Narrow Framing in Risk Aversion Experiments: Further Evidence from a Wide Replication

Ariel Gu
ariel.gu@uea.ac.uk
School of Economics
University of East Anglia

Matthew J. Walker
matt.walker@ncl.ac.uk
Economics Subject Group,
Newcastle University

Hong Il Yoo
h.i.yoo@lboro.ac.uk
Loughborough Business School
Loughborough University

July 30, 2025

Abstract

When evaluating risky options in experimental settings, do individuals integrate background finances with experimental earnings? Andersen et al. (2018, REStat) combine experimental data on lottery choices and administrative data on personal wealth in Denmark to show that individuals evaluate experimental payoffs in isolation. We replicate this finding using data from three experiments and survey-based measures of background finances for a representative Dutch sample. We show that the finding based on personal wealth extends to household wealth, personal income, and household income. The finding is also robust to different elicitation instruments, incentive structures, stake sizes, and interpersonal behavioral heterogeneity.

Keywords: expected utility; asset integration; narrow framing; choice under uncertainty; structural estimation; replication.

JEL Classification: C25; C91; C93; D81.

Acknowledgement: In this paper, we make use of data from the LISS panel (Longitudinal Internet Studies for the Social Sciences) managed by the non-profit research institute Centerdata (Tilburg University, the Netherlands). We are grateful to Centerdata for providing access to the data, and Jens Prüfer for bringing the LISS panel to our attention. We thank Heather Anderson, two anonymous reviewers, and members of the Newcastle University Applied Micro research group for helpful comments.

1 Introduction and Background

Data from lottery choice experiments are widely used to quantify economic measures of risk preferences, such as the Arrow-Pratt coefficient under expected utility theory (EUT). Typically, these experiments involve choosing between two lotteries that represent alternative probability distributions over prizes. The empirical analysis almost always specifies the prizes offered as the sole argument of the utility function, implicitly assuming a form of narrow framing whereby the decision maker evaluates those prizes separately from their background finances. Yet many non-experimental applications of EUT define the utility function over the decision maker's final monetary outcome, which includes background wealth or income. This would imply that the sum of net wealth (or non-experimental income) and lottery prizes is the correct argument. Axioms of EUT are silent on this question, admitting both perspectives. Empirical evaluation of whether the decision maker integrates their wealth and income has been limited because detailed measures of background finances are typically not collected as part of experiments.

Andersen et al. (2018) conduct a unique study which directly evaluates whether individuals integrate background finances with lottery prizes.[1] They combine data from a lottery choice experiment involving a general adult population in Denmark, with personal wealth measures derived from administrative data on assets and liabilities. They specify and estimate a multi-parameter utility function, which includes the extent of wealth integration as a parameter to be estimated. The results show that the participants integrate an economically and statistically insignificant fraction of their wealth, consistent with narrow framing. The remarkable combination of the data used means that a narrow replication of their work is difficult because access to the administrative records is restricted to researchers affiliated

[1]Their study is motivated by a broader theoretical issue compared to our direct focus on narrow framing. If the argument of the utility function is final wealth levels, some patterns of risk aversion commonly observed in small-stakes experiments imply implausible risk attitudes at large stakes (Hansson, 1988; Rabin, 2000). If the argument is changes in final wealth (*i.e.*, stakes themselves), or some non-additive aggregation of wealth and stakes, EUT is no longer subject to this type of payoff calibration critique (Cox and Sadiraj, 2006; Rubinstein, 2006).

with appropriate institutions in Denmark.[2]

Using publicly available data from the LISS panel in the Netherlands, we attempt a wide replication of their finding that individuals do not perfectly integrate background wealth with experimental prizes. Core survey modules in the panel collect a rich set of personal and household characteristics for a representative sample of Dutch individuals, including financial data pertaining to net wealth and income. Importantly, these modules can be readily merged with lottery choice experiments on risk preferences of the panel members, previously conducted by Noussair et al. (2014), Drerup et al. (2017), and Charness et al. (2020).

Besides involving a different population, our study distinguishes itself from Andersen et al. (2018) along several dimensions. First, our wealth measure is based on self-reports by individuals about their assets and liabilities around the experiment date. Although less accurate for tax purposes than administrative records, these self-reports are no less relevant for our empirical investigation: Due to the cognitive difficulties involved in computing one's exact financial position at any given point in time, choice behavior is just as likely to depend on what a person believes their net wealth to be as it is to depend on the objective wealth level.

Second, we complement the analysis based on wealth by testing for integration of background income. Economic studies and textbook examples often define the expected utility function on wealth (*e.g.*, Varian, 1992, §11; Rabin, 2000). However, there is also a long-standing and parallel tradition of considering income as the argument (*e.g.*, Friedman and Savage, 1952). For example, background income has been used as the argument in empirical studies estimating risk aversion using insurance data (*e.g.*, Cicchetti and Dubin, 1994; Cohen and Einav, 2007) or state-dependent variations of utility with respect to health (*e.g.*, Viscusi and Evans, 1990;

[2]The replication package for Andersen et al. (2018) includes a hybrid data file, which merges the actual experimental data with simulated wealth data. This enables us to check the computational validity of their programs written in the high-level user interface of *Stata*, despite the confidentiality of the original wealth data. As summarized at the beginning of Section 4 and detailed in the Online Appendix (Section A2), we successfully conduct a narrow replication of their results for this hybrid data, using our own programs written in a compiled language (*Mata*).

Evans and Viscusi, 1991; Gerking et al., 2017).[3]

Third, we study three experiments conducted in different years and with mostly non-overlapping participants. Lottery designs, incentive structures, and monetary stakes also varied within or across these experiments. These variations allow us to evaluate the robustness of findings to sample composition and survey instruments.

Finally, all three experiments can be linked to questionnaire data in the LISS panel, with two involving considerably more participants than the Danish study. We leverage these features to investigate whether the extent of integration varies with a measure of financial autonomy or across wealth (income) deciles, and whether it exhibits substantial latent heterogeneity across individuals.

Overall, we find that the results of Andersen et al. (2018) are widely replicable. Participants integrate only a tiny fraction of their wealth, which also tends to be statistically insignificant. Similarly, they integrate a very small fraction of their income, although this tends to be statistically significant. This finding holds across personal- and household-level measures of wealth and income. It is robust to the incentive structure, and there is no systematic evidence that it depends on the elicitation instrument, stake size, or behavioral heterogeneity across individuals.

2 Data

Andersen et al. (2018) use choice behavior in experimental tasks and administrative data on assets and liabilities maintained by Statistics Denmark to evaluate asset integration for 442 members of the Danish adult population in February and March 2015. For our wide replication, we make use of data from the LISS panel (Longitudinal Internet Studies for the Social Sciences) maintained by the non-profit research institute Centerdata (Tilburg University, the Netherlands). The panel provides a representative sample of Dutch individuals who participate in monthly online surveys. Households that could not otherwise participate are provided with a computer and Internet connection.

[3]Similarly, in finance, both investment returns per period and final wealth are accepted as valid arguments of the utility function (*e.g.*, Markowitz, 1991, 2014).

We collate data from three lottery choice experiments that elicited the risk preferences of a subset of panel members. Study 1 originates from Noussair et al. (2014), whose experiment in December 2009 invited a random sample of panel members aged 16 or above (data file *bl09a*). Each task in their experiment presented a choice between two equiprobable lotteries or between an equiprobable lottery and a varying sure payment. We focus on six tasks, documented as "Riskav 1-5" and "Ra_EU1" (p. 350), which are designed to elicit the usual notion of risk aversion that is equivalent to a concave utility function under EUT. Participants were randomly assigned to complete these tasks under one of three conditions, which we call *Real-Norm*, *Hypo-Norm*, and *Hypo-High*. Under the first two conditions, the prizes varied from €5 to €65; each participant in *Real-Norm* had a 1/10 chance of receiving a prize,[4] while *Hypo-Norm* presented hypothetical prizes. *Hypo-High* presented hypothetical prizes, scaled up by a factor of 150 (€750 to €9,750).

Study 2 originates from Charness et al. (2020), whose experiment in February 2012 also invited a random sample of panel members (data file *ga12a*). Participants were randomly assigned to one of five conditions, which differed by the risk preference elicitation format used; only two formats lend themselves to structural estimation of risk aversion and so are considered here. One format (p. 106) followed the design of Holt and Laury (2002), and presented 10 choices between two lotteries which varied on the probability scale, with fixed prizes ranging from €0.40 to €15.40. The other format (p. 107) was based on the design of Tanaka et al. (2010), and presented 28 choices between two lotteries which varied on the outcome scale, with prizes ranging from €1 to €340. In each condition, one of the participant's 10 or 28 choices was randomly selected to determine the prize received as payment.

Study 3 originates from Drerup et al. (2017), whose experiment in September 2013 invited a sample of panel members (one person per household) most involved in the financial administration of the household (data file *jy13a*). They admin-

[4]The actual prize was determined by randomly playing out one of 17 decision-making tasks, including the six second-order risk aversion tasks, six prudence tasks, and five temperance tasks.

istered a series of 5 interdependent hypothetical tasks to elicit a single certainty equivalent (p. 15 in their online appendix) as part of a larger experiment using real monetary rewards to elicit subjective beliefs. Each task presented a choice between an equiprobable lottery of €300 versus €0 and a varying sure payment; the sure payment iterated up or down from a starting value of €160 in the first task, depending on whether the participant accepted or rejected the lottery.

