Preference Elicitation: Common Methods and Potential Pitfalls*

Jonathan Chapman
University of Bologna
jonathan.chapman@unibo.it
jnchapman.com

Geoffrey Fisher
Cornell University
gwf25@cornell.edu
geoffreywfisher.github.io

May 28, 2025

Forthcoming: Chapter 2 in the Handbook of Experimental Methodology, Volume 1 (eds. Erik Snowberg and Leeat Yariv). Approx word count: 25,000 words

Abstract

This chapter reviews methods for eliciting risk, time, and social preferences. We first outline the major conceptual issues in preference elicitation, including how preferences are defined and formed, and the significance and potential challenges of eliciting them. Next, we present a "toolkit" of preference elicitation techniques, highlighting the strengths and limitations of prominent methods and providing an overview of important econometric techniques. The discussion covers both classic elicitation methods that have formed the foundation for insights into preferences, including binary choices and matching techniques, and more recent methods, such as multiple price lists and convex time budgets. We then consider two newer tools that offer promise for future research—dynamic optimal adaptive methods, such as DOSE, and experimentally-validated preference modules. We provide a step-by-step guide to best practices when choosing an elicitation method and implementing preference elicitations in various research contexts. This includes strategies for managing common concerns, including how best to deal with measurement error, whether incentivized measures are needed, the potential for framing effects, and other issues. The chapter concludes by identifying future directions for methodological research and possible applications of preference measures.

JEL Classifications: C81, C9, D03, D81, D9
Keywords: Risk Preferences, Social Preferences, Time Preferences, Elicitation Methods

*We thank Colin Camerer, Erik Snowberg, Charlie Sprenger, Leeat Yariv, Qijun Wang, and participants of the IDEE Meeting at the University of Bologna for helpful conversations and suggestions.

Contents

1 Introduction

Preferences are the basic building blocks of economic analysis. Under standard microeconomic theory, an individual's preferences underpin their ranking of potential consumption bundles and hence determine their choices. Accurate measurement of preferences thus offers the potential to better predict choices, understand the causes of individual behaviors, and evaluate possible policy interventions. The goal of measuring fundamental preferences dates back to Bentham's hedonic calculus for evaluating pleasure and pain (Bentham, 1789), but empirical efforts to do so became widespread only after the development of experimental techniques in economic laboratories in the 1960s. The growth of online experiments, the spread of lab-in-the-field techniques, and the incorporation of experimental procedures into large-scale surveys have since made preference elicitation central to a broad range of empirical research.

Yet while measuring preferences has become increasingly common, there is little consensus on the best way to elicit them. A huge range of preference elicitation techniques have been proposed and often it is unclear which is most appropriate in particular setting. The choice of method, and how it is implemented, may play a significant role in determining empirical results (or the lack of them). However, the methodological literature on elicitation methods can be quite inaccessible to non-specialist researchers because it is typically structured around specific research topics—e.g., the study of risk preferences—rather than providing general methodological lessons. This chapter, in contrast, draws together lessons from the study of different preferences and, in doing so, aims to provide researchers with some shortcuts through this thicket.

The chapter provides researchers with a guide to best practices in the experimental elicitation of preferences, highlighting both traditional tools and recent methodological advances. We discuss the methods used to elicit risk, time, and social preferences, as these are the foundation for many of the decisions of interest to economists. Preference elicitation is complicated due, in part, to problems of inconsistent choice and strategic behavior. Moreover, researchers face constraints in research budgets and experimental time. We provide guidance on how to navigate these challenges effectively.

We start, in Section 2, with a brief discussion of the major conceptual issues in the study of preferences—how they are defined and formed, the approaches used by economists to elicit them, and the issues associated with doing so. Section 3 discusses the "toolkit" used to elicit preferences, covering the advantages and drawbacks of prominent elicitation methods and providing an overview of important econometric techniques. Section 4 provides a step-by-step guide to eliciting preferences. This section focuses on how to use the preference elicitation toolkit in practice, emphasizing the major design choices that a researcher is likely to face. We discuss, among other topics, where to elicit preferences, whether to incentivize elicitations, and how to reduce measurement error. We conclude, in Section 5, by discussing what we see as fruitful directions for future research in preference elicitation, both in methodological improvements and applications in the field.

2 Conceptual Issues

This section introduces major conceptual issues underpinning the field of preference elicitation. The first subsection introduces standard utility frameworks that economists use to study risk, time, and social preferences. We then discuss research as to the basis of those preferences, drawing on literature from across the social sciences, before covering the main reasons economists have sought to measure preferences. The final two subsections then turn to measurement issues, laying the groundwork for the detailed discussion of elicitation techniques in Section 3.

2.1 Preferences in Economics

Preferences in economics are generally studied through the lens of utility maximization.[1] An individual's preferences can, under straightforward assumptions, be represented by a utility function. The economic tradition takes preferences as given objects, which are stable over time and, in particular, not affected by the set of choices an individual faces or the way the choice is presented.[2] The choices an individual makes in a given setting can thus "reveal" their

[1]This is a key distinction from work in behavioral economics focusing on identifying "mistakes", "heuristics", or "fallacies". For instance, Stango and Zinman (2023) study seventeen "behavioral biases", classifying many as relating to either beliefs or decision-making, rather than preferences. Psychologists, in contrast, tend to focus on personality "traits", which are increasingly being incorporated into economic models (see Becker et al., 2012, for a review).

[2]This is exemplified by how economists initially responded to findings that challenge this view from psychologists, such as preference reversals. For example, Grether and Plott (1979) note the following about preference reversal findings from psychology (e.g., Lichtenstein and Slovic, 1971), "taken at face value the data are simply inconsistent with preference theory."

underlying preference, and so predict their behavior in other, possibly unrelated, contexts.

We discuss the methods used to elicit the three sets of preferences most studied by economists.[3] *Risk* preferences refer to "the tendency to engage in behaviors or activities that involve higher variance in returns" (Mata et al., 2018, p.156), encompassing both classical measures of risk aversion as well as concepts from behavioral models (such as probability weighting and loss aversion). *Time* preferences consider preferences over payoffs at different points in time, incorporating different models of discounting consumption in the future. Finally, we also consider *Social* preferences, in which payoffs (or behavior) of other economic agents are included in an individual's utility function, and encapsulating a wide variety of "motives such as altruism, reciprocity, intrinsic pleasure in helping others, inequity aversion, ethical commitments, and other motives that induce people to help others more than would an own-material-payoff maximizing individual" (Bowles and Polania-Reyes, 2012, p.370).

Risk and time preferences can be represented by the parameters of an intertemporal utility function, such as the following widely-used "power" utility with exponential discounting:

$$U_t(c) = \delta^t c^\gamma \tag{1}$$

where c represents consumption and t represents time (e.g., days from the present). The parameter γ captures the curvature of the utility function, and hence risk preferences, while δ is a discount factor, representing the weighting an individual places on payoffs in the future. This utility framework can easily be modified to incorporate payoffs being evaluated relative to a reference point, as in prospect theory (Kahneman and Tversky, 1979), and hence loss aversion. Alternative discounting functions could allow for quasi-hyperbolic discounting (β–δ preferences).

There is more variation in the utility frameworks used to estimate social preferences, reflecting the heterogeneity of the preferences being studied. A useful framework is provided by Charness and Rabin (2002) (extended by Bruhin et al., 2019a), whose model embeds many

[3]For reviews focused on each specific preference type we refer readers to Charness et al. (2013) for risk preferences, Cohen et al. (2020) for time preferences, and Fehr and Charness (2023) for social preferences.

earlier theories:

$$U_B(\pi_A, \pi_B) = (\rho \cdot r + \sigma \cdot s + \theta \cdot q) \cdot \pi_A + (1 - \rho \cdot r - \sigma \cdot s - \theta \cdot q) \cdot \pi_B \tag{2}$$

In contrast to (1), the utility of an individual B is determined both by their own payoff (π_B) and those of an "other" (π_A). $r = 1$ if $\pi_B \geq \pi_A$, and $r = 0$ otherwise; $s = 1$ if $\pi_B < \pi_A$, and $s = 0$ otherwise; $q = -1$ if A has "misbehaved", and $q = 0$ otherwise. B's payoff is thus a weighted average of their own payoff (π_B) and their opponent's (π_A). The weights capture both distributional preferences—attitudes to both advantageous inequality (ρ) and disadvantageous inequality (σ)—and notions of reciprocity (θ). The model thus encompasses both ideas of "inequity aversion" (Fehr and Schmidt, 1999) and the "social-welfare preferences" introduced by Andreoni and Miller (2002), whereby players prefer higher payoffs of all players (but particularly the one with the lowest payoffs).

The study of social preferences is particularly complex because social behavior is determined by concerns about prevailing social norms as well as an individual's own preferences. In recent years, a growing literature has investigated the importance of "social image" concerns—an individual's desire to appear to act prosocially or according to some norm of fairness (Dana et al., 2007; Andreoni and Bernheim, 2009). Another strand of literature has attempted to disentangle the norms or "fairness ideals" that people use in making decisions (e.g., Cappelen et al., 2025).

2.2 Where Do Preferences Come From?

The concept of "preferences" or "tastes" is used across the social sciences, including in political science, marketing, psychology, and economics. Broadly speaking, preferences capture attitudes toward, or evaluations of, objects, often connected to particular behaviors and potentially underpinning choices. Understanding the factors that shape preferences is thus a central question in the quest to explain human behavior, and one that has drawn attention across several research fields.

Certainly, some of our preferences are carried in memory and can be easily recalled when an individual faces a choice. These include innate preferences we hold, such as favoring the consumption of sugary or fatty foods, and others that are genetically passed down (Cesarini et al., 2009). Furthermore, we can learn preferences through our experiences and recall them at later points. These preferences commonly include a variety of experiences we gain as children (Fehr et al., 2008; Chowdhury et al., 2022; Kosse et al., 2020), but such learning also continues into adulthood and can be altered by extreme events (Hanaoka et al., 2018). Some of these preferences can be general, such as exhibiting a tendency to help others, but other preferences may be simpler and more easily recallable from memory. In fact, Bronnenberg et al. (2012) demonstrates that preferences over consumer brands are a function of our early life history and, once formed, decay slowly. Several studies have identified correlations between measures of cognitive ability and preferences (Dohmen et al., 2010; Benjamin et al., 2013; Chapman et al., 2024), suggesting that the two may develop together.

However, many of the preferences researchers are interested in are not readily available and must be constructed at the time of choice. This means that when making a decision, individuals must integrate pieces of information to form a preference, such that their evaluation does not solely rely on memory (Slovic, 1995; Bettman et al., 1998). This idea is also in line with random utility frameworks that argue preferences have a degree of stochasticity (Luce, 1959; McFadden, 1973). Naturally, various contextual factors are likely to influence this process. For example, under time pressure individuals may shift to simpler strategies that do not consider all available information (Payne et al., 1993). This highlights the importance of considering how a specific technique used to elicit preferences may affect the data a researcher plans to analyze, and hence the estimates obtained.

2.3 Why Elicit Preferences?

There are three main reasons economists elicit preferences. First, preferences are a fundamental object of study and there is interest in investigating their magnitude, existence, and determinants. Second, preferences provide an explanation for behavior, and may be used to test or calibrate theoretical models. Third, heterogeneity in preferences may confound the attempt to isolate the effect of other causal variables of interest.

Most fundamentally, measuring preferences is key to understanding how they are formed, and how they may vary across different populations. The previous subsection has highlighted the range of research on how preferences may form and change. Preference elicitations have also been used to shed light on differences between various groups both within and across countries. A significant literature has investigated gender differences in risk preferences (Croson and Gneezy, 2009; Filippin and Crosetto, 2016), which may contribute to gender pay gaps (Cortés et al., 2023). Another strand of research has compared preferences across countries, particularly through the Global Preferences Survey (Falk et al., 2018), which collected measures of time, risk, and social preferences from representative samples across 76 countries (see Section 3.3 for a detailed discussion). A range of studies have drawn on this dataset to investigate the relationships between preferences and country-level factors such as geography, history, culture, and economic development (Becker et al., 2020; Sunde et al., 2022; Hanushek et al., 2022).

Alternatively, researchers may be interested in preferences as an explanatory factor for a measured outcome—for instance, to test a theory or calibrate macroeconomic models. Accurate measurement of preferences can play a central role in understanding and predicting individual behavior. For example, prior literature has investigated the correlations between individual-level measures of preferences and decision-making in other domains. Risk preferences have been linked with behaviors such as smoking and drinking alcohol, as well as investment choices (Barsky et al., 1997; Dohmen et al., 2011). Estimates of time preferences predict criminal behavior (Åkerlund et al., 2016), investment in human capital (Cadena and Keys, 2015), lower lifetime income (Golsteyn et al., 2014), and many more life outcomes (Bartels et al., 2023). Social preferences have been connected to experiencing vote-buying (Finan and Schechter, 2012), political preferences (Kerschbamer and Müller, 2020), management of common property resources (Rustagi et al., 2010) and occupational choice (Schneider et al., forthcoming). Such empirical research provides tests of theoretical mechanisms and is relevant to assess the predictive validity of the preferences researchers often measure in laboratory settings. Moreover, this work has implications for policy development—interventions may be less justifiable if choices represent preferences than "mistakes".

Finally, preferences may be a confounding factor even when they are not the primary focus of a research question, making preference elicitation valuable by generating important control variables. For example, prior work has proposed a myriad of reasons for the factors that influence the age when older Americans decide to claim Social Security Administration benefits.

Although time preferences have been investigated as one factor to explain the heterogeneity in claims (Gustman and Steinmeier, 2005), other "psychological" factors, such as perceived ownership, have also been proposed (Shu and Payne, 2016). Hence, researchers interested in establishing a link between such psychological factors and Social Security claims must also measure and control for preferences even if they are not of primary interest (Greenberg et al., 2023).

2.4 Preference Elicitations

Preference elicitation uses choice data to estimate the preferences discussed above. Participants are asked to make at least one choice, but usually a series of choices, and their answers are used as a measure of their preferences, either using parametric estimation or non-parametric statistics.[4] Theoretically, the "elicitation method" used to present these choices is unimportant, as the decision environment should not affect how a participant chooses. However, a long line of research has demonstrated that this is not the case in practice: different elicitation methods lead to different quantitative and even qualitative estimates of individual preferences. Moreover, in practice, preference elicitation is challenging because participants may make choices that do not reflect their true preferences, either due to a deliberate attempt to mislead or simply as a result of mistakes.

The challenges associated with preference elicitation have been well-recognized since the earliest attempts to elicit preferences experimentally. In one of the earliest such studies, Thurstone (1931) seeks to elicit an individual's indifference curve. He considers, but quickly dismisses, the idea of using (in later terminology, see Section 3.1.1) a "matching task", in which participants fill in the blank in choices such as "eight hats and eight pairs of shoes or six hats and ____ pairs of shoes." However, he concludes that such an approach would be undermined by unstable choice and participants' attempts to seek numerical consistency. Instead, he introduces an early version of a Multiple Price List (see Section 3.2.1), in which binary choices are presented

[4] Underpinning this approach is an (often implicit) "narrow bracketing" assumption, under which participants do not incorporate their financial situation outside of the experimental setting into their decision-making within the experiment. In other words, participants' background consumption is taken as given—experimental payments are not, for instance, used to hedge financial risks in day-to-day life. Such an assumption is typically justified by the relatively small size of experimental payments compared to income. However, this may not always be the case, particularly in low-income economies (Dean and Sautmann, 2021).

within a list.[5]

Many preference elicitation methods have been developed to study risk preferences, growing out of early experimental investigations testing expected utility theory. Attempts to understand behavior under risk led to experimental designs based around presenting participants with choices involving lotteries, and/or eliciting valuations for those lotteries. One key outcome of this line of research was the development of the Becker-DeGroot-Marschak incentive-compatible mechanism for eliciting participants' true valuations (Becker et al., 1964).

One—somewhat disappointing—finding from this literature is that choice data often exhibit "preference reversals", in which a participant's choice varies according to how a decision is presented—that is, they violate procedural invariance. In a classic study, Lichtenstein and Slovic (1971) offered participants two risky lotteries: a P-bet with a high probability of a moderate reward (e.g., a 95% chance of winning \$40, or otherwise losing \$10) and a \$-bet with a lower chance of a high reward (e.g., a 15% chance of winning \$160 or otherwise losing \$15). Participants frequently preferred the P-bet when asked to choose between the two, but stated that the \$-bet had a higher monetary value—clearly incompatible with the idea that individuals choose objects with higher value.[6] Other work has shown that choices may be affected by question wording, framing, or other context effects—issues we discuss further in the context of specific elicitation methods below.

