UNIVERSITY of HAWAI'I®
MĀNOA

University of Hawai`i at Mānoa Department of Economics Working Paper Series

Saunders Hall 542, 2424 Maile Way,
Honolulu, HI 96822
Phone: (808) 956 -8496
www.economics.hawaii.edu

Working Paper No. 26-03

Behavioral Measures for Laboratory Market Experiments: A Survey

By
Binierose Cacho
Katerina Sherstyuk

October 2026

Behavioral Measures for Laboratory Market Experiments: A Survey*

Binierose Cacho† and Katerina Sherstyuk‡

October 2026

Abstract

Measuring individual behavioral traits may help explain observed deviations from predicted behavior and differences among participant actions in many experimental market and auction settings. We discuss existing measures of risk and time preferences, cognitive abilities, financial literacy, regret and other emotions, overconfidence, and competitiveness. We focus on short tasks and survey items that may be used as simple add-ons to main laboratory market tasks. We further touch upon methodological issues such as monetary incentives, presentation options, repeat measurements and task timing. Finally, we catalog existing laboratory market experiments that employ behavioral measures.

JEL classification code: C83, C91, D40, D91
Key words: economic experiments, behavioral measures, markets, auctions

*We are grateful to Stephen Burks, Colin Camerer, Monica Capra, John Duffy, Kianté Fernandez, Antonio Filippin, Dan Friedman, Sascha Füllbrunn, David Gill, Max Grossmann, David Hardisty, Paul J. Healy, Elif Incekara Hafalir, Nicolas Jacquemet, Duncan James, Ian Krajbich, Charles Noussair, Levent Neyse, Jan Potters, Erik Snowberg, Cong Tao, Diemo Urbig and Peter Wakker for pointing us to relevant studies, and for insightful comments and suggestions.

†University of California Irvine. Email: bdcacho@uci.edu. This work was largely carried out during studies at the University of Hawai'i at Mānoa.

‡University of Hawai'i at Mānoa. Email: katyas@hawaii.edu

1 Introduction

Measuring behavioral preferences has become common in economics experiments for at least two reasons. First, individual behavioral features may account for deviations in the observed behavior from what is predicted by standard economic theory. Second, they may help explain differences in behavior among participants. This paper surveys methods that can be used to measure participants' behavioral preferences and features relevant to market and auction experiments. We specifically focus on short pre- or post-market tasks and surveys that can be easily added to the main part of a laboratory market session. We discuss measuring risk and time preferences, cognitive abilities, financial literacy, regret, overconfidence and competitiveness, as these features may correlate with behaviors observed in market and auction settings. We do not discuss measuring social preferences (see, e.g., Cooper and Kagel, 2016) or personality traits (e.g., Almlund et al., 2011), as our focus is on competitive environments where such features are less likely to play a significant role. We also do not discuss belief elicitation methods; see Schotter and Trevino (2014); Schlag and van der Weele (2015); Danz et al. (2022) for comprehensive surveys.

In market experiments, researchers often seek to compare individual preferences and attitudes of participants in order to correlate them with behavior, rather than solicit the preferences per se. Hence it is often sufficient to obtain simple ordinal measures within the subject population studied, rather than elicit the utility functional form, or estimate the utility function parameters assuming a given functional form (such as measuring concavity of the CRRA utility function). We will focus on such simple instruments that require a short amount of time, but would allow to compare preferences or attitudes across individuals and then correlate them with participant behavior in experimental markets or auctions.

Traditional experimental economics methodology is based on the induced valuation principle and requires the use of (typically financial) incentives to strengthen internal validity and reduce noise in the data (Smith, 1976; Voslinsky and Azar, 2021). However, the need and effect of financial incentives on behavior has been discussed since at least Camerer and Hogarth (1999), who review a set of experiments and report that incentives typically have no effect on mean performance but improve data quality in other respects. Some studies show that non-incentivized surveys and self-report measures can provide a cost-effective and time-efficient alternative to incentivized tasks, with several validations finding that survey measures predict behavior in incentivized experiments and outside the lab (Dohmen et al., 2011; Falk et al., 2018). We discuss both incentivized and non-incentivized methods that have been used in the literature, and further report which non-incentivized instruments have been validated using incentivized methods. We further touch upon such methodological issues as repeat measurements, task presentation options, and timing of tasks, among others.[1]

The paper is organized as follows. We discuss measuring risk preferences in Section 2; time preferences in Section 3; and cognitive abilities and financial literacy in Sections 4 and 5, respectively. Section 6 discusses measuring regret and other emotions; Section 7 – overconfidence, and Section 8 – competitive preferences. We briefly overview existing market and auction laboratory experiments that employ behavioral measures in Section 9. Section 10 concludes.

[1] See Snowberg and Yariv (2025) for an in-depth discussion of experimental designs.

2 Risk preferences

The importance of risk preferences in economic behavior has been widely documented (Eckel and Grossman, 2002; Anderson and Mellor, 2008; Dohmen et al., 2010; Burks et al., 2009, 2012, 2015). In laboratory market-like settings, risk aversion has been studied extensively in relation to behavior in auctions (Cox et al., 1985; Brunner et al., 2014; Holt and Sherman, 2014; Giebe et al., 2024) and asset markets (Gneezy et al., 2003a; Fellner and Maciejovsky, 2007; Breaban and Noussair, 2015; Cohn et al., 2015; Michailova et al., 2017; Aragó et al., 2022). Risk aversion is one of the leading explanations of overbidding in first price auctions, with experimental laboratory used to calibrate risk aversion parameters based on participant auction behavior (Cox et al., 1985, 1988; Holt and Sherman, 2014). Here we focus on measurements of risk preferences that are not inferred from auction or market behavior, but are obtained using separate simple instruments. These measurements can be then employed as explanatory variables or controls for participant behavior in laboratory markets and auctions, as is done in numerous studies such as Brunner et al. (2014); Breaban and Noussair (2015); Michailova et al. (2017); Karmeliuk et al. (2022); Giebe et al. (2024).

Harrison and Rutström (2008) present an in-depth analysis of risk preference elicitation devices used in experimental economics. Charness et al. (2013) provide a shorter comparative review of popular risk elicitation methods in experimental laboratory, discussing pros and cons of each elicitation method.[2] They explicitly point out that applicability of each method depends on researchers' objectives, with more sophisticated methods potentially providing a more accurate estimation of participant risk attitudes but at the same time requiring higher comprehension and possibly leading to more errors. Below we review popular risk elicitation procedures focusing on shorter ones that are applicable as add-ons to the main experimental market task, while also noting their limitations as discussed in the literature. We do not discuss some other instruments, such as the random lottery pairs design (Hey and Orme, 1994; Harrison and Rutström, 2009) and the gamble trade-off method (Wakker and Deneffe, 1996) that are widely used to evaluate generalizations of the expected utility theory and typically require a considerable amount of time to implement; see Harrison and Rutström (2008) for a detailed discussion.

2.1 Incentivized lottery choices

In experimental economics, incentivized lotteries are the standard practice to elicit risk attitudes (Harrison and Rutström, 2008). Two popular lottery-choice-based instruments are developed in Holt and Laury (2002) and Eckel and Grossman (2002, 2008).

Multiple Price Lists – Holt and Laury (2002) Holt and Laury (2002) (HL) popularized the use of multiple price list lotteries to measure risk aversion. Participants are tasked to make choices between ten pairs of lotteries, with each pair involving a choice between a lottery with less variable payoffs (Option A) and a lottery with more variable payoffs (Option B); see Table 1. The probabilities of low and high outcomes in each lottery change in a consistent manner across pairs, leading to changes in expected values, thus allowing

[2] See also Mata et al. (2018).

researchers to solicit each participant's coefficient of risk aversion assuming a constant relative risk aversion (CRRA) utility function. The task is incentivized, with one of the choices randomly picked for payment.

HL procedure has been widely used in a variety of lab and field experiments (e.g., Harrison et al. (2007); Andersen et al. (2008); Brañas-Garza et al. (2021)). It is arguably the most widely-used method to measure risk attitudes in auction and market experiments (Anderson et al., 2012; Brunner et al., 2014; Breaban and Noussair, 2015; Cueva and Rustichini, 2015; Michailova and Schmidt, 2016; Corgnet et al., 2018; Bosch-Rosa et al., 2018; Giebe et al., 2024).

The task can be used for ordinal comparison of risk attitudes across participants by comparing their number of safer lottery choices over the riskier ones, with an earlier switch from a safe to a risky lottery implying a greater tolerance of risk.

Table 1: Multiple Price List from Holt and Laury (2002)

Option A	Option B	Expected Payoff Difference
1/10 of \$2.00, 9/10 of \$1.60	1/10 of \$3.85, 9/10 of \$0.10	\$1.17
2/10 of \$2.00, 8/10 of \$1.60	2/10 of \$3.85, 8/10 of \$0.10	\$0.83
3/10 of \$2.00, 7/10 of \$1.60	3/10 of \$3.85, 7/10 of \$0.10	\$0.50
4/10 of \$2.00, 6/10 of \$1.60	4/10 of \$3.85, 6/10 of \$0.10	\$0.16
5/10 of \$2.00, 5/10 of \$1.60	5/10 of \$3.85, 5/10 of \$0.10	-\$0.18
6/10 of \$2.00, 4/10 of \$1.60	6/10 of \$3.85, 4/10 of \$0.10	-\$0.51
7/10 of \$2.00, 3/10 of \$1.60	7/10 of \$3.85, 3/10 of \$0.10	-\$0.85
8/10 of \$2.00, 2/10 of \$1.60	8/10 of \$3.85, 2/10 of \$0.10	-\$1.18
9/10 of \$2.00, 1/10 of \$1.60	9/10 of \$3.85, 1/10 of \$0.10	-\$1.52
10/10 of \$2.00, 0/10 of \$1.60	10/10 of \$3.85, 0/10 of \$0.10	-\$1.85

Ordered Lottery System This type of instrument was first used by Binswanger (1980, 1981) to solicit risk attitudes in a field setting, and later popularized by Eckel and Grossman (2002, 2008) (EG) as a simpler alternative to HL task. Participants are asked to select one from the list of several 50/50 risky gambles, as illustrated in Table 2. The first gamble provides a sure payoff, and the others increase the expected payoff and risk, so that extremely risk averse individuals would choose gamble 1, moderately risk averse individuals choose gambles 2-4, and risk neutral individuals choose gamble 5 or 6. The sixth gamble, added in Dave et al. (2010), has the same expected payoff as gamble 5 but a higher variance, so only risk-loving individuals strictly prefer it. The task therefore allows for a straightforward ordinal comparison of risk attitudes among individuals.

EG method of simple choice lottery has been widely used by experimental economists; see, for example, Crosetto and Filippin (2013); Charness et al. (2020); Cassar and Rigdon

(2021). It is typically incentivized: each individual is paid according to realization of their chosen gamble.

Dave et al. (2010) compare HL and EG risk elicitation tasks, and find that the use of EG simple choice lotteries requires fewer numeracy skills, while complex choice lotteries as in Holt and Laury (2002) may provide finer estimates but are also subject to more noise in estimation. While numeracy skills are typically not an issue in laboratory experiments conducted with college students, simpler procedures may have an added benefit of taking less time to implement or saving the participants' mental efforts for the main market task.

Table 2: Eckel and Grossman Ordered Lottery Task

Gamble	Low payoff	High payoff	Expected payoff	Std. deviation
1	28	28	28	0
2	24	36	30	6
3	20	44	32	12
4	16	52	34	18
5	12	60	36	24
6	2	70	36	34

Note: Each gamble pays the low or high amount with equal probability. Gamble 6 screens for risk-loving preferences. Reproduced from Dave et al. (2010), extending Eckel and Grossman (2002, 2008).

Lottery-Panel Task Sabater-Grande and Georgantzis (2002) elicit risk aversion using a lottery-panel task, building on earlier work of Murnighan et al. (1988) and Millner and Pratt (1991). In each panel (row), participants choose one of ten binary lotteries, to gain payoff X with probability p and receive nothing otherwise; see Table 3. Within a panel, moving from left to right, the winning probability p decreases but the payoff X increases, such that the expected value pX also rises. Thus more rightward choices indicate a greater willingness to accept higher risk in exchange for a higher expected value.

Each subject makes one choice in each of four panels. Because the payoff scale increases across panels for any given p, the authors impose a consistency restriction of monotonic preferences. An individual with monotonic preferences would choose a lottery at or to the right of the choice in the preceding row; participants who select a lottery to the left of the preceding row violate the monotonic preference assumption and are discarded. Participants are classified as high, medium, or low in risk aversion with $p > 0.5, p = 0.5, p < 0.5$, respectively, according to their choices in the second panel, the one that produced the greatest variability in choices among consistent participants. The task is incentivized, with one of the four panels selected for payment.

Aragó et al. (2022) use this lottery-panel method as an add-on to an asset market experiment and report a significant relationship between risk aversion and asset market trading.[3]

[3]In a field setting, Proestakis et al. (2013) use the lottery-panel task to elicit risk preferences from participants at varying levels of blood alcohol content.

Table 3: Lottery-Panel Risk Task from Sabater-Grande and Georgantzis (2002)

Probability (p)	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
Lottery 1	100	112	127	147	173	210	265	356	540	1090
Lottery 2	100	120	150	190	230	300	400	570	900	1900
Lottery 3	100	166	250	357	500	700	1000	1500	2500	5500
Lottery 4	100	220	380	570	830	1200	1750	2670	4500	10000

Note: Table entries are monetary values X corresponding to lotteries that pay X with probability p (column) and zero otherwise. For example, Lottery 1 in column marked "0.1" pays 1090 with probability $p = 0.1$ and has expected value 109. Subjects choose one lottery per panel; more rightward choices indicate greater risk tolerance.

2.2 Certainty Equivalence Methods

Static Certainty Equivalence (CE) Method Cohen et al. (1987) use a series of multi-item questionnaires, presented in a tabular form, where in each row a subject is choosing between a risky lottery and a guaranteed sure payment, with the magnitude of the sure payment increasing across questions. The subject's certainty equivalent is inferred from the switching decision from risky to a safe choice.[4] Alternatively, the certainty equivalent can be elicited directly, by asking subjects to state the sure amount at which they are indifferent to a given lottery (e.g., Kirchler and Maciejovsky, 2002).

CE method is widely used (Kirchler and Maciejovsky, 2002; Dohmen et al., 2011; Burks et al., 2009, 2015). In particular, Dohmen et al. (2011) use CE method to validate the qualitative risk assessment to be discussed below. In most commonly adopted variations of CE method, each choice on the list is between a 50/50 lottery and a sure payment. This makes it simpler and less numeracy intensive than the HL multiple choice list. The method is typically incentivized with probabilistic payments (Cohen et al., 1987; Dohmen et al., 2011).

CE method is a static version of the Iterative Staircase method that we discuss next.

Iterative Staircase Methods The iterative staircase model modifies the CE method by using unfurling brackets or iterative staircases to reduce the number of choices made by the participants; an example is provided in Section 2 of Appendix A.1. This method originates from the psychology literature (Cornsweet, 1962) and can be applied to solicit both risk and time preferences (Andersen et al., 2006; Dohmen et al., 2010; Falk et al., 2018).

Falk et al. (2023) employ the iterative staircase method along with a qualitative measurement (discussed below) in their Preference Survey Module which uses non-incentivized instruments validated by incentivized lotteries. In their setting, the use of the iterative staircase method allows to reduce the number of decisions from the set of potential 31 choices between sure amounts and lotteries to five sequential decisions. As in the CE method, the risk preferences are assessed by the certainty equivalent from the switching point between a

[4] Murnighan et al. (1988) give participants a series of choices between a sure payment and a risky lottery to sort participants by their risk-aversion. In their setting, the sure payment stays the same, but the probability of higher outcome gradually increases; the participants are sorted on how frequently they chose the sure payment.

lottery and a sure payment. The iterative staircase CE method is easy to comprehend, takes a short amount of time, and is validated by incentivized lotteries to use with hypothetical payments. It allows for ordinal comparison of risk attitudes between subjects.[5]

A drawback of the above staircase method is that it imposes choice consistency and may lead to inaccurate preference estimates if participants make mistakes. Several studies propose dynamic adaptive Bayesian designs that explicitly account for response error (Toubia et al., 2013; Chapman et al., 2024). The Dynamically Optimized Sequential Experimentation method (DOSE), introduced by Chapman et al. (2024), constructs personalized staircase steps for each participant based on their previous answers, accounts for the possibility of mistakes, and allows to evaluate individual participant choice consistency.[6] Chapman et al. (2025b) apply DOSE along the MPL procedures to elicit risk preferences in the representative sample of US population.

2.3 Continuous choice methods

Gillen et al. (2019) distinguish methods that elicit certainty equivalents for lotteries, and those that elicit allocations of assets between safe and risky options. The latter category typically allows for a continuous (or almost-continuous) decision variable, as we now discuss.

Risky Project - Gneezy and Potters (GP) Gneezy and Potters (1997) elicit risk preferences by asking participants to choose the amount x out of a given endowment to invest in a risky project. The risky project has probability p of losing the amount invested and probability $(1-p)$ of earning k times the amount invested. The participant choice of x is used to elicit risk preferences of participants, with higher amount indicating a higher tolerance of risk. Unlike multiple choice lists, the decision space here is continuous, and yet the decision is simple, with the method allowing for ordinal comparisons among individuals. The GP risky project method is discussed in Charness et al. (2013), and is employed in many studies, including Crosetto and Filippin (2013); Agranov and Yariv (2018); Gillen et al. (2019). Karmeliuk et al. (2022) use this method to measure participant risk attitudes in an auction experiment.

Balloon Analogue Risk Task (BART) This method, due to Lejuez et al. (2002), involves participants pumping air into a virtual balloon on a computer screen, with each additional pump increasing potential earnings, but also increasing (unknown to participants) probability that the balloon explodes and no earnings are collected. A higher number of pumps chosen characterizes more risk tolerance. This task is reviewed in detail in Charness et al. (2013). Lejuez et al. (2002) report that the performance on BART is correlated with self-reported risky behaviors in the field.

[5] Andersen et al. (2006) provide evidence that in some subject pools, the iterative method may elicit higher degree of risk aversion than the corresponding non-iterative multiple price list method; however, it still allows for ordinal comparison among individuals.

[6] Chapman et al. (2024) claim that their Bayesian adaptive staircase design is a substantial improvement across settings, with typically two to three times as much information gained as in a multiple price list or random question protocol, in risky choice (Chapman et al., 2025b; Ray et al., 2019), social preferences (Camerer et al., 2020), and intertemporal choice (Imai and Camerer, 2018).

Bomb Risk Elicitation Task (BRET) Crosetto and Filippin (2013) provide another visual alternative, the "bomb" risk elicitation task. Participants must collect boxes out of 100, one of which contains a bomb. Earnings are increased by 0.10 for each box collected and reduced to zero if the bomb is collected. A higher number of boxes collected indicates higher tolerance of risk. Both static and dynamic versions are possible, but the authors recommend the dynamic version as more visual. This task appears similar to BART; yet, in BRET the uncertainty is not realized (the bomb does not "explode") until after the participant decides on the number of boxes to collect, thus allowing for untruncated observations of decisions. Further, BRET provides sufficient information to participants to infer the exact probability of the bomb exploding associated with each number of boxes collected.

Crosetto and Filippin (2013) identify several advantages of the BRET method which is also applicable to BART: (1) it does not require numerical skills; (2) measurements are not influenced by loss aversion, as the task focuses on potential gains rather than losses; (3) the dynamic nature of the task creates a setting appropriate to measure risk attitudes in decisions that entail a time dimension, including trading and dynamic auctions.

Compared to other elicitation methods, BRET yields a higher share of participants whose decisions are consistent with risk-neutral or risk-loving attitudes (Crosetto and Filippin, 2013). This may be due to the dynamic implementation of the task that has participants continue to collect boxes as the default action; one has to actively drop out to stop collecting. However, the method should still allow for ordinal comparison of risk attitudes among individuals, unless different individuals react to the default option differently.

Several studies use the BRET task in their experiments (Füllbrunn et al., 2019; Angrisani et al., 2020; Mol et al., 2020; Hermanns and Kokot, 2023). Füllbrunn et al. (2019) find that risk preferences elicited using the BRET task are significant explanatory variables for overbidding in first price sealed bid auctions. Angrisani et al. (2020) use the BRET task to measure risk attitudes along with measuring cognitive abilities and conducting personality tests in professional traders and students.

2.4 Qualitative Assessments

The qualitative self assessment of risk preferences is a simple Likert scale test of individual perceptions of willingness to take risk. Participants are asked on an 11-point (0 to 10) scale "How willing are you to take risks, in general?" Dohmen et al. (2011) validate this qualitative instrument using an incentivized certainty equivalent lottery choice method discussed above. They report that this simple survey instrument is significantly correlated with actual risk-taking behavior of individuals. Additionally, Dohmen et al. (2011) investigate the qualitative assessment across different domains—driving, finance, sports and leisure, career, and health—and find consistent correlations across different domains. Falk et al. (2018, 2023) further employ the qualitative risk assessment along with the iterative staircase CE task to solicit risk attitudes among wide populations in their Preference Survey Module.

Increasingly, qualitative risk assessments are employed concurrently with other quantitative risk elicitation methods as they are easy to implement (Crosetto and Filippin, 2013; Falk et al., 2023; Snowberg and Yariv, 2021; Karmeliuk et al., 2022). Using survey data contained in American Life Panel, Tasoff and Zhang (2022) compare 36 risk elicitation methods and report that qualitative measures based on self-reports predict behaviors better than iterative

lottery choices (titration), MPL of Holt and Laury (2002), and Eckel and Grossman (2002, 2008) measures. However, Zambrano-Curcio et al. (2023) demonstrate that qualitative self-reported measurements reflect not only individual risk aversion but also loss aversion. Chapman et al. (2025a) caution that despite correlations with incentivized tasks and behaviors outside the lab, qualitative self-reports are not interchangeable measures of the same latent construct. Participants may use simple heuristics, and disproportionately select focal points (0, 5, 10) on numerical Likert scales in surveys, introducing systematic biases and leading to skewed estimations of correlations of estimated parameters with individual characteristics.

2.5 Testing for loss aversion

Although the discussion above focuses on risk aversion, loss aversion, or the concept that "losses loom larger than gains" (Kahneman and Tversky, 1979)[7] also plays an important role in market settings such as financial markets (e.g., Breaban and Noussair, 2015). Testing for loss aversion involves risky choices with a possibility of loss.

Most of the methods discussed above can be adopted to include loss domains.[8] Cohen et al. (1987) explicitly compare participant choices between gambles and sure payments in the loss and gain domain, and find substantial differences. Eckel and Grossman (2008) employ simple choice lotteries in the no-loss and loss frames using a between-subject design, and find no evidence of loss aversion. Crosetto and Filippin (2013) use a loss frame and report that women are no different than men in the risk domain but are more loss averse.

All such instruments involving choices with the possibility of loss use either hypothetical payments, or include a start-up bonus that would cover losses in case the loss is realized. Cohen et al. (1987) refer to Kahneman and Tversky (1979) isolation effect to justify the bonus, claiming that the participants would ignore the bonus and react to the actual binary choices offered in the loss domain "as in case of real payoffs" (p. 3).