The experimental data is merged with personal and household characteristics available in core modules of the LISS panel. Noussair et al. (2014, p. 345) derive a personal net wealth measure by combining information on assets (savings, insurance policies, risky investments, and real estate investments) and liabilities (mortgages and other loans, credits, and debts) from the *Assets* and *Housing* modules.[5] We follow their approach and construct an analogous wealth variable using individual data from the closest available date to each experiment.[6] To approximate household-level wealth, we also aggregate the personal wealth measure within households.[7] This aggregated measure is noisy as not every household member responded to both component modules; nevertheless, it is interesting to consider whether individuals integrate pooled household finances. Personal and household income variables are available directly from the *Background Variables* module.

Table 1 summarizes the key participant characteristics in each study. These statistics are calculated over participants for whom we observe at least one of the four wealth and income measures. As the sampling frame of each study is small relative to the full panel, only 79 individuals are observed in all three studies.[8]

[5]Noussair et al. (2014) do not consider wealth integration.

[6]Wealth data is reported as of 31 December 2009 for Study 1; 31 December 2011 for Study 2; and 31 December 2013 for Study 3. Andersen et al. (2018) also use wealth data as of the year end closest to their experiment (31 December 2014).

[7]The *Assets* module asked only the person primarily responsible for managing household finances to report assets jointly owned by partners. This approach helps us avoid double counting of wealth within households, although it also necessitates a *caveat* that personal wealth in our analysis is better understood as assets over which the person has primary control, rather than assets which one personally owns. As reported below, we find no heterogeneity in wealth integration across individuals with and without financial control within any of the studies. We also find qualitatively similar results between Study 3, where nearly all respondents had financial control, and the other two studies, where around 60% had control.

[8]At the time of Study 1 (December 2009), the panel included 13,412 individuals. The correspond-

Table 1: Descriptive Statistics

Variable	Definition	Study	Mean	SD	N_{ind}
Wealth: Personal	Personal net wealth in €1,000s	1	44.15	1,180.79	2,835
		2	38.04	182.92	388
		3	44.37	260.28	1,916
Wealth: Household	Household net wealth in €1,000s	1	84.56	1,608.11	2,855
		2	71.67	248.28	392
		3	55.79	272.78	1,918
Income: Personal	Personal net monthly income in €1,000s	1	1.43	2.21	3,300
		2	1.38	0.99	415
		3	1.83	4.12	2,030
Income: Household	Household net monthly income in €1,000s	1	2.82	2.40	3,209
		2	2.78	1.30	402
		3	2.81	4.23	1,981
Female	=1 for women; =0 otherwise	1	0.53	0.50	3,428
		2	0.51	0.50	433
		3	0.48	0.50	2,119
Financial Control	=1 take care of finance in household matters; =0 otherwise	1	0.58	0.49	2,840
		2	0.61	0.49	388
		3	0.98	0.13	1,920
Household Head	=1 for head of household; =0 otherwise	1	0.57	0.50	3,428
		2	0.55	0.50	433
		3	0.74	0.44	2,119

Notes: Column Mean (SD) reports the sample mean of each variable across N_{ind} individuals. Study 1 refers to the lottery choice experiment designed to elicit risk aversion as reported in Noussair et al. (2014). Study 2 refers to the risk aversion measures "list of paired lotteries" and "multiple lists of paired lotteries" as reported in Charness et al. (2020). Study 3 refers to the hypothetical binary lottery choice experiment as reported in Drerup et al. (2017).

Net wealth is negative for 7.29% (9.22%) of observations on personal (household) wealth in Study 1; 6.93% (8.55%) in Study 2; and 6.70% (6.98%) in Study 3. Following Andersen et al. (2018), we recode all negative wealth values to 0 to avoid a negative argument of the utility function. Net income measures are already non-negative and do not require corresponding adjustments. The three study samples are balanced on gender. In Studies 1 and 2, about 60% of individuals take care of household finances and about 56% are household heads. These shares are higher in Study 3, which deliberately recruited members in charge of household finances.[9]

ing figure is 11,616 for Study 2 (February 2012) and 9,639 for Study 3 (September 2013). Pairwise comparisons also reveal a limited overlap of participants. We can match 191 individuals between Study 1 and Study 2; 970 between Study 1 and Study 3; and 156 between Study 2 and Study 3.

[9]The mean of the financial control variable is less than 100% in Study 3 because we use an updated measure associated with assets information as of December 2013.

3 Model Specification

Consider structural estimation of the expected utility model, using the following utility function of monetary outcome M

$$u[M] = \frac{M^{(1-\theta)}}{(1-\theta)}, \tag{1}$$

where $\theta \in (-\infty, +\infty)$ is the coefficient of relative risk aversion. Typically, M is directly equated with a prize offered in the choice task (*e.g.*, €5 or €65 when evaluating the expected utility of an even chance of either prize). This assumption would be inappropriate if the decision maker is concerned with their final wealth or income, such that the argument of $u[\cdot]$ should be modelled as $(W + M)$, where W represents their background wealth or income external to the experiment.

The key insight of Andersen et al. (2018) is that one can empirically evaluate the extent to which the decision maker integrates their real-life W with experimental M by adopting a more general utility function

$$U[W, M] = u[\omega W + M], \tag{2}$$

where $u[\cdot]$ is as defined earlier, and $\omega \in [0, \infty)$ is an unknown parameter to be estimated alongside θ. We code both W and M in €1,000s, comparable to 10,000s of Danish kroner in the original study. If $\omega = 0$, the decision maker would focus on the experimental task in isolation, as narrow framing predicts. *Partial* $(0 < \omega < 1)$ and *full* $(\omega = 1)$ integration of background finances are nested as special cases.

Andersen et al. (2018) further generalize equation (2) by allowing for the notion that W and M act as imperfect substitutes. This is achieved by applying a constant elasticity of substitution (CES) aggregator as follows

$$\begin{aligned} U[W, M] &= u\big[\kappa[W, M]\big], \\ \text{where} \quad \kappa[W, M] &= (\omega W^{\rho} + M^{\rho})^{(1/\rho)}, \end{aligned} \tag{3}$$

where $\rho \in (-\infty, 1)$ is an unknown parameter to be estimated along with θ and ω, and the ES between W and M is given by $1/(1-\rho)$. Identification of the ρ parameter, however, is expected to become empirically fragile as the estimate of ω falls into a neighborhood of 0: Theoretically, if $\omega = 0$, any value of ρ would be consistent with observed choice behavior and result in $(M^{\rho})^{1/\rho} = M$. We adopt the more restrictive but also more easily identified equation (2) as our baseline specification, and consider (3) for robustness checks.

We estimate the model parameters using the method of maximum likelihood. The likelihood function is specified as a non-linear index logit model, where the index function for each choice task is proportional to the expected utility difference between the two options presented in the task. This difference is scaled by a latent error parameter, denoted μ, which accounts for the decision maker's evaluative noise and is also an unknown parameter to be estimated. We provide a full discussion of our likelihood specification in the Online Appendix (Section A1).

4 Results

4.1 Narrow Replication

We implemented our likelihood evaluators in *Mata*, a compiled programming language integrated with *Stata*. A narrow replication of empirical results reported by Andersen et al. (2018) is difficult because it requires administrative data on personal assets and liabilities in Denmark, which are not readily accessible to research teams based outside of the country. Nevertheless, their replication package on the *Harvard Dataverse* includes a hybrid dataset, consisting of actual experimental data and simulated wealth data that one can use to evaluate the functionality of their programs, written in the regular, high-level *Stata* user interface. As detailed in the Online Appendix (Section A2), our programs produce materially the same results as Andersen et al. (2018), except for discrepancies concerning the ρ parameter in equation (3), symptomatic of its fragile identification in a neighborhood of $\omega = 0$.

4.2 Wide Replication: Baseline Findings

In the upper panel of Table 2, we use our *Mata* programs for equation (2) to estimate parameters measuring utility curvature (θ) and behavioral noise (μ), along with the weight on background wealth (ω). In Study 1, ω is estimated to be significantly greater than 0 at the 5% level, regardless of whether we consider personal or household wealth.[10] Qualitatively, this suggests that individuals integrate their wealth with experimental earnings. Quantitatively, however, each point estimate is smaller than 0.001, indicating that individuals integrate less than €1 for every €1,000 of wealth. In fact, these estimates have five leading zeros, indicating integration of less than €0.10 for every €1,000. In Study 2 and Study 3, the estimates of ω are not significantly greater than 0, and include 6 to 12 leading zeroes. On balance, our results are thus consistent with narrow framing ($\omega = 0$), whereby individuals evaluate risk aversion tasks in isolation from their background wealth.

The lower panel of Table 2 presents parallel results based on income. Statistically, we find support for partial integration across all columns, with ω estimated to be greater than 0 at the 7% level (Study 3, household income) or 5% level (all other cases). Quantitatively, however, the extent of income integration is limited and closely approximated by the hypothesis of narrow framing. The point estimate of 0.007 (0.003) in Study 1 indicates that individuals integrate €7 (€3) for every €1,000 in their personal (household) monthly income. The corresponding amount is less than €0.01 in Study 2 and less than €0.44 in Study 3.

