncompounded) answer of $240. Nearly all other underestimates provide an answer that fails to account for even simple interest; the most common reply in this range is "$220." Only 9% of our sample overestimates the FV, with small mass points at 244, 250, 400, and 440.

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

Goda et al. (2017) and Levy and Tasoff (2016) measure asset-side EGB using more difficult questions in their representative samples. They find that 9% and 11% overestimate FVs, while 69% and 85% underestimate. We do not construct an EGB parameter to compare to theirs,

[23] 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.

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

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

because our questions lack their richness and yield heavy mass points at unbiased and linear-biased responses.

The only prior paper we know looking directly at links between a measure of asset-side EGB and field outcomes is Goda et al., who use data on fewer behavioral factors. They find significant negative correlations between asset-side EGB and retirement savings.

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 2017; Kőszegi and Szeidl 2013; Schwartzstein 2014).

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

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

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

Ours is the first work we know of estimating conditional correlations between field outcomes and directly elicited measures of limited attention/memory in a broad sample.

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

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.

3. Survey Formatting and Non-classical Measurement Error

Data Appendix Table 3 provides reassurance that, *a priori*, there is little reason to think that low survey effort *per se* could contribute to a mechanical correlation between worse financial condition and more behavioral biases. A necessary condition for that confound is that it is somehow easier, from a survey effort perspective, to indicate worse than better financial condition. The table shows that this is unlikely to be the case, given how questions are scripted and response options are arrayed.

Data Appendix Table 4 provides some additional descriptive reassurance with data, showing a lack of systematic relationship between survey time spent (across all questions for both Round 1 modules) and financial condition responses, with the possible exception of the lowest time spent quintile.

As the main text details, we deal with this potential confound formally, by controlling flexibly for survey effort in both survey rounds with flexible controls for non-response and for survey time spent, and by dropping those in the lowest decile of time spent as a robustness check.

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. 2017. "Memory, Attention, and Choice." *NBER Working Paper #23256*.

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.

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

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

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

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

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

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

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

Dean, Mark, and Pietro Ortoleva. 2018. "Is It All Connected? A Testing Ground for Unified Theories of Behavioral Economics Phenomena."

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

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

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

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

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

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

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

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

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2016. "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences."

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

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. 2017. "Predicting Retirement Savings Using Survey Measures of Exponential-Growth Bias and Present Bias."