Abdellaoui et al. (2007) and Abdellaoui et al. (2008) develop techniques for measuring loss aversion based on elicitation of certainty equivalents for two-outcome prospects for gains and losses. While these methods are commonly used in decision-making experiments, they required a considerable amount of time to implement.[9] Chapman et al. (2024) estimate loss aversion along with risk aversion using the DOSE staircase method. Muehlfeld et al. (2013) and Breaban and Noussair (2015, 2018) employ a simple loss aversion elicitation procedure, which originates from Fehr and Goette (2007)[10], and use it as an add-on to their asset market experiments. Participants are presented, in a price list format, with several lotteries, each having an equal chance of a given positive gain and a varying amount of loss, and have to indicate whether they accept or reject each lottery. The number of lotteries not accepted is used as measure of loss aversion. Upon conclusion, one of the decisions, chosen randomly, is

[7] See also Wakker (2010).

[8] The standard settings of BART and BRET are formulated in gains domain only; Crosetto and Filippin (2013) modify the reference point in the BRET task to include the possibility of a loss.

[9] Abdellaoui et al. (2007, 2013) use a fairly complex iterative choice-based procedure with hypothetical payoffs for utility elicitation that would allow for loss aversion. Tanaka et al. (2010) use a variation of the HL task using prospect theory frameworks instead of expected utility frameworks; see also Campos-Vazquez and Cuilty (2014). Their task is more complicated than the HL task as it requires several multiple price lists to estimate both the coefficients of loss aversion and risk aversion.

[10] See also Goette et al. (2004); Gächter et al. (2007).

used for payment, with no payment made if the corresponding lottery was rejected. As the task is an add-on to the main market task, negative earnings are subtracted from the earnings made in other parts of the experiment. Breaban and Noussair (2015) report that this measure of loss aversion significantly correlates with participant trading activity in their experimental asset market; furthermore, participants' loss aversion does not significantly correlate with risk aversion. Gächter et al. (2022) propose and test two loss aversion elicitation procedures, one based on riskless choice, and the other on risky (lottery) choice.[11] Darriet et al. (2020) employ their lottery-based loss aversion elicitation procedure in the context of money illusion experiment.

2.6 Discussion and summary

Comparison of methods The choice of risk elicitation method depends on the experimenters needs. As an add-on to the main market task, the risk elicitation method should be simple, short, and still provide for ordinal comparison among individual participants. Holt and Laury (2002) continue to be the most common method to measure risk preferences in market and auctions experiments (Corgnet et al., 2018; Cueva and Rustichini, 2015; Giebe et al., 2024; Brunner et al., 2014; Michailova et al., 2017; Kirchkamp and Mill, 2021). Some studies use risky projects (Cohn et al., 2015) and BRET (Füllbrunn et al., 2019) and also measure loss aversion (Breaban and Noussair, 2015).

Several researchers point out that a concerning issue with laboratory risk attitude solicitation procedures is their apparent lack of external and sometimes internal validity (Eckel and Grossman, 2002; Charness et al., 2020). Charness et al. (2020) compare several popular risk elicitation devices to their performance in the prediction of outside the laboratory behavior finding inconsistencies across risk tasks. While the risk tasks accurately predicted risk taking behavior in the lab, the study found little explanatory power of these measures in financial risk domains outside the lab. Crosetto and Filippin (2016) show that the use of different elicitation methods leads to different estimation results, partly due to the nature of tasks themselves; however, they use a between-subject design, which does not allow to check the consistency of methods in ordinal comparisons between subjects. Likewise, Tasoff and Zhang (2022) report low correlation in the risk elicitation results obtained using different methods, with an exception of qualitative self-reports of risk attitudes. However, these findings may be due to the differences in survey procedures and subject pools.

Other studies provide a more optimistic view on the consistency of different risk elicitation methods. Gillen et al. (2019) compare risk attitudes solicited using Gneezy and Potters (1997) risky projects, multiple price lists (similar to Holt and Laury (2002) but representing lotteries as betting on a certain color of a ball to be drawn from an urn, rather than tabular representation), Eckel and Grossman (2002) lottery choice task, and a qualitative assessment as in Falk et al. (2018). They report that correlations in risk attitudes elicited using different methods are significantly increased after correcting for measurement error, and recommend using duplicate elicitations with different numerical values to be able to correct for the error; they further suggest that Gneezy and Potters (1997) risky projects and Eckel and Grossman (2002) lottery choice results are more correlated with other methods than the MPL method.

[11] Gächter et al. (2022) is the published version of the frequently cited Gächter et al. (2007) paper.

Holzmeister and Stefan (2021) show that measurement of individual attitudes varies across different risk elicitation devices, but these measurements are not necessarily inconsistent. Friedman et al. (2022) demonstrate that revealed risk preferences vary across methods not only because of the measurement error but also because of variations in how the tasks are presented (text or spatial presentation, discrete or continuous choice). They recommend choosing a task that matches the attributes of an environment that researchers seek to study; in addition, in line with Gillen et al. (2019), they recommend using multiple trials of the task with varying task parameters. Chapman et al. (2024) report that their Bayesian adaptive staircase method (DOSE) allows to better assess risk attitudes in a shorter amount of time compared to multiple price list and random question protocols; DOSE estimates of loss aversion further show significant correlation with behaviors outside the lab (Chapman et al., 2025b).

Software implementation Many methods have now been implemented in standard experimental economics software with programs freely available. Reuben (2017) offers a zTree (Fischbacher, 2007) program for a simple risk elicitation task. Holzmeister and Pfurtscheller (2016) provide oTree (Chen et al., 2016) code, details, and STATA packages to set up BRET for computerized laboratory experiments. Holzmeister (2017) describes oTree code that implements multiple price lists of Holt and Laury (2002), certainty equivalent method (Cohen et al., 1987; Dohmen et al., 2011), single choice method of Eckel and Grossman (2002), and the staircase method (Andersen et al., 2006; Dohmen et al., 2011).

Presentation options Holzmeister (2017) writes that oTree allows for different lottery representations, using pie charts as well as traditional lists. An interesting and understudied question is whether task representation affects results. Friedman et al. (2022) compare several lottery representation techniques including the budget line, text and bar chart representations, and demonstrate that the results are highly variable across different representation methods. Brown and Healy (2018) demonstrate that presenting multiple decisions in a single list may create issues with incentive compatibility; instead, options should be presented separately in random order, with one task randomly chosen for payment (see also Azrieli et al., 2018; Healy and Leo, 2025). However, the latter increases the inconsistency of choices.

Field experiments traditionally use easy-to-comprehend representations of tasks; e.g., Binswanger (1980) uses coin flips in the description of their ordered lottery method; other popular representations include displaying piles of money or pie charts (see Harrison and Rutström, 2008). Some complex elicitation procedures in the laboratory also use visual representations such as pie charts (e.g., Hey and Orme, 1994; Abdellaoui et al., 2008; Harrison and Rutström, 2009) to reduce the cognitive load on participants. Decision research suggests that as complexity of a task increases, participants are more likely to use heuristics in their choices (e.g., Fox et al., 2015). Yet many economic experiments with college students describe lotteries in non-pictorial tabular form, or using balls drawn from urns; it is implicitly assumed that students comprehend abstract lottery representations well. It would be interesting to know how the results from these methods would differ if more visual representations were used. Advances in technology and availability of options in experimental software packages offer all the tools needed to explore this issue.

Summary There are several conclusions that we can draw. First, the choice of method should match the market application one is targeting to study. Holt and Laury (2002) MPL remain the most widely-used method overall. Eckel and Grossman (2002) and the staircase CE methods are appealing in their simplicity and low cognitive loads. Gillen et al. (2019) suggest that the method that elicits allocations of assets between safe and risky options (e.g., Gneezy and Potters, 1997) exhibits greater correlations with other measures, and more stability over time. For dynamic settings, methods where choices are made dynamically, such as BRET or BART, may be appropriate.

Second, duplicate elicitations with different parameters may be advisable (Gillen et al., 2019; Friedman et al., 2022),[12] even if one seeks to obtain only ordinal comparison among individuals. This may not increase the time required dramatically as current software packages allow for time-efficient implementations with more visual representations that were not available just a few years back.

Third, non-incentivized validated methods, complemented by the qualitative assessment, may provide a viable alternative to incentivized tasks. While non-incentivized measurements may attenuate the risk estimates because of higher noise in the data, they may be sufficient for ordinal comparisons among individuals, with the benefit of low operational costs.

Finally, the effect of methods representation pointed out in Friedman et al. (2022) is largely under-explored. Current software capabilities allow for options with more visual, intuitive representation of risky alternatives, and may result in better measurements.

3 Time Preferences

Preferences over inter-temporal trade-offs can affect both real-time and long-term financial and market decision-making.[13] Frederick et al. (2002); Andersen et al. (2008) and Cohen et al. (2020) provide comprehensive surveys of the time preference literature; see also Cheung (2016). Two distinct ways to measure time preferences are to use either financial flows, or time-dated consumption of primary goods or effort (Cohen et al., 2020). Theoretically, the classical discounted utility model (Samuelson, 1937) defines time preferences over consumption rather than financial flows. Practically, however, choices over financial flows, or "money earlier or later" (MEL) tasks, are often used to compare impatience of individuals,[14] as there is evidence that choices in MEL tasks are correlated with decisions outside the lab, especially in financial domains (Wölbert and Riedl, 2013; Cohen et al., 2020; Bartels et al., 2023). MEL tasks are easier to institute, and appear more suitable as add-ons to the main experimental task compared to time choices over consumption. On the other hand, time-dated consumption choices over primary goods or effort are better at capturing the present

[12] Duplicate elicitations have been employed from early on to reduce noise in experimental data; see, for example, Cohen et al. (1987).

[13] Broader economics literature indicates that time preferences correlate with health, education behavior, job-related outcomes, savings and debt accumulation (Chabris et al., 2008; Meier and Sprenger, 2010; Burks et al., 2012; Sutter et al., 2013; Epper et al., 2020); see also Cohen et al. (2020). Evidence from field experiments demonstrates that revisions of financial commitments towards the present are positively associated with present bias (Giné et al., 2018).

[14] "While one function of MEL studies is to measure the discount function, another is simply to predict behavior by classifying individuals as relatively patient or impatient" (Cohen et al., 2020, p. 335).

bias[15] and factors such as impulsivity and self-control, which are considered to be among the psychological foundations of time preferences (Frederick et al., 2002; Cohen et al., 2020). These factors may affect financial decisions and real-time market behavior such as bidding in dynamic auction or trading in stock market. Below we review several typical MEL tasks that could be used as add-ons to market experiments, and further briefly discuss tasks over time-yoked primary goods consumption or effort.

3.1 Money Earlier or Later (MEL) Tasks

Experimental designs typically use MEL tasks (more than 60% of studies as per Cohen et al. (2020)) in which individuals choose from a set of two or more time-dated choices, a smaller sum earlier or a larger sum later. The switch from one choice to another is used to find the indifference point, subsequently, the required rate of return (RRR).[16]

There are several standard designs of MEL tasks: multiple price lists (MPL), randomized binary choices, iterative staircase model, and matching (Hardisty et al., 2013; Cohen et al., 2020). They are similar to the corresponding risk elicitation tasks but explore inter-temporal tradeoffs.

Multiple Price Lists (MPL) Time preference MPLs are similar to risk elicitation MPLs, in that individuals are given a list of binary choices, but here the choices are inter-temporal choice lotteries; see Table 4. The required rate of return is elicited from MPLs by using the induced switching point between choices with higher and lower RRR. Many studies use MPL in their experiments (Kirby et al., 1999; Harrison et al., 2002; Andersen et al., 2008; Snowberg and Yariv, 2021). A drawback of MPL elicitation is that when choices are presented in a single list, this method may induce anchoring effects as individuals may select options improving upon initial choices made (Kahneman and Tversky, 1979; Frederick et al., 2002).

Random Binary Choice To address potential anchoring effects, Random Binary Choice method follows the same MPL questions, however choice questions are presented one at a time, and out of order.[17] A concern with this method is that while it corrects for the anchoring effects, it may increase noise and inconsistencies in responses (Harrison et al., 2005).

Iterative Staircase Methods The iterative staircase method (Cornsweet, 1962) is analogous to the one discussed in Section 2.2 for the risk preference solicitation, but applied to the

[15] Present bias is a well-documented behavioral tendency to "prefer a smaller immediate reward to a larger delayed reward in the present, but ... reverse ... preferences when those two alternatives are shifted into the future by the same amount of time" (Imai et al., 2021). Such time inconsistency of preferences is captured in models of hyperbolic and quasi-hyperbolic discounting (Laibson, 1997).

[16] Depending on the theoretical framework applied, the required rate of return is an approximation of the discount factor δ used in exponential discounting (Samuelson, 1937), or the δ, β parameters used in quasi-hyperbolic discounting models (Laibson, 1997).

[17] Another way to minimize anchoring is use a procedure that exposes respondents to a series of opposing anchors (Frederick et al., 2002).

Table 4: Payoff Table for 6 Month Time Horizon MPL

Task	Payoff Alternative		Annual Interest Rate (AR)	Annual Effective Rate (AER)	Preferred Payment Option
	Payment Option A (in 1 month)	Payment Option B (in 7 months)			
1	3000 DKK	3075 DKK	5.09	5.10	A or B
2	3000 DKK	3152 DKK	10	10.38	A or B
3	3000 DKK	3229 DKK	15	15.87	A or B
4	3000 DKK	3308 DKK	20	21.55	A or B
5	3000 DKK	3387 DKK	25	27.44	A or B
6	3000 DKK	3467 DKK	30	33.55	A or B
7	3000 DKK	3548 DKK	35	39.87	A or B
8	3000 DKK	3630 DKK	40	46.41	A or B
9	3000 DKK	3713 DKK	45	53.18	A or B
10	3000 DKK	3797 DKK	50	60.18	A or B

Notes: Reproduced from Andersen et al. (2008). This table format also follows Coller and Williams (1999).

time preference domain. A participant is presented with a sequence of single choices between some amount today versus some amount in the future. Each choice adjusts according to the previous answers refining upon an interval of time preferences measured. Similar to choice methods, the range of time preferences are predetermined. An example is given in Section 3 of Appendix A.1. The staircase model has been used in several papers (Harrison et al., 2005; Hardisty et al., 2013; Falk et al., 2018) and is also referred to as the dynamic MPL. This method prevents inconsistencies in responses, reduces the number of questions, and may remove framing effects by allowing participants to answer one question at a time. A further Bayesian refinement of the staircase method, DOSE (Chapman et al., 2024), selects a personalized sequence of questions for each participant, and accounts for the possibility of mistakes in responses. DOSE adopts the framework of quasi-hyperbolic discounting, and allows to estimate both the time discounting and the present bias parameters (Chapman et al., 2025b).[18]

Many studies (Harrison et al., 2005; Chapman et al., 2024) incentivize the iterative staircase method by randomly selecting one of the intertemporal choices for payment. Hardisty et al. (2013) compare hypothetical static multiple price lists and "dynamic" staircase price lists to matching methods (discussed below) and find both static and dynamic MPL to be better predictors of real-world behaviors.[19] Falk et al. (2023) validate the use of their

[18] See Yoon and Chapman (2016) for another iterative design.

[19] Hardisty et al. (2013) use two comparison groups: (1) university students inside the lab and (2) Amazon Mechanical Turk. Several of the behaviors that were better predicted by choice methods were tobacco usage,

non-incentivized staircase survey instrument with an incentivized lottery. Tasoff and Zhang (2022) report that the iterative time-preference elicitation method similar to the staircase, which they refer to as the "titration method," is the best among the 15 hypothetical survey instruments they compare in terms of predicting economically important behaviors outside the lab.

Table 5: Double Multiple Price List Method from Andersen et al. (2008)

Panel A: Intertemporal Multiple Price List

For each decision below, choose Option A or Option B by checking one box. Each option specifies a payment today and a payment in 5 weeks. For example, in Decision 1, Option A pays $19.00 today and $0 in 5 weeks, while Option B pays $0 today and $20.00 in 5 weeks.

	Option A		Option B		Choice	
Decision	*Today*	*In 5 Weeks*	*Today*	*In 5 Weeks*	A	B
1	$19.00	$0	$0	$20.00	☐	☐
2	$18.00	$0	$0	$20.00	☐	☐
3	$17.00	$0	$0	$20.00	☐	☐
4	$16.00	$0	$0	$20.00	☐	☐
5	$14.00	$0	$0	$20.00	☐	☐
6	$11.00	$0	$0	$20.00	☐	☐

Panel B: Holt–Laury Risk Multiple Price List

A ten-sided die is rolled. For each decision, check whether you prefer Option A or Option B.

	Option A				Option B			
Decision	**Choose**	**Die roll**	**$**	**Otherwise $**	**Choose**	**Die roll**	**$**	**Otherwise $**
1	☐	1	15	8.31	☐	1	20	0.52
2	☐	1–2	15	8.31	☐	1–2	20	0.52
3	☐	1–3	15	8.31	☐	1–3	20	0.52
4	☐	1–4	15	8.31	☐	1–4	20	0.52
5	☐	1–5	15	8.31	☐	1–5	20	0.52
6	☐	1–6	15	8.31	☐	1–6	20	0.52
7	☐	1–7	15	8.31	☐	1–7	20	0.52
8	☐	1–8	15	8.31	☐	1–8	20	0.52
9	☐	1–9	15	8.31	☐	1–9	20	0.52
10	☐	1–10	15	–	☐	1–10	20	–

exercise, dental checkups and flossing, and recent infidelity.

MPL with Non-Linear Utility Methods A limitation to MPL methodology is that it assumes linear utility functions (Frederick et al., 2002; Andreoni and Sprenger, 2012a). Andersen et al. (2008) use double multiple price list (DMPL) method which measures time preferences, then risk preferences using the MPL format; see Table 5. They argue that it is important to measure risk preferences because concavity of the utility function due to risk aversion can increase the appearance of impatience. Andreoni and Sprenger (2012a) develop the Convex Time Budget (CTB) method that measures both convexity and discounting by allowing choices inside the intertemporal budget constraint. They present evidence that CTB yields significantly lower discount factors than traditional MPLs. Further, increasing the riskiness of future payoffs leads to an increased preference for present time choices (Andreoni and Sprenger, 2012b).

Table 6: Convex Time Budget Task adopted from Andreoni et al. (2015)

For each decision number (1 to 6) below, decide the AMOUNTS you would like for sure today AND in 5 weeks by checking the corresponding box. *Example:* In Decision 1, if you wanted $19.00 today and $0 in five weeks, you would check the Option A box. Remember to check only one box per decision!

	Option A			**Option B**			**Option C**			**Option D**			**Option E**			**Option F**		
Decision	*Today*	*in 5 weeks*	*Check*	*Today*	*in 5 weeks*	*Check*	*Today*	*in 5 weeks*	*Check*	*Today*	*in 5 weeks*	*Check*	*Today*	*in 5 weeks*	*Check*	*Today*	*in 5 weeks*	*Check*
1	$19.00	$0	☐	$15.20	$4.00	☐	$11.40	$8.00	☐	$7.60	$12.00	☐	$3.80	$16.00	☐	$0	$20.00	☐
2	$18.00	$0	☐	$14.40	$4.00	☐	$10.80	$8.00	☐	$7.20	$12.00	☐	$3.60	$16.00	☐	$0	$20.00	☐
3	$17.00	$0	☐	$13.60	$4.00	☐	$10.20	$8.00	☐	$6.80	$12.00	☐	$3.40	$16.00	☐	$0	$20.00	☐
4	$16.00	$0	☐	$12.80	$4.00	☐	$9.60	$8.00	☐	$6.40	$12.00	☐	$3.20	$16.00	☐	$0	$20.00	☐
5	$14.00	$0	☐	$11.20	$4.00	☐	$8.40	$8.00	☐	$5.60	$12.00	☐	$2.80	$16.00	☐	$0	$20.00	☐
6	$11.00	$0	☐	$8.80	$4.00	☐	$6.60	$8.00	☐	$4.40	$12.00	☐	$2.20	$16.00	☐	$0	$20.00	☐

DMPL and CTB methods may appear more involved than MPL; in a meta-analysis, Imai et al. (2021) report that on average, a CTB study uses 22 questions. Andreoni et al. (2015) provide a discrete version of a CTB that contains only six decisions (see Table 6), although they use four different CTBs to identify time and risk preferences. Epper et al. (2020) employ a discretized CTB method using a graphical interface to rank patience of individuals in a large scale incentivized survey in Denmark.

DMPL and CTB may be preferred to simpler MPL if one wants to account for convexity of preferences, and may also be employed to estimate the present bias. In terms of the time required, the tasks appear comparable to a simple MPL for use as an add-on to the main experimental task. Rose and Rose (2019) provide oTree implementations of both methods.

Direct Method Attema et al. (2016) propose a direct method (DM) to measure temporal discounting without any assumptions on the utility function. The method is similar to MPL, but instead of choosing between a given gain at an earlier date and a larger one-time gain at a later date, each choice on the list is between a gain at an earlier date, and a sequence (a discretized flow) of gains starting at a later date, with the varying delays implying varying lengths of delayed sequences. As with MPL, the switching point between the delayed stream of payments and the early one-time payment is used to calibrate the discounting parameter and the present bias. Attema et al. (2016) provide experimental

evidence to confirm the implementability and validity of the method.[20] They claim that the method is not distorted by utility misspecifications, is easy to comprehend for subjects, requires fewer measurements and takes little time to administer in the lab. The method appears particularly useful if a researcher seeks to evaluate participant preferences over streams of temporal payments, rather than over one-time delayed payments. Rose and Rose (2019) provide oTree implementation.

Matching The matching method provides an alternative to choice tasks by directly eliciting the indifference point of a present monetary gain to some future gain. Benhabib et al. (2010) utilize this method in two frames: (1) present-oriented, soliciting the amount of money that, if paid at an earlier date, is considered as indifferent to a given amount of money paid at a later date; and (2) future-oriented, soliciting the amount of money that, if paid at a given future date, is considered indifferent to a fixed payment at an earlier date. The given amounts of later (respectively, earlier) payments, and the time intervals, varied. To provide incentives for truthful reporting, one of the questions is randomly chosen for payment, and the payment is determined using Becker–DeGroot–Marschak (BDM) mechanism (Becker et al., 1964).[21] The results showed evidence of present bias and were robust under several frames and utility specifications.

Frederick et al. (2002) cite two advantages of the matching method over choice experiments: (1) exact discount rates can be computed from a single response and (2) the anchoring effect no longer becomes an issue. However, they also discuss how the framing of the matching question can elicit different indifference points. Additionally, participants may use simple choice rules, such as multiplying future outcomes by 2, which may not accurately measure true time preferences.

Ifcher and Zarghamee (2011) use the present oriented matching method question and BDM mechanism to incentivize participants. Abstruse values were used in eliciting indifference across a specified time period in order to remove anchoring effects. The study asked several iterations of the matching question and used random ordering to remove any anchoring effects. Andreoni et al. (2015) use both present-oriented and future oriented hypothetical matching questions to test out-of-sample predictions of DMPL and CTB methods.

3.2 Qualitative Assessments

In addition to Money Earlier or Later (MEL) elicitation devices, qualitative assessments have been used to measure time preferences. Burks et al. (2012) use six self-report questions from

[20] However, they report that the separability condition that the method relies upon was rejected in one of the two tests; yet the condition held approximately for a majority of participants.