The Online Appendix reports all items numbered with the *A* prefix hereafter. In Table A3, we estimate equation (3) which allows for imperfect substitutability between background finances and experimental earnings, captured by the estimated parameter ρ. In all but one case, two-sided tests fail to reject either $H_0 : \rho = 1$, under which the model simplifies to equation (2), or $H_0 : \omega = 0$, under which the identification of ρ is questionable. The exceptional case (Study 2, household

[10] For one-sided inferences, we halve the two-sided p-values reported in our tables.

Table 2: Baseline Specifications

Panel A. Wealth

	Study 1		Study 2		Study 3	
	(1) Personal	(2) Household	(3) Personal	(4) Household	(5) Personal	(6) Household
ω	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	(0.052)	(0.080)	(0.151)	(0.417)	(0.324)	(0.440)
θ	0.888	0.880	0.504	0.503	0.761	0.758
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
μ	0.180	0.180	0.297	0.298	1.619	1.622
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
N_{cho}	16,991	17,111	7,390	7,448	9,580	9,590
N_{ind}	2,835	2,855	388	392	1,916	1,918
$logL$	−10,335	−10,421	−4,771	−4,808	−6,389	−6,397

Panel B. Income

	Study 1		Study 2		Study 3	
	(1) Personal	(2) Household	(3) Personal	(4) Household	(5) Personal	(6) Household
ω	0.007	0.003	<0.001	<0.001	<0.001	<0.001
	(<0.001)	(<0.001)	(0.011)	(<0.001)	(0.042)	(0.125)
θ	1.073	1.085	0.505	0.494	0.831	0.832
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
μ	0.179	0.182	0.285	0.287	1.811	1.756
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
N_{cho}	19,761	19,216	7,822	7,548	10,148	9,903
N_{ind}	3,300	3,209	415	402	2,030	1,981
$logL$	−11,894	−11,637	−5,019	−4,846	−6,766	−6,608

Notes: <0.001 indicates a positive number smaller than 0.001. The results in parentheses are two-sided *p*-values, adjusted for clustering at the individual level; the null hypothesis assumes that the corresponding parameter is equal to zero. N_{cho} (N_{ind}) is the number of choice observations (individuals). $logL$ is the maximized log-likelihood.

income) converged to an implausible local maximum which implies convex utility and background income weighted almost twice as much as experimental earnings.

Our Study 1 sample pools choice tasks across three conditions that vary by the incentive structure. In Table A4, we estimate the baseline model in equation (2) separately for each condition. The use of real or hypothetical incentives *per se* does not influence our findings: For the *Real-Norm* and *Hypo-Norm* conditions (prizes from €5 to €65), we continue to find that individuals integrate a trivial fraction of their background finances with their experimental earnings. We find some evidence that presenting large stakes might prompt more integration: For the *Hypo-High* condition (prizes from €750 to €9,750), ω on personal income is estimated at

0.295, suggesting that individuals integrate €295 out of every €1,000. Even under this condition, however, the coefficients on household income and either measure of wealth remain small, suggesting integration of €2 or less per €1,000.

Our Study 2 sample pools choice tasks based on the designs of Holt and Laury (2002) and Tanaka et al. (2010). Table A5 reports results for each design-specific subsample. In both subsamples, ω is estimated to include at least 8 leading zeroes, regardless of the background finance measure used.

4.3 Wide Replication: Population Heterogeneity

We conclude by investigating behavioral heterogeneity across individuals. First, we consider observed heterogeneity. We hypothesize that the extent of wealth or income integration may depend on the individual's familiarity with or control over household finances. In Table A6, the ω parameter is allowed to vary by whether the individual takes care of financial matters in the household or is the head of household.[11] We do not find any meaningful variation.

We further examine whether the extent of integration varies across individuals with different observed levels of background finances. Those with no wealth or income do not contribute to the identification of ω, except indirectly by improving the statistical precision of other parameter estimates; by extension, the identification of ω largely relies on those with substantive finances to integrate. Our results could be biased if the risk aversion parameter θ covaries with ω across levels of wealth or income.[12] We therefore estimate a separate model for individuals with positive wealth or income, dividing each subsample into deciles of the corresponding financial measure. As Figure 1 illustrates for Study 1, the estimates of ω remain small across all deciles, showing no robust pattern of correlation with θ.[13]

[11] We encountered convergence issues with richer models that incorporate a wider range of characteristics considered by Noussair et al. (2014) or allow for demographic heterogeneity in θ and μ along with ω. The results for Study 2 in Table A6 illustrate qualitatively similar issues: For personal wealth, non-trivial p-values are repeated across multiple demographic coefficients, and for personal income, some p-values cannot be computed due to the nearly singular Hessian.

[12] We thank an anonymous reviewer for encouraging us to consider this aspect.

[13] We find qualitatively similar results for Study 2 and Study 3 (see Figure A1). The only exception

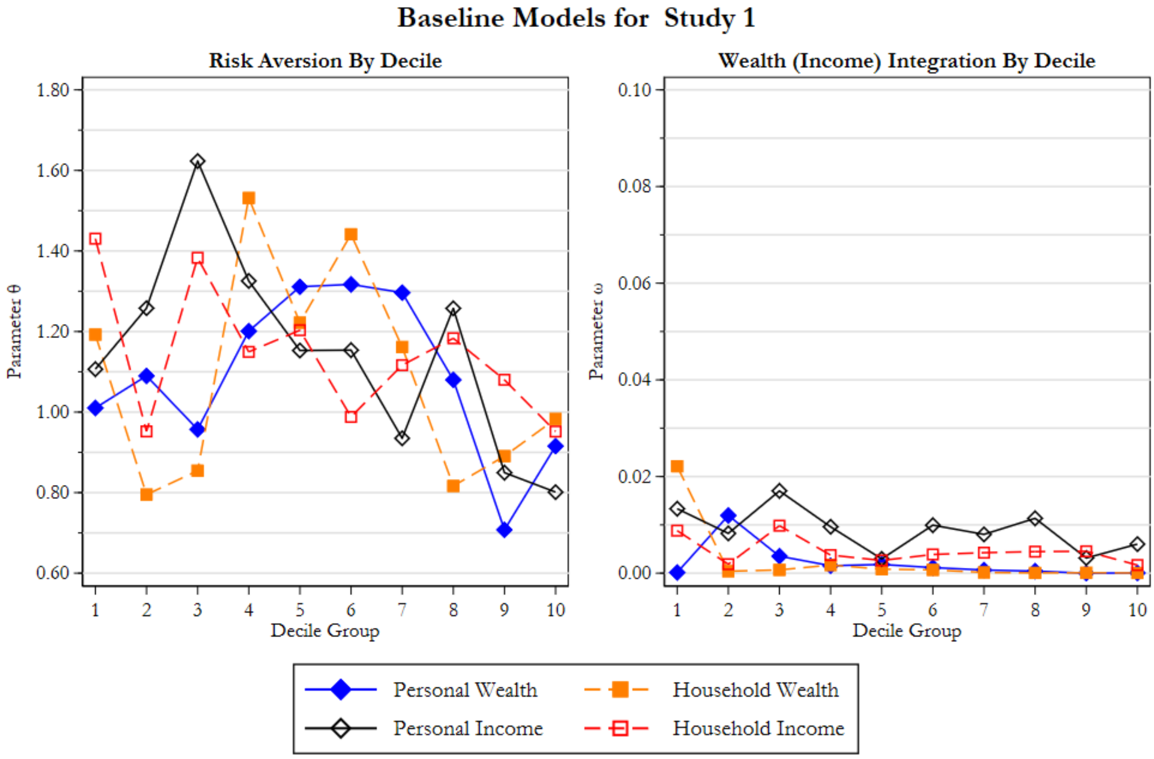

Figure 1: Preference Parameters by Wealth or Income Decile

Notes: Based on separate estimation of equation (2) for each decile group. Deciles are ordered from lowest to highest wealth or income, with group 1 (10) representing the least (most) well-off.

Finally, we complement these analyses of observed heterogeneity by estimating a random parameter model that accounts for unobserved heterogeneity. As detailed in Section A4, an overwhelming majority of participants in each study exhibit choice behavior consistent with small values of ω, despite wide variation in θ capturing both risk-seeking and risk-averse individuals.

In summary, our wide replication supports the finding of Andersen et al. (2018) that individuals integrate a very small, if any, fraction of their background finances with experimental earnings. Our results hold across various estimation samples and model configurations. This provides an important empirical validation of the common assumption of narrow framing in risk aversion experiments.

concerns three deciles of personal income in Study 3, where ω is estimated at around 0.2 (third and eighth) or 1 (sixth). This is likely due to numerical blow-up indicative of weak empirical identification: Estimates of θ for these deciles—-about 7 (third and eighth) or 30 (sixth)—-are also unusually large, considering that the implied utility function is effectively flat over the entire range of monetary stimuli in the study (€0 to €300) and that the baseline estimate is 0.831 (Table 2). No similar anomalies appear for the deciles of other financial measures.