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

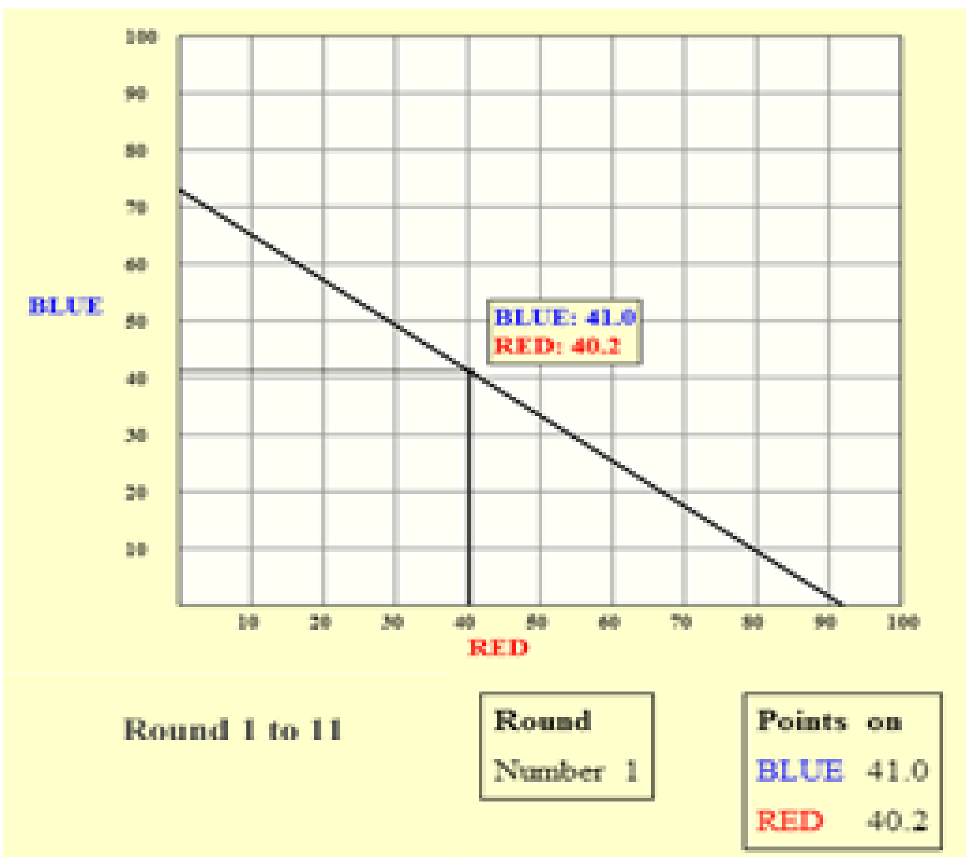
Graham, Liam, and Dennis J Snower. 2013. “Hyperbolic Discounting and Positive Optimal Inflation.” *Macroeconomic Dynamics* 17 (03): 591–620.
Grubb, Michael D. 2015. “Overconfident Consumers in the Marketplace.” *Journal of Economic Perspectives* 29 (4): 9–36.
Gul, Faruk. 1991. “A Theory of Disappointment Aversion.” *Econometrica: Journal of the Econometric Society* 59 (3): 667–86.
Gustman, Alan L, Thomas L Steinmeier, and Nahid Tabatabai. 2012. “Financial Knowledge and Financial Literacy at the Household Level.” *American Economic Review* 102 (3): 309–13.
Harris, Christopher, and David Laibson. 2013. “Instantaneous Gratification.” *The Quarterly Journal of Economics* 128 (1): 205–48.
Hwang, In Do. 2016. “Prospect Theory and Insurance Demand.”
İmrohoroğlu, Ayşe, Selahattin İmrohoroğlu, and Douglas H. Joines. 2003. “Time-Inconsistent Preferences and Social Security.” *The Quarterly Journal of Economics* 118 (2): 745–84.
Kahneman, D, and A Tversky. 1972. “Subjective Probability: A Judgement of Representativeness.” *Cognitive Psychology* 3: 430–54.
Kahneman, Daniel, and Amos Tversky. 1979. “Prospect Theory: An Analysis of Decision Under Risk.” *Econometrica* 47 (2): 263–91.
Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman. 2016. “Getting to the Top of Mind: How Reminders Increase Saving.” *Management Science* 62 (12): 3393–3411.
Kartashova, Katya. 2014. “Private Equity Premium Puzzle Revisited.” *American Economic Review* 104 (10): 3297–3334.
Kőszegi, Botond, and Matthew Rabin. 2009. “Reference-Dependent Consumption Plans.” *American Economic Review* 99 (3): 909–36.
Kőszegi, Botond, and Adam Szeidl. 2013. “A Model of Focusing in Economic Choice.” *The Quarterly Journal of Economics* 128 (1): 53–104.
Laibson, David. 1997. “Golden Eggs and Hyperbolic Discounting.” *Quarterly Journal of Economics* 112 (2): 443–77.
Larrick, Richard P, Katherine A Burson, and Jack B Soll. 2007. “Social Comparison and Confidence: When Thinking You’re Better than Average Predicts Overconfidence (and When It Does Not).” *Organizational Behavior and Human Decision Processes* 102 (1): 76–94.
Levy, Matthew, and Joshua Tasoff. 2016. “Exponential-Growth Bias and Lifecycle Consumption.” *Journal of the European Economic Association* 14 (3): 545–83.
Li, Ye, Jie Gao, A. Zeynep Enkavi, Lisa Zaval, Elke U. Weber, and Eric J. Johnson. 2015. “Sound Credit Scores and Financial Decisions despite Cognitive Aging.” *Proceedings of the National Academy of Sciences* 112 (1): 65–69.
Loomes, Graham, and Robert Sugden. 1986. “Disappointment and Dynamic Consistency in Choice under Uncertainty.” *The Review of Economic Studies* 53 (2): 271–82.
Meier, Stephan, and Charles Sprenger. 2010. “Present-Biased Preferences and Credit Card Borrowing.” *American Economic Journal: Applied Economics* 2 (1): 193–210.
Moore, Don A., and Paul J. Healy. 2008. “The Trouble with Overconfidence.” *Psychological Review* 115 (2): 502–17.
Moskowitz, T.J., and A. Vissing-Jorgensen. 2002. “The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle.” *American Economic Review* 92 (4): 745–78.