Preference elicitations also tend to perform poorly according to many of the concepts of validity commonly used to evaluate measures.[7] Often (but not always), preference elicitations are constructed to connect tightly to theoretical constructs, and so are likely to have strong construct validity. However, the existence of preference reversals hints at weak "convergent stability"—whether different measures of a construct capture a common underlying characteristic—as do a range of studies showing at best weak correlations between different measures of preferences. Preference elicitations also tend to have weak "predictive validity", in that they only weakly predict behavior outside of an experimental environment (Charness et

[5]Roth (1993) provides an interesting discussion of the early experimental literature investigating individual choice, including the responses to Thurstone. In particular, Wallis and Friedman (1942) criticized the use of hypothetical choices—an issue that remains contested today (see Section 4). In response, later work developed questions about choices over breakfast menus, incentivized by expecting participants to eat their choices.

[6]Seidl (2002) reviews the extensive literature on preference reversals in psychology and economics.

[7]See Chapter 1 for a discussion of criteria for evaluating a measure in experimental or behavioral economics. Mata et al. (2018) discusses how concepts of validity used in psychology may be applied within economics.

al., 2013; Cohen et al., 2020). Finally, although the temporal stability—measured through test-retest correlations—of various psychological factors can be quite large—around 0.6–0.7 for the big-5 personality traits (Anusic and Schimmack, 2016; Preuss, 2021)—the stability of economic preferences is generally lower. In particular, over-time correlations are typically around 0.2–0.3 for incentivized measures and around 0.5 for self-assessed measures (Meier and Sprenger, 2015; Mata et al., 2018; Chapman et al., 2024), although with considerable heterogeneity in reported stability across studies (see, for instance, Kirby, 2009).

These findings pose a number of challenges for preference elicitation. One response has been to question the assumption of "fixed" preferences, and consider the potential for preferences to vary over contexts or change over an individual's lifespan—reflected in the literature on preference formation discussed above. The second, which forms the basis for this chapter, is to question the methodology, and seek out better elicitation techniques. As detailed below, newer elicitation technologies tend to demonstrate better temporal stability and predictive validity.

2.5 Inconsistent Choice and Measurement Error

Measurement error is a serious concern in preference elicitation, but until recently has received relatively little attention from experimental economists. In standard elicitations, measurement error has been estimated to be 30–50% of the variance of the elicited measure (Gillen et al., 2019; Perez et al., 2021). Yet researchers have not focused on this issue for three main reasons. First, the baseline values of a parameter are somewhat irrelevant for certain research questions—if the preference is an outcome variable, then the more relevant concern is often whether it is affected by an experimental manipulation. Second, much existing work attempts to test for correlations between parameters and other metrics, often using preference estimates as independent variables in a regression. In such cases, identifying the correct ranking of participants' preferences may be more important to researchers than the accuracy of the estimates themselves.[8] Third, a standard argument is that classical measurement error—that is, any error that is independent of the true value—leads to attenuation bias in regression analysis and, hence, biases against finding a statistically significant result, meaning any findings can be viewed as conservative.

[8]Of course, measurement error also leads to noisy estimates of rankings—simulations by Chapman et al. (2024) suggest that inconsistent choice may lead correlations between true and estimated preference parameters to be as low as 0.28.

However, measurement error should not be ignored for several reasons. In many situations, the magnitude of a correlation is a central component of a research question—for instance, when evaluating the predictive validity of experimental preference measures by testing correlations between those measures and field behaviors or when assessing the stability of intertemporal correlations over time. In these cases, failing to account for measurement error may lead to erroneous conclusions. For instance, Wang and Navarro-Martinez (2023) find that using Gillen et al. (2019)'s "Obviously Related Instrumental Variables" (ORIV) technique to adjust for measurement error uncovers strong correlations between experimental measures of social preferences and prosociality in day-to-day behavior. Similarly, Beauchamp et al. (2017) find that adjusting for measurement error leads to larger estimates of correlations between risk attitudes and field measures of investment behavior, entrepreneurship, alcohol consumption, and smoking.

Measurement error can also generate false positive findings, a point discussed in detail in Gillen et al. (2019). In many cases, elicitations are used to control for potential confounding factors in a regression. However, if the elicitation is estimated with error the control will be ineffective, potentially leading to false positives on other explanatory variables.[9]

One particular source of measurement error that has received increasing attention is the complexity of elicitation procedures. Many elicitation techniques involve questions that are likely unfamiliar to many participants (particularly outside of economics undergraduates), may require detailed instructions, and often depend on mathematical calculation. This complexity may lead to participants using heuristics and, potentially, making choices that do not reflect their true preferences (Payne et al., 1993). Such mistakes may lead participants to falsely appear to display behavioral biases such as hyperbolic discounting (Enke et al., 2023), small-stakes risk aversion (Khaw et al., 2021), or systematic biases in assessing probabilities (Enke and Graeber, 2023). Further, difficulty in understanding may vary across participants in ways that are correlated with variables of interest. In particular, inconsistent choice tends to be lower amongst those of higher cognitive ability (Dave et al., 2010; Chapman et al., 2024).

[9]Gillen et al. (2019) present a simple simulation regressing $Y = \alpha D + \beta X + \epsilon$, where D is the main explanatory variable of interest, and X represents a control for risk attitudes. X is assumed to drive both D and Y, but there is no independent relationship between D and Y (i.e., $\alpha = 0$). Yet if X is measured with measurement error, then there is a high rate of false positives (i.e., estimated $\alpha > 0$)—in around 30% of simulations using a small sample and approaching 100% of the simulations when using a large sample.

Economists often argue that measurement error can be reduced through the use of monetary incentives in preference elicitation. Paying participants according to their choices, the argument runs, should induce effort, focus attention, and aid truthful revelation. Consequently, considerable effort has been invested in developing elicitation methods that are "incentive compatible"—i.e., mechanisms under which participants maximize their reward by acting according to their true preference. However, the importance of incentivization and incentive compatibility to accurate preference elicitation remains an open question, as we discuss further in Section 4.

3 The Preference Elicitation Toolkit

We now turn to the tools that are commonly used to elicit economic preferences. We start by discussing the most important classic elicitation techniques—matching methods, binary choices, and experimental games. We then discuss several elicitation techniques that have been developed more recently and that attempt to resolve potential concerns regarding some of these methods—multiple price lists, choice from convex budget sets, and a variety of approaches for eliciting risk preferences.[10] Next, we discuss two frontier elicitation approaches—experimentally-validated preference modules, and adaptive elicitation methods. Finally, we briefly discuss the major econometric approaches used to transform choice data into measures of an individual's preferences. Our discussion does not propose that one method is always preferred to another; rather, each may have advantages depending on the researcher's objectives.

3.1 Classic Elicitation Methods

We start with a discussion of three methods used in early preference elicitation studies, but still often used today. First we discuss two methods—matching tasks and binary choices—that serve as the basis for many more advanced elicitation technologies. We then overview the use of experimental games, which have been used to elicit a wide range of social preferences.

[10] We organize the techniques to give a rough indication of the order in which they became common in the economic literature. However, it is important to note that the categorization of techniques is not clear cut—for instance, as discussed previously, the idea of presenting choices in a list has been used since at least Thurstone (1931).

3.1.1 Matching Tasks

In a matching task, participants are presented with a choice and asked to "fill in the blank" using an open response question. Panel A of Figure 1 displays an example of such a task for the elicitation of risk preferences. Here, participants are asked to state their certainty equivalent for a given lottery. Similarly, to elicit time preferences, participants may be asked the minimum amount $\$x$ they would accept today rather than receiving a fixed amount of $\$y$ on a future date. In theory, this method directly identifies participants' indifference points and hence can be directly translated into preference parameters. Further, responses can be made incentive compatible by implementing the Becker-DeGroot-Marschak (BDM; Becker et al., 1964) valuation mechanism.[11]

Although matching tasks are convenient for researchers who desire a fast elicitation procedure, there are drawbacks to this technique. Mainly, the method typically has higher rates of choice inconsistencies than when participants are presented with a series of choices (Bostic et al., 1990b; Attema and Brouwer, 2013). Given this, many researchers have moved to provide more structure in elicitation procedures, providing participants with a fixed number of choices—often two—rather than open-response questions, which have been viewed as less realistic than choices (Huber et al., 2002).

3.1.2 Binary Choices

An alternative approach involves presenting participants with a series of binary choices, such as those in Panel B of Figure 1 . For example, Bari et al. (2024) study financing for micro-enterprises in Pakistan and elicit risk attitudes by presenting participants with a series of 30 choices between a guaranteed amount of money or a lottery between a "bad" outcome (payoff of zero) or a good outcome (payoff of 1,000 rupees). Similarly, time preferences can be elicited by asking participants to make a series of choices over smaller-sooner versus larger-later outcomes (e.g., \$15 today versus \$20 in 10 days). Different choices within a sequence can then

[11] For instance, a "value" v for a lottery can be drawn from a given distribution. If a participant's stated certainty equivalent is above v, they receive v for sure. Otherwise, the lottery is run, and they are paid accordingly. This mechanism is equivalent to a sealed-bid second-price auction with an unknown bidder, and hence is incentive compatible under Expected Utility. However, it is relatively complicated to explain, and individuals unfamiliar with the procedure often bid sub-optimally (Plott and Zeiler, 2005; Cason and Plott, 2014). See Berry et al. (2020, especially Appendix K) for a discussion of the trade-offs involved in using BDM to elicit Willingness-to-Pay in the field.

Figure 1: Classic Elicitation Methods for Eliciting Risk Preferences

Panel A: Matching Task

Consider a lottery where you can either receive $20 or receive $0, each with probability 50%.

What amount of money, if paid to you for sure, would make you indifferent to playing this lottery? $*13*

Panel B: Binary Choice

Which of the following options do you prefer?

- ● A lottery where you can either receive $20 or receive $0, each with probability 50%;
- ○ Receiving $12 for certain.

Notes: Panel A displays an example of a matching task used to elicit risk preferences, with the figure in red italics displaying an example participant's submission to the prompt. Panel B displays an example of a binary choice, in which participants are asked to choose one of the two options. In both cases, participants are typically asked a sequence of these tasks, involving different lotteries and/or sure amounts.

vary different attribute levels of the options offered—such as the monetary amounts in each questions, dates of payment, or the probabilities involved in a lottery.

One challenge with this method is that participants may make a choice in one question that is inconsistent with their response to another question. To take the time preference example above, a participant may prefer $15 today over $20 in 10 days in one trial and also prefer $20 in 15 days over $15 today in a different question. How should researchers account for this? One technique is to construct an index that counts the frequency of one type of response (for example, a lottery rather than a safe option for risk preferences or larger-later rather than a smaller-sooner option for time preferences) is chosen (Fisher, 2021; Bari et al., 2024). A second option is to estimate a model that captures the degree of "noise" in a participant's responses (Chabris et al., 2008; Sokol-Hessner et al., 2009). We discuss these issues further in Section 3.4.

3.1.3 Experimental Games

The standard approach to eliciting social preferences has been to implement experimental games, either in the laboratory or the field. In particular, a set of workhorse games have emerged to capture different components of "other-regarding" behavior although exactly which preferences these games capture is not always clear-cut. Camerer and Fehr (2004) identify seven major games that are of particular relevance to social scientists; we focus on three to highlight the broad approach: the ultimatum game, the dictator game, and the trust game.[12]

The ultimatum game (Güth et al., 1982) has two players: a proposer (P) and a receiver (R). The proposer is given some endowment, E, and is asked to propose a split of E with the receiver—for instance, sharing it equally between the two players. R then has a choice of whether to accept the proposed split, or reject it—in the latter case, both players receive a payoff of zero. In this case, R has a dominant strategy to accept any proposal in which they receive a strictly positive amount. Hence, if they use backwards induction, P should offer the smallest possible positive amount to R. However, in practice, many positive offers are rejected, and proposers frequently allocate significant shares of the endowment to the receiver. A large literature has investigated the sources of this behavior, and the willingness to reject an offer has been used as a measure of an individual's willingness to engage in "punishment".

The dictator game (Forsythe et al., 1994) seeks to isolate whether the high offers made by proposers in the ultimatum game represent a form of social preference, or whether they are a strategic response to the fact that low offers are rejected. This game again includes two players, and an endowment. However the game now consists solely of the proposer's choice of how to split the endowment—the second player makes no decision at all. In this case, a purely self-interested proposer has a dominant strategy to keep the whole endowment for themselves. However, a robust finding is that many individuals will choose to give away a fraction of their endowment (Engel, 2011) (albeit less than in the ultimatum game, suggesting both social preferences and strategic responses contribute to ultimatum game behavior). The amount sent in the dictator game has been used as a measure of "altruism" (see, for instance, Chapman et al., 2023a). However, the framework in Equation (2) makes it clear that this interpretation is not straightforward. Giving in the standard dictator game implies $\rho > 0$, but does not tell us anything about σ (or θ). As such, without additional information, we cannot distinguish a general preference for raising others' welfare from a dislike of being behind.

[12]The full set of games also includes the Prisoner's Dilemma, Public Goods, Gift Exchange, and Third Party Punishment games. Readers are referred to that paper for a fuller description of each game, including behavioral regularities in each. Chaudhuri (2008) provides an accessible discussion of experimental work on social preferences. Cooper and Kagel (2016) provide a more recent summary of experimental results in this area, and Bartoš and Levely (forthcoming) discuss the application of these methods in developing economies.

The situation becomes even more complicated when attempts are made to measure reciprocity using, for instance, the trust game (Berg et al., 1995). The trust game starts with a similar set-up to the previous two games, with a sender (S) and a receiver (R) and an initial endowment. First, S chooses how much of the endowment to send to R. Any amount they send is typically tripled, so that if S sends X, then R receives 3X. R then chooses how much of the amount they receive to return to S. In this case, as in the ultimatum game, S's incentive to choose $X > 0$ may be influenced by their beliefs over R's behavior.

In a standard interpretation of the trust game, the amount sent by S captures "trust"—their beliefs over R's reciprocity—and the amount sent back by R captures "positive reciprocity" or "trustworthiness" (θ). Yet, referring again to Equation (2), it is straightforward to see that this approach does not disentangle these factors from considerations of inequity aversion or social welfare concerns. Specifically, if σ is high enough, a sender may send money even if they don't expect to anything to be returned. This is particularly true given that any amount sent is tripled, meaning that there is an efficiency maximizing motive for sending money too.

3.2 Recent Elicitation Methods

Although the classic methods discussed in the previous subsection are still widely used today, researchers have built on these original techniques in many ways, both to address concerns raised above, and to facilitate preference elicitation outside of the laboratory. We discuss two of these methods—the "multiple price list" and choices from "convex budget sets"—in some detail, then briefly overview the wide range of techniques used to elicit risk preferences.

3.2.1 Multiple Price Lists

One of the most established elicitation methods is the multiple price list (henceforth "MPL"), which has been used to elicit a wide range of preferences, including risk preferences, time preferences, inequality aversion, ambiguity aversion, and loss aversion.[13] In this method, participants are offered several binary choices, commonly presented simultaneously as a vertical list. Proponents of the method argue that it is relatively simple to explain, has a transparent incentive structure, and makes the ranking of options clear to participants. Consequently, it is

[13] MPLs are also an important technique for belief elicitation—see Chapter 3. Many of the issues discussed here also apply in that context.

argued MPLs can reduce the amount of measurement error in preference estimates, compared to presenting binary choices sequentially.

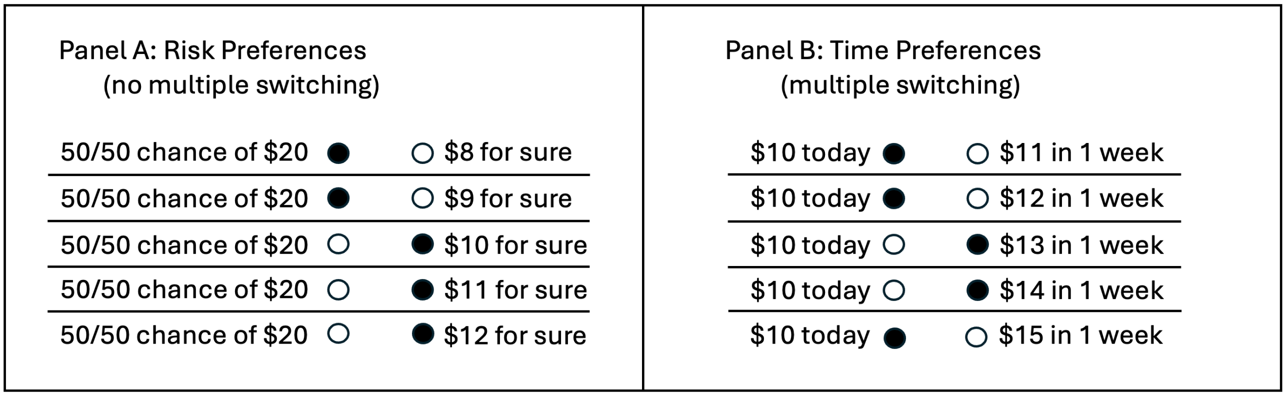
Figure 2 displays examples of MPLs typically used to elicit risk (Panel A) and time (Panel B) preferences. For risk preferences, the MPL offers a series of choices between a lottery (fixed on all rows) over \$0 and \$20 (left-hand side of Panel A), each with 50% probability, and different amounts of money for sure (right-hand side of Panel A). For time preferences, each choice involves a fixed payment of \$10 today (left-hand side of Panel B), versus different amounts of money in a week (right-hand side of Panel B). Many variations on these typical examples have been used. For instance, the MPLs used by Holt and Laury (2002) to elicit risk preferences present participants with two non-degenerate lotteries. The outcomes of these lotteries are fixed in each row, and the probability of receiving the higher prize in each lottery increases monotonically moving down the list. For time preferences, alternative versions could keep the amount fixed, but vary the length of the delay between later and sooner payments.