References

Andersen, S., J. C. Cox, G. W. Harrison, M. I. Lau, E. E. Rutström, and V. Sadiraj (2018). Asset integration and attitudes toward risk: Theory and evidence. *Review of Economics and Statistics 100*(5), 816–830.

Charness, G., T. Garcia, T. Offerman, and M. C. Villeval (2020). Do measures of risk attitude in the laboratory predict behavior under risk in and outside of the laboratory? *Journal of Risk and Uncertainty 60*, 99–123.

Cicchetti, C. J. and J. A. Dubin (1994). A microeconometric analysis of risk aversion and the decision to self-insure. *Journal of Political Economy 102*(1), 169–186.

Cohen, A. and L. Einav (2007). Estimating risk preferences from deductible choice. *American Economic Review 97*(3), 745–788.

Cox, J. C. and V. Sadiraj (2006). Small- and large-stakes risk aversion: Implications of concavity calibration for decision theory. *Games and Economic Behavior 56*(1), 45–60.

Drerup, T., B. Enke, and H.-M. Von Gaudecker (2017). The precision of subjective data and the explanatory power of economic models. *Journal of Econometrics 200*(2), 378–389.

Evans, W. N. and W. K. Viscusi (1991). Estimation of state-dependent utility functions using survey data. *Review of Economics and Statistics 73*(1), 94–104.

Friedman, M. and L. J. Savage (1952). The expected-utility hypothesis and the measurability of utility. *Journal of Political Economy 60*(6), 463–474.

Gerking, S., W. Adamowicz, M. Dickie, and M. Veronesi (2017). Baseline risk and marginal willingness to pay for health risk reduction. *Journal of Risk and Uncertainty 55*, 177–202.

Hansson, B. (1988). Risk aversion as a problem of conjoint measurement. In P. Gardenfors and N.-E. Sahlin (Eds.), *Decisions, Probability, and Utility*. Cambridge: Cambridge University Press.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American Economic Review 92*(5), 1644–1655.

Markowitz, H. (1991). Foundations of portfolio theory. *Journal of Finance 46*(2), 469–477.

Markowitz, H. (2014). Mean–variance approximations to expected utility. *European Journal of Operational Research 234*(2), 346–355.

Noussair, C. N., S. T. Trautmann, and G. Van de Kuilen (2014). Higher order risk attitudes, demographics, and financial decisions. *Review of Economic Studies 81*(1), 325–355.

Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem. *Econometrica 68*(5), 1281–1292.

Rubinstein, A. (2006). Dilemmas of an economic theorist. *Econometrica 74*(4), 865–883.

Tanaka, T., C. F. Camerer, and Q. Nguyen (2010). Risk and time preferences: Linking experimental and household survey data from Vietnam. *American Economic Review 100*(1), 557–71.

Varian, H. (1992). *Microeconomic Analysis* (3 ed.). Norton New York.

Viscusi, W. K. and W. N. Evans (1990). Utility functions that depend on health status: Estimates and economic implications. *American Economic Review 80*(3), 353–374.

Online Appendix

Narrow Framing in Risk Aversion Experiments: Further Evidence from a Wide Replication

Ariel Gu

School of Economics, University of East Anglia, UK

Matthew J. Walker

Economics Subject Group, Newcastle University, UK

and

Hong Il Yoo

Loughborough Business School, Loughborough University, UK

July 30, 2025

A.1 Likelihood Specification

Each study sample that we analyze consists of observations on pairwise choices made by individuals in experimental decision tasks. Let $n \in \{1, 2, \cdots, N\}$ be the index of individuals and $t \in \{1, 2, \cdots, T\}$ be that of decision tasks. Suppose that in task t, individual n evaluates an option $j \in \{A, B\}$ which consists of a lottery that pays prize M^j_{1nt} with probability p^j_{1nt}, and an alternative prize M^j_{2nt} with the complementary probability $(1 - p^j_{1nt})$. This formulation nests sure payment as a special case of $p^j_{1nt} = 1$.

Define the expected utility of option j as

$$EU^j_{nt}[\omega, \theta] = p^j_{1nt} U[W_n, M^j_{1nt}] + (1 - p^j_{1nt}) U[W_n, M^j_{2nt}] \tag{A1}$$

where W_n is a measure of the individual's background wealth or income, and $U[\cdot]$ refers to the utility function in equation (2), which implicitly depends on the unknown preference parameters ω and θ. Equation (A1) explicitly states the resulting dependence of $EU[\cdot]$ on these parameters to facilitate further discussion.

Let y_{nt} denote a binary choice indicator which is equal to 1 if individual n chooses option A in task t, and 0 if their choice is B instead. A rigid theoretical prediction would be that the individual chooses A over B if the expected utility criterion favors the former: $y_{nt} = 1\big[\Delta EU_{nt}[\omega, \theta] > 0\big]$ in short, where $1[\cdot]$ is the indicator function and $\Delta EU_{nt}[\cdot] := EU^A_{nt}[\cdot] - EU^B_{nt}[\cdot]$.

To allow for errors in decision making, however, we cast the individual's choice problem in the latent variable framework that underpins the econometric analysis of binary choice data. Specifically, consider a latent variable y^*_{nt} that combines the EU criterion with evaluative noise ϵ_{nt}

$$y^*_{nt} = \Delta EU_{nt}[\omega, \theta] + \epsilon_{nt} \tag{A2}$$

where the noise term follows a zero-mean logistic distribution with an unknown scale parameter.[A1] Following Andersen et al. (2018), we assume that this scale is specified as the Contextual Utility model of Wilcox (2011), which exhibits heteroskedasticity over decision tasks depending on the range of the monetary prizes offered. Specifically, the scale of ϵ_{nt} is specified as $\mu(U[W_n, M_{nt}^{max}] - U[W_n, M_{nt}^{min}])$, where μ is a baseline noise parameter to be estimated, and M_{nt}^{max} and M_{nt}^{min} denote the largest and smallest prizes offered by the decision task, respectively.

Suppose that the observed y_{nt} relates to the latent y_{nt}^* by the usual observation rule: $y_{nt} = 1[y_{nt}^* > 0]$. Then, our assumptions thus far collectively imply that, conditional on parameters ω, θ, and μ, the likelihood of observing y_{nt} takes the form of a non-linear index logit model. To be specific, the conditional likelihood of y_{nt} is given by

$$P_{nt}[\omega, \theta, \mu] = \Lambda\Big[\frac{(2y_{nt} - 1)\Delta \mathrm{EU}_{nt}[\omega, \theta]}{\mu(U[W_n, M_{nt}^{max}] - U[W_n, M_{nt}^{min}])}\Big] \tag{A3}$$

where $\Lambda[z] = \exp[z]/(1 + \exp[z])$ denotes the standard logistic distribution function. To complete the likelihood specification, we assume that the error terms are independently distributed between individuals and across tasks, and specify the sample likelihood function as

$$L[\omega, \theta, \mu] = \prod_{n=1}^{N} \Big(\prod_{t=1}^{T} P_{nt}[\omega, \theta, \mu] \Big). \tag{A4}$$

In the case where the the utility function follows a more general functional form that permits W and M to be imperfect substitutes as in equation (3), the sample likelihood function can be constructed in substantively the same manner, by including ρ as an extra parameter to be estimated along with ω, θ, and μ.

Most of our estimates are computed by numerically solving for the values of the parameters ω, θ, and μ that maximize equation (A4), or its variant that additionally includes the parameter ρ. The exception concerns the results for the Holt-and-

[A1] The standard deviation of ϵ_{nt} is equal to $\pi/\sqrt{3}$ times this scale.

Laury subsample of Study 2 and for Study 3. In these cases, the Contextual Utility model fails to achieve convergence, presumably due to limited variations in the range of available monetary stimuli; we therefore adopt the Fechner error model popularized by Hey and Orme (1994), which assumes that the logistic error term has a constant scale of μ. The sample likelihood function under the Fechner error model has the same algebraic form as equation (A4), except that the denominator of the index function is simply equal to μ.

Additional References

Hey, J. D. and C. Orme (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica 62*(6), 1291–1326.

Wilcox, N. T. (2011). 'Stochastically more risk averse:' A contextual theory of stochastic discrete choice under risk. *Journal of Econometrics 162*(1), 89–104.

A.2 Narrow Replication Results for Hybrid Data

Andersen et al. (2018) study data from an experiment with adult residents of Greater Copenhagen, Denmark, conducted in February and March 2015. To test for narrow framing, they combine the experimental data with personal wealth data derived from administrative records on assets and liabilities, obtained from the Danish Civil Registration Office and the Danish Tax Authorities (pp. 820-821). These records are current as of the year end closest to the experiment (31 December 2014), similarly as the survey-based measures in our study. Access to the administrative data is granted upon application by researchers affiliated with appropriate Danish institutions, as outlined in the online appendix (Appendix E) to their study.

The data access requirements make it difficult for research teams based outside of Denmark, such as ours, to attempt a narrow replication of empirical findings reported by Andersen et al. (2018). Nevertheless, their replication code package in the *Harvard Dataverse* includes a hybrid data file (*Data.dta*) which can be used to test the functionality of their estimation programs.[A2] The data file merges actual experimental data with simulated data for personal wealth which preserve the key moments of the administrative data. This package can be accessed at doi.org/10.7910/DVN/SWRYPL.