[21] The BDM mechanism works as follows. For the question chosen randomly for payment, to determine whether or not a participant is paid today or at the specified future date, a random number from a range between zero and the given delayed payment is drawn. If the random number is less than the reported indifference amount, then the participant would have to wait for the given delayed payment. If the random number is greater than their reported indifference amount then the participant is paid the number drawn today; see Benhabib et al. (2010). The BDM method is widely used; yet, some researchers argue that the mechanism may not be incentive compatible, especially for probabilistic payments (Horowitz, 2006), or that it causes participant confusion (Cason and Plott, 2014).

the World Health Organization for adults to measure short-term patience.[22] The participants respond to 5-point Likert scale (never, rarely, sometimes, often, very often). The results of the analysis find consistency in participant responses and cite evidence of a common factor (some underlying time preference) driving the results. They further find weak but significant correlations between the survey responses and the present-bias and discount factors elicited using MEL tasks. Further analysis however finds little predictive power in behavioral outcomes typically associated with impatience (i.e high BMI index, smoking, credit score, and job success). Vischer et al. (2013) validate qualitative non-incentivized ultra-short survey measure of patience using the question from the German Socio-Economic Panel (SOEP). The survey asks participants "How patient are you in general?" with respondents providing a rating on a scale from 0 (not at all patient) to 10 (very patient). The study reports a strong correlation between this qualitative assessment and the behavior in an incentivized time preference experiment. Falk et al. (2023) use the qualitative question "In comparison to others, are you a person who is generally willing to give up something today in order to benefit from that in the future?" and find a strong correlation of the answers with elicited experimental time preference parameters. These studies suggest that qualitative assessment may be a viable alternative to incentivized methods to quickly assess participant time preferences. In contrast, Chapman et al. (2025a) argue that qualitative self-assessments are likely to introduce non-classical measurement error and lead to biased preference estimates.

3.3 Time-Yoked Consumption and Impulsivity

Time-yoked consumption Measuring time preferences in Money Earlier or Later (MEL) tasks involves receiving income or monetary payments at different times. Time preferences over financial flows may be subject to consumption smoothing, in which despite receiving a monetary payment at a later time, some of the monetary payment is used at a still later time. To circumvent the problem of consumption smoothing over financial flows, researchers use time-yoked consumption tasks, in which participants choose the time of consumption of goods or services for which a further delay in consumption (i.e consumption smoothing) is not possible. The findings using time-yoked consumption tasks highlight the present bias (dynamically inconsistent preferences) and greater discount rates on primary rewards such as food as compared to monetary flows (Cohen et al., 2020).

Time-yoked studies are domain specific as they require the choice over specific goods. Mischel et al. (1989) marshmallow experiment is arguably the best-known time-yoked consumption task, in which children were asked to choose between consuming a smaller amount of marshmallows immediately versus consuming a larger amount at a later time. Using the Marshmallow Task, Mischel et al. (1989) find that delayed gratification in children is correlated with cognitive abilities and positive outcomes outside the laboratory such as social

[22]The World Health Organization questions are as follows: (i) How often do you have trouble wrapping up the final details of a project, once the challenging parts have been done? (ii) How often do you have difficulty getting things in order when you have to do a task that requires organization? (iii) How often do you have problems remembering appointments or obligations? (iv) When you have a task that requires a lot of thought, how often do you avoid or delay getting started? (v) How often do you fidget or squirm with your hands or feet when you have to sit down for a long time? (vi) How often do you feel overly active and compelled to do things, like driven by a motor?

competency and greater scholastic achievement. Fehr and Leibbrandt (2011) use a similar task with pralines to measure impatience in a field experiment with fisherman and find significant correlation between impatience and the fishermen field behavior.

Impulsivity task Time preferences, especially those modeled using hyperbolic discounting, are often associated with impulsivity and self-control problems (Laibson, 1997; Frederick et al., 2002; Kirby and Finch, 2010). Outside economics, impulsivity is often measured with the self-reported Barratt Impulsiveness Scale (Patton et al., 1995). Burks et al. (2012) propose an extension of the Marshmallow Task and devise an impulsivity task for adults using monetary reward. Participants are asked to wait for 10 minutes and are rewarded some monetary amount ($5) if they wait patiently during the task. They are not allowed to do anything during their task aside from sitting. Participants are provided an option to reduce their waiting time by some cost of the additional monetary reward by pressing a red button. The number of clicks by a participant acts as a proxy for impulsivity. Burks et al. (2012) find individuals with less present bias as measured using a MPL task are less likely to click the button, indicating a significant correlation between subjects' present bias and impulsivity.

3.4 Discussion and summary

Incentivized vs Non-Incentivized Instruments Several comprehensive time preference surveys find little evidence of the effect of incentives on measured time preferences (Cohen et al., 2020; Matousek et al., 2022). Brañas-Garza et al. (2021) conduct a study in the UK, Spain and Nigeria using laboratory, field, and online experiments to test the effect of incentives in intertemporal choice MPL responses and find little to no significance of incentives on the measured time preference parameters. On the other hand, influential field studies that correlate time preference with behavior outside the lab typically use either incentivized choice experiments (Chabris et al., 2008; Meier and Sprenger, 2010; Sutter et al., 2013; Epper et al., 2020), or non-incentivized surveys validated in the lab (Falk et al., 2018). Many laboratory and lab-in-the-field studies incentivize the participants, typically with probabilistic payments (Andersen et al., 2008; Benhabib et al., 2010; Andreoni and Sprenger, 2012a; Chapman et al., 2025b).

Transaction costs and medium of payment Transaction costs are the added logistical expenses of receiving a payment later.[23] These costs are higher with delayed payments and can introduce noise in measuring true time preferences due to the extra effort needed to collect the payment, or uncertainty associated with it. One way to equalize transaction costs between earlier and later dates is to have both payment dates in the future (Coller and Williams, 1999; Andersen et al., 2008). Another way is to use procedural measures to equalize transaction costs and ensure payment reliability (Ifcher and Zarghamee, 2011; Andreoni and Sprenger, 2012a; Andreoni et al., 2015). In field experiments, payments are often guaranteed

[23] In addition, Cohen et al. (2020) note that "it is possible that other factors, such as mistrust, temptation, inattention, confusion, context, and beliefs explain why humans behave as if they devalue promised future payments in a wide range of settings" (p. 388).

by trustworthy financial institutions to equalize transaction costs of different payment dates (Andersen et al., 2008; Meier and Sprenger, 2010).

Online or mobile forms of payments such as bank transfers or Venmo/cash apps have recently emerged as an alternative to cash payments to experimental participants (Harrison et al., 2022; Hoel et al., 2022; Kim, 2023). Several researchers point out that such cashless payment methods can be used to equalize transaction costs associated with receiving payments at different times and minimize future payment uncertainty; e.g., Wölbert and Riedl (2013); Hoel et al. (2022); Kim (2023).[24] Hoel et al. (2022) show that, for a student subject pool who have Venmo accounts, using Venmo instead of cash payments does not significantly affect elicited risk preferences and time discounting.[25] Epper et al. (2020) use electronic bank transfers to individual accounts to implement payments in a large incentivized survey using a CTB method in Denmark.

Comparison of elicitation methods Several studies provide comparison of methods in terms of their internal and external validity, demonstrating different advantages of different methods. Burks et al. (2012) compare four commonly used economic and psychology measures of time preferences and correlate them to real-world behaviors using a large sample of truck driver trainees. The methods considered are two incentivized MEL economic measures derived from Frederick et al. (2002) to find the discount rate and present bias, correspondingly; an impulsivity measure adopted from Mischel et al. (1989) for lab experiments ("the red button," as discussed above); and a self-reported survey on short-term impatience using the 5-point Likert scale. All four measures are validated as predictive measures for the real-world behaviors; however, the authors conclude that the economic approach of MEL for estimating the discount rate and the present bias parameters is the most robust predictor of outcomes among the four time preference measures considered.

Hardisty et al. (2013) compare matching and choice methods across loss and gain domains, reporting that matching methods provide more accurate theoretical predictions of hyperbolic discounting. Andreoni et al. (2015) find that both methods that control for risk preferences in time preference elicitation, DMPL and CTB, produce consistent estimates, but CTB is better for out-of-sample predictions compared to DMPL. Freeman et al. (2016) compare the choice method of MPL with the matching methods using the BDM procedure and the second price auction, and report that the discount rates elicited using the matching method are significantly lower than those elicited using MPL. Tasoff and Zhang (2022) find, based on a large survey results database, that the iterative time preference elicitation method is the best in terms of predicting the behaviors outside the economic laboratory. Chapman et al. (2024) argue that MPL and especially double MPL procedures may lead to measurement errors and biased preference estimates due to noise in participant responses, while DOSE procedure leads to more accurate individual estimates that are stable and strongly correlate with cognitive ability and demographics.

Summary With online payments, it has become easier to incentivize time-preference elicitation tasks and equalize transaction costs across payment dates. Rich graphical interfaces

[24] We are grateful to Elif Incekara Hafalir for pointing out this literature.

[25] All payments in their study are delayed; they do not elicit the present bias.

now available in experimental software such as oTree apps (Rose and Rose, 2019) allow for visual representation of seemingly elaborate tasks, such as CTB and the direct method by Attema et al. (2016), which should help participant comprehension and reduce time required to implement these methods in the lab. In addition, as with risk-preference elicitation tasks, duplicate elicitations with varying values are desirable to reduce the measurement error (Gillen et al., 2019). Yet, Li et al. (2022) warn that with too many repetitions participants may adapt to similarly structured questions in ways that reduce their external validity. Standard choice tasks may be supplemented with qualitative assessment.

In application to laboratory market and auction experiments, studies that correlate time preferences with market and auction behavior are relatively scarce. Deck and Jahedi (2015b,c) find that time delays in payments have a significant effect on behavior in first price and second price auctions, but not in all-pay auctions. It would be interesting to correlate participant time preferences, especially their short-term impatience and impulsivity, with trading behavior in laboratory markets and auctions. Insomuch as impulsivity relates to self-control, a possibly relevant study is Kocher et al. (2019), who show that depleting participant self-control leads to less rational trading in experimental asset markets.[26] It is also an open question whether preferences over temporal flows, rather than temporal one-time payments, would correlate better with behaviors in certain markets, such as asset markets.

4 Cognitive Ability

Cognitive abilities have been shown to correlate with strategic behavior and cooperation in the economic laboratory, and with economic behavior outside the lab (e.g., Burks et al., 2009, 2015; Fe et al., 2022; Gill and Rosokha, 2024). They have been reported to correlate with individual preferences towards time and risk, and with personality features (Burks et al., 2009; Dohmen et al., 2010; Brañas-Garza and Smith, 2016; Rustichini et al., 2016). Sofianos (2025) reviews commonly used cognitive ability measures to understand intelligence in decision-making in experimental settings. Bosch-Rosa and Corgnet (2022); Eber et al. (2024) and Sofianos (2025) further review the connection of cognitive abilities to risk, patience, behavioral biases, error-prone behavior and social preferences.

In laboratory market and auction experiments, higher cognitive abilities are linked to more rational (or more successful) performance in experimental common value auction (Casari et al., 2007); first- and second-price private-value auctions (Bartling and Netzer, 2016; Lee et al., 2020; Karmeliuk et al., 2022; Giebe et al., 2024); all-pay auctions (Breaban et al., 2020); asset markets (Breaban and Noussair, 2015; Cueva and Rustichini, 2015; Noussair et al., 2016; Corgnet et al., 2018), and double auction markets (Tai et al., 2018). On the market level, Bosch-Rosa et al. (2018) report that asset markets with cognitively sophisticated participants are not prone to bubbles, whereas markets with participants with low sophistication exhibit bubbles and crashes commonly observed in the Smith et al. (1988) setup. Corgnet et al. (2021) show that information aggregation in Plott and Sunder (1988) asset markets correlates with cognitive abilities of the market participants. Bosch-Rosa and

[26] Deck and Jahedi (2015a) show that increasing cognitive load leads to more impatient choices in tasks with real stakes. Yet Cobb-Clark et al. (2026) propose that impulsivity and self-control are distinctive personality traits that may not be related to time preferences or present bias; see also Huang et al. (2024).

Corgnet (2022) and Eber et al. (2024) review the relation of cognitive ability to financial decision making and trading in experimental financial markets.

Here we review the most common cognitive ability measures and their use in market experiments.

4.1 Cognitive Reflection Tests

The cognitive reflection test (CRT), introduced by Frederick (2005), is probably the most popular simple measure of cognitive ability among experimental economists. It tests the participants' ability in deliberate slow thinking. The original short 3-itemized test consists of the following questions:

- A bat and a ball cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost?
- If it takes 5 machines 5 min to make 5 widgets, how long would it take 100 machines to make 100 widgets?
- In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Each question has an intuitive but incorrect answer; however, more effortful thought is required to find the correct answer. For example, the intuitive answer for the bat-and-ball question is $0.10; however, the correct answer is $0.05.

The CRT is validated through cross-correlation with other well known validated aptitude tests such as the SAT, ACT, the Wonderlic Personnel Test, and the Need For Cognition (NFC) scale (Frederick, 2005). Further studies demonstrate that CRT is a strong predictor of performance on rational thinking tasks (Toplak et al., 2014), and it correlates with risk, time and social preferences (Oechssler et al., 2009). In a meta-analysis, Otero et al. (2022) show that cognitive reflection correlates substantially with all cognitive abilities and skills. Performance on CRT tests significantly correlates with successful performance in experimental markets (Breaban and Noussair, 2015; Cueva et al., 2016; Tai et al., 2018; Corgnet et al., 2018).

Brañas-Garza et al. (2019) provide a meta-analysis of 118 CRT studies between 2006 and 2015. They report that women do consistently worse than men; performance on the test is lower if the test is instituted after, rather than before, the experiment; and that performance on the test is not significantly different with and without monetary incentives.

One issue with the CRT test is the increasing familiarity of its three questions, which may decrease the test's efficacy. To account for popularity of the test, several studies suggest new test questions that can be used instead of, or in combination with the original items (Toplak et al., 2014; Thomson and Oppenheimer, 2016; Miklánek and Zajíček, 2020). Another simple way to address the familiarity of CRT questions is to modify the questions while preserving their logic. This approach that has been adopted by a number of studies (Stagnaro et al., 2018; Gillen et al., 2019; Breaban et al., 2020). For example, Gillen et al. (2019) turn the

"bat-and-ball" question into the "computer-and-keyboard" question,[27] and Breaban et al. (2020) turn it into "a shirt and a hat" question.

Brañas-Garza et al. (2019) find little effect of visibility of the test over time, if studies on Amazon Mechanical Turk are excluded.[28] Stagnaro et al. (2018) analyze a panel data aggregated from eleven studies on Amazon Mechanical Turk, and find significant correlation between CRT scores across time. While these results document stability, they do not directly address the issue of familiarity, as the study considers only two observations per individual, which are separated in time by over a year on average. Familiarity may become a bigger issue with fast developing internet resources, especially in online experiments.

Another important shortcoming of the CRT test is that it requires numerical skills. To address these issue, researchers created CRT items that rely significantly less on numeracy (Thomson and Oppenheimer, 2016), or use only word problems (e.g., Sirota et al., 2021). Otero et al. (2022) present evidence that all three groups of CRT tests: the original 3-item of CRT, other numerical CRT tests, and verbal CRT tests, significantly correlate with cognitive intelligence, when correcting for measurement error.

Raven's matrices, an alternative cognitive abilities test, avoids the numeracy issue by presenting problems that do not require numeracy or literacy.

4.2 Raven Matrices

Raven matrices are used frequently in psychology and economics to measure cognitive skills. The Raven Progressive Matrices Test (Raven et al., 1989) uses a non-verbal format to measure analytic intelligence, or the ability to solve problems; it is also often referred to as 'fluid intelligence" (Carpenter et al., 1990). The tests consist of a 3 x 3 visual matrix with each square containing a specific combination of geometric shapes and the bottom right box being empty or blank. The task is to identify from a list of six choices which geometric pattern best suits the empty space using the eight remaining boxes as references to deduce the correct answer. The progressive matrix tests are developed so that each item would increase in difficulty and progressively get harder.[29] Matrix reasoning also correlates particularly strongly with g, a single summary measure of cognitive ability (Gill et al., 2025). Due to the popularity of the test, multiple analyses have been run on the robustness of the test and have been repeated using the data from many countries. The studies show that the test maintains validity across different socio-economic groups, ethnic groups within the United States, and other countries, although there are significant variation in the mean scores across groups (Raven, 2008).

Condon and Revelle (2014) develop and evaluate a public domain measure of cognitive ability, the International Cognitive Ability Resource (ICAR). The test includes matrix reasoning items similar to Raven's Progressive Matrices, along with other categories such as

[27] "A monitor and a keyboard cost $350 in total. The monitor costs $300 more than the keyboard. How much does the keyboard cost?" (Gillen et al., 2019, Online Supplementary Material).

[28] Brañas-Garza et al. (2019) report a stronger visibility effect if Amazon Mechanical Turk studies are included in the analysis. Thomson and Oppenheimer (2016) report very high prior exposure rates to CRT questions for Amazon Mechanical Turk subjects.

[29] There are several iterations of the Raven Progressive Matrices: Standard Progressive Matrices, Standard Progressive Matrices Plus, Colored Progressive Matrices, and Advanced Progressive Matrices.

three dimensional rotation.[30] The authors evaluate the measure for its reliability and validity in internet-based untimed and unproctored studies conducted on a large international sample. ICAR test results are further shown to correlate with (self-reported) ACT and SAT scores, and with the performance on the commercially developed Shipley-2 test of cognitive functioning. ICAR test questions are accessible upon registration at https://icar-project.com, and have been used in Gillen et al. (2019); Snowberg and Yariv (2021); Chapman et al. (2024); Gill and Rosokha (2024); Cacho (2024).

Raven's matrices tests have been widely used by experimental economists as a measure of analytical non-verbal skills (e.g., Burks et al., 2009). They have been applied to measure cognitive abilities in market and auction experiments (Rustichini et al., 2016; Corgnet et al., 2018; Cueva and Rustichini, 2015; Karmeliuk et al., 2022; Giebe et al., 2024).

4.3 Other Cognitive Ability Tests

Gneezy et al. (2010) study how students learn in sequential "Race Game," a backward induction task that requires a participant to think backward from a given numerical goal to determine the best plan of action. The task is also referred to in Burks et al. (2009, 2015) as "Hit15",[31] which they use as a measure of planning ability. They show that performance on this task correlates positively with the ability to plan effectively in financial and economic decisions, collegiate success, job performance and persistence in employment. Cueva and Rustichini (2015) employ the Hit15 task, along with Raven's matrices, to measure participants' cognitive abilities in an asset market experiment. Bosch-Rosa et al. (2018) employ the Race Game task, along with CRT and the guessing game (Nagel, 1995) to study cognitive bubbles in an experimental asset market.

Tai et al. (2018) measure Working Memory Capacity (WMC), viewed in psychology as a cognitive resource for different tasks; they use five different tasks to measure WMC and find that WMC correlates with market performance when trading in laboratory double auctions with robot traders. Breaban et al. (2020) use probability calibration test along with CRT to correlate with performance in all-pay auctions, on the grounds that overbidding may be due to probability distortion. Roger et al. (2022) employ a number line estimation task to measure the acuity of the Approximate Number System (ANS), a cognitive function that allows to perform approximate calculations. They find that subjects who perform better on this task earn more in an experimental asset market.

Experimental economists also employ more broadly used measures of cognitive abilities. Casari et al. (2007) correlate student SAT/ACT scores with their performance in a laboratory common value auction. Dohmen et al. (2010) measure cognitive ability using questions similar to those of the Wechsler Adult Intelligence Scale (WAIS) and show that higher cognitive ability correlates with less risk aversion and more patience. Burks et al. (2009, 2015) employ a part of quantitative literacy test published by Educational Testing Service to measure numeracy, in addition to using Raven's matrices and Hit15 tasks.

[30] The test consists of 60 items, including 9-items letter and number series, 11-items matrix reasoning, 16-items verbal reasoning, and 24-items three dimensional rotation (Condon and Revelle, 2014).

[31] The game is played against a computer with the goal to be the first to add to a specific point total, such as 15, with players taking turns to add 1, 2 or 3 points. The game is solvable through backward induction.

4.4 Discussion and summary

Comparison of Tests CRT and Raven's Matrices are two most common methods to measure cognitive abilities in economic experiments. Which test is preferable depends on the type of cognitive ability that is relevant. The CRT questions are designed specifically to test for reflective reasoning, and to distinguish the use of intuitive reasoning (system 1) and reflective reasoning (system 2) (e.g., Noussair et al., 2016). For the asset market experiments, most studies employ CRT, referring to cognitive reflection as an important skill that is needed for rational trading (Noussair et al., 2016; Breaban and Noussair, 2015; Bosch-Rosa et al., 2018). Laboratory auction experiments often use Raven's matrices Bartling and Netzer (2016); Karmeliuk et al. (2022); Giebe et al. (2024) to test analytical abilities to solve problems.

Using a laboratory asset market setting with dispersed information as in Plott and Sunder (1988), Corgnet et al. (2018) argue that successful traders need a variety of cognitive skills: fluid intelligence (as measured by Raven's test) to make statistical inference to learn the true value of the asset; cognitive reflection skills (as measured by CRT) to avoid behavioral biases; and an ability to infer other traders' intentions, a skill measured by Theory of Mind eye gazing test (Bruguier et al., 2010). Corgnet et al. (2018) administer the three test, and demonstrate that all three positively correlate with trader earnings in their experimental laboratory asset market, suggesting that these tests may be complementary. Corgnet et al. (2022) further show that fluid intelligence as measured by Raven's matrices, as well as the Theory of Mind and the manipulative personality trait help explain forecasting performance in experimental asset markets.

Implementation Implementation details of both CRT and Raven's tests vary across studies. Some studies incentivize correct answers (Gillen et al., 2019; Snowberg and Yariv, 2021), whereas others do not (Gill and Rosokha, 2024). Another issue is time limits. Condon and Revelle (2014) test ICAR using unlimited time for all questions, including Raven's matrices. Gill and Rosokha (2024) give seven minutes to complete 11 questions on progressive matrices from ICAR, with the clock ticking down. Snowberg and Yariv (2021) allow 30 seconds to answer each question on both the Raven's matrices and the CRT tests, and report lower average performance on five Raven's matrices taken from ICAR compared to the mean scores reported in Condon and Revelle (2014).[32] This suggests that an excessively stringent time constraint may affect the performance on Raven's matrix tasks.

Brañas-Garza et al. (2019) report that the performance on the tests is lower if the test is instituted after the experiment, which suggests it may be better to conduct these test before the main experiment (as is done in e.g., Noussair et al. (2016); Giebe et al. (2024)). However, some studies include cognitive ability tests in the after-experiment questionnaire

[32] Condon and Revelle (2014), Table 2, report mean frequencies of correct answers between 0.77 and 0.28 for each of the 11 Raven's matrices, with the average of 0.52 among all 11 questions, and 0.41 among the last five of these progressively difficult questions. Snowberg and Yariv (2021), Table 1, report the average number correct, out of 5, as 1.2 (U.S. Representative Sample), 1.3 (MTurk), 1.8 (Caltech online study) and 1.3 (UBC), or 0.24-0.36 out of 1, which is lower than the averages reported in Condon and Revelle (2014). Cacho (2024) administer five first (easiest) matrix questions from the ICAR test with a one-minute limit per question in a lab university environment, and obtain the mean score of 0.65 per question which is very comparable to 0.63 reported in (Condon and Revelle, 2014, Table 2, corresponding items MR.43-MR.47).

(Cueva and Rustichini, 2015; Corgnet et al., 2018, 2021; Karmeliuk et al., 2022). Others conduct behavioral measurements in a separate session (Bosch-Rosa et al., 2018; Corgnet et al., 2018).