Neilson, William S. 1992. “Some Mixed Results on Boundary Effects.” *Economics Letters* 39 (3): 275–78.
Pérez Kakabadse, Alonso, and Ignacio Palacios Huerta. 2013. “Consumption and Portfolio Rules with Stochastic Hyperbolic Discounting.”
Rabin, Matthew. 2000. “Risk Aversion and Expected-Utility Theory: A Calibration Theorem.” *Econometrica* 68 (5): 1281–92.
Rabin, Matthew, and Dimitri Vayanos. 2010. “The Gambler’s and Hot-Hand Fallacies: Theory and Applications.” *Review of Economic Studies* 77 (2): 730–78.
Rabin, Matthew, and Georg Weizsäcker. 2009. “Narrow Bracketing and Dominated Choices.” *American Economic Review* 99 (4): 1508–43.
Read, Daniel, and Barbara van Leeuwen. 1998. “Predicting Hunger: The Effects of Appetite and Delay on Choice.” *Organizational Behavior and Human Decision Processes* 76 (2): 189–205.
Schmidt, Ulrich. 1998. “A Measurement of the Certainty Effect.” *Journal of Mathematical Psychology* 42 (1): 32–47.
Schwartzstein, Joshua. 2014. “Selective Attention and Learning.” *Journal of the European Economic Association* 12 (6): 1423–52.
Stango, Victor, and Jonathan Zinman. 2009. “Exponential Growth Bias and Household Finance.” *The Journal of Finance* 64 (6): 2807–49.
———. 2011. “Fuzzy Math, Disclosure Regulation, and Credit Market Outcomes: Evidence from Truth-in-Lending Reform.” *Review of Financial Studies* 24 (2): 506–34.
———. 2014. “Limited and Varying Consumer Attention: Evidence from Shocks to the Salience of Bank Overdraft Fees.” *Review of Financial Studies* 27 (4): 990–1030.
———. 2016. “Borrowing High vs. Borrowing Higher: Price Dispersion and Shopping Behavior in the U.S. Credit Card Market.” *Review of Financial Studies* 29 (4): 979–1006.
Sutter, Matthias, Martin G Kocher, Daniela Glätzle-Rützler, and Stefan T Trautmann. 2013. “Impatience and Uncertainty: Experimental Decisions Predict Adolescents’ Field Behavior.” *American Economic Review* 103 (1): 510–31.
Toubia, Olivier, Eric Johnson, Theodoros Evgeniou, and Philippe Delquié. 2013. “Dynamic Experiments for Estimating Preferences: An Adaptive Method of Eliciting Time and Risk Parameters.” *Management Science* 59 (3): 613–40.
Tversky, A., and D. Kahneman. 1981. “The Framing of Decisions and the Psychology of Choice.” *Science* 211 (4481): 453–58.
Tversky, Amos, and Daniel Kahneman. 1992. “Advances in Prospect Theory: Cumulative Representation of Uncertainty.” *Journal of Risk and Uncertainty* 5 (4): 297–323.
Von Gaudecker, Hans-Martin, Arthur Van Soest, and Erik Wengström. 2011. “Heterogeneity in Risky Choice Behavior in a Broad Population.” *The American Economic Review* 101 (2): 664–94.
Zinman, Jonathan. 2014. “Consumer Credit: Too Much or Too Little (or Just Right)?” *Journal of Legal Studies* 43 (S2 Special Issue on Benefit-Cost Analysis of Financial Regulation): S209–37.

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

	Token value 5 weeks from today	Token value 14 weeks from today	Decision: How many of the 100 tokens would you like to allocate to the **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

Data Appendix Table 2. Estimated distributions of individual-level present bias parameter from our money discounting data

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

Money discounting measured using 24 choices from Convex Time Budgets (Data Appendix I-A). All models estimated using the nonlinear least squares version of the model in Andreoni and Sprenger (2012).

Background consumption:

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

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

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

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

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

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

Data Appendix Table 3. Survey formatting should not bias toward worse financial condition reporting

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

Variables here are the components of our objective and subjective financial condition indices; see Appendix Table 3 for more details.

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

Data Appendix Table 4. Survey response time and financial condition components

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

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