In this section, we summarize our narrow replication results for their hybrid dataset. Compared to their likelihood evaluators programmed in the regular high-level user interface of *Stata*, our likelihood evaluators are directly written in *Mata*, a compiled programming language integrated with *Stata*. We compare maximum likelihood estimates obtained by applying the two sets of likelihood evaluators.

The first two columns of Table A1 report the results for a model which imposes *full* integration of background finances, corresponding to the likelihood evaluator *ML_crra_FAI* in the replication package. In the context of equations presented in

[A2] The hybrid data file includes 454 individuals, compared to 442 individuals reported in their empirical study, presumably because the administrative wealth data are not available for a small subset of participants in the experiment.

Table A1: Narrow Replication Results for Hybrid Data

	Full Integration		Partial Integration		
	(1) Ours	(2) Original	(3) Ours	(4) Original A	(5) Original B
ω	1.000	1.000	<0.001	<0.001	<0.001
	(const.)	(const.)	(0.797)	(0.120)	(0.799)
ρ	1.000	1.000	0.735	<0.001	0.731
	(const.)	(const.)	(0.252)	(<0.001)	(0.255)
θ	0.390	0.390	0.654	0.619	0.654
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
μ	0.193	0.193	0.082	0.080	0.082
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
N_{cho}	27,240	27,240	27,240	27,240	27,240
N_{ind}	454	454	454	454	454
$logL$	−17,878	−17,878	−17,476	−17,482	−17,476

Notes: Original A (column 4) uses the original code with original starting values for numerical maximization. Original B (column 5) uses the original code with starting values equal to our solution in column 3 instead. <0.001 indicates a positive number smaller than 0.001. (const.) indicates that the parameter is constrained to the value above. Other results in parentheses are two-sided p-values, adjusted for clustering at the individual level; the corresponding parameter is assumed to be zero under the null hypothesis, except for ρ which is assumed to be equal to unity. N_{cho} (N_{ind}) is the number of choice observations (individuals). $logL$ is the maximized log-likelihood.

our main text, this is a special case of (2) with $\omega = 1$, or alternatively that of (3) with $\omega = 1$ and $\rho = 1$. In this case, we obtain numerically identical estimates of the risk aversion parameter θ and the noise parameter μ regardless of whether we use our own code (column 1) or the original code (column 2).

The remaining columns of Table A1 report the results based on equation (3), corresponding to the likelihood evaluator *ML_crra_PAI* in the replication package. This specification includes both ω and ρ as free parameters to be estimated, thereby allowing for *partial* integration of background finances ($0 < \omega < 1$), as well as imperfect substitution between background finances and experimental earnings ($\rho \neq 1$). As we note in the main text, a potential challenge to estimating this general functional form is that the theoretical identification of the substitutability parameter ρ fails if the integration parameter ω is equal to 0. One may, therefore, expect the empirical identification of ρ to be fragile if the estimate of ω falls into a neighborhood of 0. Our replication results illustrate this point.

We first consider the estimates of ω in column 3 (our code) and column 4 (original code). Substantively, both estimates suggest that virtually no integration is taking place, with the estimate carrying 5 (column 3) or 4 (column 4) leading zeroes. Nevertheless, these estimates display discrepancies in terms of numerical values and statistical significance. With our code, ω is estimated to be 8.91×10^-5, and statistically indistinguishable from 0 at any conventional significance level: That is, we cannot reject the hypothesis of narrow framing, $H_0 : \omega = 0$. With the original code, it is estimated to be 1.36×10^{-4} and significantly greater than 0 at the 1% level: Therefore, we find evidence of partial integration in a narrow statistical sense, albeit the practical extent of it is close to no integration.

In comparison, the estimates of ρ in these columns lead to both substantively and statistically different conclusions. With our code, ρ is estimated to be 0.735, and we *cannot reject* the hypothesis of perfect substitution ($H_0 : \rho = 1$) at any conventional significance level. With the original code, ρ is estimated to be much smaller at 2.36×10^{-6}, implying a unit elasticity of substitution, and we *reject* the hypothesis of perfect substitution at the 1% level. Additionally, we also observe some numerical differences in the estimates for θ (0.654 with our code versus 0.619 with the original code) and μ (0.082 versus 0.080).

The differences between the two sets of estimates, however, can be readily reconciled by noticing that they have converged to different local maxima. The log-likelihood at the results found using our code is $-17,476$ (column 3), which is slightly better than $-17,482$ found using the original code (column 4). In the fifth and final column of Table A1, we apply the original code, setting our solution in column 3 as starting values for numerical maximization. In this case, the original code fails to improve on the starting log-likelihood of $-17,476$, and declares convergence at the same maximum: This new solution retains practically the same numerical estimates of ω, θ, and μ as column 3, but displays a discrepancy in ρ which is noticeable at the third decimal point (0.735 in column 3 versus 0.731 in column 5), further illustrating the fragile identification of the ρ parameter.

In the mirrored case where we start our code using the original code's solution in column 4 as starting values, the estimation run fails to achieve convergence

Table A2: Narrow Replication Results for Hybrid Data at $\rho = 1$

	(1) Ours	(2) Original
ω	<0.001	<0.001
	(0.452)	(0.450)
θ	0.648	0.648
	(<0.001)	(<0.001)
μ	0.083	0.083
	(<0.001)	(<0.001)
N_{cho}	27,240	27,240
N_{ind}	454	454
$logL$	-17,476	-17,476

Notes: <0.001 indicates a positive number smaller than 0.001. The results in parentheses are two-sided p-values, adjusted for clustering at the individual level; the null hypothesis assumes that the corresponding parameter is equal to zero. N_{cho} (N_{ind}) is the number of choice observations (individuals). $logL$ is the maximized log-likelihood.

as it encounters a flat region of the log-likelihood function. This flat region occurs at $-17,481$, which is slightly better compared to the starting log-likelihood of $-17,482$. Together with the results in column 5, this suggests that the original code has prematurely declared convergence in column 4, identifying a relatively flat region of the log-likelihood as a maximum.[A3]

Equation (2), which forms the basis for our analysis in the main text, allows for partial asset integration under the assumption that the two sources of money are perfect substitutes. It therefore represents an intermediate case between the full and partial integration models in Table A1. Andersen et al. (2018) do not directly implement this specification, but their original code *ML_crra_PAI* can be readily revised to estimate it by imposing the constraint $\rho = 1$. In Table A2, we report results using our own code and the constrained version of the original code. As the model is no longer subject to the potential identification failure due to the ρ parameter, both codes find solutions converging to the same maximum, yielding materially the same parameter estimates.

[A3]In both columns 3 and 4, we have used the starting values available in the original replication package. Therefore, their discrepancy cannot be attributed to the use of different starting values.