In terms of software implementation of the tests, a simple option, implemented, for example, in Cacho (2024), is to design the survey in Qualtrics, which offers a rich and user-friendly environment for questionnaire design, and then embed links to the survey into the experimental software (such as oTree).

Fast growing access to electronic resources and artificial intelligence tools presents new challenges to soliciting genuine participant responses, especially in online experiments. Condon and Revelle (2014) report that ICAR test administered unproctored online does not produce uniformly high scores with insufficient variance. Likewise, Snowberg and Yariv (2021) find little difference in performance when Raven's matrices and CRT test are administered either online or in the lab. Yet Brañas-Garza et al. (2019) note that answers to CRT questions are easily available through internet search, and that the performance on the CRT test improves for Amazon Mechanical Turk studies as the test's exposure increases. More recent behavioral research conducted online warns of threats to data quality from easily available artificial intelligence agents that are able to process audio-visual, as well as verbal information (Westwood, 2025). Inquiries into effective safeguards against these new threats are only starting to emerge (e.g., Fernandez et al., 2026).

Summary CRT, Raven's matrices, and, possibly Hit15 task, are common ways to measure cognitive abilities in market and auction experiments. Raven's matrices are often used in environments where the participants may benefit from analytical thinking in general, whereas CRT is often applied in dynamic settings where reflective reasoning is required. Both tests are short, can be administered as add-ons to the main experimental task, and are easy to embed into the experimental software. There is no consensus in the literature on whether the tests should be incentivized or not, and whether they should be given before or after the main experimental task. Imposing reasonable time limits, such as one minute per question, does not seem to inhibit participant performance on the tests. If using CRT, due to an increasing familiarity of its questions, it appears beneficial to modify the questions while preserving their logic. If conducted online, care should be taken to ensure genuine participant responses.

5 Financial literacy

Remund (2010) defines financial literacy through five key dimensions: knowledge of concepts, ability to communicate concepts, skill in managing personal finances, skill in financial decision making and confidence in financial planning. Empirical studies find that financial literacy correlates with demographic factors, such as gender, age, academic major, and socio-economic status (Chen and Volpe, 1998; Burke and Manz, 2014; Almenberg and Dreber, 2015; Henager and Cude, 2016; Bucher-Koenen et al., 2021) and with financial behaviors, such as saving habits and stock investment and participation (Lusardi and Mitchell, 2011; Andarsari and Ningtyas, 2019; Lusardi, 2019; Kaiser and Lusardi, 2024). These studies consistently report that higher literacy predicts better financial behaviors and outcomes.

The relationship between financial literacy and cognitive abilities has been well documented: large survey studies have found numeracy skills, which are more closely related to cognitive ability, to be a strong predictor of financial literacy (Skagerlund et al., 2018; Almenberg and Widmark, 2011). Darriet et al. (2021) review the literature on the relationship of numeracy and financial literacy in detail. Some studies suggest that numeracy and cognitive abilities are a stronger predictor of market behavior than financial literacy (e.g., Muñoz-Murillo et al., 2020). Given this overlap, many researchers opt to measure cognitive ability, which captures domain-general effects, rather than financial literacy, which is domain-specific. However, Burke and Manz (2014) and Darriet et al. (2020) report that numeracy alone does not correlate with more sophisticated financial decision making, while financial and economic literacy do. Here we briefly present some popular measures of financial literacy, and give a short overview of the evidence of relationship between financial literacy and behavior in experiments.

5.1 Measuring Financial Literacy

Survey Questionnaires Lusardi and Mitchell (2011) popularized a short method of measuring financial literacy. The three-item questionnaire focuses on measuring interest compounding, inflation, and risk diversification and has been applied in a wide variety of contexts, also known as "the Big Three". The survey is given below:

1. 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? (a) More than $102 (b) Exactly $102 (c) Less than $102; (d) Do not know; (e) Refuse to answer.

2. 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? (a) More than today; (b) Exactly the same; (c) Less than today; (d) Do not know; (e) Refuse to answer.

3. 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." (a) True (b) False (c) Do not know (d) Refuse to answer

The authors cite simplicity, relevance, brevity, and capacity to differentiate, as the main guiding principles in the design of their survey. The Big Three has been used in experimental studies to measure and control for financial literacy (e.g., Darriet et al., 2020).

Other experimental studies use different surveys to assess financial literacy, often combining items from widely recognized financial literacy measures. Almenberg and Dreber (2015) employ a 6-item survey that includes questions testing not only financial knowledge but also numeracy and cognitive abilities. Muñoz-Murillo et al. (2020) use a 6-item instrument focused on concepts such as interest compounding, inflation and debt literacy, using questions adapted from the OECD/INFE Toolkit for Measuring Financial Literacy (OECD, 2022). Corgnet et al. (2018) employ a 13-item financial literacy questionnaire adopted from Fernandes et al. (2014).

In line with the experimental economics methods, some papers choose to incentivize their financial literacy questionnaires. Muñoz-Murillo et al. (2020) use subjective belief elicitation procedure (Di Girolamo et al., 2015) to elicit beliefs of participants over possible answers and use a quadratic scoring rule to incentivize reporting the beliefs truthfully. Others choose not to pay individuals for participating in the survey (Almenberg and Dreber, 2015).

In addition to typical achievement tests measuring participants financial knowledge, survey questions often explore other dimensions of financial literacy such as confidence, attitudes, and skills. Lusardi and Mitchell (2014) discuss typical measure of self-assessment of financial literacy as given by the following question: "On a scale from 1 to 7, where 1 means very low and 7 means very high, how would you assess your overall financial knowledge?" They further discuss the mismatch in self-assessed knowledge and actual financial literacy scores in survey participants, suggesting overconfidence (for application see Mudzingiri (2024)). Longer surveys—such as the 18-item instrument proposed by Stella et al. (2020)—add dedicated measures of financial skills and attitudes, combining performance tests with self-reports. The study finds financial knowledge predicts financial skills, yet many individuals display strong skills despite lower knowledge. Moreover, attitudes are uncorrelated with knowledge but do align with skills. These findings underscore distinct subdomains of financial literacy, and raise the question of which dimension (knowledge, skill, or attitude) most strongly drives financial decision making.

5.2 Discussion

Relation to other behavioral features and experimental market behavior Financial literacy is strongly associated with cognitive ability, while other features such as risk aversion and patience exhibit weaker relationships. In Muñoz-Murillo et al. (2020), cognitive ability, as measured by the cognitive reflection task (CRT), robustly predicts financial literacy, while patience and risk aversion are not strongly correlated to financial literacy once controls are included. Beyond objective financial literacy, financial literacy overconfidence is associated with misalignment between subjective and elicited measures of financial risk preferences (Mudzingiri, 2024). These findings suggest that confidence can shape how individuals interpret and act on their risk preferences in the financial domain.

In market experiments, individuals with higher financial literacy perform more sophisticated stock trading, produce more accurate forecasts of inflation rates in simulation experiments, and are less affected by money illusion bias (Burke and Manz, 2014; Darriet et al., 2020). Yet Corgnet et al. (2018) show that financial literacy does not significantly affect trader performance in experimental assets market when controlling for cognitive abilities and personality traits.

Summary There is a variety of short financial literacy instruments. The most widely used is Lusardi and Mitchell (2011) "Big Three", which has been shown to predict stronger performance in financial and market experiments. A number of alternative surveys are available, each emphasizing different dimensions, such as numeracy, debt understanding, retirement planning. While financial literacy is shown to correlate with numeracy and cognitive abilities, in some cases it may be useful to distinguish it from these concepts, with financial literacy encompassing financial knowledge, skill, ability, and attitudes.

6 Regret and other emotions

Regret Many psychological and behavioral factors can impact the decisions made within competitive laboratory settings. One of these potential factors is regret aversion which Schwartz et al. (2002) suggest may be a more potent force in decision-making than loss aversion of Kahneman and Tversky (1979). Regret theory was introduced by Fishburn (1982); Bell (1982) and Loomes and Sugden (1982), with a key insight that people tend to compare their actual outcomes with the alternatives they would be in if they were to make different choices in the past (Loomes and Sugden, 1983); see Bleichrodt and Wakker (2015) for a survey. Posey et al. (2026) use a laboratory experiment to show coexistence of loss aversion and regret preferences.

There are several ways to measure regret that may be used in economic experiments. Bleichrodt et al. (2010) use a series of binary-choice tasks to elicit participants' indifference curves for utility functions that incorporate regret. Schwartz et al. (2002) develop a general methodology scale to measure maximizing, satisficing and regret by means of surveys. The scale includes five questions on a 7-point Likert-type scale that measure general tendencies for regret, such as "Whenever I make a choice, I'm curious about what would have happened if I had chosen differently." The measurement is validated through face validity, and principal-component factor analysis. Their study reports that people with a stronger desire to maximize are more sensitive to regret and are less satisfied in the ultimatum bargaining game.

Camille et al. (2004) measure regret using affect state solicitation method, which employs a small gambling task similar to risk preference elicitation, and a "what if" scenario comparing the two outcomes. After the task is realized and a participant sees both outcomes, they report their affective states on a slider scale from "extremely sad" to "extremely happy." These affective state measurements, though not incentivized, are validated by a physiological indicator, the skin conductance report. The study confirms that regret emotions, as indicated by a more negative affective state, are manifested in scenarios when a participant has rejected a gamble which resulted in a more favorable outcome in favor of a different gamble with a resulting lower payoff outcome.

Closely related emotional responses, discussed as possible explanations of overbidding in auctions, are the "joy of winning"[33] and "frustration of losing." Astor et al. (2013) provide physiological data (measuring skin conductance response and heart rate) to assess "frustration of losing" and "joy of winning" in auctions, and report the joy of winning in first price sealed bid auctions results in stronger reactions than the frustration of losing. They also provide an overview of how regret is measured in auctions, indicating that survey responses have been used successfully.

Hafalir et al. (2026) employ a post-auction survey soliciting affective states based on the questions from Camille et al. (2004) to demonstrate that a hybrid auction format used in some perishable good markets leads to lower loser regret than the standard Dutch auction (see also Cacho, 2024).

[33] Andreoni et al. (2007) and Goeree and Offerman (2003) reject overbidding due to the joy of winning in auction settings in favor of risk or loss aversion.

Emotions A number of studies consider the association of participant emotions and their behavior in auctions (Bosman and Riedl, 2003; Capra et al., 2010; Astor et al., 2013) or asset markets (Hargreaves Heap and Zizzo, 2011; Andrade et al., 2016; Breaban and Noussair, 2018; Kocher et al., 2019).[34] Studies differ in whether they induce emotional states or only measure them. Induction protocols use recollection or video clips of sad or happy events; participants are randomly assigned to emotional conditions, allowing to identify the causal effect of emotions on behavior (Capra et al., 2010; Andrade et al., 2016; Lahav and Meer, 2022). Measurement protocols such as Likert-scale self-reports (Hargreaves Heap and Zizzo, 2011; Kocher et al., 2019) and face-reading software (Breaban and Noussair, 2018) track emotions as they arise and establish correlations between emotions and behavior. Most asset market studies report that positive emotions such as excitement are associated with higher prices and overpricing, while negative emotions such as fear are associated with lower market prices (Hargreaves Heap and Zizzo, 2011; Andrade et al., 2016; Breaban and Noussair, 2018; Lahav and Meer, 2022). Kocher et al. (2019) find that traders with depleted self-control report experiencing more intense emotions.[35] However, Huber et al. (2024) report that they did not replicate the findings on the effect of emotions and self-control on asset market bubble formation.

Summary Regret and affective states may be measured using simple survey instruments that, depending on the purpose, may be administered before, during or after the main experimental task. These measurements can be used both to study the association of participant affective states with their market and auction behavior, and to compare participants' affect under different institutions.

7 Overconfidence

7.1 Types

Overconfidence has been studied extensively in both psychology and economics; see, e.g., Skala (2008) and Moore and Healy (2008) for overview of the literature. In decision-making contexts, overconfidence may lead to suboptimal outcomes driven by biased or miscalibrated beliefs (Biais et al., 2005; Eckel and Füllbrunn, 2015; Binnendyk and Pennycook, 2024). Overconfidence in experimental markets has been documented to correlate with reduced trading profits (Kirchler and Maciejovsky, 2002; Biais et al., 2005), increased trading activity (Deaves et al., 2009), and asset market bubbles (Eckel and Füllbrunn, 2015; Michailova and Schmidt, 2016).

Moore and Healy (2008) point out that the literature defines three different types of overconfidence: overprecision, overestimation, and overplacement. Overprecision, or excessive certainty in one's beliefs, implies that individuals underestimate uncertainty of outcomes,

[34] We do not review studies that use instruments to measure heart rate (e.g., Smith and Dickhaut, 2005) or brain activity (Dickhaut and Rustichini, 2008) to assess emotions.

[35] Muehlfeld et al. (2013) study investor trading behavior in a laboratory asset market depending on individual sensitivity of their basic neurophysiological systems, and find that traders with different sensitivities react differently to different asset market shocks.

underweight new information and update their beliefs too little when conditions change; it is often relevant in behavioral finance, as traders face uncertain fundamentals (such as asset values). Evidence from laboratory experiments shows that greater overprecision is associated with lower trading profits and related market inefficiencies (Kirchler and Maciejovsky, 2002; Biais et al., 2005). Overestimation refers to overconfidence in one's subjective performance being better than the objective truth, which—contrary to overprecision—provides a reference point to assess one's perception of their ability relative to their performance (Pikulina et al., 2017). Overestimation has been shown to correlate with excessive trading volumes and larger price bubbles in experimental asset markets of Smith et al. (1988) type, or overinvestment decisions (Michailova and Schmidt, 2016; Pikulina et al., 2017). Overplacement, or overconfidence in one's own ranking relative to others, is also relevant in competitive environments, including financial markets, where inflated beliefs about one's relative performance can lead to phenomena such as excessive market entry and trading volume (Camerer and Lovallo, 1999; Glaser and Weber, 2007; Deaves et al., 2009).

Moore and Healy (2008) argue that the three types of overconfidence stem from distinct phenomena, but are often treated as if they were the same, leading to empirical inconsistencies and methodological issues. Here we briefly review methods used to measure overconfidence in economic experiments, and point out the Moore and Healy (2008) critique as appropriate. As in other sections, we highlight the methods that could be used as short add-ons to market experiments. We constrain the discussion to self-assessment-based approaches to measure overconfidence, rather than inferring overconfidence from participant behavior.[36]

7.2 Methods

Klayman et al. (1999) refer to overconfidence as a systematic difference between subjective confidence judgment and observed accuracy. They discuss and compare two common methods to measure overconfidence, the two-choice method and the confidence interval method.[37]

Two-choice method Under the two-choice method, participants are presented with sets of questions, each with two alternative answers. For each question, the participant is asked to choose which answer they think is more likely to be right, and indicate their confidence that they have chosen correctly, on the scale of 50% to 100%. Overconfidence is then measured as the difference between the average reported confidence and the actual accuracy (percentage of questions answered correctly). Pikulina et al. (2017) implement the two-choice overconfidence method using financial literacy knowledge tests and subjective confidence. Michailova and Schmidt (2016) modify the measure by using the three-choice questions and confidence ratings between 33% and 100%; they demonstrate that higher overconfidence of market

[36] Alternatively, overconfidence can be inferred from participant actions; e.g., Camerer and Lovallo (1999) measure overconfidence by comparing market entry rates to skill-rank performance; see also Niederle and Vesterlund (2007). Odean (1999) uses gender differences in trading behavior as evidence of overconfidence.

[37] Klayman et al. (1999) draw the following analogs of these methods with real-world decision-making: "We are often called upon to make a choice between two alternatives, and then our subjective confidence in that choice determines how much we commit to one course, how much we seek further information, and how much we hedge our bets. Confidence-range estimates may seem less natural, but they also have everyday counterparts. Confidence ranges are implied whenever one plans a 'margin of error.'"

participants is associated with higher price expectations and larger asset market bubbles.

While the above method appears to measure excessive belief in own performance, i.e., overestimation, Moore and Healy (2008) note that item-based confidence measurements conflate overprecision and overestimation, since confidence in a single answer reflects both excessive certainty in one's information and an inflated belief in overall performance.[38]

Confidence interval method Under the confidence interval method, a participant is given a set of questions with a single numerical answer; for each question, they are asked to provide a range (a high and a low estimate) corresponding to a given level of confidence (usually 90 percent) that the participant believes the answer is in that range (Lichtenstein and Fischhoff, 1977; Russo and Schoemaker, 1992; Klayman et al., 1999).[39] Overconfidence is then measured by considering the frequency with which the participant's correct answers fall inside their confidence ranges (Kirchler and Maciejovsky, 2002; Biais et al., 2005; Glaser and Weber, 2007; Deaves et al., 2009). For example, Klayman et al. (1999) report that in their study, the correct answer fell inside the participants' 90 percent confidence ranges 43 percent of the time, indicating overconfidence. This type of overconfidence indicates overprecision (confidence intervals being too narrow), and is also referred to as miscalibration of beliefs (e.g., Biais et al., 2005).

Numerous studies indicate that the confidence interval method leads to higher measures of overconfidence compared to other methods, suggesting that it could be an artifact of the method itself (Klayman et al., 1999; Cesarini et al., 2006). Cesarini et al. (2006) attribute this phenomenon to subjects making a trade-off between informativeness and accuracy, and go as far as suggesting that "confidence interval estimation tasks will remain flawed as a measure of overconfidence " (p. 465).

Task-level assessment methods Moore and Healy (2008) use quizzes just like in the two-choice method, but measure overconfidence at the quiz level, rather than at the item level, by eliciting participants' subjective probability distribution (SPD) over their quiz scores. For each possible score of a ten-item quiz (0–10), a participant assigns a probability that it will be their actual score. Expected score is then calculated from SPD, and overestimation is defined as the difference between the expected score and the actual accuracy. Moore and Healy (2008) further use the SPD to construct the narrowest 90 percent confidence interval and measure how often the participant's actual score falls within it. A key advantage of this approach is that it disentangles overestimation from overprecision.

A related method for eliciting overconfidence on a task (rather than an item) level is to institute a multi-item knowledge or ability task, and then ask the participant to rate how many of the items they got correct. Cesarini et al. (2006) refer to this method as the frequency assessment, and apply it to measure overconfidence of economics students on a ten-item economics quiz; they document a significantly lower overconfidence assessed by the

[38] Suppose an individual answered a two-item question incorrectly but scored their probability greater than 50%. The probability provided would both indicate one's overconfidence in the precision of their information and one's performance.

[39] For example, one of the questions used in Biais et al. (2005) is to provide high and low estimates of Martin Luther King's age at death, such that the participant is 90% sure that the actual value falls within the range specified.

frequency method compared to the confidence interval method. Gillen et al. (2019) and Snowberg and Yariv (2021) measure overestimation of one's performance as the difference between self-assessed and actual performance on two five-item cognitive abilities tasks (CRT and Raven's Matrices).

Likert-scale measurements In a study of overconfidence in political behavior, Ortoleva and Snowberg (2015) measure overconfidence (in the sense of overprecision) by asking participants to answer four economic and political factual questions and report their confidence on a 1–6 Likert scale (from "not confident at all" to "certain"). Overconfidence is then assessed as the residual from regressing confidence on accuracy of one's answers. The authors report that robustness checks using four general knowledge questions yield substantively identical results, suggesting no domain-specific differences. McCannon et al. (2016) use 16-item Likert-scale measurement of overconfidence in a financial decisions experiment. Snowberg and Yariv (2021) implement a non-context-specific version of the task where participants guess the number of jellybeans in a pictured jar (1–1000) and report their confidence on the same 1–6 scale. The task is repeated three times and the average confidence is used in the measurement of overprecision.[40]

Methods to measure overplacement To assess overplacement, a participant is asked to rank their performance relative to others on a task. Overplacement is then measured as the difference between the reported subjective rank and the objective rank relative to a group. For example, Niederle and Vesterlund (2007) inform participants about their absolute performance in a real effort task, and then ask them to report their subjective ordinal rank in a four-person group.

In many studies, overplacement is assessed along with (over)-estimation of one's performance. Moore and Healy (2008) solicit participant beliefs about their own performance on quizzes, and ask them to compare their own performance (better or worse) relative to a randomly selected (previous) participant. Likewise, after having participants answer cognitive ability questions, Gillen et al. (2019) and Snowberg and Yariv (2021) ask them both "How many questions do you think you got correct?" (soliciting overestimation), and "Out of 100 other randomly picked survey participants, what percentage do you think answered more questions correctly than you?" (soliciting overplacement).

A special case of overplacement is Better-Than-Average (BTA) belief, measured by asking a participant to compare own performance to the average on a task[41] (e.g., Deaves et al., 2009; Pikulina et al., 2017; Glaser and Weber, 2007). While this approach is appealing due

[40] Likert-scale measurements have been also used to measure the illusion of control, a phenomenon which refers to one's self-assessed probability of success being greater than the objective probability (Langer, 1975). This phenomenon is closely related to overestimation. Deaves et al. (2009) measure the illusion of control using a simple two-item Likert-scale asking participants to rate from a scale of 1 to 5 (where 5 = totally agree and 1 = totally disagree) the following: (1) I never buy securities or funds that will underperform in the future; (2) I am not able to identify securities or funds with above-average performance in the future. The second question is reverse-coded as a simple measure to elicit illusion of control. Other experimental studies that measure the illusion of control include Fellner (2009); Filippin and Crosetto (2016); Sloof and von Siemens (2017).

[41] For a broad psychology literature review on BTA see Zell et al. (2020).

to its simplicity, Moore and Healy (2008) recommend a comparison of own performance to a randomly selected other participant instead. The random peer performance has a much higher variance than the average statistic, and therefore comparing own performance to a random peer provides a cleaner measure of relative belief.

7.3 Discussion and summary

Hard-easy effect, questions domain, and incentives There are several issues that should be taken into consideration when measuring overconfidence. When using knowledge or ability tasks, the tasks should be balanced for difficulty due to the hard–easy effect: people tend to overestimate own absolute performance on hard tasks and underestimate it on easy tasks; conversely, overplacement is higher in easy domains and lower in hard ones (Moore and Healy, 2008; Grieco and Hogarth, 2009). Further, confidence is correlated with underlying competence, which leads some researchers to use questions in a specific field, such as economics or finance (e.g., Michailova and Katter, 2014), while others recommend general knowledge questions (Moore and Healy, 2008). Li and Noussair (2025) show that participant overplacement relative to their peers is decreased in more selective groups, attributing it to a sorting bias. There are further considerations in avoiding other biases, such as gender or cultural biases, in the choice of tasks (e.g., Michailova and Katter, 2014).

Psychology studies traditionally do not incentivize confidence elicitations, while many economics studies do. Among the latter, Niederle and Vesterlund (2007) paid participants $1 for each correct guess of their rank. Moore and Healy (2008) and Li and Noussair (2025) incentivize their overconfidence elicitation tasks using quadratic scoring rules, which reward accurate assessment. The confidence interval method is not incentivized, although incentivized interval methods to measure beliefs have been proposed (Schlag and van der Weele, 2015).[42] Cesarini et al. (2006) report positive but insignificant effects of monetary incentives on increasing accuracy of self-assessment using the frequency method. Gillen et al. (2019); Snowberg and Yariv (2021) report paying for correct answers in all tasks.