The typical MPL design orders choices so that the value of the right-hand side increases monotonically as one moves down the list.[14] In Panel A of Figure 2, for example, the sure amount increases in value moving down the list. This design means that once a participant chooses an option on the right-hand side of an MPL they should continue to choose the right-hand side on all rows further down. The point at which participants switch from the left- to the right-hand side then identifies a range of possible certainty equivalents for the lottery. In this example, the participant switches in the third row of the MPL, implying that their certainty equivalent falls in the interval $(\$9, \$10]$. This interval can then be mapped to a range of associated parameter values, if functional forms over utility are assumed. More generally, the row in which a participant switches in an MPL identifies the ordered category into which the quantity of interest falls, and this can be analyzed with a variety of econometric techniques, such as interval regression, that capture the censored data features (Andersen et al., 2006).

While commonly used for risk elicitation, MPLs have also been used to elicit a wide range of other preferences, as well as beliefs. For time preferences, a typical design—-as in Panel B of

[14]See Andersen et al. (2006) for a discussion of design options for the MPL, including the idea of an "iterative MPL". This method first offers participants an MPL with a relatively coarse set of options, and identifies a wide interval of possible values of the participant's preference. A subsequent MPL then offers options only within that interval, more precisely pinning down their preference. For instance, for the lottery in Figure 2, the first question might include sure amounts at \$4 intervals between \$0 and \$20. If a participant chose to switch on the row offering \$12, then the following question would offer sure amounts between \$8 and \$12.

Figure 2: Examples of Multiple Price Lists to Elicit Risk and Time Preferences



Notes: The figure displays examples of short MPLs used to elicit risk preferences (left-hand panel) and time preferences (right-hand panel). In each MPL, participants are asked to choose either the left- or right- hand option on each row, with their choice indicated by a black circle. In Panel A, the participant switches between columns only once (in the third row) indicating that their certainty equivalent for the lottery falls in the interval [\$9, \$10]. In Panel B, in contrast, the participant switches between columns multiple times (in rows three and five)—a pattern which violates standard assumptions over preferences. For examples of actual wording used in a range of MPLs (and other elicitation methods), see the design document available in Chapman et al. (2023a)'s replication package (Chapman et al., 2022).

Figure 2—offers a choice between x (fixed on all rows) at a sooner date, and a series of higher payments at a later date. The switching row in this case identifies a range of possible discount rates for the individual. Dynamic inconsistency can be identified by offering participants two MPLs which both offer payments x days apart, but vary in how close to the present the payments are offered. Chapman et al. (2023a) use MPLs to elicit a wide range of preferences, including time preferences, inequality aversion, ambiguity aversion, and loss aversion. MPLs have also been used to elicit Willingness-to-Pay, such as in Kahneman et al. (1990)'s classic paper on the endowment effect.

Combinations of MPLs can be used to elicit several preference parameters at once. Andersen et al. (2008) use a "double MPL" method, in which they present each participant with MPLs for both risk and time preferences (as in Figure 2). Using the choices in both MPLs, they then estimate both utility curvature (risk aversion) and discount rates using structural maximum likelihood. By doing so, they avoid issues associated with assuming linear utility, which may inflate estimated discount rates if participants are, in fact, risk averse. Similarly, Tanaka et al. (2010) use three MPLs to elicit the parameters of a prospect theory utility function. Von Gaudecker et al. (2011) use a series of MPLs to estimate a structural model of the distribution of risk aversion, loss aversion, and preference for early resolution of uncertainty.

In the domain of social preferences, Kerschbamer and Müller (2020) use multiple MPLs to carry out an Equality Equivalence Test ("EET") that discriminates between nine different "archetypes" of distributional preferences. Archetypes are defined by their preferences over the split of an endowment between themselves and another person. A selfish person, for instance, always prefers to increase their own payoff and is indifferent to the payoff of others, whereas an altruistic (spiteful) person gains utility from an increase (decrease) in the payoff of others. The EET identifies which of these types a participant belongs to using two MPLs—one offering choices over distributions when the decision-maker earns less than their opponent (disadvantageous inequality) and the other offering choices where the decision-maker earns more than their opponent (advantageous inequality). The method thus offers a relatively simple non-parametric way of classifying distributional preferences, and has become an increasingly popular method to elicit social preferences in both the laboratory (e.g., Kittel et al., 2017) and in surveys (e.g., Kerschbamer and Müller, 2020; Chapman et al., 2023a; Holmén et al., 2023).[15]

One common concern with MPLs is that, in practice, participants often switch multiple times between the left-hand and right-hand side of an MPL—a set of choices that is inconsistent under standard revealed preference assumptions. In Panel B of Figure 2, for example, the participant indicates that they prefer \$14 in a week to \$10 today, and also prefer \$10 today to \$15 after a week. This combination of choices violates monotonicity, and means that the MPL does not identify a clear interval for the participant's indifference point. Such behavior is, unfortunately, quite common: a meta-analysis of 63 studies using the Holt-Laury design to estimate risk preferences found that 16% of subjects made inconsistent choices, predominantly involving such "multiple switchovers" (Filippin and Crosetto, 2016, Table 7).

The literature has proposed three major ways of dealing with the problem of multiple switchovers. First, one might opt to drop those with multiple switches on the basis that these are simply mistakes. The second approach avoids the multiple switching points issue altogether by preventing participants from making such choices—i.e., using an experimental design that forces the participant to make consistent choices. For instance, participants can be asked to choose only the row in which they switch, rather than making a choice in each and every row, with the remaining choices then auto-filled by a computer (Morag and Loewenstein, 2023).[16] A

[15]Krawczyk and Le Lec (2021) test an extended version of the EET using adaptive MPL techniques (discussed further below), and find that the assumptions underlying the simpler method are largely justified.

[16]Note that in such designs participants can still alter their switching row after the auto-fill is completed—they

softer version of this approach allows participants to make inconsistent choices, but then asks them whether they wish to reconsider their choice before proceeding. Yu et al. (2021) introduce such a "nudge" and find that it reduces the number of multiple switchovers in an MPL from 31% to 10%.

The third method takes an alternative approach by analyzing all the available data, even the observations containing inconsistent choice. Both methods in the previous paragraph simplify the analysis by excluding inconsistent choices entirely—either during data collection or ex post—but may lose valuable information in the process. Dropping participants reduces sample size and potentially biases the sample by removing those most likely to make such mistakes. Forcing a single switchover may prevent participants from expressing a genuine preference—multiple-switching may represent indifference between several options (Andersen et al., 2006), or in the case of risk, could capture departures from expected utility theory (Chew et al., 2022). As an alternative, inconsistent choices can be included in the analysis either through structural modeling that allows for inconsistency (see Section 3.4), or by utilizing non-parametric measures such as the total number of safe choices made across the MPL.

A further potential disadvantage of MPLs is that the presentation of choices within a list may create framing effects that bias preference estimates. One possibility is that participants are drawn toward values presented at the top or center of a list, with potentially serious consequences. Andersson et al. (2016, 2020) show that altering the position of risk neutrality in an MPL can reverse the estimated correlation between cognitive ability and risk aversion. Andersen et al. (2006) find that a "skewed" MPL affects the average elicitation of risk aversion, but not discount rates. A more extreme example is provided by Von Gaudecker et al. (2011) who use a series of MPLs to estimate loss aversion (and other preferences) in a general population sample. Nearly all the choices they offer participants imply some level of loss aversion, likely biasing estimates of loss aversion upwards (Chapman et al., forthcoming). These results suggest MPLs should, when possible, be designed to be symmetric around a neutral "mid-point".

A harder problem to solve is that MPLs may create reference points. Sprenger (2015) suggests that the structure of MPLs may induce reference point effects, whereby the fixed element of the MPL acts as an endowment. In particular, this work finds that participants appear more risk averse when completing an MPL that elicits probability equivalents for a fixed

are simply precluded from switching multiple times.

sure amount than an MPL that elicits certainty equivalents for a fixed lottery, consistent with the theoretical predictions of Kőszegi and Rabin (2006, 2007). Consistent with this hypothesis, Chapman et al. (2023b) find that MPL-based measures of risk preferences cluster according to whether the fixed item is a lottery or a sure amount of money. These results suggest that MPL elicitations are sensitive to small changes in presentation, and researchers should pay close attention to design before using an MPL.

3.2.2 Convex Budget Sets

While MPLs offer participants a series of (ordered) binary choices, an alternative approach offers participants a convex choice set that allows the estimation of individual parameter values. An early version of this approach is found in the investment game of Gneezy and Potters (1997) where participants were asked how much of an endowment they preferred to bet on a risky lottery (with known objective probabilities). They provided participants with a fixed stock of points, and allowed them to "invest" in a lottery that offered either (1) a return of three times the investment or (2) zero.[17]

Convex budget sets have become particularly popular in estimating discount rates, stimulated by Andreoni and Sprenger (2012a)'s introduction of "Convex Time Budgets" (CTBs)—see Figure 3 for an example. In this approach, participants are given a series of choices regarding how much of a points-based budget to allocate between a sooner and later date. Choices vary in the relative price of allocating points to the future versus today and/or the payment dates. In Figure 3, participants choose how to allocate 100 tokens between a payment to be received today versus in one week. Tokens allocated to the later date have a value of \$0.20, and the value at the earlier date varies on different rows. Participants are often presented with a number of such budgets, with different timings of payment in each choice. CTBs can also be presented discretely—in Andreoni et al. (2015), for example, participants are given a choice of six options along each intertemporal budget constraint—or graphically (Imai and Camerer, 2018; Epper et al., 2020).

A key advantage of CTBs is that they allow both discounting and utility curvature to be elicited in a single experimental instrument. Variation in the interest rate between sooner and

[17]The original investment game has the limitation that participants cannot display risk-seeking behavior because even a risk-neutral individual should invest the full amount in the risky asset. However, a more general design overcomes this issue by providing participants with multiple possible securities available at different prices (Choi et al., 2007).

Figure 3: An Example Convex Time Budget

						today	1 week
Allocate 100 tokens:	*74*	at $0.20 today	*26*	at $0.20 in 1 week		*$14.80*	*$5.20*
Allocate 100 tokens:	*70*	at $0.19 today	*30*	at $0.20 in 1 week		*$13.30*	*$6.00*
Allocate 100 tokens:	*50*	at $0.18 today	*50*	at $0.20 in 1 week		*$9.00*	*$10.00*
Allocate 100 tokens:	*45*	at $0.16 today	*55*	at $0.20 in 1 week		*$7.20*	*$11.00*
Allocate 100 tokens:	*39*	at $0.13 today	*61*	at $0.20 in 1 week		*$5.07*	*$12.20*

Notes: The figure displays an example Convex Time Budget. In each row, participants are asked to allocate 100 tokens between today and one week—figures in red italics indicate choices—at different prices (interest rates). The final two columns indicate the associated payoffs at the two different points in time. Variation in the interest rate across rows allows the curvature of the utility function to be identified separately from discounting. A typical design would present multiple budgets, with variation in the delay between payment dates.

later dates allows identification of curvature, while variation in the length of delay identifies discounting. The choice data are then used to estimate both risk aversion and discounting parameters using non-linear least squares. Consequently, the method overcomes a problem noted above: failing to account for utility curvature may bias estimated discount rates upward. Further, in contrast to the "double MPL" method, CTBs elicit curvature within intertemporal decision making. This may explain the higher out-of-sample predictive power of CTB estimates (Andreoni et al., 2015). Furthermore, although a series of binary choices can also identify utility curvature, it may take more trials to do so than in a CTB.

In an influential study, Augenblick et al. (2015a) use CTBs to understand participants' preferences over the timing of effort, rather than monetary consumption. In their design participants are asked to complete a number of onerous tasks, such as transliterating meaningless Greek text. Rather than determining the timing of payment—which occurs on a fixed date—the CTB is used to determine when the tasks must be completed. The experiment shows that participants demonstrated significant levels of present bias over effort—participants allocated around 9% less work to a date when making the choice on the date itself rather than in advance—but not for money (also elicited using a CTB).

Convex budget sets have also been used to elicit social preferences. Andreoni and Miller (2002) presented participants a series of "modified dictator games", involving splitting an endowment between themselves and another player at various price ratios. This set-up allows them to identify trade-offs between total payoffs and distributional outcomes. This approach was developed further by Fisman et al. (2007, 2017) who presented participants with a series of 50 graphically-represented budget sets splitting an endowment between themselves and another player. The responses were then used to estimate the parameters of a constant elasticity of substitution (CES) utility function for each individual. A further treatment involves three-person budget sets, allowing the analysis to differentiate between concerns over self versus others, concerns over the distribution between others, and efficiency (the combined payoffs of all players). They observe considerable heterogeneity in preferences, with one sizable group of participants prioritizing social welfare, another prioritizing reducing differences, and a third group acting selfishly. More recently, Bruhin et al. (2019b) use convex budgets that incorporate reciprocity concerns by allowing responders to make a "kind" or "unkind" act before the dictator makes an allocation.

There still remain, however, a number of open questions about the effectiveness of convex budget sets in elicitation. One striking fact is that, despite the convexification of the choices, many participants tend to choose corner solutions—i.e., allocating all their endowment to the present or the future. For instance, Andreoni and Sprenger (2012a) report that 37% of participants report corner solutions in every choice, and the remaining participants choose corner solutions at least half the time. This raises the question of how much information is truly gained from convexifying the budget sets, as it is not straightforward to estimate parameter estimates that account for such choices.[18]

Further, many criticisms of MPLs could equally be applied to CTBs. The classic CTB elicitation, such as Figure 3, presents many choices together (see also Andreoni and Sprenger (2012b, Figure 1)), potentially creating framing effects. Little research has been carried out into how participants perceive CTBs, and whether they are indeed "simple". Chakraborty et al. (2017) investigate the choice data in both Andreoni and Sprenger (2012a) and Augenblick et al. (2015b) and identify a large volume of inconsistent choices, both when moving between CTBs and MPLs, and within CTB elicitations. In particular, a number of participants in Andreoni and Sprenger (2012a) (particularly those choosing corner solutions) violate the Weak Axiom of

[18]See Harrison et al. (2013) for a critique of the methods proposed by Andreoni and Sprenger (2012a). However, a meta-analysis of CTBs suggests that the econometric strategy makes little difference to estimated present bias (Imai et al., 2021).

Revealed Preference when moving between the CTB and MPL.[19] Further, many participants (more than half of those choosing interior solutions) violate monotonicity conditions—such as (weakly) reducing earlier consumption as interest rates rise. Thus, there is considerable scope to scrutinize behavior in CTBs, and identify possible solutions, as thoroughly as in MPLs.

3.2.3 Alternative Approaches to Eliciting Risk Preferences

The elicitation of risk preferences has been extensively studied, reflecting their central role in economic theory. A host of alternative risk elicitation methods have been developed, and we do not aim to cover all of them here. However we summarize some of the more popular approaches used in the literature.[20]

One popular approach, pioneered by Binswanger (1980) and popularized by Eckel and Grossman (2002, 2008), asks individuals to make a single choice from a "menu" of possible options. Eckel and Grossman (2008) offer participants five lotteries, each with a 50/50 chance of a high or low prize. One option is a degenerate lottery (i.e., the two possible prizes are identical), and hence provides a safe option. The remaining lotteries then become increasingly risky (in the sense of variance of the prizes), such that a risk neutral individual would choose the riskiest option.[21] This procedure can be useful in quickly producing a coarse classification of individuals but is of limited use in settings where fine-grained preference estimates are required.

A concern with all the approaches above is that they present choices between risky prospects, which rely on some numerical skills to properly evaluate. One method that avoids this issue the Balloon Analogue Risk Task ("BART"; Lejuez et al., 2002), where participants are shown a computer simulation of a balloon growing in size as it is pumped full of air. Participants must decide when the air stops, with the payoff determined by the size of the balloon when they do so. However, as the balloon grows, it has a greater chance of bursting—in which case the participant receives nothing. This approach has been used in a wide range of contexts, including in neuroscience—Rao et al. (2008) introduce a version of the BART modified for use

[19]The Weak Axiom of Revealed Preference is a classic axiom in revealed preference theory. It states that if alternative a is chosen over alternative b, then b should not be chosen over a when both are available alternatives.