A.3 Additional Wide Replication Results for LISS Panel

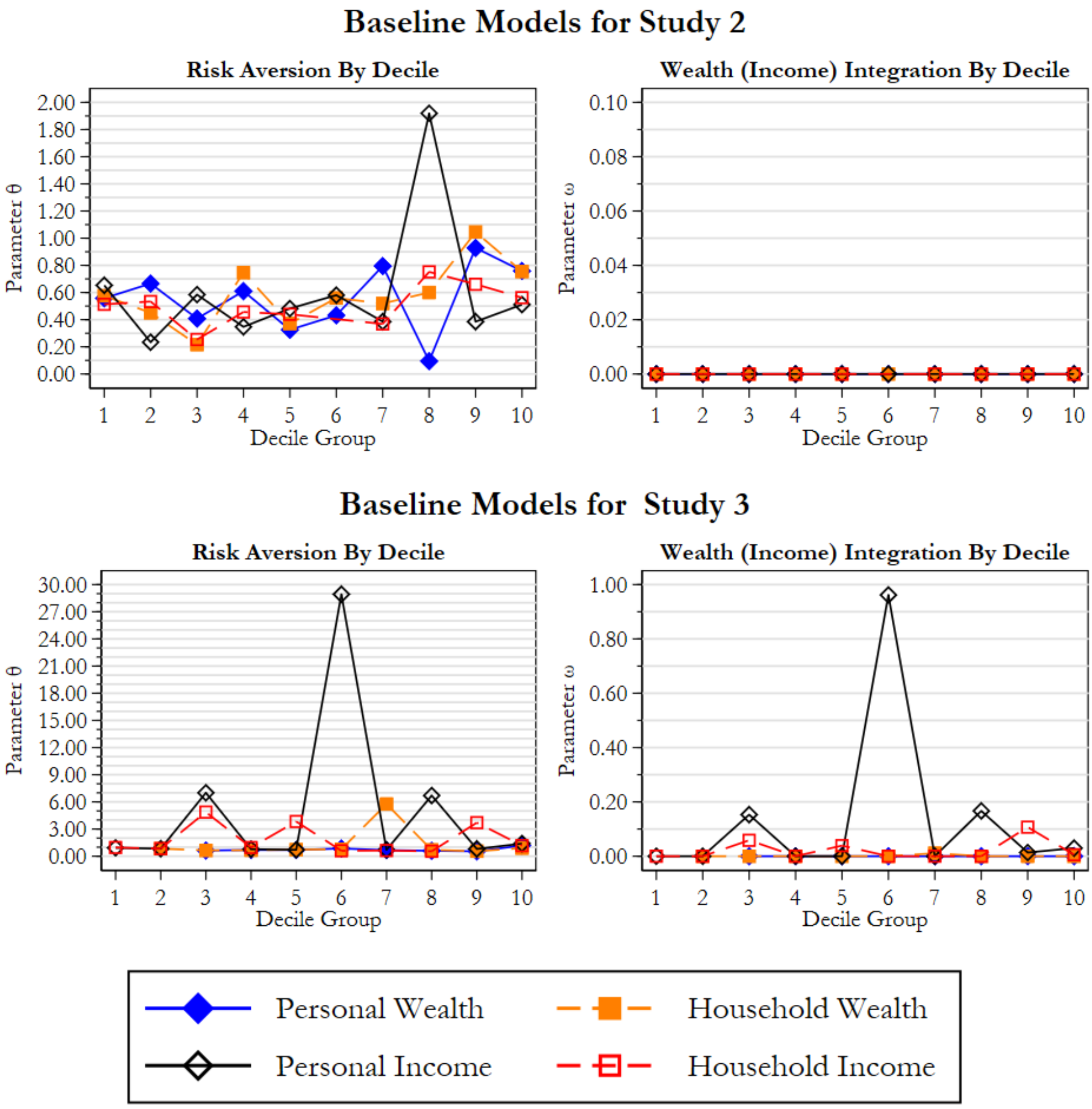

Figure A1: Preference Parameters by Wealth or Income Decile

Notes: Based on separate estimation of equation (2) for each decile group. Deciles are ordered from lowest to highest wealth or income, with group 1 (10) representing the least (most) well-off.

Table A3: Models Allowing for Imperfect Substitution Between W and M

Panel A. Wealth

	Study 1		Study 2		Study 3	
	(1) Personal	(2) Household	(3) Personal	(4) Household	(5) Personal	(6) Household
ω	0.023	0.019	0.110	0.143	<0.001	<0.001
	(0.088)	(0.135)	(0.366)	(0.190)	(0.308)	(0.436)
ρ	0.533	0.484	0.270	0.275	0.987	0.993
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.002)	(0.028)
θ	0.923	0.917	0.407	0.328	0.760	0.758
	(<0.001)	(<0.001)	(<0.001)	(0.001)	(<0.001)	(<0.001)
μ	0.178	0.179	0.296	0.293	1.609	1.621
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
N_{cho}	16,991	17,111	7,390	7,448	9,580	9,590
N_{ind}	2,835	2,855	388	392	1,916	1918
$logL$	−10,315	−10,404	−4,763	−4,791	−6,389	−6,397

Panel B. Income

	Study 1		Study 2		Study 3	
	(1) Personal	(2) Household	(3) Personal	(4) Household	(5) Personal	(6) Household
ω	0.007	0.051	0.127	1.750	0.001	0.001
	(<0.001)	(0.186)	(0.294)	(0.042)	(0.017)	(0.061)
ρ	1.000	0.636	0.342	0.290	0.953	0.920
	(0.758)	(0.567)	(<0.001)	(<0.001)	(0.378)	(0.024)
θ	1.073	1.134	0.395	-1.882	0.831	0.831
	(<0.001)	(<0.001)	(0.001)	(<0.001)	(<0.001)	(<0.001)
μ	0.179	0.180	0.285	0.283	1.808	1.749
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
N_{obs}	19,761	19,216	7,822	7,548	10,148	9,903
N_{ind}	3,300	3,209	415	402	2,030	1,981
$logL$	−11,894	−11,629	−5,016	−4,826	−6,766	−6,608

Notes: <0.001 indicates a positive number smaller than 0.001. The results in parentheses are two-sided p-values, adjusted for clustering at the individual level; the corresponding parameter is assumed to be zero under the null hypothesis, except for ρ which is assumed to be equal to unity. N_{cho} (N_{ind}) is the number of choice observations (individuals). $logL$ is the maximized log-likelihood.

Table A4: Study 1 — Subsamples Facing Different Incentives

Panel A. Wealth

	Real-Norm		Hypo-Norm		Hypo-High	
	(1) Personal	(2) Household	(3) Personal	(4) Household	(5) Personal	(6) Household
ω	<0.001	<0.001	<0.001	<0.001	0.002	0.001
	(0.080)	(0.421)	(0.301)	(0.316)	(0.231)	(0.160)
θ	0.771	0.757	0.693	0.697	1.246	1.248
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
μ	0.172	0.173	0.165	0.165	0.200	0.200
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
N_{cho}	6,826	6,874	5,263	5,305	4,902	4,932
N_{ind}	1,139	1,147	878	885	1,003	823
$logL$	−4,227	−4,268	−3,273	−3,299	−2,730	−2,745

Panel B. Income

	Real-Norm		Hypo-Norm		Hypo-High	
	(1) Personal	(2) Household	(3) Personal	(4) Household	(5) Personal	(6) Household
ω	0.005	0.001	0.004	<0.001	0.295	0.002
	(<0.001)	(0.139)	(0.006)	(0.850)	(0.042)	(0.931)
θ	0.976	0.866	0.872	0.678	1.415	1.216
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
μ	0.174	0.176	0.164	0.161	0.207	0.205
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
N_{cho}	7,946	7,701	6,142	6,004	5,673	5511
N_{ind}	1,326	1,285	1,026	1,003	948	921
$logL$	−4,890	−4,776	−3,800	−3,730	−3,177	−3,106

Notes: <0.001 indicates a positive number smaller than 0.001. The results in parentheses are two-sided *p*-values, adjusted for clustering at the individual level; the null hypothesis assumes that the corresponding parameter is equal to zero. N_{cho} (N_{ind}) is the number of choice observations (individuals). $logL$ is the maximized log-likelihood. Standard errors reported in parentheses, adjusted for clustering at the individual level.

Table A5: Study 2 — Subsamples Facing Different Types of Choice Tasks

Panel A. Wealth

	Holt and Laury		Tanaka et al.	
	(1) Personal	(2) Household	(3) Personal	(4) Household
ω	<0.001	<0.001	<0.001	<0.001
	(0.241)	(0.418)	(0.716)	(0.025)
θ	0.759	0.740	0.431	0.425
	(<0.001)	(<0.001)	(<0.001)	(<0.001)
μ	0.238	0.209	0.184	0.184
	(0.056)	(0.037)	(<0.001)	(<0.001)
N_{cho}	1,930	1,960	5,460	5,488
N_{ind}	193	196	195	196
$logL$	−1,095	−1,109	−3,524	−3,543

Panel B. Income

	Holt and Laury		Tanaka et al.	
	(1) Personal	(2) Household	(3) Personal	(4) Household
ω	<0.001	<0.001	<0.001	<0.001
	(0.122)	(0.046)	(0.248)	(0.102)
θ	0.708	0.722	0.433	0.413
	(<0.001)	(<0.001)	(<0.001)	(<0.001)
μ	0.161	0.176	0.187	0.187
	(0.022)	(0.028)	(<0.001)	(<0.001)
N_{cho}	2,110	2,060	5,712	5,488
N_{ind}	211	206	204	196
$logL$	−1,169	−1,143	−3,696	−3,547

Notes: <0.001 indicates a positive number smaller than 0.001. The results in parentheses are two-sided *p*-values, adjusted for clustering at the individual level; the null hypothesis assumes that the corresponding parameter is equal to zero. N_{cho} (N_{ind}) is the number of choice observations (individuals). $logL$ is the maximized log-likelihood. Standard errors reported in parentheses, adjusted for clustering at the individual level.

Table A6: Demographic Specifications

Panel A. Wealth

		Study 1		Study 2		Study 3	
		(1) Personal	(2) Household	(3) Personal	(4) Household	(5) Personal	(6) Household
ω:	constant	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
		(0.449)	(0.712)	(0.984)	(0.581)	(0.759)	(0.838)
	female	$<\underline{0.001}$	$<\underline{0.001}$	$<\underline{0.001}$	<0.001	$<\underline{0.001}$	$<\underline{0.001}$
		(0.449)	(0.712)	(0.984)	(0.581)	(0.581)	(0.838)
	financial control	$<\underline{0.001}$	<0.001	$<\underline{0.001}$	<0.001	$<\underline{0.001}$	$<\underline{0.001}$
		(0.864)	(0.563)	(<0.001)	(0.315)	(0.760)	(0.840)
	household head	$<\underline{0.001}$	<0.001	$<\underline{0.001}$	<0.001	<0.001	$<\underline{0.001}$
		(0.668)	(0.809)	(0.984)	(0.100)	(0.786)	(0.950)
θ		0.890	0.886	0.504	0.503	0.756	0.756
		(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
μ		0.180	0.180	0.297	0.297	1.587	1.588
		(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
N_{cho}		16,991	16,997	7,390	7,390	9,580	9,590
N_{ind}		2,835	2,836	388	388	1,916	1,918
$logL$		−10,319	−10,331	−4,771	−4,767	−6,383	−6,390