Task length and compact measurements Many methods discussed above (the two-item method, the confidence interval method and the SPD) rely on multiple questions and quizzes that need to balance the hard-easy effect and the areas of general knowledge, among other aspects; this could make the instrument time-consuming. For example, Moore and Healy (2008) use a lengthy battery of tests, with the total of 18 different quizzes (180 questions). Such length has some justification, as evidence suggests that individual confidence estimates stabilize after roughly 50 item confidence reports (Bouleau et al., 2025).

Some studies offer compact measurements of overconfidence. Gillen et al. (2019) and Snowberg and Yariv (2021) measure all three types of overconfidence with a short over-precision task (a "jellybeans in a jar" item), coupled with likert-scale subjective confidence elicitation, and add overestimation/overplacement questions to the cognitive ability tasks.

[42] Schlag and van der Weele (2015) propose an incentivized "most likely interval method" to elicit beliefs. The participants are asked for the interval to which the answer belongs most likely; the method rewards accuracy but imposes penalty for reporting a wider interval. However, the probabilities of the answer falling into the interval are not explicitly stated, making measuring overconfidence problematic.

Summary The type of overconfidence that is most relevant depends on the contents of the study; e.g., overprecision is relevant in the presence of uncertainty where new information calls for belief updating about some aspects of the market; whereas overestimation and overplacement may play a role in decisions on how much to engage in trading or on whether and how to bid in an auction. Short measurements such as those employed in Gillen et al. (2019) and Snowberg and Yariv (2021) may be easily adopted as add-ons to the main experiments, especially if cognitive abilities are also measured. Such measurements avoid the need to balance hard and easy questions or to choose the questions domain, and avoid biases in the choice of tasks. As with other behavioral instruments, repeating the overconfidence tasks several times improves measurement. Most tasks, aside from those using confidence intervals, can be easily incentivized.

8 Competitiveness and propensity to speculate

Competitiveness is defined as "an individual's general tendency to select into competitive environments" (Bönte et al., 2017). Theoretical frameworks have cited competitiveness as an important driver of productivity (e.g., Lazear and Rosen, 1981). Empirical research shows that competitiveness correlates with market and labor outcomes such as income, occupation, completed level of education, and field of study (Buser et al., 2024). Laboratory studies demonstrate relationship between competitiveness and behavior in experimental games (Griffith and Rust, 1997; Niederle and Vesterlund, 2007). This section gives a brief overview of two popular types of methods to measure competitiveness: methods based on a real-effort tournament self-selection task (Niederle and Vesterlund, 2007), and survey methods.[43]

8.1 Measures of competitiveness

Self-selection into tournaments Niederle and Vesterlund (2007) is the most popular method to measure competitiveness based on the revealed preference paradigm. Participants engage in a real effort task of adding five two-digit numbers. In the first round of the experiment, participants are paid using the piece-rate scheme (a fixed payment per correct answer) and receive feedback on their performance (the number of correct answers). In the second round, participants are grouped into tournaments in which only the highest performer receives a higher prize per correct answer, while others receive nothing. Participants do not receive feedback on the outcome of the tournament until the end of the experiment. In the final round, participants choose between piece-rate and tournament-payment schemes. The self-selection of participants into the tournament is used as a proxy for competitiveness.

This experimental framework has been highly influential, forming a basis for numerous subsequent studies in both laboratory and field settings (e.g., Leibbrandt et al., 2013; Bönte et al., 2017; Saccardo et al., 2018; Fallucchi et al., 2020; Gillen et al., 2019; Snowberg and Yariv, 2021). While many studies continue to use the original real-effort arithmetic task, others adopt alternatives, such as a ball-throwing task (Leibbrandt et al., 2013; Saccardo et al., 2018) or a trivia quiz (Bönte et al., 2017). A valid methodological concern is that observed

[43] Chapter 2 of Haran and Bereby-Meyer (2024) provides a detailed overview of methods to measure competitiveness in psychology and economics.

differences in tournament entry may not only reflect underlying preferences for competition but also task-specific ability, confidence and risk preference (Gneezy et al., 2003b; Niederle and Vesterlund, 2007; Saccardo et al., 2018). Bönte et al. (2018) modify the task by asking participants to compete against their own past performance, and report that the competitiveness gap between men and women can be mainly explained by gender differences in risk preferences. Miklánek and Zajíček (2020) attempt to disentangle competitiveness from risk preferences by making participants choose between a tournament and a lottery, arguing that this design is robust to variation in risk preferences. Urbig et al. (2020) remove the ability from the task and use choice to compete in a chance task based on dice rolling to measure competitiveness.

Saccardo et al. (2018) convexify the competitiveness decision by allowing participants to allocate a percentage of their payoff between piece-rate and tournament schemes, rather than choosing exclusively between the two; see also Fallucchi et al. (2020). This allows to measure competitiveness not only at the extensive margin (the binary choice to compete or not) but also at the intensive margin (the degree to which individuals prefer competition).

While self-selection into the tournament is arguably the most popular way to measure competitiveness, it may require a considerable amount of time. In comparison, survey measures may be a quicker option more suitable as an add-on to the main market experiment. Some of these measures have been validated using the tournament task.

Survey measures Unlike the incentivized tournament-entry design of Niederle and Vesterlund (2007), survey measures of competitiveness rely on self-perception. This makes them easy to implement, but raises concerns about self-report bias and construct validity. Fallucchi et al. (2020) suggest a single-item survey question to measure competitiveness. They report that out of eight considered survey items, "Competition brings out the best in me" (1 - "Not at all like me" to 7 - "Exactly like me"), best predicts self-selection into tournaments. The results are robust after controlling for ability, risk attitudes, and beliefs of winning the tournament. Additionally, they assert the item's external validity finding significant group differences between competitive (professional athletes) and non-competitive (non-athletes) individuals. Other studies have implemented multiple-item surveys to measure competitiveness. Eckel and Füllbrunn (2015) measure competitiveness in an asset market experiment using a multi-item Type-A personality survey adopted from Friedman (1996). They report that their measure of competitiveness is positively correlated with the incidence of severe market bubbles. Bönte et al. (2017) use a subjective four-item survey measure of competitiveness assessed on a Likert scale, and report it to positively correlate with the tournament entry measure. Buser et al. (2026) use a general question on competitiveness based on Dohmen et al. (2011) qualitative measure of risk preference, and report that it correlates well with the incentivized measure of Niederle and Vesterlund (2007) and with education and labor market performance.

Some survey-based measurements further disaggregate competitiveness into distinct underlying motivations, rather than treating it as a single construct. Bönte et al. (2017) distinguish between competitiveness motivated by personal development, and competitiveness not motivated by personal development, finding that only the latter predicts tournament entry. Similarly, Urbig et al. (2021) identify multiple dimensions of competitiveness and show that

gender differences vary across these dimensions; see also Buser and Oosterbeek (2023).

8.2 Propensity to speculate

Aside from competitiveness, asset market researchers may measure individual propensity to speculate as a potential contributing factor to market bubbles. Janssen et al. (2019) employ the Speculation Elicitation Task (SET), based on the bubble game by Moinas and Pouget (2013). Three traders sequentially buy and sell an asset commonly known to have zero fundamental value, each hoping to resell it at a higher price to the next trader in line. Traders do not know their position in the sequence, but a higher offer price implies a higher probability of being last, and hence of being left holding a worthless asset. Under common knowledge of rationality and finite price caps, backward induction implies no trader should ever buy, and a bubble should never form. However, Moinas and Pouget (2013) observe substantial trading, consistent with models of bounded rationality. Janssen et al. (2019) adopt the game to solicit participants decision to buy at every possible price, with the number of buy-decisions constituting the SET score. They find that this SET measure strongly correlates with individual's belief about others' propensity to speculate, and further report that asset markets with higher trader SET scores exhibit significantly higher overpricing. Miklánek and Zajíček (2020) also find that SET scores are linked to trading activity and trader profit.

8.3 Summary

The most widely-adopted method to measure competitiveness remains the tournament entry design of Niederle and Vesterlund (2007). As an alternative, recent studies have introduced validated survey instruments; both those based on a single item (Fallucchi et al., 2020) and those based on multiple items (Bönte et al., 2017; Eckel and Füllbrunn, 2015) have been consistently shown to correlate strongly with the Niederle and Vesterlund (2007) measure. The survey methods' main advantages are that they are inexpensive, easy to administer on a large scale, and—unlike behavioral tasks—can also be designed to disaggregate competitiveness into distinct dimensions for researchers interested in exploring its underlying components.

For markets where speculative behavior may play a role, individual propensity to speculate may be measured using the SET score (Moinas and Pouget, 2013; Janssen et al., 2019). When measuring competitiveness, related traits such as ability, confidence, and risk preferences should be controlled for, as they have been found to correlate with competitiveness and can confound interpretation if left uncontrolled (Saccardo et al., 2018; Snowberg and Yariv, 2021). In addition, a large literature documents robust gender differences in competitiveness (Niederle and Vesterlund, 2007, 2011), although these effects depend on task design and experimental context (Gneezy et al., 2003b; Datta Gupta et al., 2013).

Competitiveness and speculation measures have been successfully incorporated as add-ons to the main experimental tasks, such as laboratory asset markets. Eckel and Füllbrunn (2015) use a survey-based method, while Sutter et al. (2020) and Miklánek and Zajíček (2020) employ the tournament-self-selection method to measure competitiveness, suggesting that both simple survey items and short self-selection tasks may be a practical way to embed competitiveness measures into market experiments. Janssen et al. (2019) and Miklánek and

Zajíček (2020) adopt the SET scores to measure individual speculative tendencies. Taken together, these approaches provide a flexible toolkit for researchers interested in measuring these features while accounting for its key correlates and contextual dependencies.

9 Existing market studies with behavioral tests

In this section, we give a short overview of existing laboratory market experiments that employ behavioral measures. Table 7 presents a non-exhaustive list of studies grouped by the type of market institution studied: asset markets, commodity markets, auctions, and other markets and related financial decision tasks. Many of these studies are already mentioned in the sections above in discussing specific behavioral measures. While some studies focus on measuring a single feature, others use batteries of instruments to measure several features of interest.

Asset markets Among auction and market experiments, behavioral measures have been most widely employed in experimental asset markets, both those studying information aggregation (following the framework of Plott and Sunder, 1988), and those investigating irrational pricing and asset market bubbles and crashes (following Smith et al., 1988).

Asset market experiments studying information aggregation are among the first to incorporate behavioral measurements in experimental design. The early studies mostly focus on the role of overconfidence in trader performance, as overconfidence may affect how traders process new information. Consistent with this, studies report that overconfidence leads to lower trader earnings (Kirchler and Maciejovsky, 2002; Biais et al., 2005); see also Deaves et al. (2009). Risk preferences are measured either to investigate their correlation with trading activity (Fellner and Maciejovsky, 2007), or as controls to account for trader heterogeneity (Kirchler and Maciejovsky, 2002; Muehlfeld et al., 2013). Later studies investigate the role of cognitive abilities (as measured by either CRT or Raven's matrices) and find their positive association with trader earnings and market efficiency (Corgnet et al., 2018, 2021) and forecasting performance (Corgnet et al., 2022). Some studies also consider the role of financial literacy (Corgnet et al., 2018).

Asset market experiments that study the persistent phenomena of price bubbles and crashes, as pioneered by Smith et al. (1988), often consider the relationship between cognitive abilities and mispricing. Cognitive abilities are most frequently (but not exclusively) measured using CRT as cognitive reflection and analytical skills assessed by this task are necessary to determine the asset's fundamental value (Breaban and Noussair, 2015; Levine et al., 2017). The studies find that higher trader cognitive skills are associated with smaller bubbles, less mispricing and lower market volatility (Breaban and Noussair, 2015; Noussair et al., 2016; Levine et al., 2017; Bosch-Rosa et al., 2018; Kocher et al., 2019; Miklánek and Zajíček, 2020; Roger et al., 2022); see Bosch-Rosa and Corgnet (2022) and Eber et al. (2024) for excellent surveys. Risk preferences are also considered, with some studies finding that higher risk tolerance leads to higher trading intensity (Breaban and Noussair, 2015) and larger bubbles (Eckel and Füllbrunn, 2015; Kiss et al., 2022); while others find no significant relationship between risk aversion and trading behavior (Michailova et al., 2017). In a one-period asset market setting, Aragó et al. (2022) find that more risk averse participants are

less likely to participate in high volatility markets altogether. Risk preferences are also used as controls, usually motivated by concerns that individual differences in risk attitudes could confound the effects of other variable of interest (Michailova and Schmidt, 2016; Miklánek and Zajíček, 2020). Some studies also measure competitive preferences (Eckel and Füllbrunn, 2015) and overconfidence (Michailova and Schmidt, 2016), with both features reported to be positively associated with larger price bubbles.

Commodity markets In comparison to the asset market experiments, behavioral measures are less frequently employed in commodity markets, most likely because commodity markets, especially those operated as double auctions, are known to exhibit high efficiency and robust convergence to the competitive equilibrium (Smith, 1962). Among a few commodity markets studies that consider behavioral features, Tai et al. (2018) investigate the role of cognitive abilities measured by Working Memory Capacity test and find a positive association with trader profits. In a double auction setting where trades generate negative externalities, Sutter et al. (2020) measure participant risk aversion and competitiveness, among other characteristics, and find no significant association between these features and the number of trades. One could envision more scope for the role of behavioral features in more complex environments, such as markets with dynamic structures or other interdependent multi-market systems; however, to the best of our knowledge, trader behavioral characteristics have not been measured in such settings.

Auctions Auction experiments have considered the role of risk preferences, cognitive abilities, emotions and regret in bidder behavior.

Risk aversion is a common explanation for overbidding in auctions (Cox et al., 1985, 1988). Correspondingly, several studies assess bidder risk preferences in laboratory auctions experiments. A variety of instruments have been employed, including Gneezy and Potters (1997) risky projects (Agranov and Yariv, 2018; Karmeliuk et al., 2022), BRET (Füllbrunn et al., 2019), and HL multiple price lists (Kirchkamp and Mill, 2021). While some studies find that, consistent with the theory, risk aversion is associated with higher bids (Füllbrunn et al., 2019; Karmeliuk et al., 2022), others find no statistically significant effect (Agranov and Yariv, 2018). Kirchkamp and Mill (2021) report that in second-price all-pay auctions, spiteful preferences play a more important role than risk aversion in explaining observed bids.

A number of studies consider the role of cognitive abilities as measured by Raven's matrices. Karmeliuk et al. (2022) write that in auctions, bidders must understand best responses, so that higher cognitive abilities are associated with more rational behavior in the game-theoretical sense. Indeed, they show that higher cognitive abilities are associated with lower bids and higher profits in first-price auctions, and more rational bidding in second-price auctions. Giebe et al. (2024) further find that higher cognitive ability is associated with less overbidding in second-price auctions.

Emotions in auctions have been considered since the early 2000-s, with studies showing that emotional states influence bidding behavior (Bosman and Riedl, 2003; Capra et al., 2010). Conversely, some recent experimental studies (Cacho, 2024; Hafalir et al., 2026) indicate that auction formats may differ in the positive affect and loser regret that they

induce on bidders, suggesting that some unusual auction formats could have evolved to increase bidder satisfaction.

Cacho (2024) studies bidder behavioral features, including cognitive abilities measured by Raven's matrices (Condon and Revelle, 2014), staircase risk and time preference measurements and qualitative assessments of risk and time (Falk et al., 2023), qualitative measures of competitiveness (Fallucchi et al., 2020) and affective states and regret (Camille et al., 2004; Schwartz et al., 2002), in Dutch and hybrid auctions investigated by Hafalir et al. (2026). The complete survey is included in Appendix A. Aside from different auction formats inducing different degrees of happiness and regret on participants, cognitive abilities are found to correlate with bidder earnings in Dutch auctions, with other behavioral measurements proving insignificant.

Other Market and Financial Experiments The role of behavioral features has also been considered in oligopoly and financial experiments. Anderson et al. (2012) study the impact of risk preferences on strategic behavior, specifically collusion, in a repeated Bertrand duopoly setting. Pikulina et al. (2017) examine investment decisions and find that overconfident individuals tend to over-invest, while less confident individuals under-invest. These patterns remain robust even after controlling for risk preferences and cognitive ability (as measured by CRT). Darriet et al. (2020) investigate the role of financial literacy and numeracy in shaping money illusion, i.e., the inability to correctly account for inflation and deflation in financial contexts. Their results show that higher financial literacy reduces these effects, whereas numeracy alone does not have a significant impact.

Table 7: Market experiments with behavioral measures, by market setting

Paper	Market setting	Behavioral measures	Task timing	Behavioral findings
Information Aggregation Asset Market setting (Plott and Sunder, 1988)				
Kirchler and Maciejovsky (2002)	Multi-period information aggregation market	CE risk and lotteries; two measures of overconfidence	Before market (risk); at the start of each trading period (overconfidence)	Traders do not generally show overconfidence; classification depends on measure. Risk is used as control
Biais et al. (2005)	Asset Market (PS Market 7, Series C)	Subjective Confidence Intervals (SCI)	Separate session (prior)	Overconfidence lowers trading performance by increasing susceptibility to winner's curse
Fellner and Maciejovsky (2007)	Several different asset markets	CE risk	Before market	High risk averse traders submit fewer offers in markets and engage less often in trades
Deaves et al. (2009)	Noisy-signal single period asset market	3 measures of overconfidence: Calibration Based Overconfidence (CBO), Better-Than-Average (BTA), Illusion of Control (IoC)	Before and after market	Overconfidence is strongest as measured by CBO, which is positively associated with trading activity; BTA is also positive, while IoC is not significant
Muehlfeld et al. (2013)	Risky single-period asset market	HL risk; loss aversion; overconfidence	After market	Measures used as controls for main analyses
Corgnet et al. (2018)	Asset Market (PS Market 9, Series C)	Cognitive abilities: Raven matrices, CRT, Fin. Lit; ToM; HL Risk; Big Five	After market	Cognitive abilities: fluid intelligence, CRT, and ToM are positively associated with trader performance
Corgnet et al. (2021)	Asset Market (PS Market 9, Series C)	CRT	Separate session (prior) and after market	Information aggregation is enhanced when participants have high cognitive abilities, and this is common knowledge among them

Continued on next page

Paper	Market setting	Behavioral measures	Task timing	Behavioral findings
Corgnet et al. (2022)	Asset Market	Raven matrices, BRET (risk), ToM, manipulativeness	After market	Forecaster performance depends on intelligence, manipulativeness, and theory of mind; which skills matter depends on the level of market mispricing and information display
Bubble-Prone Asset Markets: Smith et al. (1988) setting (SSW)				
Breaban and Noussair (2015)	SSW asset market	Loss aversion (Fehr and Goette, 2007); CRT; HL risk	Before market	Higher CRT score predicts prices closer to fundamental values, greater risk aversion – lower prices and greater loss aversion – lower quantity traded
Cueva and Rustichini (2015)	Constant-fundamental-value asset market	HL risk; Hit-15; Raven's matrices; confidence (overestimation and overplacement)	Risk before; cognitive skills and confidence after market	Higher average cognitive skills of traders are associated with lower market volatility
Eckel and Füllbrunn (2015)	SSW asset market	EG risk; Competitiveness (Type A); Personality Traits	Not reported	Markets with more competitive participants have larger bubbles; markets with more risk-averse participants have shorter and smaller bubbles
Noussair et al. (2016)	SSW market with futures	CRT	Before market	Higher CRT is associated with higher earnings and better price discovery
Michailova and Schmidt (2016)	SSW asset market	Overconfidence battery	Separate session (prior)	Markets with more overconfident participants have higher trading volume, higher magnitude bubbles and higher asset price forecasting errors

Continued on next page

Paper	Market setting	Behavioral measures	Task timing	Behavioral findings
Andrade et al. (2016)	SSW market	Emotions induced through video clips	Before market	Excitement results in larger bubbles
Michailova et al. (2017)	SSW asset market	Overconfidence; HL risk	Separate sessions before and after	Overconfidence affected trading for females; risk had no influence on outcomes
Levine et al. (2017)	SSW asset market	Fin Lit; Race to 100, CRT, Need for Cognition; Guessing Game	Separate session (prior)	Analytical intelligence and strategic intelligence lead to better trading performance
Bosch-Rosa et al. (2018)	Asset market	HL risk; CRT; Guessing Game; Race to 60	Separate session (prior)	Markets with cognitively sophisticated traders do not exhibit bubbles and crashes
Janssen et al. (2019)	Asset market	HL risk; Non-incentivized Dohmen risk; CRT; SET	Before market	Individual speculative behavior fuels overpricing in (experimental) asset markets. Other measured behaviors are used as controls
Kocher et al. (2019)	SSW with Stroop task	CE risk; CRT; emotions	Before market	Reduced self-control increases market overpricing
Miklánek and Zajíček (2020)	SSW constant-value market	Speculation task; CE risk; extended CRT; overconfidence; NV competitiveness	Separate session (prior)	Speculative preferences are linked to trading activity. Other traits not significant
Roger et al. (2022)	Asset market with fluctuating fundamental value	Number Line Estimation (NLE) task; self-reported risk	NLE task unclear; risk after market	Higher NLE scores are associated with higher earnings; risk insignificant
Kiss et al. (2022)	SSW asset market	GP Risk; Stag-Hunt; CRT	Before Market	Markets with the most risk-tolerant traders exhibit larger overpricing than markets with the most risk-averse traders

Continued on next page

Paper	Market setting	Behavioral measures	Task timing	Behavioral findings
Aragó et al. (2022)	One period asset markets with high and low risk regimes	SGG risk	Pre- and post-market	Market risk aversion varies between high and low volatility periods due to participation effect
Huber et al. (2024)	SSW with Stroop task (replication study)	Emotions; CRT; Raven's matrices; ToM; CE Risk	Same session	Cognitive skills explain trading success; results on emotions and self-control are not replicated
Commodity Markets				
Tai et al. (2018)	Double auction commodity market	Working Memory Capacity (WMC)	Separate session (after)	Participants with high-WMC make more profits than low-WMC and also underbid more than low-WMC
Sutter et al. (2020)	Double auction commodity market with negative spillovers from trades	CE risk; NV competitiveness; Other socio-demographic features	After market	Risk and competitiveness are not significantly associated with trading
Auctions				
Bosman and Riedl (2003)	First-price auction	14-item emotion Likert scale	Between market parts	Economic shocks impact emotions. Participants in negative emotional state place higher bids; positive emotional states do not affect bidding
Capra et al. (2010)	Random *n-th* price auctions	3-item mood self-report Likert	Before market	Treatment of inducing mood in participants was significant; positive mood led to upward bidding bias
Brunner et al. (2014)	English premium auctions (EPA)	HL Risk	Separate online session (prior)	Revenues in the EPA are significantly higher when bidders are risk loving rather than risk averse

Continued on next page

Paper	Market setting	Behavioral measures	Task timing	Behavioral findings
Agranov and Yariv (2018)	First- and second- price auctions	GP Risk	After market	No significant effect of risk on bidding strategies
Füllbrunn et al. (2019)	First-price sealed-bid auction	BRET risk	Before market	Risk aversion is associated with overbidding and overpricing
Kirchkamp and Mill (2021)	Second-price all-pay auction	HL risk; Spite preferences	Before market	More spiteful bidders overbid while risk-averse bidders underbid. Spite preferences may explain more than risk
Karmeliuk et al. (2022)	Team bidding in first-price, second-price, and English auctions	GP risk; Raven matrices	After market	Individuals with higher cognitive ability and those who are less risk averse bid lower and make more profits in first-price auctions
Giebe et al. (2024)	First-price auctions (FPA) and second-price auctions (SPA)	Raven Matrices	Before market	Experience in FPA reduces overbidding in SPA for lower cognitive ability bidders; experience in SPA reduces bid shading in FPA for higher ability bidders
Hafalir et al. (2026)	Dutch and Hybrid auctions	Regret and Affect	After market	Participants reported less regret and greater happiness in hybrid auctions than in Dutch auctions
Cacho (2024)	Hafalir et al. (2026) auctions	Raven Matrices; Risk and Time Staircase; Regret; Competitiveness	Before and after market	Cognitive ability is associated with higher earnings in Dutch but not hybrid auctions. Regret aversion may explain the hybrid auction adoption, while other features insignificant
Financial Decision and Other Markets				

Continued on next page

Paper	Market setting	Behavioral measures	Task timing	Behavioral findings
Fellner (2009)	Illusion-of-control in investment portfolio choice task	HL Risk	Before market	Higher risk aversion is associated with more investments in less risky assets. Subjects show illusion of control
Anderson et al. (2012)	Tacit collusion in Bertrand duopoly	HL Risk	Before market	Less risk averse subjects price higher than more risk averse subjects in markets where actions are strategic substitutes
Burke and Manz (2014)	Inflation forecasting	Financial literacy	After main task	More financially literate participants have smaller forecasting errors, and more accurate perceptions of actual US inflation
Pikulina et al. (2017)	Individual investment task	HL Risk; CRT; overconfidence	Before main task	Miscalibrated confidence in financial skills predicts investment: results are robust to risk aversion, project riskiness, and incentives
Darriet et al. (2020)	Financial task	Financial literacy; money illusion; Numeracy; HL Risk; Loss aversion	Intervened with and after main task	Individual with financial knowledge are less sensitive to money illusion; no evidence of impact of numeracy

Notes: Task timing: "Before/After market" refers to behavioral tasks administered in the same session as the market task. "Separate session" indicates the measurements were elicited on a different day. *Market settings*: SSW: Smith et al. (1988); PS: Plott and Sunder (1988). *Behavioral measures: Risk*: HL: Holt and Laury (2002); EG: Eckel and Grossman (2002); Dave et al. (2010); SGG: Sabater-Grande and Georgantzis (2002); CE: Certainty Equivalence (Kirchler and Maciejovsky, 2002; Dohmen et al., 2011); BRET: Bomb Risk Elicitation Task (Crosetto and Filippin, 2013); GP: Gneezy and Potters (1997). *Cognitive measures*: CRT: Cognitive Reflection Test; Fin. Lit: Financial Literacy; ToM: Theory of Mind; WMC: Working Memory Capacity. *Overconfidence:* SCI: Subjective Confidence Intervals; CBO: Calibration Based Overconfidence; BTA: Better-Than-Average; IoC: Illusion of Control. *Competitiveness:* NV: Niederle and Vesterlund (2007).