[20]See Zhou and Hey (2018) for a thorough list of risk elicitation techniques and a summary of experimental investigations comparing different risk elicitation methods. Trautmann and van de Kuilen (2018) review experimental evidence regarding higher-order risk attitudes (relating to the third or higher derivatives of the utility function).

[21]Dave et al. (2010) use a modified version of this design in which a sixth lottery offers the same expected value as the fifth lottery, but with higher variance—thus separating risk-seeking from risk-neutral participants.

in fMRI—the study of risky behaviors (Harmon et al., 2021), and psychology (Hunt et al., 2005).[22] Compared to other experimental elicitations of risk attitudes, BART demonstrates better predictive validity (Charness et al., 2013; Lejuez et al., 2003; Ju and Wallraven, 2023). However, the simplicity of this design comes at a cost of lower construct validity, as BART does not allow the delineation between risk (where probabilities are objective) and uncertainty (where they are subjective).

These types of intuitive elicitation procedures may be particularly valuable when studying groups, such as children or animals, that may struggle to understand more complicated or longer tasks. In a classic study, Slovic (1966) assessed children's (aged 6–16) risk aversion by placing them in front of ten switches, and telling them that nine were "safe"—if pulled, the child received M&M candies placed into a bowl—while the remaining was a "disaster" switch, leading to all candies being lost and the game ending. The child then chose how many switches to pull, with the number pulled providing a measure of risk-taking tendency. More recently, Andreoni et al. (2020), use the "pencil task" to investigate the existence of a gender gap in risk attitudes amongst children. In this task, participants are shown a set of five pencils in a jar, and told they can keep any pencils they take from the jar. However, if a pencil had a red mark on the bottom, the participant would have to give up all pencils.[23] Mischel and Ebbesen (1970) famously asked children to wait a period of time to receive more of a desirable reward—a marshmallow. A large literature has also studied risky decision-making in nonhumans, typically by utilizing primary rewards (i.e., food items) rather than money as stimuli, and found that many animals display risk aversion and context-dependence in risky choice (for a review, see Platt and Huettel, 2008).

Various additional methods have been proposed to isolate other components of risk preferences, such as loss aversion or probability weights. Estimating these aspects of preferences can be complex as it necessitates estimating multiple parameters simultaneously. For instance, to accurately estimate loss aversion, one must also identify risk aversion for gains. Doing so generally necessitates asking participants many questions, leading inevitably to problems associated

[22]fMRI (an abbreviation for functional Magnetic Resonance Imaging) is an indirect method to measure neural activity that works by detecting changes in blood flow in the brain. See Chapter 8 for a discussion of its use in economics.

[23]Unlike in Slovic (1966), where decision-making was sequential, here children were asked to make a single choice of how many pencils to choose to take from the jar. Andreoni et al. (2020) also present the pencil task to adolescents, but allow them to exchange pencils for money.

with inconsistent choice. In turn, this leads to difficulty estimating individual-level parameters as estimates of noise in the choice process must be addressed. Hence, some studies may only report population preference parameters (e.g., Rieger et al., 2017) rather than providing individual-level estimates. Recent advances in optimal adaptive elicitation procedures, which we discuss in detail in the following subsection, offer one possible solution to this problem.

3.3 Frontier Elicitation Methods

The majority of the elicitation methods detailed above have been standard techniques utilized for years, if not decades, across countless experiments. Yet, several recent developments have led to the introduction of newer elicitation methods that potentially resolve some of the concerns regarding the elicitation methods noted above. First, since many studies are now computerized, as opposed to paper-and-pencil, there is more opportunity to exploit computing power when interacting with participants. In turn, this has led to the development of more sophisticated adaptive procedures. Second, given the increased interest in the external validity of laboratory findings, researchers have become interested in adopting techniques that can be used when measuring preferences in more natural settings (i.e., lab-in-the-field).

3.3.1 Adaptive Elicitation Methods

While classic elicitation technologies use a set of questions that are fixed across participants, recent advances in computing power have led to a move towards more flexible elicitations that are tailored to different respondents. There exists a large literature in statistics and computer science on "optimal experimental design", but this has received relatively little attention in economics. In recent years several adaptive designs have emerged and, although none have yet become widespread, they offer the potential for future improvements in elicitation.

The most common adaptive approach has been the use of fixed "staircase", "titration", or "bisection" designs (Andersen et al., 2010; Falk et al., 2018). Under this approach participants are first offered a binary choice, and subsequent questions then seek to gain more precision.[24] For instance, a participant may first be offered the choice between a risky and safe option. If the safe option is chosen, then the next choice offers a choice between the same risky option, but a

[24] Adaptive methods are not limited to binary choices—for instance see the iterative MPL method discussed above.

smaller safe option. If the risky option is chosen, then the next choice involves a slightly higher safe option. A more complicated variant allows the size of the step to depend on the sequence of previous answers (e.g., Bostic et al., 1990a). This adaptive approach offers two potential advantages over "simple" MPL methods. First, it can obtain a similar level of information in a smaller number of questions. Second, it may avoid the issues associated with presenting choices in a list format discussed above. However, this approach is vulnerable to participant mistakes, as a failure to answer correctly on the first question of the sequence can create significant bias in the eventual preference estimate.

More advanced "optimal adaptive elicitations", such as Dynamically Optimized Sequential Experimentation ("DOSE"; Chapman et al., 2024) improve on these methods by selecting informationally-optimal questions, and directly accounting for individual mistakes when eliciting preferences.[25] DOSE methods estimate preference parameters by dynamically selecting personalized question sequences for each participant. Specifically, the procedure starts with a utility specification, such as Equation (1), a Bayesian prior over the parameters of that utility function, and a (large) set of possible questions. After each question, the priors are updated using Bayes' law, and the next question is selected according to a criterion chosen by the researcher.[26] Further, such procedures can account for participant mistakes by modeling the process by which utility is mapped to choice using, for instance, a logit function.

Dynamic preference elicitation procedures offer particular promise in capturing heterogeneity in preferences, and in disentangling preferences from one another. In particular, personalized question sequences can pin-point each participant's preferences in fewer questions than prior methods (and hence reduce experimental time), and reduce potential measurement error by directly accounting for choice inconsistency. They can also be implemented with binary choices, potentially avoiding concerns relating to framing effects associated with methods such as MPLs. Moreover, in contrast to more established methods, they do not require researchers

[25]DOSE is the optimal adaptive procedure that has received the most attention in economics, while the closely-related Dynamic Experiments for Estimating Preferences ("DEEP"; Toubia et al., 2013) has been more popular in marketing. For simplicity, we focus on one method, but similar points apply to other procedures.

[26]The DOSE procedure implemented in Chapman et al. (forthcoming), for instance, selects questions based on the expected Kullback-Leibler (KL) divergence between the prior and possible posteriors associated with each answer—that is, to maximize the highest expected information gain. Alternative information criteria that have been used in the adaptive elicitation literature include Equivalence Class Edge Cutting (EC^2) criterion Imai and Camerer (2018), and maximizing the expected norm of the Hessian of the posterior distribution at its mode (Toubia et al., 2013). Ryan et al. (2016) provides a useful discussion of the optimal design literature in statistics.

to specify questions in advance—question banks can thus include questions that seem, ex ante, unlikely to be useful—important when carrying out research in previously understudied populations where little information about plausible parameter values is available. In addition, both DOSE and DEEP are explicitly designed to distinguish between different parameters of utility functions—such as disentangling risk from time preferences, or separating loss aversion from risk aversion—leading to, at least in principle, high construct validity. Moreover, dynamic procedures produce accurate preference parameters at individual-level—as opposed to population parameters—a feature that could be valuable in identifying correlations with other variables (reducing measurement error and increasing predictive validity) or that could be used directly to write, for example, personalized contracts (Andreoni et al., 2023).

Chapman et al. (forthcoming) use DOSE to study loss aversion in the U.S. population, highlighting how optimal adaptive elicitation procedures can uncover substantively important findings. Despite its central role in prospect theory (Kahneman and Tversky, 1979), few studies have measured loss aversion outside the laboratory—and those that have done so have run into the problems associated with estimating multi-parameter models discussed above. Using DOSE, the study identifies considerable heterogeneity in loss aversion and finds that around 50% of the U.S. population is "loss tolerant", meaning that they weight losses less than gains. The finding of widespread loss tolerance appears to be explained by the general population sample: only around 30% of an undergraduate student sample are classified as loss-tolerant by DOSE, in line with previous laboratory studies and the predictions of an expert survey (DellaVigna et al., 2019). Further, the paper provides evidence that DOSE-elicited measures of loss aversion exhibit relatively high temporal stability and predictive validity—they are correlated with cognitive ability, self-reported gambling, exposure to financial shocks, and stock market investments. These findings demonstrate how optimal adaptive procedures can identify unexpected heterogeneity in preferences—widespread loss tolerance was unanticipated at survey launch, and was only discovered due to the large question bank that DOSE was able to draw upon. Further, the results demonstrate how reducing noise in preference estimates can uncover patterns of correlations that might otherwise have been missed—DOSE preference measures were more strongly correlated with other individual characteristics than other elicitations, consistent with the adaptive procedure being less susceptible to measurement error and consequent attenuation bias.

Despite these advantages, the take-up of dynamic procedures has been quite slow to date, due to two major concerns. First, they have been seen as difficult to implement in practice due to the need for computational power to work in "real time", and because "off-the-shelf" implementations have not been available. Such issues have begun to wane in recent years, as computational power has increased, and as authors have made software publicly available.[27] Moreover, as Chapman et al. (forthcoming) show, dynamic procedures can be used to pre-program a question tree for survey implementation—avoiding the need for real time question selection. Second, adaptive procedures have been criticized as being subject to possible manipulation, particularly if the standard approach of paying one randomly-selected question is used. In principle, a participant could make choices early in a question sequence in order to receive higher possible payoffs later on. However, such behavior seems unlikely given the difficulty of knowing how to manipulate a question sequence, and laboratory evidence suggests that such strategic behavior is not common. Moreover, it is possible to design payment schemes such that adaptive procedures are incentive compatible—although with the trade-off of making the procedures more complicated to explain.[28] A better understanding of such trade-offs, particularly in field settings, is an open area for future research.

3.3.2 Experimentally-Validated Question Modules

The standard approach in economics is to utilize incentivized preference measures, but in recent years the use of hypothetical measures has become more popular, particularly from the Global Preferences Survey ("GPS") (Falk et al., 2018). The use of qualitative survey questions to measure preferences is not new, but has largely been used by non-economists.[29] However, attitudes

[27]Fidanoski and Johnson (2023) provide Z-tree code that implements the DEEP procedure for risk and time preferences, while Toubia et al. (2013) make available a pre-programmed 12-question module on their institutional website. Drake et al. (2024) provide software to generate menus of choices such as, in their example implementation, to identify the range Willingness-to-Pay for workplace flexibility.

[28]For example, Johnson et al. (2021) suggest the "PRINCE" incentive system whereby the question to be paid is randomly selected from *all* possible questions at the end of the experiment. If that question has already been answered, the answer determines the payment. If not, the participant answers it, and this answer determines the payoff. Ray et al. (2012) tests the importance of strategic manipulation experimentally by explicitly telling participants that they may benefit from manipulating an adaptive procedure. Few participants even tried to act strategically, based either on self-reports or observed patterns of behavior—and those that did try to do so were unable to find an effective method. See Chapman et al. (2024) for further discussion of this issue.

[29]For example, the Domain-Specific Risk-Taking ("DOSPERT") Scale (Weber et al., 2002; Blais and Weber, 2006) asks participants 30–40 questions regarding their likelihood of risk-taking and perceived riskiness across five domains—ethical, financial, health/safety, social, and recreational decisions.

have changed due to a series of papers "experimentally validating" hypothetical measures by demonstrating that they predict choices in established incentivized elicitations.

Falk et al. (2023) introduce an experimentally-validated "preference survey module" in which survey items are explicitly selected based on their ability to predict incentivized choices in the laboratory.[30] In particular, they investigate the correlations between standard experimental measures of preferences—MPLs for risk and time preferences, and experimental games to capture trust, altruism, and (positive and negative) reciprocity—and almost 200 candidate survey items. They then identify the survey items that best predict each of the incentivized measures, identifying two survey items for each of the six preferences they examine. A single measure of each preference is then constructed by weighting the two items according to the OLS coefficients from regressing the incentivized measure on the survey measures (details are reported in Falk et al., 2023, Table 3).

The survey items included in the preference module are summarized in Table 1. Notably, the selected measures tend to combine a hypothetical version of standard lab measures, similar to the incentivized measures, and a qualitative self-assessment item. Performance on the incentivized risk MPL, for example, is best predicted by i) a hypothetical MPL and ii) the question "how willing are you to take risks". Similarly, the amount sent in an incentivized trust game is best predicted by i) the amount sent in a hypothetical trust game and ii) answers to the question "As long as I am not convinced otherwise, I assume that people have only the best intentions." These results suggest that qualitative self-assessments add additional information over hypothetical monetary versions of the risk measures, even in predicting an incentivized question.

A modified version of the Falk et al. (2023) module was implemented in representative samples in 76 countries as part of the Global Preferences Survey ("GPS"; Falk et al., 2018), performed as part of the Gallup World Poll, largely by telephone. The GPS version of the survey module excludes the more complicated hypothetical experimental measures, replacing them with simpler quantitative measures (for risk and time), alternative hypothetical scenarios with quantitative answers (altruism or positive reciprocity), an additional survey question (negative

[30] In an earlier study, Dohmen et al. (2011) find that responses to a survey question about willingness to take risks" is positively associated with the certainty equivalent elicited from an incentivized MPL in a laboratory, and then implement the qualitative question in a large scale survey.

Table 1: Falk et al. (2023)'s Preference Survey Module

Preference	Item Description
Risk Aversion	MPL between a lottery and a safe option. Self-assessment: willingness to take risks in general.
Time Discounting	List of 25 choices between payment today or in 12 months. Self-assessment: willingness to give something up today for future benefit.
Trust	Hypothetical trust game: first mover behavior. Self-assessment: people have only the best intentions.
Altruism	Hypothetical scenario: how much would you donate to charity? Self-assessment: willingness to share with others.
Positive Reciprocity	Hypothetical trust game: second mover behavior. Hypothetical scenario: which of six bottles of wine to give as a thank-you gift?
Negative Reciprocity	Minimum acceptable offer in hypothetical ultimatum game. Self-assessment: willingness to punish unfair behavior.

Notes: The table presents the survey items included in Falk et al. (2023)'s Preference Survey Module, adapted from their Table 3—see that paper for full item descriptions, and the weights placed on each item when constructing preference estimates. All experimental items, such as the MPL used to elicit for risk aversion, were implemented with hypothetical payment.

reciprocity), or leaving as just one measure (trust). Extensive piloting also led to wording changes, such as removing the word "lottery" for "random draw".

The GPS dataset has opened a range of questions regarding the correlates and, consequently, causes of economic preferences. By linking to the Gallup World Poll dataset, it is possible to connect individual preference estimates to a much broader set of individual characteristics and attitudes. Falk et al. (2018) identify associations between preferences and gender, self-reported maths skills, and age, as well as country-level factors such as geographic or cultural variables. Follow up work has investigated gender differences in preferences (Falk and Hermle, 2018), the role of patience in economic development (Sunde et al., 2022), the historic roots of cross-country differences in preferences (Becker et al., 2020), and the effect of national preferences on skill development (Hanushek et al., 2022).

A recent study raises important concerns about the use of experimental validation to select preference measures. Experimental validations, such as those used to select the GPS survey questions, select measures based on correlations with choices in associated incentivized preference elicitations. Chapman et al. (2025) show theoretically that survey questions selected this way may fail to be reliable proxies for incentivized elicitations. They then implement both incentivized elicitations and qualitative self-assessments from Falk et al. (2018, 2023) in a representative U.S. sample. The two types of measures identify very different relationships between preferences and other individual attributes, including cognitive ability, demographics, and self-reported behaviors. In fact, the relationships between qualitative self-assessments and these other attributes appear largely unrelated to the variation in incentivized elicitations. These findings suggest that qualitative self-assessments, and hence the preference indices used in the GPS, may have poor construct validity—they may capture differences in understanding, self-perceptions, or behavior, rather than economic preferences.

The use of qualitative self-assessments to elicit preferences may be particularly problematic because they are difficult for participants to interpret. Economists often consider such questions "simple" on the basis that they do not require numerical calculations. However, they do require participants to interpret the definition of the behavior in the question (e.g., what is meant by "risks" or "punishment"?), consider how that behavior applies to their own life, and then map that behavior onto a numerical scale.[31] Such a process may be cognitively

[31]Krosnick and Presser (2010) provides a detailed discussion of the theoretical issues and empirical evidence

challenging, be implemented heterogeneously across participants, lead to the use of heuristics, and confound relationships with variables that are correlated with the ease of answering such questions (Barrington-Leigh, 2024). Charness and Viceisza (2016) report that participants in a low-income sample had difficulty in understanding the "willingness-to-take-risks", leading to extremely high rates of measured risk tolerance. Studying a representative sample, Chapman et al. (2024, 2025) observe high levels of focal value response (choices of salient numbers such as 0, 5, or 10) to self-assessments, with particularly high rates among participants with lower cognitive ability. Such patterns suggest that people may vary in their interpretation and/or understanding of the numerical scales used in self-assessments, complicating the interpretation of the responses received.