Panel B. Income

		Study 1		Study 2		Study 3	
		(1) Personal	(2) Household	(3) Personal	(4) Household	(5) Personal	(6) Household
ω:	constant	0.013	0.004	<0.001	<0.001	0.001	<0.001
		(0.016)	(0.011)	(N/A)	(0.983)	(0.603)	(0.558)
	female	−0.009	−0.004	<0.001	$<\underline{0.001}$	−0.001	$<\underline{0.001}$
		(0.070)	(0.011)	(N/A)	(>0.999)	(0.614)	(0.575)
	financial control	0.002	0.005	<0.001	$<\underline{0.001}$	$<\underline{0.001}$	<0.001
		(0.573)	(0.219)	(N/A)	(0.999)	(0.876)	(0.904)
	household head	−0.004	−0.001	$<\underline{0.001}$	$<\underline{0.001}$	$<\underline{0.001}$	<0.001
		(0.450)	(0.575)	(N/A)	(>0.999)	(0.909)	(0.547)
θ		1.077	0.982	0.525	0.512	0.827	0.844
		(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
μ		0.178	0.178	0.296	0.297	1.714	1.684
		(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
N_{cho}		16,277	15,863	6,976	6,712	9,155	8,930
N_{ind}		2,716	2,647	370	358	1,831	1,786
$logL$		−9,731	−9,532	−4,491	−4,323	−6,079	−5,928

Notes: <0.001 indicates a positive number smaller than 0.001; >0.999 indicates a positive number which is greater than 0.999 but smaller than 1; and finally $<\underline{0.001}$ indicates a negative number whose absolute value is smaller than 0.001. The results in parentheses are two-sided p-values, adjusted for clustering at the individual level; the null hypothesis assumes that the corresponding parameter is equal to zero; (N/A) indicates non-calculable. N_{cho} (N_{ind}) is the number of choice observations (individuals). $logL$ is the maximized log-likelihood.

A.4 "Wider" Replication Results for LISS Panel

Our wide replication supports the finding of Andersen et al. (2018) that individuals integrate a very small, if any, fraction of their personal wealth with experimental earnings. In particular, we have demonstrated that this finding is robust to (1) a different population; (2) three different samples of mostly non-overlapping individuals; (3) alternative measures of wealth, as well as income, at the personal and household levels; (4) different choice set designs, incentive structures, and stake sizes; (5) controlling for gender, financial control, and household headship; and (6) behavioral heterogeneity across deciles of the income and wealth distributions. Despite its broad scope, our replication is based on the same type of structural model used in the original study, which limits us to capturing unobserved heterogeneity across individuals only indirectly, through subsample analyses such as we have conducted to address (4) and (6).

In this appendix, we widen the scope of our replication further to incorporate an alternative model which directly accounts for unobserved heterogeneity across individuals. Each structural parameter is indexed by the individual subscript n henceforth, to denote the individual-specific values of the wealth or income integration parameter (ω_n), the risk aversion parameter (θ_n), and the behavioral noise parameter (μ_n). Two of our three samples include considerably more participants than the original study, providing a suitable opportunity for exploring the empirical identification of population heterogeneity.[A4] Specifically, the study by Andersen et al. (2018) is based on 442 individuals. Depending on the background finance measure, the number of individuals in our estimation sample ranges from 2,835 to 3,300 in Study 1, 388 to 415 in Study 2, and 1,916 to 2,030 in Study 3.

We now specify the random parameter model of interpersonal heterogeneity (McFadden and Train, 2000) that we estimate by applying the method of simulated likelihood. We assume that the population consists of three types—or classes, in

[A4]We thank an anonymous referee for pointing this out.

the terminology of latent class modeling (*e.g.* Doiron and Yoo, 2020)—of individuals: Class 1 includes narrow framers ($\omega_n = 0$), class 2 includes individuals who integrate €300 for every €1,000 in wealth or income ($\omega_n = 0.3$), and finally class 3 includes individuals whose integration parameter $\bar{\omega}$ is to be estimated from the data ($\omega_n = \bar{\omega}$). The value of ω_n for class 2 is the highest that we have plausibly estimated in our replication results so far.[A5] Let $\pi_c \in (0, 1)$ denote the population share of each class $c \in \{1, 2, 3\}$, where $\sum_{c=1}^{3} \pi_c = 1$; we estimate π_1 and π_2 from the data, and derive π_3 from this add-up restriction.[A6] We further assume that the population distribution of risk aversion parameters θ_n and log behavioral noise parameters $\ln[\mu_n]$ is multivariate normal. Let $f\left[\theta_n, \ln[\mu_n] \middle| \boldsymbol{m}, \boldsymbol{V}\right]$ be the density function representing this distribution, where $\boldsymbol{m}$ and $\boldsymbol{V}$ are its population mean and variance-covariance matrix that we estimate from the data. Although it is possible to further generalize this model specification, our experience suggests that empirical identification of the resulting model is likely to be fragile.[A7] In fact, as we discuss below, even estimating the current specification proved challenging in the cases of Study 2 and Study 3, requiring us to impose an additional constraint to enable the numerical optimizer to achieve convergence.

Besides the new assumptions concerning population heterogeneity, we maintain the same modeling assumptions as earlier. Thus, conditional on ω_n, θ_n, and μ_n, the sample likelihood function takes the same form as $L[\cdot]$ in equation (A4). By integrating out the random parameters from each individual's contribution to this

[A5]We have found this for the "Hypo-High" subsample of Study 1 in Table A4. The only higher estimate in our analysis is for the sixth of personal income deciles in Study 3 (Figure A1), where the ω parameter is almost equal to 1. As we discuss in the final footnote to the main text, however, we consider this as a numerical blow-up symptomatic of weak empirical identification: We do not find a similarly large estimate of ω using other measures of background finances, and furthermore this estimate is paired with an unusually large estimate of the risk aversion parameter—namely θ is almost equal to 30—which implies that the utility function is effectively flat.

[A6]Similar to the choice of the base group for a categorical variable in linear regression models, the choice of which two shares to estimate is inconsequential to substantive estimation results.

[A7]For example, the numerical optimizer failed to achieve convergence when we specified the ω_n parameters of class 1 and class 2 as free parameters, instead of pre-specifying them as narrows framers and partial integrators. Similarly, we encountered convergence failures when we modeled heterogeneity in ω_n using a log-normal or logit-normal distribution to allow for a continuum of individual types, rather than the discrete three-point distribution on $\{0, 0.3, \bar{\omega}\}$.

conditional likelihood, we obtain the unconditional likelihood function $G[\cdot]$ used in the model estimation

$$G[\boldsymbol{\gamma}] = \prod_{n=1}^{N} \left(\sum_{c=1}^{3} \pi_c \Big(\int \Big(\prod_{t=1}^{T} P_{nt}[\omega_c, \theta_n, \mu_n] \Big) f\left[\theta_n, \ln[\mu_n] \middle| \boldsymbol{m}, \boldsymbol{V}\right] d\theta_n d\mu_n \Big) \right) \qquad \text{(A5)}$$

where $\boldsymbol{\gamma} = \{\bar{\omega}; \pi_1, \pi_2; \boldsymbol{m}, \boldsymbol{V}\}$ collects the population-level parameters, and $\{\omega_1, \omega_2, \omega_3\}$ is set to $\{0, 0.3, \bar{\omega}\}$. As with mixed logit models (McFadden and Train, 2000; Keane and Wasi, 2013), the individual-level integrals above do not have analytic expressions. We therefore simulate $G[.]$ by using shuffled Halton sequences to generate 500 draws from $f[\cdot]$ per individual.

In Table A7, we estimate this model of population heterogeneity for Study 1, using each of the four wealth and income measures in turn. Consider first the risk aversion parameter θ_n. Our results are qualitatively similar across all four specifications, and to existing findings on preference heterogeneity in a general population (von Gaudecker et al., 2011; Harrison et al., 2020). Although the average decision maker is risk-averse, there is considerable variation in the degree of risk aversion across individuals: we find a positive population mean, along with a standard deviation which is comparable in magnitude. Combined with the assumed marginal distribution of θ_n, these mean and standard deviation values suggest that some 17% of the population are risk seekers.[A8]

As may be expected given our earlier estimates for various subsamples, we find limited heterogeneity in the integration parameter ω_n, reinforcing our conclusion that narrow framing is a useful modeling assumption. The estimated population shares indicate that narrow framers ($\omega_n = 0$) make up over 75% (wealth specifications) or over 65% (income specifications) of the population. Almost all of the rest belong to the class with the unconstrained integration parameter, $\omega_n = \bar{\omega}$: this parameter is estimated to be small fractions, especially in personal wealth ($\bar{\omega} = 0.001$)

[A8]Since risk seekers have $\theta_n < 0$, their population share is equal to $\Phi[(0 - \text{EV})/\text{SD}]$, where $\Phi[\cdot]$ denotes the standard normal distribution function; and EV and SD are the mean and standard deviation of θ_n reported in Table A7.