10 Conclusion

We discussed behavioral elicitation tasks that can be adopted as simple add-ons to market and auction experiments. We reviewed measures of risk, time, cognitive ability, financial literacy, regret and emotions, overconfidence and competitiveness, focusing on crude

measurements and ordinal comparisons that could be applied to help explain variations in performances on both the individual and – sometimes – on market levels.

Among the discussed behavioral features, cognitive ability is among the most widely studied, and is shown to correlate robustly with individual performance success and with "rational" market outcomes in a variety of settings. Evidence on the effects of other behavioral traits remains more context-specific and often debated. An emerging theme in the literature is that participants with lower cognitive abilities make more errors, leading to noisier measurements, and possibly biases in estimated correlations of these features and behavior (Bosch-Rosa and Corgnet, 2022; Chapman et al., 2024).

Time preference, especially short-term impatience, impulsivity and individual perception of time, appear understudied in the context of market experiments. It is possible that these features may provide clues to variations in individual behavior in dynamic market settings that require actions in real time.

On the methodology side, we discussed the choice between using incentivized and non-incentivized measurements. While incentivized tasks remain the default within experimental economics, a number of validation studies have shown that short non-incentivized survey items often correlate with their incentivized counterparts, both in and outside the lab (Dohmen et al., 2011; Vischer et al., 2013; Falk et al., 2023). This is consistent with the evidence that incentives often have small or inconclusive effects on measured preferences (Cesarini et al., 2006; Brañas-Garza et al., 2019, 2021; Cohen et al., 2020); see also Tasoff and Zhang (2022). Yet some researchers caution that despite correlations with incentivized tasks, qualitative self-reports may fail to produce reliable measures (Chapman et al., 2025a).

Task presentation formats (e.g., tabular vs visual presentations) may also matter for measurement. Theoretical studies further recommend presenting options one at a time, rather than in lists, and paying for one of the options chosen randomly, to ensure incentive compatibility (e.g., Brown and Healy, 2018). Exploring the role of presentation modes in measuring behavioral features appears an interesting and under-explored issue, especially in view of fast-growing options for alternative task representations becoming available in standard experimental software packages such zTree (Fischbacher, 2007) and oTree (Chen et al., 2016). Further, repeat or multiple measurements are desirable to reduce measurement error (Gillen et al., 2019). While multiple measurements may increase the required task time, such increase may be minimal with availability of effective software options for task implementations.

Although we focus on short instruments, allowing enough time to complete the task may be important to solicit thoughtful and less noisy responses. Survey timing—instituting the survey before or after the main market experiment—may also affect the results. Some studies indicate that performance on cognitive ability tests is higher if the test is instituted before, rather than after, the experiment (Brañas-Garza et al., 2019). However, there is no consensus and a large variability of survey timing across the existing studies; see Table 7.

A related important issue is ensuring data quality. Traditional tools to detect inattentive or careless human responders include attention check[44] and memory test items and scale reversal (Curran, 2016). Conducting experiments online and using online subject pools, along with the advancement of artificial intelligence (AI), creates new challenges to data quality.

[44]For an example of an attention check item, see Question 2 in the Appendix survey A.3.

Westwood (2025) reports that sophisticated synthetic agents easily pass traditional data quality checks designed for humans, can be prompted to behave like humans and make mistakes consistent with their demographics, and, in addition, can correctly identify researchers' presumed hypotheses and mirror the experimenter demand effect in human subjects. Zhang et al. (2026) compare the effectiveness of several AI detection tests, including typing metrics, tasks that require processing of audio/visual information, "reverse shibboleth" (questions that require superhuman knowledge), and reCAPTCHA, among others. Affonso (2026) propose visual cognitive traps tests that are easily passable by humans but fail by AI agents; Fernandez et al. (2026) develop webcam eye tracking for AI detection in online surveys and economic experiments. These studies emphasize that capabilities of AI agents are fast evolving, and their costs per use are low, opening doors for financially motivated fraud. Conducting experiments online and employing participants from online subject pools increasingly requires ongoing participant validation and quality checks (Westwood, 2025).

This survey is intended to document and discuss the existing measures of individual preferences and behavioral features in the context of market behavior. The connection of these features to market performance remains a deep and underexplored question. Experimental and behavioral economists have been studying for decades why some market institutions, such as traditional double auction commodity markets, are robust to individual behavioral variations, while others, such as financial markets, are sensitive to individual features and biases, with their volatility increasing with deviations in individual behavior (Plott, 1986; Smith, 1991; Camerer, 1995; Friedman and Kelley, 2006). We leave this intriguing and still unresolved question for further exploration.

References

Abdellaoui, M., Bleichrodt, H., and Kammoun, H. (2013). Do financial professionals behave according to prospect theory? an experimental study. *Theory and Decision*, 74:411–429.

Abdellaoui, M., Bleichrodt, H., and l'Haridon, O. (2008). A tractable method to measure utility and loss aversion under prospect theory. *Journal of Risk and uncertainty*, 36(3):245–266.

Abdellaoui, M., Bleichrodt, H., and Paraschiv, C. (2007). Loss aversion under prospect theory: A parameter-free measurement. *Management science*, 53(10):1659–1674.

Affonso, F. M. (2026). Brief commentary: A framework for detecting ai agents in online research. *Journal of Consumer Research*, page ucag006.

Agranov, M. and Yariv, L. (2018). Collusion through communication in auctions. *Games and Economic Behavior*, 107:93–108.

Almenberg, J. and Dreber, A. (2015). Gender, stock market participation and financial literacy. *Economics letters*, 137:140–142.

Almenberg, J. and Widmark, O. (2011). Numeracy, financial literacy and participation in asset markets. *Available at SSRN 1756674*.

Almlund, M., Duckworth, A. L., Heckman, J., and Kautz, T. (2011). Personality psychology and economics. In *Handbook of the Economics of Education*, volume 4, pages 1–181. Elsevier.

Andarsari, P. R. and Ningtyas, M. N. (2019). The role of financial literacy on financial behavior. *Journal of accounting and business education*, 4(1):24–33.

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, 9:383–405.

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2008). Eliciting risk and time preferences. *Econometrica*, 76(3):583–618.

Anderson, L. R., Freeborn, B. A., and Hulbert, J. P. (2012). Risk aversion and tacit collusion in a bertrand duopoly experiment. *Review of Industrial Organization*, 40(1):37–50.

Anderson, L. R. and Mellor, J. M. (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of health economics*, 27(5):1260–1274.

Andrade, E. B., Odean, T., and Lin, S. (2016). Bubbling with excitement: An experiment. *Review of Finance*, 20(2):447–466.

Andreoni, J., Che, Y.-K., and Kim, J. (2007). Asymmetric information about rivals' types in standard auctions: An experiment. *Games and Economic Behavior*, 59(2):240–259.

Andreoni, J., Kuhn, M. A., and Sprenger, C. (2015). Measuring time preferences: A comparison of experimental methods. *Journal of Economic Behavior & Organization*, 116:451–464.

Andreoni, J. and Sprenger, C. (2012a). Estimating time preferences from convex budgets. *American Economic Review*, 102(7):3333–56.

Andreoni, J. and Sprenger, C. (2012b). Risk preferences are not time preferences. *American Economic Review*, 102(7):3357–3376.

Angrisani, M., Cipriani, M., Guarino, A., Kendall, R., and Ortiz de Zarate, J. (2020). Risk preferences at the time of covid-19: An experiment with professional traders and students. *FRB of New York Staff Report*, (927).

Aragó, V., Barreda-Tarrazona, I., Breaban, A., Matallín, J. C., and Salvador, E. (2022). Market risk aversion under volatility shifts: An experimental study. *International Review of Economics & Finance*, 80:552–568.

Astor, P. J., Adam, M. T., Jähnig, C., and Seifert, S. (2013). The joy of winning and the frustration of losing: A psychophysiological analysis of emotions in first-price sealed-bid auctions. *Journal of Neuroscience, Psychology, and Economics*, 6(1):14.

Attema, A. E., Bleichrodt, H., Gao, Y., Huang, Z., and Wakker, P. P. (2016). Measuring discounting without measuring utility. *American Economic Review*, 106(6):1476–1494.

Azrieli, Y., Chambers, C. P., and Healy, P. J. (2018). Incentives in experiments: A theoretical analysis. *Journal of Political Economy*, 126(4):1472–1503.

Bartels, D. M., Li, Y., and Bharti, S. (2023). How well do laboratory-derived estimates of time preference predict real-world behaviors? comparisons to four benchmarks. *Journal of Experimental Psychology: General*, 152(9):2651.

Bartling, B. and Netzer, N. (2016). An externality-robust auction: Theory and experimental evidence. *Games and Economic Behavior*, 97:186–204.

Becker, G. M., DeGroot, M. H., and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral science*, 9(3):226–232.

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations research*, 30(5):961–981.

Benhabib, J., Bisin, A., and Schotter, A. (2010). Present-bias, quasi-hyperbolic discounting, and fixed costs. *Games and economic behavior*, 69(2):205–223.

Biais, B., Hilton, D., Mazurier, K., and Pouget, S. (2005). Judgemental overconfidence, self-monitoring, and trading performance in an experimental financial market. *The Review of economic studies*, 72(2):287–312.

Binnendyk, J. and Pennycook, G. (2024). Individual differences in overconfidence: A new measurement approach. *Judgment and Decision Making*, 19:e28.

Binswanger, H. P. (1980). Attitudes toward risk: Experimental measurement in rural india. *American journal of agricultural economics*, 62(3):395–407.

Binswanger, H. P. (1981). Attitudes toward risk: Theoretical implications of an experiment in rural india. *The Economic Journal*, 91(364):867–890.

Bleichrodt, H., Cillo, A., and Diecidue, E. (2010). A quantitative measurement of regret theory. *Management Science*, 56(1):161–175.

Bleichrodt, H. and Wakker, P. P. (2015). Regret theory: A bold alternative to the alternatives. *The Economic Journal*, 125(583):493–532.

Bönte, W., Lombardo, S., and Urbig, D. (2017). Economics meets psychology: Experimental and self-reported measures of individual competitiveness. *Personality and Individual Differences*, 116:179–185.

Bosch-Rosa, C. and Corgnet, B. (2022). Cognitive finance. In *Handbook of Experimental Finance*, pages 73–88. Edward Elgar Publishing.

Bosch-Rosa, C., Meissner, T., and Bosch-Domènech, A. (2018). Cognitive bubbles. *Experimental Economics*, 21:132–153.

Bosman, R. and Riedl, A. (2003). Emotions and economic shocks in a first-price auction. Technical report, Tinbergen Institute discussion paper.

Bouleau, C., Lebreton, M., and Jacquemet, N. (2025). Large-scale experimental investigation of the reliability of confidence measures. *Communications Psychology*, 3(1):159.

Brañas-Garza, P., Estepa-Mohedano, L., Jorrat, D., Orozco, V., and Rascón-Ramírez, E. (2021). To pay or not to pay: Measuring risk preferences in lab and field. *Judgment and Decision Making*, 16(5):1290–1313.

Brañas-Garza, P., Kujal, P., and Lenkei, B. (2019). Cognitive reflection test: Whom, how, when. *Journal of Behavioral and Experimental Economics*, 82:101455.

Brañas-Garza, P. and Smith, J. (2016). Cognitive abilities and economic behavior. *Journal of Behavioral and Experimental Economics*, 64:1–4.

Breaban, A. and Noussair, C. N. (2015). Trader characteristics and fundamental value trajectories in an asset market experiment. *Journal of Behavioral and Experimental Finance*, 8:1–17.

Breaban, A. and Noussair, C. N. (2018). Emotional state and market behavior. *Review of Finance*, 22(1):279–309.

Breaban, A., Noussair, C. N., and Popescu, A. V. (2020). Contests with money and time: Experimental evidence on overbidding in all-pay auctions. *Journal of Economic Behavior Organization*, 171:391–405.

Brown, A. L. and Healy, P. J. (2018). Separated decisions. *European Economic Review*, 101:20–34.

Bruguier, A. J., Quartz, S. R., and Bossaerts, P. (2010). Exploring the nature of "trader intuition". *The Journal of Finance*, 65(5):1703–1723.

Brunner, C., Hu, A., and Oechssler, J. (2014). Premium auctions and risk preferences: an experimental study. *Games and Economic Behavior*, 87:467–484.

Bucher-Koenen, T., Alessie, R. J., Lusardi, A., and Van Rooij, M. (2021). Fearless woman: Financial literacy and stock market participation. Technical report, National Bureau of Economic Research.

Burke, M. A. and Manz, M. (2014). Economic literacy and inflation expectations: Evidence from a laboratory experiment. *Journal of Money, Credit and Banking*, 46(7):1421–1456.

Burks, S., Carpenter, J., Götte, L., and Rustichini, A. (2012). Which measures of time preference best predict outcomes: Evidence from a large-scale field experiment. *Journal of Economic Behavior & Organization*, 84(1):308–320.

Burks, S. V., Carpenter, J. P., Goette, L., and Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences*, 106(19):7745–7750.

Burks, S. V., Lewis, C., Kivi, P. A., Wiener, A., Anderson, J. E., Götte, L., DeYoung, C. G., and Rustichini, A. (2015). Cognitive skills, personality, and economic preferences in collegiate success. *Journal of Economic Behavior & Organization*, 115:30–44.

Buser, T., Niederle, M., and Oosterbeek, H. (2024). Can competitiveness predict education and labor market outcomes? evidence from incentivized choice and survey measures. *Review of Economics and Statistics*, pages 1–45.

Buser, T., Niederle, M., and Oosterbeek, H. (2026). Can competitiveness predict education and labor market outcomes? evidence from incentivized choice and survey measures. *Review of Economics and Statistics*, 108(3):755–773.

Buser, T. and Oosterbeek, H. (2023). The anatomy of competitiveness. IZA Discussion Papers 16224, IZA Network @ LISER.

Bönte, W., Procher, V., and Urbig, D. (2018). Gender differences in selection into self-competition. *Applied Economics Letters*, 25(8):539–543.

Cacho, B. (2024). Impacts of behavioral preferences in experimental auctions. *Horizons: The Undergraduate Research Journal of the University of Hawai'i at Mānoa*, 9(1).

Camerer, C. (1995). Individual decision making. *The handbook of experimental economics*, 1:587–704.

Camerer, C., Bose, D., Daviet, R., and Imai, T. (2020). Social preference estimation using adaptive experimental design. *Available at SSRN 4899381*.

Camerer, C. and Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American economic review*, 89(1):306–318.

Camerer, C. F. and Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of risk and uncertainty*, 19(1):7–42.

Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J.-R., and Sirigu, A. (2004). The involvement of the orbitofrontal cortex in the experience of regret. *Science*, 304(5674):1167–1170.

Campos-Vazquez, R. M. and Cuilty, E. (2014). The role of emotions on risk aversion: A prospect theory experiment. *Journal of Behavioral and Experimental Economics*, 50:1–9.

Capra, C. M., Lanier, K. F., and Meer, S. (2010). The effects of induced mood on bidding in random nth-price auctions. *Journal of Economic Behavior & Organization*, 75(2):223–234.

Carpenter, P. A., Just, M. A., and Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the raven progressive matrices test. *Psychological review*, 97(3):404.

Casari, M., Ham, J. C., and Kagel, J. H. (2007). Selection bias, demographic effects, and ability effects in common value auction experiments. *American Economic Review*, 97(4):1278–1304.

Cason, T. N. and Plott, C. R. (2014). Misconceptions and game form recognition: Challenges to theories of revealed preference and framing. *Journal of Political Economy*, 122(6):1235–1270.

Cassar, A. and Rigdon, M. L. (2021). Prosocial option increases women's entry into competition. *Proceedings of the National Academy of Sciences*, 118(45):e2111943118.

Cesarini, D., Sandewall, Ö., and Johannesson, M. (2006). Confidence interval estimation tasks and the economics of overconfidence. *Journal of Economic Behavior & Organization*, 61(3):453–470.

Chabris, C. F., Laibson, D., Morris, C. L., Schuldt, J. P., and Taubinsky, D. (2008). Individual laboratory-measured discount rates predict field behavior. *Journal of risk and uncertainty*, 37(2):237–269.

Chapman, J., Ortoleva, P., Snowberg, E., Yariv, L., and Camerer, C. (2025a). Reassessing qualitative self-assessments and experimental validation. Technical Report No. w33520, National Bureau of Economic Research.

Chapman, J., Snowberg, E., Wang, S. W., and Camerer, C. (2025b). Looming large or seeming small? attitudes towards losses in a representative sample. *Review of Economic Studies*, 92(5):2893–2922.

Chapman, J., Snowberg, E., Wang, S. W., and Camerer, C. F. (2024). Dynamically optimized sequential experimentation (DOSE) for estimating economic preference parameters. NBER Working Paper 33013, National Bureau of Economic Research.

Charness, G., Garcia, T., Offerman, T., and Villeval, M. C. (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.

Charness, G., Gneezy, U., and Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of economic behavior & organization*, 87:43–51.

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Chen, H. and Volpe, R. P. (1998). An analysis of personal financial literacy among college students. *Financial services review*, 7(2):107–128.

Cheung, S. L. (2016). Recent developments in the experimental elicitation of time preference. *Journal of Behavioral and Experimental Finance*, 11:1–8.

Cobb-Clark, D. A., Silva-Goncalves, J., Tymula, A., and Wang, X. (2026). The conceptual foundations of self-control and its link to impulsivity and attention. Available at SSRN: https://ssrn.com/abstract=6301058 or http://dx.doi.org/10.2139/ssrn.6301058.

Cohen, J., Ericson, K. M., Laibson, D., and White, J. M. (2020). Measuring time preferences. *Journal of Economic Literature*, 58(2):299–347.

Cohen, M., Jaffray, J.-Y., and Said, T. (1987). Experimental comparison of individual behavior under risk and under uncertainty for gains and for losses. *Organizational behavior and human decision processes*, 39(1):1–22.

Cohn, A., Engelmann, J., Fehr, E., and Maréchal, M. A. (2015). Evidence for countercyclical risk aversion: An experiment with financial professionals. *American Economic Review*, 105(2):860–885.

Coller, M. and Williams, M. B. (1999). Eliciting individual discount rates. *Experimental economics*, 2:107–127.

Condon, D. M. and Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, 43:52–64.

Cooper, D. J. and Kagel, J. H. (2016). Other-regarding preferences. *The handbook of experimental economics*, 2:217.

Corgnet, B., Deck, C., DeSantis, M., and Porter, D. (2022). Forecasting skills in experimental markets: illusion or reality? *Management Science*, 68(7):5216–5232.

Corgnet, B., Desantis, M., and Porter, D. (2018). What makes a good trader? on the role of intuition and reflection on trader performance. *The Journal of Finance*, 73(3):1113–1137.

Corgnet, B., DeSantis, M., and Porter, D. (2021). Information aggregation and the cognitive make-up of market participants. *European Economic Review*, 133:103667.

Cornsweet, T. N. (1962). The staircase-method in psychophysics. *The American journal of psychology*, 75(3):485–491.

Cox, J. C., Smith, V. L., and Walker, J. M. (1985). Experimental development of sealed-bid auction theory; calibrating controls for risk aversion. *The American Economic Review*, 75(2):160–165.

Cox, J. C., Smith, V. L., and Walker, J. M. (1988). Theory and individual behavior of first-price auctions. *Journal of Risk and uncertainty*, 1(1):61–99.

Crosetto, P. and Filippin, A. (2013). The "bomb" risk elicitation task. *Journal of risk and uncertainty*, 47(1):31–65.

Crosetto, P. and Filippin, A. (2016). A theoretical and experimental appraisal of four risk elicitation methods. *Experimental Economics*, 19:613–641.

Cueva, C., Iturbe-Ormaetxe, I., Mata-Pérez, E., Ponti, G., Sartarelli, M., Yu, H., and Zhukova, V. (2016). Cognitive (ir)reflection: New experimental evidence. *Journal of Behavioral and Experimental Economics*, 64:81–93. Cognitive Abilities and Economic Behavior.

Cueva, C. and Rustichini, A. (2015). Is financial instability male-driven? gender and cognitive skills in experimental asset markets. *Journal of Economic Behavior & Organization*, 119:330–344.

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, 66:4–19.

Danz, D., Vesterlund, L., and Wilson, A. J. (2022). Belief elicitation and behavioral incentive compatibility. *American Economic Review*, 112(9):2851–2883.

Darriet, E., Guille, M., and Vergnaud, J.-C. (2021). Financial literacy and numeracy. HAL archives ouvertes. ffhalshs-03461252.

Darriet, E., Guille, M., Vergnaud, J.-C., and Shimizu, M. (2020). Money illusion, financial literacy and numeracy: Experimental evidence. *Journal of Economic Psychology*, 76:102211.