There is, therefore, a need for the continued development of experimentally-validated survey modules. More research is required to see whether the issues identified by Chapman et al. (2025) also apply to the quantitative measures in the GPS. The latter appear more similar to incentivized elicitations, and hence may be more reliable proxies. Qualitative self-assessments may also be useful, but more work is needed to pin down the underlying constructs they capture—the extensive literature on the development of question scales in psychology may offer valuable insights. Future validation exercises should be carried out within the population to be studied. For instance, if the goal is to study the general population, then validation should be carried out using a representative sample.[32] Further, survey questions should be selected based on a specific research question of interest, with careful consideration of potential confounding factors—it may not be feasible to identify a single survey module that is valid for all purposes.

3.4 Econometric Techniques

After data collection, the choice data will need to be transformed into an estimate of an individual's preferences, either parametrically or non-parametrically. The best way to do so will

underlying rating scales. Vieider et al. (2015) report an average correlation between incentivized and self-assessed risk attitudes of around 0.2 across samples from 30 countries, but with considerable variation across countries—possibly reflecting sampling variation, but also consistent with different interpretations of the rating scale.

[32] Bauer et al. (2020); Kosfeld and Sharafi (2024) investigate the generalizability of Falk et al. (2018)'s experimental validation (which was undertaken on a pool of German undergraduate students). The results suggest that the correlation between the GPS' quantitative measures and incentivized elicitations replicates better than the correlations between incentivized elicitations and qualitative self-assessments.

depend on the reasons for constructing preference estimates in the first place and, in particular, the level of precision required. Preference measures are often collected to be used in regressions, as either explanatory or dependent variables, in which case identifying accurate individual estimates may be less important than ranking individuals correctly. In this case, it is often easiest to use non-parametric measures. In other cases, however, linking elicitations more closely to theoretical models may be important, and it may be desirable to elicit parametric estimates. In turn, the estimation method used will determine the quantity and quality of data required, and hence it is important to determine this when designing an experiment. Here we point to the main directions researchers may wish to consider when designing their elicitation.

Non-Parametric Measures The most straightforward approach constructs non-parametric measures from the choice data. For example, a participant's risk aversion could be estimated using their reported certainty equivalent for a lottery or their estimated risk premium. The amount given in a dictator game is often used as a measure of altruism. More ad hoc measures include the number of rows with lotteries chosen in an MPL or the number of lotteries chosen in a series of binary choices. The advantage of such an approach is that it can side-step problems in obtaining parametric estimates due to inconsistent choice. However, this comes at the cost of lower construct validity—such measures do not identify the parameters of a utility function—and may lead to high measurement error.

Structural Maximum Likelihood Estimation Parametric estimates can be produced using maximum likelihood estimation (MLE) techniques.[33] In this approach, researchers start by specifying a utility function they wish to estimate, such as Equation (1), and a "link function" that maps this utility function to choice. A simple link function could allow for some probability that an individual makes a mistake and chooses an option that does not maximize their utility. Alternatively, it is often assumed that the likelihood that individuals make mistakes depends on the *difference* in utility between the two parameters. For instance, a common logit (or softmax) link function for a binary choice between options o_1 and o_2 with associated utilities

[33]For examples of this approach, see Andreoni et al. (2015) for time preferences, and Fisman et al. (2017) for social preferences. DellaVigna (2018) provides a more general overview of structural estimation in behavioral economics.

such that $U(o_1) > U(o_2)$ is as follows[34]:

$$\text{Prob}[o_1] = \frac{1}{1 + e^{-\mu(U(o_1) - U(o_2))}} \tag{3}$$

The probability of choosing the higher utility option, o_1, is increasing in the difference in utility of the two different options. $\mu \in \mathbb{R}^+$ represents the sensitivity of choice to the difference in utilities between the two options—higher μ means that an individual is more likely to make a utility-maximizing choice (all else being equal). Having chosen a utility specification and a link function, it is then straightforward to write down a likelihood function and hence estimate preference parameters via maximum likelihood using statistical software—see Harrison and Rutström (2008) and Harrison (2008) for a step-by-step guide and sample STATA code. Various statistical tools, from simple grid searches to far more complex algorithms, can be used to search the parameter space to solve the maximization problem.

Researchers using MLE methods should carefully consider the functional form assumptions underlying their results and, where possible, check robustness to alternative specifications for both utility and participant errors. For instance, Apesteguia and Ballester (2018) point out that choice models such as those in Equation (3) can suffer from non-monotonicities when estimating risk and time preferences, and instead propose using a Random Parameter Model (RPM), in which errors affect the preference parameter an individual uses when making a choice. However, such concerns are not necessarily an issue in practice and the RPM may also be problematic (Holzmeister and Stefan, 2021). Generally, comparing estimates obtained under different specifications can provide reassurance that a researcher's assumptions are not critical to any results, or highlight potential issues that require further consideration.

A major drawback with MLE methods is that they may require significant amounts of data. With small amounts of data, numerical optimization procedures may fail to converge, meaning that preference parameters will not be obtained. This drawback may be particularly limiting when seeking to obtain individual-level preference estimates—rather than obtaining population- or group-level parameters—in which case individual choices can be pooled across participants.

[34]The logit function has been widely used in economics and psychophysics because it is closely connected to the random utility model: choice probabilities will be logit if the errors in the random utility model have an Extreme Value Type I distribution (see McFadden, 2001, for a broader discussion). See Harrison and Rutström (2008) and Apesteguia and Ballester (2018) for comparisons of estimates under alternative link functions.

Papers using MLE to produce individual-level estimates have thus tended to ask a large amount of questions (e.g., Sokol-Hessner et al. (2009) ask individuals a series of 120 binary choices), which may be complicated in some research settings.[35]

Bayesian Estimation Parameter estimates can also be obtained using Bayesian estimates, although this has been relatively uncommon in economics.[36] In this approach, as with MLE, researchers must choose a utility function to be estimated and an error specification, such as in Equation (3). In addition, a prior distribution over each parameter value must be chosen. This prior is then updated using Bayes rule depending on the choices made by participants, and the mean (or median) of the resultant posterior provides a preference estimate. In contrast to MLE, this approach provides parameter estimates for all participants, even with limited amounts of data.[37]

A key concern in any Bayesian analysis is the choice of the initial prior over parameter values, and understanding how that prior affects the final estimates. The simplest possibility is to use a uniform (or diffuse) prior for each individual, in which each possible parameter value is seen as equally likely—that is, the researcher has no information ex ante.[38] A similar approach is used by Chapman et al. (forthcoming), discussed above, and may be particularly appropriate in previously understudied populations—for instance, running an experiment in a developing country rather than among undergraduate students in the U.S. In well-studied populations, datasets from earlier experimental work could be used to construct priors for future experiments. Alternatively, Gao et al. (2023) use a hierarchical prior which describes a distribution both for each individual, and also for the distribution of individuals across the sampled population—an approach which may be particularly useful in situations with a clear group structure. Further research is needed to understand which approach is most appropriate

[35]Andreoni et al. (2023) take a different approach, taking advantage of improvements in computational power to implement the Method of Simulated Likelihood, and hence avoid difficulties with MLE due to nonlinearities in the choice data.

[36]Bland (2023) provides a useful practical guide to implementing Bayesian techniques in experimental economics, including an application to estimating risk preferences.

[37]Gao et al. (2023) report that using MLE they are unable to estimate preference parameters for between 8% and 28% of observations even after asking participants 80 binary choice questions. With Bayesian methods, in contrast, they obtain estimates for all participants and, importantly, those estimates appear to be informative, as they diverge from the mean of the initial prior used for estimation.

[38]Note that even if a uniform prior is used, some judgment must be used in selecting the minimum and maximum possible parameter values.

in different settings; however a simple solution for a researcher is to check sensitivity of results to alternative prior specifications.

Cluster Analysis An alternative approach, which has become increasingly popular in the study of social preferences, involves classifying individuals into "types", rather than producing individual-level parameters. Bruhin et al. (2019b) apply a finite mixture model to the study of social preferences and endogenously classify participants into three groups, without making any initial assumptions regarding the existence of particular (e.g., altruistic) types. This method has the advantage of producing a parsimonious categorization of individuals, but requires a large amount of choice data (each participant answered 117 binary choices), as well as an initial assumption regarding the number of types to be identified.[39] Further there is no guarantee that the resultant clusters will be intuitive, or have clear theoretical interpretations.

In a recent study Fehr et al. (2023) classify social preferences within the Swiss population using a method that avoids the need to specify the number of clusters ex ante. Similarly to Bruhin et al. (2019b) they ask participants to make a large number of choices regarding allocations between themselves and another player. They then categorize participants using the Dirichlet Process Means algorithm (Kulis and Jordan, 2012). With this algorithm the number of clusters is determined endogenously—experimenters need only specify a level of precision ex ante. In their case, three intuitive clusters emerge—a group that equalize payoffs (inequality averse), a selfish group, and a group that increase payoffs of those that are worse off but not decrease the payoffs of the better off (in contrast to the inequality averse individuals). Further, Fehr et al. (2024) show that these clusters predict support for government redistribution.

Simulation and Parameter Recovery Simulation exercises can play a useful role in understanding the likely performance of preference elicitations before data collection. The use of simulations—Monte Carlo studies—is common in the development of econometric methods (e.g., Athey et al., 2021), but such methods have received relatively little attention in the experimental literature. There is considerable scope for greater use of simulation methods in preference elicitation both when developing new methods, and making more minor, but im-

[39]Similarly, Carpenter and Robbett (2024) estimate social preferences allowing for the desire to adhere to social norms. Finite mixture models have also become popular in the study of fairness preferences (Cappelen et al., 2013).

portant, design choices that may affect results. Simulation can also reduce the need to pilot different measures, which can be problematic for both practical reasons (small sample sizes and budget constraints) and in creating researcher degrees of freedom—see Chapter 11.

Specifically, such an exercise begins by generating a large set (e.g., 10,000) of simulated participants, each with preferences defined by the parameters of a utility function such as Equation (1). Parameter values for each simulated participant are drawn probabilistically from some underlying distribution. The researcher then defines a choice procedure used by each participant to make choices—for instance, using the logit function in Equation (3) in the case of binary choices.[40] With this information the researcher can simulate the choice of each simulated participant on the preference elicitation(s) they are considering using in the experiment, and the associated preference parameters (estimated using methods such as those above).

This simulated choice data can be used to undertake a parameter recovery (or "ground truth") exercise—that is, identifying how accurately the estimates obtained by the elicitation procedure represent simulated participants' genuine individual preference parameters—which is not possible with real world data, because we can never know the "true preferences". The anticipated measurement error from alternative candidate elicitations can be compared. This could involve comparing different elicitations entirely—an MPL to a CTB, for example—or assessing tweaks to a particular elicitation method, such as reducing the number of rows of an MPL or altering the on monetary amounts or payment dates offered. Similarly, simulations could be used to test whether an elicitation is sufficiently accurate to robustly identify correlations with other variables, or to distinguish differences in preferences across experimental treatments. Most straightforwardly, for example, a simulation may identify a poorly-calibrated design that leads all individuals to make the same choice.

This discussion has highlighted the range of elicitation methods and econometric techniques used to capture participant's choices and produce preference estimates. Researchers need to think carefully before choosing a methodology and forecasting their analysis plans, with attention given to how different elicitation approaches can alter, constrain, and potentially bias,

[40] Ideally, the likely distribution of preferences and possible choice procedures would be estimated using data from earlier experiments in similar participant pools. Comparing simulated and real responses on the same elicitation allows an assessment of whether the assumptions underlying a simulation are appropriate—for instance, standard models may overlook potential framing effects and hence underestimate the number of corner choices. Note that once a simulation dataset has been created it can be re-used many times to evaluate different elicitations, meaning that the marginal cost of running additional simulation exercises can be relatively low.

their analyses. In the next section we provide a step-by-step guide to the main issues that should be considered.

4 A Step-by-Step Guide to Eliciting Preferences

How should a researcher select an elicitation method from the vast range of techniques reviewed above? There is no single answer to this question, particularly given the constraints of experimental time and research budgets. The best solution will depend on the details of the research setting, and as such, we provide some guidance by highlighting the key issues a researcher should consider.

Step 1: Where to Elicit Preferences The research environment and subject pool determine the degree of experimental control and may preclude the use of time-consuming or complex elicitation methods. Historically, preferences have largely been captured in laboratories located on university campuses, but in recent years, experimenters' attention has shifted towards collecting preferences elsewhere, such as in the field (e.g., Bigoni et al., 2016), online (e.g., via Amazon Mechanical Turk or Prolific), or in representative surveys (e.g., Dohmen et al., 2011; Von Gaudecker et al., 2011; Falk et al., 2018; Chapman et al., 2023a). This has been prompted by improvements in technology, the general growth of field experiments and randomized trials, and the growing acceptance of surveys within economics.[41] These broader research settings may limit the time available for preference elicitation (running surveys is expensive, and participants may opt to prematurely end a study if it takes too long) and may also mean more noisy estimates (Snowberg and Yariv, 2021). Moreover, in some samples, participants may have less experience with economic-style tasks. This may benefit researchers who desire an "uncontaminated" sample but, at the same time, preclude the use of more complex elicitation methods.

More advanced methodologies may require questions to be given on a computer (as opposed to paper-and-pencil) as these require various computations to determine the stimulus values that participants make decisions over (e.g., adaptive methods such as DOSE, DEEP, or staircase

[41] See Stantcheva (2023) for a discussion of the benefits of survey research and a guide on best survey practice in economics.

procedures). Researchers opting for this route may need to invest time familiarizing themselves with a particular platform that hosts their study. Some common behavioral research platforms include Qualtrics, oTree, and PsychoPy. SurveyCTO provides a data collection platform that is effective offline, and is frequently used in development settings.

Step 2: Which Preference(s) to Elicit It is critical to think carefully about the preferences that need to be captured and why before considering possible elicitation methods. Conceptually, the central question is to identify an elicitation with "content validity"—that is, that *adequately* captures the full content of the preference under consideration. How much is "adequate" will be determined by the particular research question. For instance, a researcher interested in identifying the parameters of a prospect theory utility function may need to separately identify three preference parameters—risk attitudes over gains, risk attitudes over losses, and loss aversion—whereas a researcher interested in understanding how risk preferences correlate with education may be happy to elicit a single risk aversion parameter or collect a non-parametric measure of risk preferences. Some researchers may wish to disentangle the motivations underpinning giving in a dictator game, whereas others may be happy to use the level of giving as a general measure of prosociality. It is important to consider these issues when designing an experiment to ensure that the data collected are sufficient to carry out the analysis required (see discussion in Section 3.4).

Step 3: Dealing with Noise Measurement error poses a major challenge for preference elicitation. A substantial amount of noise in the data is likely to undermine the accuracy of preference estimates, may attenuate observed relationships between collected variables, and can bias empirical findings. We identify four, non-mutually exclusive, strategies to reduce the amount of measurement error in estimated preferences.

First, measurement error can be reduced by careful design. Elicitations can be made as simple as possible by implementing them in a way that facilitates understanding by participants, taking into account the research setting. Participants may require training or practice rounds, particularly if they do not have prior experience with economics experiments. Question wording should be appropriately targeted towards the population one samples from (typically as simple as possible), clearly explained, and minimize the cognitive load on participants. Visual aids may be useful, particularly to communicate probabilities (see, for instance, Baillon et al., 2022a; Friedman et al., 2022). Piloting may often be warranted, particularly in less studied sub-populations. The discussion above has identified potential issues and biases associated with different elicitation methods, and researchers should think carefully about which are likely to be of particular concern in their case.[42] The options available in each elicitation (such as the number and spacing of rows in an MPL) will determine the coarseness of parameter estimates received, and also requires careful consideration ex ante—the simulation and parameter recovery methods discussed in Section 3.4 may be a useful tool at this point.

Second, multiple elicitations of a measure can be included in an experiment, allowing ex post adjustments to account for potential measurement error, either through simple averaging or instrumental variables techniques, such as ORIV (Gillen et al., 2019). Presenting participants with more choices increases sample size, and hence allows model parameters to be recovered more accurately. Some data can be excluded from preference estimation and used as a holdout sample for any additional testable predictions. However, some have argued that more questions come at the cost of greater external validity in that participants can utilize more task-specific decision processes that may not apply to other tasks (Li et al., 2022). Moreover, asking more questions is time consuming, and does not necessarily produce estimates for all individuals (in the case of choices of corner solutions, for instance).