Table A7: Three-class Models for Study 1

	Wealth		Income	
	(1) Personal	(2) Household	(3) Personal	(4) Household
SH: $\omega = 0$	0.764	0.773	0.671	0.654
	(<0.001)	(<0.001)	(<0.001)	(<0.001)
SH: $\omega = 0.3$	<0.001	<0.001	<0.001	<0.001
	(0.656)	(N/A)	(0.270)	(0.529)
SH: $\omega = \bar{\omega}$	0.236	0.227	0.329	0.346
	(<0.001)	(<0.001)	(<0.001)	(<0.001)
$\bar{\omega}$	0.002	0.001	0.028	0.012
	(0.002)	(0.122)	(<0.001)	(<0.001)
EV: θ	0.872	0.874	1.012	1.021
	(<0.001)	(<0.001)	(<0.001)	(<0.001)
EV: $\ln[\mu]$	-2.573	-2.573	-2.537	-2.527
	(<0.001)	(<0.001)	(<0.001)	(<0.001)
SD: θ	0.930	0.931	1.051	1.076
	(<0.001)	(<0.001)	(<0.001)	(<0.001)
SD: $\ln[\mu]$	0.711	0.709	0.682	0.667
	(<0.001)	(<0.001)	(<0.001)	(<0.001)
$\text{COR}_{\theta,\ln[\mu]}$	-0.191	-0.200	-0.141	-0.128
	(<0.001)	(<0.001)	(0.022)	(0.074)
N_{cho}	16,991	17,111	19,761	19,216
N_{ind}	2,835	2,855	3,300	3,209
$logL$	-8,528	-8,614	-9,923	-9,686

Notes: SH is the population share of decision makers whose ω is equal to the corresponding value. EV and SD are the population mean and standard deviation of the corresponding parameter. $\text{COR}_{\theta,\ln[\mu]}$ measures the population correlation between the two random parameters. <0.001 indicates a positive number smaller than 0.001. The results in parentheses are two-sided *p*-values, adjusted for clustering at the individual level; the null hypothesis assumes that the corresponding parameter is equal to zero. In column (2), the standard error—hence the *p*-value—for the population share of $\omega = 0.3$ could not be computed (N/A) because the point estimate is practically equal to zero (2.02×10^{-58}), which is at the boundary of the parametric space. N_{cho} (N_{ind}) is the number of choice observations (individuals). $logL$ is the maximized log-likelihood.

and household wealth ($\bar{\omega} = 0.002$) specifications. Finally, those with the relatively large integration parameter ($\omega_n = 0.3$) make up less than 0.1% of the population.[A9]

We encountered convergence failures when estimating this model for Study 2 and Study 3, where the number of participants is smaller by approximately 87% and 40% compared to Study 1.[A10] For all three studies, however, we are able to estimate a special case of equation (A5) that constrains π_3 to 0. That is, a restricted

[A9]These share estimates are also statistically insignificant in all but one specification. The exception concerns household wealth, where the estimated share ($2.02 \times 10^{-}58$) virtually lies at the boundary of zero, making it difficult to obtain the associated standard error and *p*-value.

[A10]These percentages refer to the personal income specification, where our estimation sample includes 3,300 individuals in Study 1,415 in Study 2, and 2,030 in Study 3. The sample sizes in the other specifications show similar percentage differences.

specification which assumes that the population consists of either narrow framers with $\omega_n = 0$ or partial integrators with $\omega_n = 0.3$, and excludes the third class whose ω_n is an estimated parameter. Nevertheless, even with this simplification, some convergence issues persist in Study 2 and Study 3, as we will describe shortly.

Table A8 reports this two-class model of population heterogeneity for all studies. In Study 1, the results largely align with our findings from the three-class model, suggesting that virtually the entire population consists of narrow framers or can be reasonably approximated as such: The estimated population share of narrow framers ranges from 89% to 98%, leaving partial integrators as a small minority. This minority share of partial integrators with $\omega_n = 0.3$ is significantly greater than 0 at the 5% level in the personal income specification only.

In Study 2, for the numerical optimizer to achieve convergence, we had to constrain the two-class model further by assuming that the behavioral noise parameter μ_n does not vary across individuals.[A11] Compared to Study 1, we find a smaller share of narrow framers, but they still comprise a vast majority of the population: Their estimated shares range from 82.2% to 91.3%. As in Study 1, the share of partial integrators with $\omega_n = 0.3$ is not significantly greater than 0 at the 5% level for either wealth measure, although it is for both personal and household income.[A12]

In Study 3, even after applying the homogeneity constraint on μ_n, we encountered convergence failures; we were able to obtain the results in Table A8 only after replacing the logit kernel used in specifying $P_{nt}[\cdot]$ with the probit kernel.[A13] Although empirical identification of the model is thus fragile, we again find a predominance of narrow framers, whose population share is estimated to range from 91% to practically 100%.[A14]

[A11]This is equivalent to assuming that $\ln[\mu_n]$ follows a degenerate population distribution with a standard deviation of zero. In the full model that allowed for three types of ω_n, imposing this constraint on $\ln[\mu_n]$ did not resolve the convergence failure.

[A12]We also find relatively a large degree of heterogeneity in risk aversion in this study. Compared to Study 1 and Study 3 where the standard deviation of θ_n is comparable to the mean, Study 2 yields a standard deviation that is nearly three times larger.

[A13]In relation to the latent variable in equation (A2), the probit kernel reflects the alternative assumption that the noise term ϵ_{nt} follows a normal distribution, rather than a logistic distribution.

[A14]The estimated population share of narrow framers in either income specification is numerically

Table A8: Two-class Models for All Studies

Panel A. Wealth

	Study 1		Study 2		Study 3	
	(1) Personal	(2) Household	(3) Personal	(4) Household	(5) Personal	(6) Household
SH: $\omega = 0$	0.969	0.983	0.892	0.913	0.908	0.911
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
SH: $\omega = 0.3$	0.031	0.017	0.108	0.087	0.092	0.089
	(0.225)	(0.471)	(0.115)	(0.187)	(0.005)	(0.007)
EV: θ	0.826	0.819	0.521	0.517	0.414	0.415
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
EV: $\ln[\mu]$	-2.583	-2.589	-1.764	-1.763	-1.856	-1.858
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
SD: θ	0.894	0.883	1.399	1.403	0.375	0.373
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
SD: $\ln[\mu]$	0.718	0.723	0.000	0.000	0.000	0.000
	(<0.001)	(<0.001)	(const.)	(const.)	(const.)	(const.)
$\text{COR}_{\theta,\ln[\mu]}$	-0.187	-0.215	0.000	0.000	0.000	0.000
	(<0.001)	(<0.001)	(const.)	(const.)	(const.)	(const.)
N_{cho}	16,991	17,111	7,390	7,448	9,580	9,590
N_{ind}	2,835	2,855	388	392	1,916	1,918
$logL$	-8,545	-8,614	-4,007	-4,033	-6,120	-6,126

Panel B. Income

	Study 1		Study 2		Study 3	
	(1) Personal	(2) Household	(3) Personal	(4) Household	(5) Personal	(6) Household
SH: $\omega = 0$	0.890	0.959	0.859	0.822	>0.999	>0.999
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(N/A)	(N/A)
SH: $\omega = 0.3$	0.110	0.041	0.141	0.178	<0.001	<0.001
	(0.001)	(0.126)	(0.036)	(0.015)	(<0.001)	(<0.001)
EV: θ	0.910	0.845	0.557	0.602	0.366	0.368
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
EV: $\ln[\mu]$	-2.540	-2.553	-1.781	-1.770	-1.843	-1.848
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
SD: θ	0.934	0.905	1.429	1.611	0.401	0.397
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
SD: $\ln[\mu]$	0.676	0.691	0.000	0.000	0.000	0.000
	(<0.001)	(<0.001)	(const.)	(const.)	(const.)	(const.)
$\text{COR}_{\theta,\ln[\mu]}$	-0.149	-0.143	0.000	0.000	0.000	0.000
	(0.394)	(0.115)	(const.)	(const.)	(const.)	(const.)
N_{cho}	19,761	19,216	7,882	7,548	10,148	9,903
N_{ind}	3,300	3,209	415	402	2,030	1,918
$logL$	-9,960	-9,721	-4,233	-4,071	-6,497	-6,339

Notes: >0.999 indicates a positive number which is greater than 0.999 but less than 1. (const.) indicates that the the parameter is constrained to the value above. In columns (5) and (6), the standard error—hence the p-value—for the population share of $\omega = 0$ could not be computed because, given machine precision, the point estimate is numerically equal to unity, which is at the boundary of the parametric space. Unlike other estimation results in this paper, columns (5) and (6) are based on the probit kernel instead of the logit kernel. All other information remains the same as explained in the notes to Table A7.

equal to 1 in machine precision. Accordingly, we cannot compute standard errors and p-values for these estimates.

Additional References

Doiron, D. and H. I. Yoo (2020). Stated preferences over job characteristics: A panel study. *Canadian Journal of Economics/Revue canadienne d'économique 53*(1), 43–82.

Harrison, G. W., M. I. Lau, and H. I. Yoo (2020). Risk attitudes, sample selection, and attrition in a longitudinal field experiment. *Review of Economics and Statistics 102*(3), 552–568.

Keane, M. and N. Wasi (2013). Comparing alternative models of heterogeneity in consumer choice behavior. *Journal of Applied Econometrics 28*(6), 1018–1045.

McFadden, D. and K. Train (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics 15*(5), 447–470.

von Gaudecker, H.-M., A. Van Soest, and E. Wengström (2011). Heterogeneity in risky choice behavior in a broad population. *American Economic Review 101*(2), 664–694.