Datta Gupta, N., Poulsen, A., and Villeval, M. C. (2013). Gender matching and competitiveness: Experimental evidence. *Economic Inquiry*, 51(1):816–835.

Dave, C., Eckel, C. C., Johnson, C. A., and Rojas, C. (2010). Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty*, 41(3):219–243.

Deaves, R., Lüders, E., and Luo, G. Y. (2009). An experimental test of the impact of overconfidence and gender on trading activity. *Review of finance*, 13(3):555–575.

Deck, C. and Jahedi, S. (2015a). The effect of cognitive load on economic decision making: A survey and new experiments. *European Economic Review*, 78:97–119.

Deck, C. and Jahedi, S. (2015b). An experimental investigation of time discounting in strategic settings. *Journal of Behavioral and Experimental Economics*, 54:95–104.

Deck, C. and Jahedi, S. (2015c). Time discounting in strategic contests. *Journal of Economics & Management Strategy*, 24(1):151–164.

Di Girolamo, A., Harrison, G. W., Lau, M. I., and Swarthout, J. T. (2015). Subjective belief distributions and the characterization of economic literacy. *Journal of Behavioral and Experimental Economics*, 59:1–12.

Dickhaut, J. and Rustichini, A. (2008). Neuroeconomics. In *The New Palgrave Dictionary of Economics*, pages 1–9. Springer.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, 100(3):1238–1260.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the european economic association*, 9(3):522–550.

Eber, N., Roger, P., and Roger, T. (2024). Finance and intelligence: An overview of the literature. *Journal of Economic Surveys*, 38(2):503–554.

Eckel, C. C. and Füllbrunn, S. C. (2015). Thar she blows? gender, competition, and bubbles in experimental asset markets. *American Economic Review*, 105(2):906–920.

Eckel, C. C. and Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and human behavior*, 23(4):281–295.

Eckel, C. C. and Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, 68(1):1–17.

Epper, T., Fehr, E., Fehr-Duda, H., Kreiner, C. T., Lassen, D. D., Leth-Petersen, S., and Rasmussen, G. N. (2020). Time discounting and wealth inequality. *American Economic Review*, 110(4):1177–1205.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, 133(4):1645–1692.

Falk, A., Becker, A., Dohmen, T., Huffman, D., and Sunde, U. (2023). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science*, 69(4):1935–1950.

Fallucchi, F., Nosenzo, D., and Reuben, E. (2020). Measuring preferences for competition with experimentally-validated survey questions. *Journal of Economic Behavior & Organization*, 178:402–423.

Fe, E., Gill, D., and Prowse, V. (2022). Cognitive skills, strategic sophistication, and life outcomes. *Journal of Political Economy*, 130(10):2643–2704.

Fehr, E. and Goette, L. (2007). Do workers work more if wages are high? evidence from a randomized field experiment. *American Economic Review*, 97(1):298–317.

Fehr, E. and Leibbrandt, A. (2011). A field study on cooperativeness and impatience in the tragedy of the commons. *Journal of public economics*, 95(9-10):1144–1155.

Fellner, G. (2009). Illusion of control as a source of poor diversification: Experimental evidence. *The Journal of Behavioral Finance*, 10(1):55–67.

Fellner, G. and Maciejovsky, B. (2007). Risk attitude and market behavior: Evidence from experimental asset markets. *Journal of Economic Psychology*, 28(3):338–350.

Fernandes, D., Lynch Jr, J. G., and Netemeyer, R. G. (2014). Financial literacy, financial education, and downstream financial behaviors. *Management science*, 60(8):1861–1883.

Fernandez, K., Yang, X., Smith, S. M., and Krajbich, I. (2026). Defending against ai agents with webcam eye tracking. https://osf.io/preprints/psyarxiv/ecrdw_v1, Last accessed on 2026-10-16.

Filippin, A. and Crosetto, P. (2016). Click 'n'roll: no evidence of illusion of control. *De Economist*, 164(3):281–295.

Filiz-Ozbay, E. and Ozbay, E. Y. (2007). Auctions with anticipated regret: Theory and experiment. *American Economic Review*, 97(4):1407–1418.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178.

Fishburn, P. C. (1982). Nontransitive measurable utility. *Journal of Mathematical Psychology*, 26(1):31–67.

Fox, C., Erner, C., and Walters, D. (2015). Decision under risk: From the field to the lab and back. In Keren, G. and Wu, G., editors, *Blackwell Handbook of Judgment and Decision Making*. Blackwell.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, 19(4):25–42.

Frederick, S., Loewenstein, G., and O'donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of economic literature*, 40(2):351–401.

Freeman, D., Manzini, P., Mariotti, M., and Mittone, L. (2016). Procedures for eliciting time preferences. *Journal of Economic Behavior Organization*, 126:235–242.

Friedman, D., Habib, S., James, D., and Williams, B. (2022). Varieties of risk preference elicitation. *Games and Economic Behavior*, 133:58–76.

Friedman, D. and Kelley, H. M. (2006). *Asset Market Experiments*. John Wiley Sons, Ltd.

Friedman, M. (1996). *Type A behavior: Its diagnosis and treatment*. Springer Science & Business Media.

Füllbrunn, S., Janssen, D.-J., and Weitzel, U. (2019). Risk aversion and overbidding in first price sealed bid auctions: new experimental evidence. *Economic Inquiry*, 57(1):631–647.

Gächter, S., Johnson, E. J., and Herrmann, A. (2007). Individual-level loss aversion in riskless and risky choices.

Gächter, S., Johnson, E. J., and Herrmann, A. (2022). Individual-level loss aversion in riskless and risky choices. *Theory and Decision*, 92(3):599–624.

Giebe, T., Ivanova-Stenzel, R., Kocher, M. G., and Schudy, S. (2024). Cross-game learning and cognitive ability in auctions. *Experimental Economics*, 27(1):80–108.

Gill, D., Knepper, Z., Prowse, V., and Zhou, J. (2025). How cognitive skills affect strategic behavior: Cognitive ability, fluid intelligence and judgment. *Games and Economic Behavior*, 149:82–95.

Gill, D. and Rosokha, Y. (2024). Beliefs, learning, and personality in the indefinitely repeated prisoner's dilemma. *American Economic Journal: Microeconomics*, 16(3):259–283.

Gillen, B., Snowberg, E., and Yariv, L. (2019). Experimenting with measurement error: Techniques with applications to the caltech cohort study. *Journal of Political Economy*, 127(4):1826–1863.

Giné, X., Goldberg, J., Silverman, D., and Yang, D. (2018). Revising commitments: Field evidence on the adjustment of prior choices. *The Economic Journal*, 128(608):159–188.

Glaser, M. and Weber, M. (2007). Overconfidence and trading volume. *The Geneva Risk and Insurance Review*, 32:1–36.

Gneezy, U., Kapteyn, A., and Potters, J. (2003a). Evaluation periods and asset prices in a market experiment. *The Journal of Finance*, 58(2):821–837.

Gneezy, U., Niederle, M., and Rustichini, A. (2003b). Performance in competitive environments: Gender differences. *The quarterly journal of economics*, 118(3):1049–1074.

Gneezy, U. and Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, 112(2):631–645.

Gneezy, U., Rustichini, A., and Vostroknutov, A. (2010). Experience and insight in the race game. *Journal of economic behavior & organization*, 75(2):144–155.

Goeree, J. K. and Offerman, T. (2003). Winner's curse without overbidding. *European Economic Review*, 47(4):625–644.

Goette, L., Huffman, D., and Fehr, E. (2004). Loss aversion and labor supply. *Journal of the European Economic Association*, 2(2-3):216–228.

Grieco, D. and Hogarth, R. M. (2009). Overconfidence in absolute and relative performance: The regression hypothesis and bayesian updating. *Journal of Economic Psychology*, 30(5):756–771.

Griffith, D. E. and Rust, R. T. (1997). The price of competitiveness in competitive pricing. *Journal of the Academy of Marketing Science*, 25(2):109–116.

Hafalir, I., Kesten, O., Sherstyuk, K., and Tao, C. (2026). When speed is of essence: perishable goods auctions. *Journal of Economic Theory*, 232:106149.

Haran, U. and Bereby-Meyer, Y. (2024). Competition in psychology and experimental economics. *The Oxford handbook of the psychology of competition*, pages 9–36.

Hardisty, D. J., Thompson, K. F., Krantz, D. H., and Weber, E. U. (2013). How to measure time preferences: An experimental comparison of three methods. *Judgment and Decision making*, 8(3):236–249.

Hargreaves Heap, S. P. and Zizzo, D. J. (2011). Emotions and chat in a financial markets experiment. *Available at SSRN 1783462*.

Harrison, G. W., Hofmeyr, A., Kincaid, H., Monroe, B., Ross, D., Schneider, M., and Swarthout, J. T. (2022). Subjective beliefs and economic preferences during the covid-19 pandemic. *Experimental economics*, 25(3):795–823.

Harrison, G. W., Igel Lau, M., Rutström, E. E., and Sullivan, M. B. (2005). Eliciting risk and time preferences using field experiments: Some methodological issues. In *Field experiments in economics*, pages 125–218. Emerald Group Publishing Limited.

Harrison, G. W., Lau, M. I., and Rutström, E. E. (2007). Estimating risk attitudes in denmark: A field experiment. *scandinavian Journal of Economics*, 109(2):341–368.

Harrison, G. W., Lau, M. I., and Williams, M. B. (2002). Estimating individual discount rates in denmark: A field experiment. *American economic review*, 92(5):1606–1617.

Harrison, G. W. and Rutström, E. E. (2008). Risk aversion in the laboratory. In *Risk aversion in experiments*. Emerald Group Publishing Limited.

Harrison, G. W. and Rutström, E. E. (2009). Expected utility theory and prospect theory: One wedding and a decent funeral. *Experimental economics*, 12(2):133–158.

Healy, P. J. and Leo, G. (2025). Belief elicitation: a user's guide. In *Handbook of Experimental Methodology*, volume 1, pages 81–162. North-Holland.

Henager, R. and Cude, B. J. (2016). Financial literacy and long-and short-term financial behavior in different age groups. *Journal of Financial Counseling and Planning*, 27(1):3–19.

Hermanns, B. and Kokot, J. (2023). Contextual framing effects on risk aversion assessed using the bomb risk elicitation task. *Economics Letters*, 229:111227.

Hey, J. D. and Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica: Journal of the Econometric Society*, pages 1291–1326.

Hoel, J. B., Jain, P., and Galaty, B. (2022). JUST VENMO ME: Does form of payment affect risk taking and intertemporal choice? *Journal of the Economic Science Association*, 8(1):16–33.

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

Holt, C. A. and Sherman, R. (2014). Risk aversion and the winner's curse. *Southern Economic Journal*, 81(1):7–22.

Holzmeister, F. (2017). otree: Ready-made apps for risk preference elicitation methods. *Journal of Behavioral and Experimental Finance*, 16:33–38.

Holzmeister, F. and Pfurtscheller, A. (2016). otree: The "bomb" risk elicitation task. *Journal of Behavioral and Experimental Finance*, 10:105–108.

Holzmeister, F. and Stefan, M. (2021). The risk elicitation puzzle revisited: Across-methods (in) consistency? *Experimental economics*, 24:593–616.

Horowitz, J. K. (2006). The becker-degroot-marschak mechanism is not necessarily incentive compatible, even for non-random goods. *Economics Letters*, 93(1):6–11.

Huang, Y., Luan, S., Wu, B., Li, Y., Wu, J., Chen, W., and Hertwig, R. (2024). Impulsivity is a stable, measurable, and predictive psychological trait. *Proceedings of the National Academy of Sciences*, 121(24):e2321758121.

Huber, C., Holzmeister, F., Johannesson, M., König-Kersting, C., Dreber, A., Huber, J., and Kirchler, M. (2024). Do experimental asset market results replicate? high-powered preregistered replications of 17 claims. Technical report, Working Papers in Economics and Statistics.

Ifcher, J. and Zarghamee, H. (2011). Happiness and time preference: The effect of positive affect in a random-assignment experiment. *American Economic Review*, 101(7):3109–3129.

Imai, T. and Camerer, C. F. (2018). Estimating time preferences from budget set choices using optimal adaptive design. *Unpublished manuscript*, pages 1788–1814.

Imai, T., Rutter, T. A., and Camerer, C. F. (2021). Meta-analysis of present-bias estimation using convex time budgets. *The Economic Journal*, 131(636):1788–1814.

Janssen, D.-J., Füllbrunn, S., and Weitzel, U. (2019). Individual speculative behavior and overpricing in experimental asset markets. *Experimental Economics*, 22(3):653–675.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, 47(2):263–291.

Kaiser, T. and Lusardi, A. (2024). Financial literacy and financial education: An overview. Working Paper 32355, National Bureau of Economic Research.

Karmeliuk, M., Kocher, M. G., and Schmidt, G. (2022). Teams and individuals in standard auction formats: Decisions and emotions. *Experimental Economics*, 25(5):1327–1348.

Kim, J. (2023). The effects of time preferences on cooperation: Experimental evidence from infinitely repeated games. *American economic journal.*, 15(1).

Kirby, K. N. and Finch, J. C. (2010). The hierarchical structure of self-reported impulsivity. *Personality and individual differences*, 48(6):704–713.

Kirby, K. N., Petry, N. M., and Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental psychology: general*, 128(1):78.

Kirchkamp, O. and Mill, W. (2021). Spite vs. risk: Explaining overbidding in the second-price all-pay auction: A theoretical and experimental investigation. *Games and Economic Behavior*, 130:616–635.

Kirchler, E. and Maciejovsky, B. (2002). Simultaneous over-and underconfidence: Evidence from experimental asset markets. *Journal of Risk and Uncertainty*, 25(1):65–85.

Kiss, H. J., Kóczy, L. , Ágnes Pintér, and Sziklai, B. R. (2022). Does risk sorting explain overpricing in experimental asset markets? *Journal of Behavioral and Experimental Economics*, 99:101882.

Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., and Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational behavior and human decision processes*, 79(3):216–247.

Kocher, M. G., Lucks, K. E., and Schindler, D. (2019). Unleashing animal spirits: Self-control and overpricing in experimental asset markets. *The Review of Financial Studies*, 32(6):2149–2178.

Lahav, Y. and Meer, S. (2022). The effect of induced mood on traders' preferences in asset markets–experimental evidence. *Review of Behavioral Finance*, 14(1):16–34.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2):443–478.

Langer, E. J. (1975). The illusion of control. *Journal of personality and social psychology*, 32(2):311.

Lazear, E. P. and Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of political Economy*, 89(5):841–864.

Lee, J. Y., Nayga Jr, R. M., Deck, C., and Drichoutis, A. C. (2020). Cognitive ability and bidding behavior in second price auctions: An experimental study. *American Journal of Agricultural Economics*, 102(5):1494–1510.

Leibbrandt, A., Gneezy, U., and List, J. A. (2013). Rise and fall of competitiveness in individualistic and collectivistic societies. *Proceedings of the National Academy of Sciences*, 110(23):9305–9308.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., and Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: the balloon analogue risk task (bart). *Journal of Experimental Psychology: Applied*, 8(2):75.

Levine, S. S., Bernard, M., and Nagel, R. (2017). Strategic intelligence: The cognitive capability to anticipate competitor behavior. *Strategic Management Journal*, 38(12):2390–2423.

Li, T. and Noussair, C. N. (2025). Competitor ability, sorting and overconfidence: An experiment. *Journal of Behavioral and Experimental Economics*, 117:102373.

Li, Y., Krefeld-Schwalb, A., Wall, D. G., Johnson, E. J., Toubia, O., and Bartels, D. M. (2022). The more you ask, the less you get: When additional questions hurt external validity. *Journal of Marketing Research*, 59(5):963–982.

Lichtenstein, S. and Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational behavior and human performance*, 20(2):159–183.

Loomes, G. and Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The economic journal*, 92(368):805–824.

Loomes, G. and Sugden, R. (1983). A rationale for preference reversal. *The American Economic Review*, 73(3):428–432.

Lusardi, A. (2019). Financial literacy and the need for financial education: evidence and implications. *Swiss journal of economics and statistics*, 155(1):1–8.

Lusardi, A. and Mitchell, O. S. (2011). Financial literacy around the world: an overview. *Journal of pension economics & finance*, 10(4):497–508.

Lusardi, A. and Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature*, 52(1):5–44.

Mata, R., Frey, R., Richter, D., Schupp, J., and Hertwig, R. (2018). Risk preference: A view from psychology. *Journal of Economic Perspectives*, 32(2):155–172.

Matousek, J., Havranek, T., and Irsova, Z. (2022). Individual discount rates: a meta-analysis of experimental evidence. *Experimental Economics*, 25(1):318–358.

McCannon, B. C., Asaad, C. T., and Wilson, M. (2016). Financial competence, overconfidence, and trusting investments: Results from an experiment. *Journal of Economics and Finance*, 40(3):590–606.

Meier, S. and Sprenger, C. (2010). Present-biased preferences and credit card borrowing. *American Economic Journal: Applied Economics*, 2(1):193–210.

Michailova, J. and Katter, J. K. (2014). Quantifying overconfidence in experimental finance. *International Journal of Behavioural Accounting and Finance*, 4(4):351–367.

Michailova, J., Mačiulis, A., and Tvaronavičienė, M. (2017). Overconfidence, risk aversion and individual financial decisions in experimental asset markets. *Economic research-Ekonomska istraživanja*, 30(1):1119–1131.

Michailova, J. and Schmidt, U. (2016). Overconfidence and bubbles in experimental asset markets. *Journal of Behavioral Finance*, 17(3):280–292.

Miklánek, T. and Zajíček, M. (2020). Personal traits and trading in an experimental asset market. *Journal of Behavioral and Experimental Economics*, 86:101538.

Millner, E. L. and Pratt, M. D. (1991). Risk aversion and rent-seeking: An extension and some experimental evidence. *Public Choice*, 69(1):81–92.

Mischel, W., Shoda, Y., and Rodriguez, M. L. (1989). Delay of gratification in children. *Science*, 244(4907):933–938.

Moinas, S. and Pouget, S. (2013). The bubble game: An experimental study of speculation. *Econometrica*, 81(4):1507–1539.

Mol, J. M., Botzen, W. W., and Blasch, J. E. (2020). Risk reduction in compulsory disaster insurance: Experimental evidence on moral hazard and financial incentives. *Journal of Behavioral and Experimental Economics*, 84:101500.

Moore, D. A. and Healy, P. J. (2008). The trouble with overconfidence. *Psychological review*, 115(2):502.

Mudzingiri, C. (2024). The effect of financial literacy confidence on financial risk preference confidence. a lab experiment approach. *SAGE Open*, 14(2):21582440241253911.

Muehlfeld, K., Weitzel, U., and van Witteloostuijn, A. (2013). Fight or freeze? individual differences in investors' motivational systems and trading in experimental asset markets. *Journal of Economic Psychology*, 34:195–209.

Muñoz-Murillo, M., Álvarez-Franco, P. B., and Restrepo-Tobón, D. A. (2020). The role of cognitive abilities on financial literacy: New experimental evidence. *Journal of Behavioral and Experimental Economics*, 84:101482.

Murnighan, J. K., Roth, A. E., and Schoumaker, F. (1988). Risk aversion in bargaining: An experimental study. *Journal of Risk and Uncertainty*, 1(1):101–124.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American economic review*, 85(5):1313–1326.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The quarterly journal of economics*, 122(3):1067–1101.

Niederle, M. and Vesterlund, L. (2011). Gender and competition. *Annu. Rev. Econ.*, 3(1):601–630.

Noussair, C. N., Tucker, S., and Xu, Y. (2016). Futures markets, cognitive ability, and mispricing in experimental asset markets. *Journal of Economic Behavior & Organization*, 130:166–179.

Odean, T. (1999). Do investors trade too much? *American economic review*, 89(5):1279–1298.

OECD (2022). *OECD/INFE Toolkit for Measuring Financial Literacy and Financial Inclusion 2022*. OECD Publishing, Paris.

Oechssler, J., Roider, A., and Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization*, 72(1):147–152.

Ortoleva, P. and Snowberg, E. (2015). Overconfidence in political behavior. *American Economic Review*, 105(2):504–535.

Otero, I., Salgado, J. F., and Moscoso, S. (2022). Cognitive reflection, cognitive intelligence, and cognitive abilities: A meta-analysis. *Intelligence.*, 90.

Patton, J. H., Stanford, M. S., and Barratt, E. S. (1995). Factor structure of the barratt impulsiveness scale. *Journal of clinical psychology*, 51(6):768–774.

Pikulina, E., Renneboog, L., and Tobler, P. N. (2017). Overconfidence and investment: An experimental approach. *Journal of Corporate Finance*, 43:175–192.

Plott, C. R. (1986). Rational choice in experimental markets. *Journal of Business*, pages S301–S327.

Plott, C. R. and Sunder, S. (1988). Rational expectations and the aggregation of diverse information in laboratory security markets. *Econometrica: Journal of the Econometric Society*, pages 1085–1118.

Posey, L. L., Kwasnica, A. M., and Geier, C. F. (2026). The coexistence of loss aversion and regret aversion in decision making under risk. *Journal of Risk and Uncertainty*, pages 1–37.

Proestakis, A., Espín, A. M., Exadaktylos, F., Cortés Aguilar, A., Oyediran, O. A., and Palacio, L. A. (2013). The separate effects of self-estimated and actual alcohol intoxication on risk taking: A field experiment. *Journal of Neuroscience, Psychology, and Economics*, 6(2):115.

Raven, J. (2008). The raven progressive matrices tests: their theoretical basis and measurement model. *Uses and abuses of Intelligence. Studies advancing Spearman and Raven's quest for non-arbitrary metrics*, pages 17–68.

Raven, J. C., Court, J. H., and Raven, J. E. (1989). *Standard progressive matrices*. Australian Council for Educational Research Limited Camberwell, Australia.

Ray, D., Golovin, D., Krause, A., and Camerer, C. (2019). Bayesian rapid optimal adaptive design (broad): Method and application distinguishing models of risky choice.

Remund, D. L. (2010). Financial literacy explicated: The case for a clearer definition in an increasingly complex economy. *Journal of consumer affairs*, 44(2):276–295.

Reuben, E. (2017). ztree programming. https://www.ereuben.net/teach/, Last accessed on 2024-10-27.

Roger, T., Roger, P., and Willinger, M. (2022). Number sense, trading decisions and mispricing: An experiment. *Journal of Economic Dynamics and Control*, 135:104293.

Rose, J. and Rose, M. (2019). Ready-made otree apps for time preference elicitation methods. *Journal of Behavioral and Experimental Finance*, 23:23–28.

Russo, J. E. and Schoemaker, P. J. (1992). Managing overconfidence. *Sloan Management Review*, 33(2):7–17.

Rustichini, A., DeYoung, C. G., Anderson, J. E., and Burks, S. V. (2016). Toward the integration of personality theory and decision theory in explaining economic behavior: An experimental investigation. *Journal of Behavioral and Experimental Economics*, 64:122–137.

Sabater-Grande, G. and Georgantzis, N. (2002). Accounting for risk aversion in repeated prisoners' dilemma games: An experimental test. *Journal of economic behavior & organization*, 48(1):37–50.

Saccardo, S., Pietrasz, A., and Gneezy, U. (2018). On the size of the gender difference in competitiveness. *Management Science*, 64(4):1541–1554.

Samuelson, P. A. (1937). A Note on Measurement of Utility. *The Review of Economic Studies*, 4(2):155–161.

Schlag, K. H. and van der Weele, J. J. (2015). A method to elicit beliefs as most likely intervals. *Judgment and Decision Making*, 10(5):456–468.