Third, data can be excluded ex post if it appears to be due to mistakes or inattention. Many studies simply exclude apparently noisy data, either in the main results or as robustness checks. Comprehension quizzes, error checks, or simply a self-reported question (Dohmen and Jagelka, 2024), may be used to screen out "noisy" participants. Similarly, participants are often excluded due to making multiple switches in MPLs or always choosing corner solutions in CTBs. This approach relies on the assumption that the cause of apparent "mistakes"—such as inattention or non-standard preferences—is orthogonal to the experimental treatment / research question of interest. Moreover, it can lead to losing a large amount of data, particularly in settings such as MTurk, and may raise concerns about "p-hacking" if implemented ex post. As such, the decision to exclude noisy participants should be accounted for in power calculations and,

[42] See Enke and Shubatt (2023) for a systematic examination of the characteristics of choices between two lotteries associated with greater complexity. Their findings demonstrate the trade-offs researchers face: the presence of negative payoffs, for example, is found to increase complexity but such questions are central to some research questions.

ideally, be clearly stated ex ante in a pre-registration.

Fourth, noise can be accounted for after data collection has occurred, using the econometric methods discussed in Section 3.4 to estimate parameters. For example, noise can be modeled explicitly using maximum likelihood or Bayesian methods, but may require participants to make many choices. Correlations and regressions can be corrected to account for classical measurement error—using, for instance, the ORIV technique—but this necessitates multiple elicitations of the same preference to be collected, and relies on specific assumptions as to the form the noise takes.

The optimal adaptive methods discussed in Section 3.3.1 may mitigate many of the issues mentioned above and hence provide more accurate preference estimates. These methods provide structural estimates, accounting for inconsistent choice, as with maximum likelihood estimation. Unlike earlier methods, however, they account for inconsistent choice during question selection, and hence reduce the amount of data needed. Further, existing implementations have used relatively simple questions, likely reducing possible biases due to framing and reference effects. However, such methods remain relatively novel and have been implemented to elicit a small range of preferences so far.

Step 4: Incentivized or Hypothetical Elicitations The decision of whether and how to incentivize participants will be shaped both by the preferences being elicited and the specifics of the research environment. As discussed above, the literature suggests that incentives may play a large role for some behaviors (e.g., rejections in the ultimatum game), but not others. Practical considerations may limit the ability to offer incentives, or mean that it is difficult to implement complicated incentive structures. There may be a trade-off between incentive compatibility and complexity, as explaining an incentive-compatible method may require complicated instructions and training rounds, which may not always be feasible outside the laboratory.

Hypothetical questions tend to lead to over-statements of Willingness-to-Pay for goods (Hausman, 2012) and appear to be associated with different patterns of brain activity (Camerer and Mobbs, 2017) compared to real choices. However, findings regarding the effect of incentives on the elicitation of preferences in general remain quite mixed.[43] Holt and Laury (2002) find

[43]See Camerer and Hogarth (1999) for a standard review of the role of incentives on performance in experiments, concluding that incentives do not increase performance on average but do reduce variance.

that higher incentives increase risk aversion; however, this finding has not been consistently replicated and may be limited to particular elicitation procedures (Hackethal et al., 2023). Brañas-Garza et al. (2023) find little difference in elicited time preferences when randomly-assigning participants to receive hypothetical or real rewards. Further, Stango and Zinman (2020) collect time preference data from a hypothetical CTB and report similar behavior to previous incentivzed elicitations, and more generally find levels of temporal stability of behavioral measures that are similar to other studies using incentives, suggesting that the hypothetical incentivization does not lead to significantly noisier estimates.

One possible explanation for the limited impact of incentivization is that the incentives in experiments are relatively weak. Incentives are generally relatively small (typically on the order of \$10-\$20 in laboratory experiments, and even less online), and any individual choice is typically paid with small probability—participants tend to act more carefully and process more information as the probability of receiving a payment increases (Yang et al., 2018). Some studies have found that behavior changes at very high stakes. For instance, Andersen et al. (2011) find that rejection rates in ultimatum games become very low as stakes become high. However, other work has still found evidence of rejections even with higher stakes (Hoffman et al., 1996; Cameron, 1999; List and Cherry, 2000). Moreover, high stakes do not appear to remove cooperative behavior (Van den Assem et al., 2012) or cognitive biases. As such, as with the role of hypothetical versus incentivized choices in general, the literature does not provide a clear answer as to when high stakes are likely to influence prior results.

Incentivizing choices is complicated by the fact that, in most experiments, participants are expected to make many decisions. The experimenter must then decide whether to pay all the choices a participant makes, or only a subset of those choices. A common approach is to use the "random–lottery incentive system", whereby one question is chosen at random for payment at the end of the experiment. This incentive system ensures that the rewards from each task are independent of each other—if multiple choices were paid, then a participant's endowment would vary depending on the choices made elsewhere in the experiment. An alternative is to pay only a fraction of the participants for their responses, allowing higher stakes (for a given budget) and reducing the experimenter's administrative burden.[44]

[44]See Charness et al. (2016) for a review of different payment systems. Holt (1986) raised a potential theoretical concern with the random-lottery incentive system: participants may treat all the choices in an experiment as

There may be occasions where the experimental environment prevents any incentivization of choices. Incentive compatibility requires a level of trust in the experimenter—for instance, to believe that a lottery will be run fairly if selected—which may be difficult in some circumstances. It may also be difficult in practice to pay participants for individual questions if they are, for instance, contacted through an online or telephone survey. For these reasons, some major recent studies in general populations have used hypothetical questions alongside a fixed incentive for participation (Falk et al., 2018; Stango and Zinman, 2023). There have been two major ways to mitigate differences between hypothetical and incentivized elicitations. First, "instrument calibration" prior to an elicitation largely attempts to help participants avoid any biases related to the hypothetical nature of the choice (Cummings and Taylor, 1999; List, 2001). Second, "statistical calibration" attempts to statistically remove any biases related to the hypothetical nature of the choice after the elicitation, but commonly will need to utilize multiple data sources, including incentivized choice (Blackburn et al., 1994; Cao and Zhang, 2021).

Step 5: How to Pay Participants If incentives are to be implemented, then experimenters need to think carefully about how participant payments will be distributed, particularly when eliciting time preferences, or when imposing losses on participants. If payment occurs at a different time from the experiment, as is common when eliciting time preferences, then participants may view future payment as uncertain (or incurring high transaction costs) and elicitations may capture risk rather than time preferences. This point has been made by Andreoni and Sprenger (2012b), who suggest that findings of present bias may be explained by payment risk. The best way to mitigate such risks will vary according to research environment. For instance, Andreoni and Sprenger (2012b) implement future payment through campus mail and a personal check from the experimenter, but such methods are unlikely to scale to field or online settings. The development of and changing preferences over electronic payments may lead to other methods becoming viable.

A separate concern is that it is difficult to impose real losses on participants for both ethical and practical reasons, complicating the estimation of key parameters of prospect theory.

a single choice problem, meaning choices are incentive compatible only under Expected Utility Theory. Evidence has tended to suggest that this is not an issue in practice (Cubitt et al., 1998; Hey and Lee, 2005), although some other work has suggested there may be problems at least in some circumstances (Freeman and Mayraz, 2019; Baillon et al., 2022b).

The standard approach has been to give participants an initial endowment, from which any losses realized during the experiment are subtracted. Participants then have “limited liability”, meaning that they cannot receive a total negative payment from an experiment. This poses the concern that participants may incorporate the (positive) endowment when evaluating payoffs, and hence not perceive “losses” as actual losses at all. Despite this concern, to date there has been limited evidence that it confounds experimental findings—Etchart-Vincent and l’Haridon (2011) investigate different methods for implementing experimental losses, and observe similar behavior when paying losses out of an endowment or out of a participant’s own pocket.

More generally, endowments may induce risk-taking through “house money” effects (Thaler and Johnson, 1990), whereby money received during the experiment is treated differently from money outside the experiment.[45] To minimize the potential for such effects, researchers can limit feedback regarding payment before the end of the experiment. Additionally, when phrasing instructions to participants, one should separate information regarding any endowment from the choices.[46]

Step 6: Framing Preference Elicitations It is important to carefully consider the details of how a preference elicitation is designed, particularly the framing and ordering of tasks assigned to participants. These features can alter baseline preference estimates, and may also interact with other characteristics of an experiment (such as the research setting) or participants (such as their cognitive ability) in ways that could confound results.

The classic economic framework considers “true” preferences as context-independent and measurable utility function parameters that can apply regardless of circumstances. The standard approach is thus to provide tasks and associated instructions that are as context-free as possible—simply defining the relevant choices, payoffs, and “rules of the game”. However, providing some context is unavoidable, and unfortunately, small adjustments in presentation can significantly alter participant behavior. For instance, cooperation in the Prisoners’ Dilemma is higher when a game is called a “Community Game” than a “Wall Street Game” (Kay and

[45]Experimental studies of the house money effect have found somewhat mixed results. Cárdenas et al. (2014) find subjects using their own money appear more risk averse than those given an endowment. Flepp and Rüdisser (2019), in contrast, find a reverse house money effect among gamblers in Switzerland.

[46]The endowment does not even need to be given within the experiment itself, but can be offered on an earlier date in order to increase the participants’ sense they are playing with their own money. See Payzan-LeNestour and Doran (2024) for an example.

Ross, 2003; Liberman et al., 2004).[47] Seemingly irrelevant cues can create "anchors" that influence stated Willingness-to-Pay for objects (Maniadis et al., 2014). The experimental set-up may also influence the "reference point" (Kahneman and Tversky, 1979; Kőszegi and Rabin, 2006) relative to which potential gains and losses are evaluated and hence affect participants' choices.[48]

In particular, several effects identified in the study of time preferences illustrate the need to consider both how questions are framed and the values of the stimuli presented to participants. While some time preference elicitations specify the number of days until a delayed reward will be received (e.g., 15 days), others will note the precise date when the delayed reward will arrive (e.g., April 13). Prior research has shown that participants use lower discount rates in the date frame compared to the delay frame (Read et al., 2005; Leboeuf, 2006). Additionally, there is evidence of a "magnitude effect", in which larger outcomes are discounted at a lower rate than smaller ones (Thaler, 1981). Researchers should, therefore, take care when selecting both the framing and numerical values of their stimuli.

A particularly pernicious form of framing effect arises when the order in which participants complete tasks significantly alters their responses.[49] Experimenters must give careful thought to the order in which their questions appear to participants, and there is no perfect solution. One option is to randomize the presentation of different parts of an experiment, so that order effects can be explicitly examined and, if necessary, controlled for. Alternatively, the order may be fixed to avoid order effects affecting the main research question. Preference elicitations, in fact, often appear late in experiments so that they do not bias experimental manipulations that are the main focus of a study. Such an approach means, of course, that the elicited preferences

[47] The theoretical implications of such context effects for the studying of preferences are not obvious. Changing behavior in the Prisoners' Dilemma as a response to social cues could represent either changes in preferences or a change in beliefs over others' behavior, if the cue acts as a coordination device. In a dictator game, where there is no interaction between participants, Dreber et al. (2013) find that social effects framing has no effect on giving in dictator games. However, giving in the dictator game is susceptible to other cues (see Krupka and Weber, 2009, for instance). Identifying the reasons for such social context effects thus remains an ongoing topic of research—See Alekseev et al. (2017) for a relatively recent review.

[48] See Marzilli Ericson and Fuster (2011) and Abeler et al. (2011) for examples of experiments deliberately manipulating participants' reference points. See O'Donoghue and Sprenger (2018) for an in-depth discussion of the literature studying reference-dependent preferences.

[49] Kessler and Meier (2014) report an interesting account of this phenomenon. Although an initial study supported a relationship between cognitive load and charitable giving, they failed to replicate the finding and proposed that the order in which a behavioral laboratory conducted other, irrelevant, tasks that occurred in the same session influenced this.

may be biased by those earlier questions.

Potential experimenter demand effects—whereby participants produce the behavior they believe the researcher is looking for—represent a significant threat to the validity of preference elicitation, particularly for social preferences. The abstract nature of experimental games runs the risk of subjects acting "unnaturally" and producing behavior that does not represent their true preferences. These concerns are particularly salient when seeking to obtain quantitative estimates of preferences, rather than simply compare across treatment conditions. In such cases, even subtle biases can meaningfully distort results. See De Quidt et al. (2019) for a discussion of best practices for reducing potential demand effects.

Summary: Choosing a Preference Elicitation In Table 2 we summarize the major elicitation methods, drawing on the discussion and evidence above. Each of the methods can be used in laboratory, field, or online settings and can be used to elicit risk, time, or social preferences.

The main features governing whether to use a method given a particular setting are covered in columns 2 through 5 of the table. Column 2 details the number of elicitations we suggest researchers use to account for likely measurement error, based on their implementation in prior studies. For example, it is possible to classify participants using a single binary choice; however, more are required to obtain sufficient information to account for inconsistent choice or produce useful parameter estimates. The range in the table reflect the need for more questions when a researcher wishes to elicit multiple parameters (e.g., separating risk and loss aversion). Further, it is worth noting that more is not always better, or possibly even feasible outside of lab settings, as long sequences risk participant inattention and boredom.

In column 3 we refer to the intuitiveness of the elicitation task, i.e. whether it is likely to be clear to participants what they should do in the task. For example, methods which are less familiar to participants—such as MPLs or Convex Budget Sets—tend to require extensive instructions or training. This necessitates additional response time, sufficient participant sophistication, and additional attention such that they can understand the instructions. Adaptive methods are potentially valuable in reducing the amount of questions and sophistication needed, as making a binary choice is fairly intuitive, but require access to a computer to route participants through an experiment, noted in column 4.

Column 5 of the table indicates whether a method can be incentive compatible. Economists have tended to prioritize incentive compatibility in order to encourage truthful revelation of preferences. However, whether incentivization is necessary remains an open question. Incentive compatibility may be particularly important for elicitations that require participants to pay a great deal of attention, or where the benefits from strategic manipulation are easy for participants to identify. On the other hand, incentives may be hard to implement in practice, and incentive compatible mechanisms—such as the BDM method—may be difficult to explain. For this reason, hypothetical survey questions have become increasingly common, particularly outside of laboratory settings. In principle, optimal adaptive methods can be incentive compatible, but further research is required to understand how effective this would be in practice.

The final two columns present our assessment of the current evidence regarding sources of possible measurement error in elicitation methods. The high rate of inconsistent choice in matching tasks has been recognized in many studies, prompting a growing focus on choice-based elicitation approaches. The structure of newer methods, such as MPLs and Convex Budget Sets, has likely reduced the rate of inconsistent choice—in some cases by force, such as requiring a single switching row in an MPL—although few attempts have been made to quantify this. Optimal adaptive elicitations directly model inconsistent choice and hence account for it when estimating preferences. In contrast, qualitative self-assessments simply avoid the issue by moving away from the choice-based approach. Further, many elicitation methods appear to suffer from the use of heuristics, such as focal value rounding, due to framing effects. This issue has received less attention from economists to date, but may be particularly relevant outside of laboratory settings.

Table 2: Comparison of Leading Methods for Eliciting Risk, Time, and Social Preferences.

	Implementation				Potential Concerns	
Example Question	Suggested # of Elicitations	Training Required	Computer Required	Incentive Compatible	Inconsistent Choice	Framing Effects
Matching Tasks *(Thaler, 1981)*						
Certainty Equivalent for a lottery	1 Task	Yes	No	Yes†	Very High	Low
Binary Choices *(Kirby and Marakovic, 1996)*						
A lottery between \$0 and \$10 or \$5 for sure	20–120 Choices	No	No	Yes	High	Low
Multiple Price Lists *(Andersen et al., 2008; Chapman et al., 2023a)*						
A lottery between \$0 or \$10 or varying amounts for sure	2–4 MPLs	Yes	No	Yes	Medium	High
Convex Budget Sets *(Andreoni and Sprenger, 2012a; Augenblick et al., 2015b; Bruhin et al., 2019a)*						
How much of endowment to keep for future	14–45 Budgets	Yes	No	Yes	Medium	Medium
Optimal Adaptive Elicitations *(Toubia et al., 2013; Chapman et al., 2024)*						
(see binary choice)	4–10 Choices	No	Yes	Yes†	Low	Low
Validated Preference Modules *(Falk et al., 2018, 2023)*						
Quantitative (e.g. CE for lottery)	1	No	Yes	No	n.a.	Low
Qualitative (e.g. Willing to take risks)	1	No	No	No	n.a.	Medium

Notes: "Suggested # of Elicitations" is the minimum number of elicitations recommended to elicit an relatively low noise estimate, with the lower bound referring to eliciting one preference, and the higher to eliciting multiple preferences. The numbers are based on use in prior studies, but researchers should plan to conduct any necessary parameter recovery exercises. "Training required" indicates that either detailed instructions or practice questions are needed. † indicates that incentive compatible implementations are either complicated to explain (matching) or largely untried in practice (optimal adaptive methods).

5 Future Directions

Improving elicitation methodologies is an active area of research, and so we conclude this chapter by discussing potential future directions for preference elicitation, both in terms of its methodology and possible applications.

Preferences, Mistakes, Complexity Better understanding how to minimize, and account for, complexity is a pressing concern for the field of preference elicitation. Complexity can lead to measurement error, and hence hinder the endeavor to understand the basis of preferences. Jagelka (2024), for example, finds that 60% of heterogeneity in risk and time preferences can be explained by cognitive ability and three of the Big-5 personality traits (extraversion, conscientiousness, and emotional stability). These relationships would be obscured, he argues, without careful accounting for measurement error and inconsistent choice. Recent work suggests that complexity may artefactually create apparent behavioral regularities (Enke et al., 2023), even in laboratory studies, where participants are likely more sophisticated than in the general population. The discussion above has highlighted the progress economists have made in addressing these issues, but also demonstrated a number of unresolved issues with existing elicitation methodologies.

In particular, future research can incorporate formal notions of complexity into the design of elicitation technologies. Studies often introduce elicitation technologies as "simple", but with little supporting evidence—or even agreement as to what such evidence would be. However, this has begun to change. Oprea (2020) introduces an experimental method for identifying complexity, while Enke and Shubatt (2023) identify characteristics of lotteries that appear associated with complexity as well as eliciting participants' own subjective assessment that they made a choice that matches their preferences. The impact of complexity can be experimentally tested by investigating and comparing behavior in elicitations to that in "mirrors" that isolate mistakes (Oprea, 2024; Banki et al., 2025), or by prompting participants to review their choices (Nielsen and Rehbeck, 2022). Further work is needed to incorporate these insights into better elicitation procedures.

One avenue for future development is to estimate models of participants' (seeming) mistakes alongside the preferences more typically studied by economists. Choice can be modelled probabilistically using, for example, the classic Luce model (Luce, 1959). An alternative approach is to consider seemingly inconsistent choice—such as multiple switching in an MPL—as a deliberate part of behavior, representing preferences that depart from the assumptions of expected utility theory (Chew et al., 2022). To date, this line of research has largely modeled decision-making as depending only on the characteristics of the choices themselves (such as the possible prizes in a lottery), but future work could incorporate elements of the choice architecture (e.g., accounting for presentation in a list format). Optimal adaptive methods offer a promising way to incorporate such behavioral models into elicitation procedures.[50]

Machine Learning Recent developments in artificial intelligence open exciting opportunities to develop better theories of decision-making that can be incorporated into elicitation methods. Decades of experimental work mean that there now exist considerable amounts of behavioral data to be explored, reflected in a growing number of meta-analyses (e.g., Imai and Camerer, 2019; Tasoff and Zhang, 2022; Brown et al., 2024). Rapid improvements in machine learning offer the potential to explore this data in new ways. Peterson et al. (2021) demonstrate how a neural network can be used to evaluate different theories of decision-making, including both standard "human-devised" theories (e.g., expected utility, prospect theory), and more complicated models that allow for contextual effects (e.g., where the attributes of choice B affect the valuation of choice A). Human-devised theories perform well but appear, in some sense, too simple—failing to account for all the interactions between different components of a choice problem.[51] Future work could use a similar algorithm to explore models that incorporate potential strategic learning, mistakes, and actual economic preferences. The most effective models can then be incorporated into preference elicitations.

Process Data Economists have traditionally exclusively used choice data to investigate individual behavior. Indeed, to estimate the models referenced in Section 2.2 one needs only choice data. However, there has recently been an increased interest in economics in using choice process data that occurs as individuals make decisions.[52] Most of this current work has integrated

[50]Chapman et al. (2024), for instance, show that a choice consistency parameter elicited through DOSE captures useful information about participants.

[51]Similarly, Peysakhovich and Naecker (2017) find that economic models perform well in explaining choice under risk—but also find that they perform less well for choice under ambiguity.

[52]See Chapter 8 for a comprehensive account of process data in economic choice.

response time (i.e., the time that it takes an individual to make a choice) with choice data to better understand the decision-making process, although other work has also made use of attention-tracking data using eye movements (Wang et al., 2010). Although such ideas have a rich history in psychology, they are relatively new to economics.

Much of this work has focused on whether response times contain information that improve predictions beyond choices alone. Prior work has demonstrated that response times can provide information on the strength of preferences (Konovalov and Krajbich, 2019), and has connected decision times to choices in settings including information cascades (Frydman and Krajbich, 2022), risky decision-making (Alós-Ferrer and Garagnani, 2024), product choice (Clithero, 2018; Alós-Ferrer et al., 2021), and coordination games with incomplete information (Schotter and Trevino, 2021). Although much of this work has occurred in laboratory settings, response times have also proved useful in the field. Chiong et al. (2023) model user decision-making utilizing response times from mobile advertisements in a decision about whether to close the ad. Cotet and Krajbich (2021) find that differences in the time at which eBay sellers respond to offers can provide information on their strength of preference. Card et al. (2024) develop a model that utilizes response time and find that the likelihood of receiving a positive decision at a top economics journal rises with response time.

Applications Advances in preference elicitation technology potentially expand the range of settings in which preference elicitations can be collected and the ways in which they can be applied. As accurate preference measures can be obtained quickly in the field, they can be used for both academic research and for policy interventions.

Measures of risk preferences are already required during the provision of financial advice. Under financial regulation established after the 2008 global financial crisis, financial institutions are mandated to provide products that account for a clients' risk tolerance and their ability and willingness to bear losses (Ranganathan and Lejarraga, 2021). Experimental elicitation methods such as those discussed here could be used to improve the quality of advice offered, particularly for "robo-advisors" which produce investment recommendations based on a small number of questions (see Tertilt and Scholz, 2018, for a discussion). van Dolder and Vandenbroucke (2022), for instance, implement an experimental measure of loss aversion into risk-profiling among investors, and find higher rates of acceptance of the final classification than with a more standard profiling method. Further advances in this area could allow the development of investment portfolios customized according to elicited preference parameters.

More broadly, better measurement opens up the potential for preferences to be used for by policymakers. For instance, in a novel paper, Andreoni et al. (2023) use the convex time budget approach of Augenblick et al. (2015b) to design customized incentive contracts for polio vaccinators in Pakistan. In particular, in a first stage they asked vaccinators to allocate vaccination targets across two different days. The resultant choices were used to estimate participants' discount factors—and the estimates then used to construct a second, tailored, incentive contract for each vaccinator with the goal of inducing an even allocation of vaccinations over time. This approach leads to behavior around 30% closer to the policy target of smoothed vaccination provision than alternatives.

Summary Understanding preferences is fundamental to a better understanding of individual decision-making. Decades of experimental work has demonstrated that measuring preferences is not a straightforward task, and findings of low predictive validity and temporal stability have led some researchers to question the value of the measures typically elicited by economists. Methodological improvements, both in data collection and econometric analysis, have begun to address these critiques and have enabled the study of preferences in a variety of settings beyond the laboratory. While many challenges remain, we have highlighted many exciting avenues for future research to further improve preference elicitations, and incorporate the findings into behavioral policy interventions.

References

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman, "Reference points and effort provision," *American Economic Review*, 2011, *101* (2), 470–492.

Åkerlund, David, Bart HH Golsteyn, Hans Grönqvist, and Lena Lindahl, "Time discounting and criminal behavior," *Proceedings of the National Academy of Sciences*, 2016, *113* (22), 6160–6165.

Alekseev, Aleksandr, Gary Charness, and Uri Gneezy, "Experimental methods: When and why contextual instructions are important," *Journal of Economic Behavior & Organization*, 2017, *134*, 48–59.

Alós-Ferrer, Carlos and Michele Garagnani, "Improving Risky-Choice Predictions Using Response Times," *Journal of Political Economy Microeconomics*, 2024, *2* (2), 335–354.

_ , Ernst Fehr, and Nick Netzer, "Time will tell: Recovering preferences when choices are noisy," *Journal of Political Economy*, 2021, *129* (6), 1828–1877.

Andersen, Steffen, Glenn W Harrison, Morten I Lau, and E Elisabet Rutström, "Eliciting Risk and Time Preferences," *Econometrica*, 2008, *76* (3), 583–618.

_ , _ , Morten Igel Lau, and E Elisabet Rutström, "Elicitation using Multiple Price List Formats," *Experimental Economics*, 2006, *9*, 383–405.

_ , _ , _ , and _ , "Preference Heterogeneity in Experiments: Comparing the Field and Laboratory," *Journal of Economic Behavior & Organization*, 2010, *73* (2), 209–224.

_ , Seda Ertaç, Uri Gneezy, Moshe Hoffman, and John A List, "Stakes matter in ultimatum games," *American Economic Review*, 2011, *101* (7), 3427–3439.

Andersson, Ola, Håkan J Holm, Jean-Robert Tyran, and Erik Wengström, "Risk aversion relates to cognitive ability: Preferences or noise?," *Journal of the European Economic Association*, 2016, *14* (5), 1129–1154.

_ , _ , _ , and _ , "Robust Inference in Risk Elicitation Tasks," *Journal of Risk and Uncertainty*, 2020, *61*, 195–209.

Andreoni, James, Amalia Di Girolamo, John A List, Claire Mackevicius, and Anya Samek, "Risk Preferences of Children and Adolescents in Relation to Gender, Cognitive Skills, Soft Skills, and Executive Functions," *Journal of Economic Behavior & Organization*, 2020, *179*, 729–742.

_ and B Douglas Bernheim, "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects," *Econometrica*, 2009, *77* (5), 1607–1636.

_ and Charles Sprenger, "Estimating Time Preferences from Convex Budgets," *American Economic Review*, 2012, *102* (7), 3333–3356.

_ and _ , "Risk preferences are not time preferences," *American Economic Review*, 2012, *102* (7), 3357–3376.

_ and John Miller, "Giving according to GARP: An experimental test of the consistency of preferences for altruism," *Econometrica*, 2002, *70* (2), 737–753.

_ , Michael A Kuhn, and Charles Sprenger, "Measuring time preferences: A comparison of experimental methods," *Journal of Economic Behavior & Organization*, 2015, *116*, 451–464.

_ , Michael Callen, Karrar Hussain, Muhammad Yasir Khan, and Charles Sprenger, "Using preference estimates to customize incentives: an application to polio vaccination drives in Pakistan," *Journal of the European Economic Association*, 2023, *21* (4), 1428–1477.

Anusic, Ivana and Ulrich Schimmack, "Stability and change of personality traits, selfesteem, and well-being: Introducing the meta-analytic stability and change model of retest correlations.," *Journal of Personality and Social Psychology*, 2016, *110* (5), 766.

Apesteguia, Jose and Miguel A Ballester, "Monotone stochastic choice models: The case of risk and time preferences," *Journal of Political Economy*, 2018, *126* (1), 74–106.

Athey, Susan, Guido W Imbens, Jonas Metzger, and Evan Munro, "UsingWasserstein generative adversarial networks for the design of Monte Carlo simulations," *Journal of Econometrics*, 2021, p. 105076.

Attema, Arthur E. and Werner B.F. Brouwer, "In search of a preferred preference elicitation method: A test of the internal consistency of choice and matching tasks," *Journal of Economic Psychology*, 2013, *39*, 126–140.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger, "Working over time: Dynamic inconsistency in real effort tasks," *The Quarterly Journal of Economics*, 2015, *130* (3), 1067–1115.

_ , _ , and _ , "Working over time: Dynamic inconsistency in real effort tasks," *The Quarterly Journal of Economics*, 2015, *130* (3), 1067–1115.

Baillon, Aurélien, Owen O'Donnell, Stella Quimbo, and Kim van Wilgenburg, "Do time preferences explain low health insurance take-up?," *Journal of Risk and Insurance*, 2022, *89* (4), 951–983.

_ , Yoram Halevy, and Chen Li, "Randomize at your own Risk: on the Observability of Ambiguity Aversion," *Econometrica*, 2022, *90* (3), 1085–1107.

Banki, Daniel, Uri Simonsohn, Robert Walatka, and George Wu, "Decisions under Risk Are Decisions under Complexity: Comment," *Available at SSRN 5127515*, 2025.

Bari, Faisal, Kashif Malik, Muhammad Meki, and Simon Quinn, "Asset-based Microfinance for Microenterprises: Evidence from Pakistan," *American Economic Review*, 2024, *114* (2), 534–574.

Barrington-Leigh, Christopher P, "The Econometrics of Happiness: Are we Underestimating the Returns to Education and Income?," *Journal of Public Economics*, 2024, *230*, 105052.

Barsky, Robert B, F Thomas Juster, Miles S Kimball, and Matthew D Shapiro, "Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study," *The Quarterly Journal of Economics*, 1997, *112* (2), 537–579.

Bartels, Daniel M., Ye Li, and Soaham Bharti, "How Well Do Laboratory-Derived Estimates of Time Preference Predict Real-World Behaviors? Comparisons to Four Benchmarks," *Journal of Experimental Psychology: General*, 2023, *152* (9), 2651—2665.

Bartoš, Vojtěch and Ian Levely, "Measuring Social Preferences in Developing Economies," in Utteeyo Dasgupta and Pushkar Maitra, eds., *Handbook of Experimental Development Economics*, Edward Elgar Publishing, forthcoming.

Bauer, Michal, Julie Chytilová, and Edward Miguel, "Using survey questions to measure preferences: Lessons from an experimental validation in Kenya," *European Economic Review*, 2020, *127*, 103493.

Beauchamp, Jonathan P, David Cesarini, and Magnus Johannesson, "The psychometric and empirical properties of measures of risk preferences," *Journal of Risk and Uncertainty*, 2017, *54*, 203–237.

Becker, Anke, Benjamin Enke, and Armin Falk, "Ancient Origins of the Global Variation in Economic Preferences," *AEA Papers and Proceedings*, 2020, *110*, 319–323.

_ , Thomas Deckers, Thomas Dohmen, Armin Falk, and Fabian Kosse, "The relationship between economic preferences and psychological personality measures," *Annual Review of Economics*, 2012, *4* (1), 453–478.

Becker, Gordon M, Morris H DeGroot, and Jacob Marschak, "Measuring utility by a single-response sequential method," *Behavioral Science*, 1964, *9* (3), 226–232.

Benjamin, Daniel J, Sebastian A Brown, and Jesse M Shapiro, "Who is 'behavioral'? Cognitive ability and anomalous preferences," *Journal of the European Economic Association*, 2013, *11* (6), 1231–1255.

Bentham, Jeremy, *An introduction to the principles of morals and legislation*, Clarendon Press, Oxford., 1789.

Berg, Joyce, John Dickhaut, and Kevin McCabe, "Trust, reciprocity, and social history," *Games and economic behavior*, 1995, *10* (1), 122–142.

Berry, James, Greg Fischer, and Raymond Guiteras, "Eliciting and utilizing willingness to pay: Evidence from field trials in Northern Ghana," *Journal of Political Economy*, 2020, *128* (4), 1436–1473.

Bettman, James R., Mary Frances Luce, and John W. Payne, "Constructive Consumer Choice Processes," *Journal of Consumer Research*, 1998, *25*, 187–217.

Bigoni, Maria, Stefania Bortolotti, Marco Casari, Diego Gambetta, and Francesca Pancotto, "Amoral familism, social capital, or trust? The behavioural foundations of the Italian North–South divide," *The Economic Journal*, 2016, *126* (594), 1318–1341.

Binswanger, Hans P, "Attitudes toward risk: Experimental measurement in rural India," *American Journal of Agricultural Economics*, 1980, *62* (3), 395–407.

Blackburn, McKinley, Glenn W Harrison, and E Elisabet Rutström, "Statistical bias functions and informative hypothetical surveys," *American Journal of Agricultural Economics*, 1994, *76* (5), 1084–1088.

Blais, Ann-Renée and Elke U Weber, "A domain-specific risk-taking (DOSPERT) scale for adult populations," *Judgment and Decision Making*, 2006, *1* (1), 33–47.

Bland, James, "Structural Bayesian Techniques for Experimental and Behavioral Economics, with applications in R and Stan," https://jamesblandecon.github.io/StructuralBayesianTechniques/section.html 2023.

Bostic, Raphael, Richard J Herrnstein, and R Duncan Luce, "The effect on the preference-reversal phenomenon of using choice indifferences," *Journal of Economic Behavior & Organization*, 1990, *13* (2), 193–212.

_ , R.J. Herrnstein, and R. Duncan Luce, "The effect on the preference-reversal phenomenon of using choice indifferences," *Journal of Economic Behavior & Organization*, 1990, *13* (2), 193–212.

Bowles, Samuel and Sandra Polania-Reyes, "Economic incentives and social preferences: substitutes or complements?," *Journal of Economic Literature*, 2012, *50* (2), 368–425.

Brañas-Garza, Pablo, Diego Jorrat, Antonio M Espín, and Angel Sánchez, "Paid and hypothetical time preferences are the same: Lab, field and online evidence," *Experimental Economics*, 2023, *26* (2), 412–434.