Schotter, A. and Trevino, I. (2014). Belief elicitation in the laboratory. *Annu. Rev. Econ.*, 6(1):103–128.

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., and Lehman, D. R. (2002). Maximizing versus satisficing: happiness is a matter of choice. *Journal of personality and social psychology*, 83(5):1178.

Sirota, M., Dewberry, C., Juanchich, M., Valuš, L., and Marshall, A. C. (2021). Measuring cognitive reflection without maths: Development and validation of the verbal cognitive reflection test. *Journal of Behavioral Decision Making*, 34(3):322–343.

Skagerlund, K., Lind, T., Strömbäck, C., Tinghög, G., and Västfjäll, D. (2018). Financial literacy and the role of numeracy–how individuals' attitude and affinity with numbers influence financial literacy. *Journal of behavioral and experimental economics*, 74:18–25.

Skala, D. (2008). Overconfidence in psychology and finance–an interdisciplinary literature review.

Sloof, R. and von Siemens, F. A. (2017). Illusion of control and the pursuit of authority. *Experimental Economics*, 20(3):556–573.

Smith, K. and Dickhaut, J. (2005). Economics and emotion: Institutions matter. *Games and Economic Behavior*, 52(2):316–335. Special Issue on Neuroeconomics.

Smith, V. L. (1962). An experimental study of competitive market behavior. *Journal of political economy*, 70(2):111–137.

Smith, V. L. (1976). Experimental economics: Induced value theory. *The American Economic Review*, 66(2):274–279.

Smith, V. L. (1991). Rational choice: The contrast between economics and psychology. *Journal of Political Economy*, 99(4):877–897.

Smith, V. L., Suchanek, G. L., and Williams, A. W. (1988). Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica: Journal of the Econometric Society*, pages 1119–1151.

Snowberg, E. and Yariv, L. (2021). Testing the waters: Behavior across participant pools. *American Economic Review*, 111(2):687–719.

Snowberg, E. and Yariv, L. (2025). Chapter 1 - evaluating experimental designs. In Snowberg, E. and Yariv, L., editors, *Handbook of Experimental Methodology*, volume 1 of *Handbook of Experimental Methodology*, pages 1–24. North-Holland.

Sofianos, A. (2025). Intelligence in experimental economics. In Banerjee, A., editor, *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press.

Stagnaro, M. N., Pennycook, G., and Rand, D. G. (2018). Performance on the cognitive reflection test is stable across time. *Judgment and Decision Making*, 13(3):260–267.

Stella, G. P., Filotto, U., Cervellati, E. M., et al. (2020). A proposal for a new financial literacy questionnaire. *International Journal of Business and Management*, 15(2):34–48.

Sutter, M., Huber, J., Kirchler, M., Stefan, M., and Walzl, M. (2020). Where to look for the morals in markets? *Experimental Economics*, 23(1):30–52.

Sutter, M., Kocher, M. G., Glätzle-Rützler, D., and Trautmann, S. T. (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *American Economic Review*, 103(1):510–31.

Tai, C.-C., Chen, S.-H., and Yang, L.-X. (2018). Cognitive ability and earnings performance: Evidence from double auction market experiments. *Journal of Economic Dynamics and Control*, 91:409–440.

Tanaka, T., Camerer, C. F., and Nguyen, Q. (2010). Risk and time preferences: Linking experimental and household survey data from vietnam. *American economic review*, 100(1):557–571.

Tasoff, J. and Zhang, W. (2022). The performance of time-preference and risk-preference measures in surveys. *Management Science*, 68(2):1149–1173.

Thomson, K. S. and Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision making*, 11(1):99–113.

Toplak, M. E., West, R. F., and Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking & Reasoning*, 20(2):147–168.

Toubia, O., Johnson, E., Evgeniou, T., and Delquié, P. (2013). Dynamic experiments for estimating preferences: An adaptive method of eliciting time and risk parameters. *Management Science*, 59(3):613–640.

Urbig, D., Bönte, W., Procher, V. D., and Lombardo, S. (2020). Entrepreneurs embrace competition: evidence from a lab-in-the-field study. *Small Business Economics*, 55(1):193–214.

Urbig, D., Bönte, W., Schmutzler, J., Curcio, A. F. Z., and Andonova, V. (2021). Diverging associations of dimensions of competitiveness with gender and personality. *Personality and Individual Differences*, 176:110775.

Vischer, T., Dohmen, T., Falk, A., Huffman, D., Schupp, J., Sunde, U., and Wagner, G. G. (2013). Validating an ultra-short survey measure of patience. *Economics Letters*, 120(2):142–145.

Voslinsky, A. and Azar, O. H. (2021). Incentives in experimental economics. *Journal of Behavioral and Experimental Economics*, 93:101706.

Wakker, P. and Deneffe, D. (1996). Eliciting von neumann-morgenstern utilities when probabilities are distorted or unknown. *Management science*, 42(8):1131–1150.

Wakker, P. P. (2010). *Prospect theory: For risk and ambiguity*. Cambridge university press.

Westwood, S. J. (2025). The potential existential threat of large language models to online survey research. *Proceedings of the National Academy of Sciences*, 122(47):e2518075122.

Wölbert, E. and Riedl, A. (2013). Measuring time and risk preferences: Reliability, stability, domain specificity.

Yoon, H. and Chapman, G. B. (2016). A closer look at the yardstick: A new discount rate measure with precision and range. *Journal of Behavioral Decision Making*, 29(5):470–480.

Zambrano-Curcio, A. F., Urbig, D., Bönte, W., and Schmutzler, J. (2023). At the risk of loss: Experimental evidence on the discriminant validity of self-reported measures of risk preferences. *Available at SSRN 4371973*.

Zell, E., Strickhouser, J. E., Sedikides, C., and Alicke, M. D. (2020). The better-than-average effect in comparative self-evaluation: A comprehensive review and meta-analysis. *Psychological bulletin*, 146(2):118.

Zhang, G., Walatka, R., Chen, S. Y., Moore, A. K., Urminsky, O., Fernandez, K., Low, A., Bogard, J. E., and Fox, C. R. (2026). Estimating the threat of ai-agent responding across online survey platforms. *Proceedings of the National Academy of Sciences*, 123(42):e2606460123.

Online Appendix

A Pre-and post-auction surveys

This appendix describes the (non-incentivized) surveys conducted in conjunction with the auction experiment described in Hafalir et al. (2026). The survey analyses performed to assess the correlations between individual behavioral features and the auction behavior is given in Cacho (2024). The survey details are as follows.

Experimental auctions were preceded by a three-part behavioral survey. Part 1 assesses participants' cognitive abilities using ICAR tests as given in Condon and Revelle (2014).[45] Parts 2 and 3 contain staircase risk and time preference tasks adopted from Falk et al. (2023).

Each auction part is followed by a short survey assessing participant affective states and their feedback on the auction institution.

Each experimental session ends with a short questionnaire containing qualitative questions on individual competitiveness Fallucchi et al. (2020), risk and time preferences Falk et al. (2023).

A.1 Pre-auction survey

Survey Section 1 (ICAR: Condon and Revelle, 2014) This section will contain five questions and you are given up to one minute to answer each question, taking approximately five minutes for this section. A 3 x 3 matrix of patterns will be presented to you and you will be asked to indicate the best answer to complete the figure.

(The five matrix reasoning questions given to participants are MR.43 – MR.47, and can be obtained through `https://icar-project.com/`. Each question defaults to "No Answer" if no response is given within the time frame of one minute.)

Survey Section 2 (Risk Preferences Staircase: Falk et al., 2023) In this section you will be asked to make a hypothetical choice between a 50/50 Chance of getting 300 dollars or nothing, and a sure payment. Please choose whether you would prefer the 50/50 chance, or the sure payment for each of the following questions.

1. What would you prefer: a draw with a 50 percent chance of receiving 300 Dollars, and the same 50 percent chance of receiving nothing, or the amount of 160 Dollars as a sure payment?
   (a) 50/50 Chance – Go to 17
   (b) Sure Payment – Go to 2
2. Would you prefer the 50/50 chance or the amount of 80 Dollars as a sure payment?
   (a) 50/50 Chance – Go to 10
   (b) Sure Payment – Go to 3

[45] In a pilot experiment, there were additional three CRT questions from Frederick (2005), not included here.

3. Would you prefer the 50/50 chance or the amount of 40 Dollars as a sure payment?
   (a) 50/50 Chance – Go to 4
   (b) Sure Payment – Go to 7
4. Would you prefer the 50/50 chance or the amount of 60 Dollars as a sure payment?
   (a) 50/50 Chance – Go to 5
   (b) Sure Payment – Go to 6
5. Would you prefer the 50/50 chance or the amount of 70 Dollars as a sure payment?
   (a) 50/50 Chance – Go to next section
   (b) Sure Payment – Go to next section
6. Would you prefer the 50/50 chance or the amount of 50 Dollars as a sure payment?
   (a) 50/50 Chance – Go to next section
   (b) Sure Payment – Go to next section
7. Would you prefer the 50/50 chance or the amount of 20 Dollars as a sure payment?
   (a) 50/50 Chance – Go to 8
   (b) Sure Payment – Go to 9
8. Would you prefer the 50/50 chance or the amount of 30 Dollars as a sure payment?
   (a) 50/50 Chance – Go to next section
   (b) Sure Payment – Go to next section
9. Would you prefer the 50/50 chance or the amount of 10 Dollars as a sure payment?
   (a) 50/50 Chance – Go to next section
   (b) Sure Payment – Go to next section
10. Would you prefer the 50/50 chance or the amount of 120 Dollars as a sure payment?
    (a) 50/50 Chance – Go to 14
    (b) Sure Payment – Go to 11
11. Would you prefer the 50/50 chance or the amount of 100 Dollars as a sure payment?
    (a) 50/50 Chance – Go to 13
    (b) Sure Payment – Go to 12
12. Would you prefer the 50/50 chance or the amount of 90 Dollars as a sure payment?
    (a) 50/50 Chance – Go to next section

(b) Sure Payment – Go to next section

13. Would you prefer the 50/50 chance or the amount of 110 Dollars as a sure payment?

    (a) 50/50 Chance – Go to next section
    (b) Sure Payment – Go to next section

14. Would you prefer the 50/50 chance or the amount of 140 Dollars as a sure payment?

    (a) 50/50 Chance – Go to 15
    (b) Sure Payment – Go to 16

15. Would you prefer the 50/50 chance or the amount of 150 Dollars as a sure payment?

    (a) 50/50 Chance – Go to next section
    (b) Sure Payment – Go to next section

16. Would you prefer the 50/50 chance or the amount of 130 Dollars as a sure payment?

    (a) 50/50 Chance – Go to next section
    (b) Sure Payment – Go to next section

17. Would you prefer the 50/50 chance or the amount of 240 Dollars as a sure payment?

    (a) 50/50 Chance – Go to 25
    (b) Sure Payment – Go to 18

18. Would you prefer the 50/50 chance or the amount of 200 Dollars as a sure payment?

    (a) 50/50 Chance – Go to 22
    (b) Sure Payment – Go to 19

19. Would you prefer the 50/50 chance or the amount of 180 Dollars as a sure payment?

    (a) 50/50 Chance – Go to 20
    (b) Sure Payment – Go to 21

20. Would you prefer the 50/50 chance or the amount of 190 Dollars as a sure payment?

    (a) 50/50 Chance – Go to next section
    (b) Sure Payment – Go to next section

21. Would you prefer the 50/50 chance or the amount of 170 Dollars as a sure payment?

    (a) 50/50 Chance – Go to next section
    (b) Sure Payment – Go to next section

22. Would you prefer the 50/50 chance or the amount of 220 Dollars as a sure payment?

(a) 50/50 Chance – Go to 23

(b) Sure Payment – Go to 24

23. Would you prefer the 50/50 chance or the amount of 230 Dollars as a sure payment?

(a) 50/50 Chance – Go to next section

(b) Sure Payment – Go to next section

24. Would you prefer the 50/50 chance or the amount of 210 Dollars as a sure payment?

(a) 50/50 Chance – Go to next section

(b) Sure Payment – Go to next section

25. Would you prefer the 50/50 chance or the amount of 280 Dollars as a sure payment?

(a) 50/50 Chance – Go to 29

(b) Sure Payment – Go to 26

26. Would you prefer the 50/50 chance or the amount of 260 Dollars as a sure payment?

(a) 50/50 Chance – Go to 27

(b) Sure Payment – Go to 28

27. Would you prefer the 50/50 chance or the amount of 270 Dollars as a sure payment?

(a) 50/50 Chance – Go to next section

(b) Sure Payment – Go to next section

28. Would you prefer the 50/50 chance or the amount of 250 Dollars as a sure payment?

(a) 50/50 Chance – Go to next section

(b) Sure Payment – Go to next section

29. Would you prefer the 50/50 chance or the amount of 300 Dollars as a sure payment?

(a) 50/50 Chance – Go to 31

(b) Sure Payment – Go to 30

30. Would you prefer the 50/50 chance or the amount of 290 Dollars as a sure payment?

(a) 50/50 Chance – Go to next section

(b) Sure Payment – Go to next section

31. Would you prefer the 50/50 chance or the amount of 310 Dollars as a sure payment?

(a) 50/50 Chance – Go to next section

(b) Sure Payment – Go to next section

Survey Section 3 (Time Preferences Staircase: Falk et al., 2023) In this section you will be asked to make several hypothetical choices between receiving 100 Dollars today, or a larger amount in 12 months. Please indicate which option you would prefer for each of the following questions.

1. Please consider the following: would you rather receive 100 Dollars today or 154 Dollars in 12 months?
   (a) Today – Go to 17
   (b) In 12 Months – Go to 2
2. Would you rather receive 100 Dollars today or 125 Dollars in 12 months?
   (a) Today – Go to 10
   (b) In 12 Months – Go to 3
3. Would you rather receive 100 Dollars today or 112 Dollars in 12 months?
   (a) Today – Go to 7
   (b) In 12 Months – Go to 4
4. Would you rather receive 100 Dollars today or 106 Dollars in 12 months?
   (a) Today – Go to 6
   (b) In 12 Months – Go to 5
5. Would you rather receive 100 Dollars today or 103 Dollars in 12 months?
   (a) Today – End of survey
   (b) In 12 Months – End of survey
6. Would you rather receive 100 Dollars today or 109 Dollars in 12 months?
   (a) Today – End of survey
   (b) In 12 Months – End of survey
7. Would you rather receive 100 Dollars today or 119 Dollars in 12 months?
   (a) Today – Go to 8
   (b) In 12 Months – Go to 9
8. Would you rather receive 100 Dollars today or 122 Dollars in 12 months?
   (a) Today – End of survey
   (b) In 12 Months – End of survey
9. Would you rather receive 100 Dollars today or 116 Dollars in 12 months?

(a) Today – End of survey

(b) In 12 Months – End of survey

10. Would you rather receive 100 Dollars today or 139 Dollars in 12 months?

    (a) Today – Go to 14

    (b) In 12 Months – Go to 11

11. Would you rather receive 100 Dollars today or 132 Dollars in 12 months?

    (a) Today – Go to 13

    (b) In 12 Months – Go to 12

12. Would you rather receive 100 Dollars today or 129 Dollars in 12 months?

    (a) Today – End of survey

    (b) In 12 Months – End of survey

13. Would you rather receive 100 Dollars today or 136 Dollars in 12 months?

    (a) Today – End of survey

    (b) In 12 Months – End of survey

14. Would you rather receive 100 Dollars today or 146 Dollars in 12 months?

    (a) Today – Go to 16

    (b) In 12 Months – Go to 15

15. Would you rather receive 100 Dollars today or 143 Dollars in 12 months?

    (a) Today – End of survey

    (b) In 12 Months – End of survey

16. Would you rather receive 100 Dollars today or 150 Dollars in 12 months?

    (a) Today – End of survey

    (b) In 12 Months – End of survey

17. Would you rather receive 100 Dollars today or 185 Dollars in 12 months?

    (a) Today – Go to 18

    (b) In 12 Months – Go to 25

18. Would you rather receive 100 Dollars today or 202 Dollars in 12 months?

    (a) Today – Go to 22

    (b) In 12 Months – Go to 19

19. Would you rather receive 100 Dollars today or 193 Dollars in 12 months?
    (a) Today – Go to 20
    (b) In 12 Months – Go to 21
20. Would you rather receive 100 Dollars today or 197 Dollars in 12 months?
    (a) Today – End of survey
    (b) In 12 Months – End of survey
21. Would you rather receive 100 Dollars today or 189 Dollars in 12 months?
    (a) Today – End of survey
    (b) In 12 Months – End of survey
22. Would you rather receive 100 Dollars today or 210 Dollars in 12 months?
    (a) Today – Go to 23
    (b) In 12 Months – Go to 24
23. Would you rather receive 100 Dollars today or 215 Dollars in 12 months?
    (a) Today – End of survey
    (b) In 12 Months – End of survey
24. Would you rather receive 100 Dollars today or 206 Dollars in 12 months?
    (a) Today – End of survey
    (b) In 12 Months – End of survey
25. Would you rather receive 100 Dollars today or 169 Dollars in 12 months?
    (a) Today – Go to 29
    (b) In 12 Months – Go to 26
26. Would you rather receive 100 Dollars today or 161 Dollars in 12 months?
    (a) Today – Go to 28
    (b) In 12 Months – Go to 27
27. Would you rather receive 100 Dollars today or 158 Dollars in 12 months?
    (a) Today – End of survey
    (b) In 12 Months – End of survey
28. Would you rather receive 100 Dollars today or 165 Dollars in 12 months?
    (a) Today – End of survey

(b) In 12 Months – End of survey

29. Would you rather receive 100 Dollars today or 177 Dollars in 12 months?

    (a) Today – Go to 31

    (b) In 12 Months – Go to 30

30. Would you rather receive 100 Dollars today or 173 Dollars in 12 months?

    (a) Today – End of survey

    (b) In 12 Months – End of survey

31. Would you rather receive 100 Dollars today or 181 Dollars in 12 months?

    (a) Today – End of survey

    (b) In 12 Months – End of survey

A.2 Feedback survey after each auction format

(This subsection provides details on the feedback questions given after each auction part of the experiment as described in Hafalir et al. (2026). Regret questions (Questions 4 and 5) are based on counterfactual thinking (Camille et al., 2004) and include winner regret (Question 4) and loser regret (Question 5)[46]. Affective states questions (Questions 6 and 7) are motivated by Camille et al. (2004).)

1. For every period in the auction, I

    (a) always bought the object – Hide 5 and 7

    (b) sometimes bought the object, and sometimes not.

    (c) never bought the object – Hide 4 and 6

2. When I bid, all I cared about was buying as fast as possible.

    (a) Completely Disagree

    (b) Disagree

    (c) Somewhat Disagree

    (d) Neither Agree Nor Disagree

    (e) Somewhat Agree

    (f) Agree

    (g) Completely Agree

3. When I bid, all I cared about was paying the lowest price possible.

[46] See also Filiz-Ozbay and Ozbay (2007) on winner and loser regret.

(a) Completely Disagree
(b) Disagree
(c) Somewhat Disagree
(d) Neither Agree Nor Disagree
(e) Somewhat Agree
(f) Agree
(g) Completely Agree

4. When I bought the object, I often thought that I could have bought it at a lower price if I bid differently.

(a) Completely Disagree
(b) Disagree
(c) Somewhat Disagree
(d) Neither Agree Nor Disagree
(e) Somewhat Agree
(f) Agree
(g) Completely Agree

5. When I did not buy the object, I often thought that I could have bought it and made a profit if I bid differently.

(a) Completely Disagree
(b) Disagree
(c) Somewhat Disagree
(d) Neither Agree Nor Disagree
(e) Somewhat Agree
(f) Agree
(g) Completely Agree

6. When I bought an object in the auction, I felt:

(a) Extremely Sad
(b) Sad
(c) Somewhat Sad
(d) Neither Happy Nor Sad
(e) Somewhat Happy
(f) Happy
(g) Extremely Happy

7. When I did not buy an object in the auction, I felt:

   (a) Extremely Sad
   (b) Sad
   (c) Somewhat Sad
   (d) Neither Happy Nor Sad
   (e) Somewhat Happy
   (f) Happy
   (g) Extremely Happy

A.3 Post-auction survey

(This subsection provides details on our post-auction qualitative survey questions, which measures competition preference (Question 1), regret (Questions 3-7), risk preference (Question 8) and time preference (Questions 9, 10).)

Please rate each statement according to the given scales.

1. "Competition brings the best out of me." (Fallucchi et al., 2020)

   (a) Not At All Like Me
   (b) Not Like Me
   (c) Somewhat Not Like Me
   (d) Neutral
   (e) Somewhat Like Me
   (f) Like Me
   (g) Exactly Like Me

2. Please rate this questions with "somewhat agree." (Attention check)

   (a) Strongly Disagree
   (b) Disagree
   (c) Somewhat Disagree
   (d) Neither Agree Nor Disagree
   (e) Somewhat Agree
   (f) Agree
   (g) Strongly Agree

3. "Whenever I make a choice, I try to get information about how the other alternatives turned out." (Schwartz et al., 2002)

   (a) Strongly Disagree

(b) Disagree

(c) Somewhat Disagree

(d) Neither Agree Nor Disagree

(e) Somewhat Agree

(f) Agree

(g) Strongly Agree

4. "If I make a choice and it turns out well, I still feel like something of a failure if I find out that another choice would have turned out better." (Schwartz et al., 2002)

   (a) Strongly Disagree

   (b) Disagree

   (c) Somewhat Disagree

   (d) Neither Agree Nor Disagree

   (e) Somewhat Agree

   (f) Agree

   (g) Strongly Agree

5. "Whenever I make a choice, I'm curious about what would have happened if I had chosen differently." (Schwartz et al., 2002)

   (a) Strongly Disagree

   (b) Disagree

   (c) Somewhat Disagree

   (d) Neither Agree Nor Disagree

   (e) Somewhat Agree

   (f) Agree

   (g) Strongly Agree

6. "When I think about how I'm doing in life, I often assess opportunities I have passed up." (Schwartz et al., 2002)

   (a) Strongly Disagree

   (b) Disagree

   (c) Somewhat Disagree

   (d) Neither Agree Nor Disagree

   (e) Somewhat Agree

   (f) Agree

   (g) Strongly Agree

7. "Once I make a decision, I don't look back." (Schwartz et al., 2002)
   (a) Strongly Disagree
   (b) Disagree
   (c) Somewhat Disagree
   (d) Neither Agree Nor Disagree
   (e) Somewhat Agree
   (f) Agree
   (g) Strongly Agree
8. How willing or unwilling you are to take risks? (Falk et al., 2023)
   (a) Completely Unwilling To Do So
   (b) Unwilling
   (c) Somewhat Unwilling
   (d) Neither Unwilling Nor Willing
   (e) Somewhat Willing
   (f) Willing
   (g) Completely willing To Do So
9. Are you generally an impatient person, or someone who always shows great patience? (Vischer et al., 2013)
   (a) Very Impatient
   (b) Impatient
   (c) Somewhat Impatient
   (d) Neither Impatient Nor Patient
   (e) Somewhat Patient
   (f) Patient
   (g) Very Patient
10. How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future? (Falk et al., 2023)
   (a) Completely Unwilling To Do So
   (b) Unwilling
   (c) Somewhat Unwilling
   (d) Neither Unwilling Nor Willing
   (e) Somewhat Willing
   (f) Willing
   (g) Completely Willing To Do So