Bronnenberg, Bart J., Jean-Pierre H. Dubé, and Matthew Gentzkow, "The Evolution of Brand Preferences: Evidence from Consumer Migration," *American Economic Review*, 2012, *102*, 2472—-2508.

Brown, Alexander L, Taisuke Imai, Ferdinand M Vieider, and Colin F Camerer, "Meta-analysis of empirical estimates of loss aversion," *Journal of Economic Literature*, 2024, *62* (2), 485–516.

Bruhin, Adrian, Ernst Fehr, and Daniel Schunk, "The many faces of human sociality: Uncovering the distribution and stability of social preferences," *Journal of the European Economic Association*, 2019, *17* (4), 1025–1069.

_ , _ , and _ , "The many faces of human sociality: Uncovering the distribution and stability of social preferences," *Journal of the European Economic Association*, 2019, *17* (4), 1025–1069.

Cadena, Brian C and Benjamin J Keys, "Human capital and the lifetime costs of impatience," *American Economic Journal: Economic Policy*, 2015, *7* (3), 126–153.

Camerer, Colin and Dean Mobbs, "Differences in behavior and brain activity during hypothetical and real choices," *Trends in Cognitive Sciences*, 2017, *21* (1), 46–56.

Camerer, Colin F and Ernst Fehr, "Measuring social norms and preferences using experimental games: A guide for social scientists," in Joseph et al. Henrich, ed., *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*, Oxford: Oxford University Press, 2004, pp. 55–95.

_ and Robin M Hogarth, "The effects of financial incentives in experiments: A review and capital-labor-production framework," *Journal of Risk and Uncertainty*, 1999, *19*, 7–42.

Cameron, Lisa A., "Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia," *Economic Inquiry*, 1999, *37* (1), 47–59.

Cao, Xinyu and Juanjuan Zhang, "Preference learning and demand forecast," *Marketing Science*, 2021, *40* (1), 62–79.

Cappelen, Alexander W, Benjamin Enke, and Bertil Tungodden, "Universalism: Global Evidence," *American Economic Review*, 2025, *115* (1), 43–76.

_ , James Konow, Erik Ø Sørensen, and Bertil Tungodden, "Just luck: An experimental study of risk-taking and fairness," *American Economic Review*, 2013, *103* (4), 1398–1413.

Card, David E., Stefano DellaVigna, Chenxi Jiang, and Dmitry Taubinsky, "Understanding Expert Choices Using Decision Time," 2024. NBER Working Paper #32,515.

Cárdenas, Juan Camilo, Nicolas De Roux, Christian R Jaramillo, and Luis Roberto Martinez, "Is it my money or not? An experiment on risk aversion and the house-money effect," *Experimental Economics*, 2014, *17*, 47–60.

Carpenter, Jeffrey P and Andrea Robbett, "Measuring Socially Appropriate Social Preferences," *Games and Economic Behavior*, 2024, *147*, 517–532.

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

Cesarini, David, Christopher T. Dawes, Magnus Johannesson, Paul Lichtenstein, and Wallace Bjorn, "Genetic Variation in Preferences for Giving and Risk Taking," *The Quarterly Journal of Economics*, 2009, *124*, 809–842.

Chabris, Christopher F., David Laibson, Carrie L. Morris, Jonathon P. Schuldt, and Dmitry Taubinsky, "Individual laboratory-measured discount rates predict field behavior," *Journal of Risk and Uncertainty*, 2008, *37* (2-3), 237–269.

Chakraborty, Anujit, Evan M Calford, Guidon Fenig, and Yoram Halevy, "External and internal consistency of choices made in convex time budgets," *Experimental Economics*, 2017, *20*, 687–706.

Chapman, Jonathan, Erik Snowberg, Stephanie Wang, and Colin Camerer, "Dynamically Optimized Sequential Experimentation (DOSE) for Estimating Economic Preference Parameters," 2024. NBER Working Paper #33,013.

— , — , — , **and Colin F. Camerer**, "Looming Large or Seeming Small? Attitudes Towards Losses in a Representative Sample," *Review of Economic Studies*, forthcoming. rdae093.

— **, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer**, "Replication Data for: 'Econographics.'," https://doi.org/10.7910/DVN/IGVOFO 2022.

— , — , — , — **, and** — , "Econographics," *Journal of Political Economy Microeconomics*, 2023, *1* (1), 115–161.

— , — , — , — **, and** — , "Willingness to Accept, Willingness to Pay, and Loss Aversion," Technical Report 2023. NBER Working Paper #30,836.

— **, Pietro Ortoleva, Erik Snowberg, Leeat Yariv, and Colin Camerer**, "Reassessing Qualitative Self-Assessments and Experimental Validation," 2025. NBER Working Paper #33,250.

Charness, Gary and Angelino Viceisza, "Three risk-elicitation methods in the field: Evidence from rural Senegal," *Review of Behavioral Economics*, 2016, *3* (2), 145–171.

— **and Matthew Rabin**, "Understanding social preferences with simple tests," *The Quarterly Journal of Economics*, 2002, *117* (3), 817–869.

— **, Uri Gneezy, and Alex Imas**, "Experimental methods: Eliciting risk preferences," *Journal of Economic Behavior & Organization*, 2013, *87*, 43–51.

— , — **, and Brianna Halladay**, "Experimental methods: Pay one or pay all," *Journal of Economic Behavior & Organization*, 2016, *131*, 141–150.

Chaudhuri, Ananish, *Experiments in economics: playing fair with money*, Routledge, 2008.

Chew, Soo Hong, Bin Miao, Qiang Shen, and Songfa Zhong, "Multiple-switching behavior in choice-list elicitation of risk preference," *Journal of Economic Theory*, 2022, *204*, 105510.

Chiong, Khai Xiang, Matthew Shum, Ryan Webb, and Richard Chen, "Combining Choice and Response Time Data: A Drift–Diffusion Model of Mobile Advertisements," *Management Science*, 2023, *70*, 1238–1257.

Choi, Syngjoo, Raymond Fisman, Douglas Gale, and Shachar Kariv, "Consistency and Heterogeneity of Individual Behavior under Uncertainty," *American Economic Review*, 2007, *97* (5), 1921–1938.

Chowdhury, Shyamal, Matthias Sutter, and Klaus F. Zimmermann, "Economic Preferences across Generations and Family Clusters: A Large-Scale Experiment in a Developing Country," *Journal of Political Economy*, 2022, *130*, 2361–2410.

Clithero, John A., "Improving out-of-sample predictions using response times and a model of the decision process," *Journal of Economic Behavior & Organization*, 2018, *148*, 344–375.

Cohen, Jonathan, Keith Marzilli Ericson, David Laibson, and John Myles White, "Measuring time preferences," *Journal of Economic Literature*, 2020, *58* (2), 299–347.

Cooper, David J and John H Kagel, "Other-regarding preferences," *The Handbook of Experimental Economics*, 2016, *2*, 217.

Cortés, Patricia, Jessica Pan, Laura Pilossoph, Ernesto Reuben, and Basit Zafar, "Gender differences in job search and the earnings gap: Evidence from the field and lab," *The Quarterly Journal of Economics*, 2023, *138* (4), 2069–2126.

Cotet, Miruna and Ian Krajbich, "Response Times in the Wild: eBay Sellers Take Hours Longer to Reject High Offers and Accept Low Offers," *SSRN Working Paper 3804578*, 2021.

Croson, Rachel and Uri Gneezy, "Gender differences in preferences," *Journal of Economic Literature*, 2009, *47* (2), 448–474.

Cubitt, Robin P, Chris Starmer, and Robert Sugden, "On the validity of the random lottery incentive system," *Experimental Economics*, 1998, *1*, 115–131.

Cummings, Ronald G and Laura O Taylor, "Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method," *American Economic Review*, 1999, *89* (3), 649–665.

Dana, Jason, Roberto A Weber, and Jason Xi Kuang, "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness," *Economic Theory*, 2007, *33*, 67–80.

Dave, Chetan, Catherine C Eckel, Cathleen A Johnson, and Christian Rojas, "Eliciting risk preferences: When is simple better?," *Journal of Risk and Uncertainty*, 2010, *41*, 219–243.

Dean, Mark and Anja Sautmann, "Credit constraints and the measurement of time preferences," *Review of Economics and Statistics*, 2021, *103* (1), 119–135.

DellaVigna, Stefano, "Structural behavioral economics," in B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *Handbook of Behavioral Economics: Applications and Foundations 1*, Vol. 1, Elsevier, 2018, pp. 613–723.

_ , Devin Pope, and Eva Vivalt, "Predict science to improve science," *Science*, 2019, *366* (6464), 428–429.

den Assem, Martijn J Van, Dennie Van Dolder, and Richard H Thaler, "Split or steal? Cooperative behavior when the stakes are large," *Management Science*, 2012, *58* (1), 2–20.

Dohmen, Thomas and Tomáš Jagelka, "Accounting for Individual-Specific Reliability of Self-Assessed Measures of Economic Preferences and Personality Traits," *Journal of Political Economy: Microeconomics*, 2024, *2* (3), 399–462.

_ , Armin Falk, David Huffman, and Uwe Sunde, "Are risk aversion and impatience related to cognitive ability?," *American Economic Review*, 2010, *100* (3), 1238–1260.

_ , _ , _ , _ , Jürgen Schupp, and Gert G Wagner, "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences," *Journal of the European Economic Association*, 2011, *9* (3), 522–550.

Drake, Marshall, Fernando Payró, Neil Thakral, and Linh T Tô, "Bayesian Adaptive Choice Experiments," 2024. Mimeo.

Dreber, Anna, Tore Ellingsen, Magnus Johannesson, and David G Rand, "Do people care about social context? Framing effects in dictator games," *Experimental Economics*, 2013, *16*, 349–371.

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

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

Engel, Christoph, "Dictator games: A meta study," *Experimental Economics*, 2011, *14*, 583–610.

Enke, Benjamin and Cassidy Shubatt, "Quantifying lottery choice complexity," 2023. NBER Working Paper #31,677.

_ and Thomas Graeber, "Cognitive Uncertainty," *The Quarterly Journal of Economics*, 2023, *138* (4), 2021–2067.

_ , _ , and Ryan Oprea, "Complexity and Hyperbolic Discounting," 2023. NBER Working Paper #31,407.

Epper, Thomas, Ernst Fehr, Helga Fehr-Duda, Claus Thustrup Kreiner,

David Dreyer Lassen, Søren Leth-Petersen, and Gregers Nytoft Rasmussen, "Time discounting and wealth inequality," *American Economic Review*, 2020, *110* (4), 1177–1205.
Ericson, Keith M Marzilli and Andreas Fuster, "Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments," *The Quarterly Journal of Economics*, 2011, *126* (4), 1879–1907.
Etchart-Vincent, Nathalie and Olivier l'Haridon, "Monetary Incentives in the Loss Domain and Behavior Toward Risk: An Experimental Comparison of Three Reward Schemes Including Real Losses," *Journal of Risk and Uncertainty*, 2011, *42* (1), 61–83.
Falk, Armin and Johannes Hermle, "Relationship of Gender Differences in Preferences to Economic Development and Gender Equality," *Science*, 2018, *362* (6412), eaas9899.
_ , Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde, "Global evidence on economic preferences," *Quarterly Journal of Economics*, 2018, *133* (4), 1645–1692.
_ , _ , _ , David Huffman, and Uwe Sunde, "The preference survey module: A validated instrument for measuring risk, time, and social preferences," *Management Science*, 2023, *69* (4), 1935–1950.
Fehr, Ernst and Gary Charness, "Social preferences: fundamental characteristics and economic consequences," *CESifo Working Paper*, 2023.
_ and Klaus M Schmidt, "A theory of fairness, competition, and cooperation," *Quarterly Journal of Economics*, 1999, *114* (3), 817–868.
_ , Helen Bernhard, and Bettina Rockenbach, "Egalitarianism in Young Children," *Nature*, 2008, *454*, 1079–1083.
_ , Thomas Epper, and Julien Senn, "The fundamental properties, stability and predictive power of distributional preferences," *CESifo Working Paper*, 2023.
_ , _ , and _ , "Social Preferences and Redistributive Politics," *Review of Economics and Statistics*, 2024, pp. 1–45.
Fidanoski, Filip and Timothy Johnson, "A z-Tree implementation of the Dynamic Experiments for Estimating Preferences [DEEP] method," *Journal of Behavioral and Experimental Finance*, 2023, *38*, 100805.
Filippin, Antonio and Paolo Crosetto, "A reconsideration of gender differences in risk attitudes," *Management Science*, 2016, *62* (11), 3138–3160.
Finan, Frederico and Laura Schechter, "Vote-buying and reciprocity," *Econometrica*, 2012, *80* (2), 863–881.
Fisher, Geoffrey, "Intertemporal Choices Are Causally Influenced by Fluctuations in Visual Attention," *Management Science*, 2021, *67* (8), 4961–4981.
Fisman, Raymond, Pamela Jakiela, and Shachar Kariv, "Distributional preferences and political behavior," *Journal of Public Economics*, 2017, *155*, 1–10.
_ , Shachar Kariv, and Daniel Markovits, "Individual preferences for giving," *American Economic Review*, 2007, *97* (5), 1858–1876.
Flepp, Raphael and Maximilian Rüdisser, "Revisiting the house money effect in the field: Evidence from casino jackpots," *Economics Letters*, 2019, *181*, 146–148.
Forsythe, Robert, Joel L. Horowitz, N.E. Savin, and Martin Sefton, "Fairness in Simple Bargaining Experiments," *Games and Economic Behavior*, 1994, *6*, 347–369.
Freeman, David J and Guy Mayraz, "Why Choice Lists Increase Risk Taking," *Experimental Economics*, 2019, *22*, 131–154.
Friedman, Daniel, Sameh Habib, Duncan James, and Brett Williams, "Varieties of

risk preference elicitation," *Games and Economic Behavior*, 2022, *133*, 58–76.

Frydman, Cary and Ian Krajbich, "Using Response Times to Infer Others' Private Information: An Application to Information Cascades," *Management Science*, 2022, *68* (4), 2377–3174.

Gao, Xiaoxue Sherry, Glenn W Harrison, and Rusty Tchernis, "Behavioral welfare economics and risk preferences: A Bayesian approach," *Experimental Economics*, 2023, *26* (2), 273–303.

Gaudecker, Hans-Martin Von, Arthur Van Soest, and Erik Wengström, "Heterogeneity in risky choice behavior in a broad population," *American Economic Review*, 2011, *101* (2), 664–694.

Gillen, Ben, Erik Snowberg, and Leeat Yariv, "Experimenting with measurement error: Techniques with applications to the Caltech Cohort Study," *Journal of Political Economy*, 2019, *127* (4), 1826–1863.

Golsteyn, Bart HH, Hans Grönqvist, and Lena Lindahl, "Adolescent time preferences predict lifetime outcomes," *The Economic Journal*, 2014, *124* (580), F739–F761.

Greenberg, Adam Eric, Hal E. Hershfield, Suzanne B. Shu, and Stephen A. Spiller, "What Motivates Social Security Claiming Age Intentions? Testing Behaviorally Informed Interventions Alongside Individual Differences," *Journal of Marketing Research*, 2023, *60* (6), 1052—1070.

Grether, David M and Charles R Plott, "Economic theory of choice and the preference reversal phenomenon," *The American Economic Review*, 1979, *69* (4), 623–638.

Gustman, Alan L. and Thomas L. Steinmeier, "The Social Security Early Entitlement Age in a Structural Model of Retirement and Wealth," *Journal of Public Economics*, 2005, *89* (2–3), 441—463.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze, "An experimental analysis of ultimatum bargaining," *Journal of Economic Behavior & Organization*, 1982, *3* (4), 367–388.

Hackethal, Andreas, Michael Kirchler, Christine Laudenbach, Michael Razen, and Annika Weber, "On the role of monetary incentives in risk preference elicitation experiments," *Journal of Risk and Uncertainty*, 2023, *66* (2), 189–213.

Hanaoka, Chie, Hitoshi Shigeoka, and Yasutora Watanabe, "Do Risk Preferences Change? Evidence from the Great East Japan Earthquake," *American Economic Journal: Applied Economics*, 2018, *10* (2), 298–330.

Hanushek, Eric A, Lavinia Kinne, Philipp Lergetporer, and Ludger Woessmann, "Patience, Risk-Taking, and Human Capital Investment Across Countries," *The Economic Journal*, 2022, *132* (646), 2290–2307.

Harmon, Daniel A, Amie L Haas, and Alex Peterkin, "Experimental tasks of behavioral risk taking in alcohol administration studies: A systematic review," *Addictive Behaviors*, 2021, *113*, 106678.

Harrison, Glenn W, "Maximum likelihood estimation of utility functions using Stata," *University of Central Florida, Working Paper*, 2008, pp. 06–12.

_ **and Elisabet Rutström**, "Risk Aversion in the Laboratory," in J.C. Cox and G.W. Harrison, eds., *Risk Aversion in Experiments (Research in Experimental Economics, Vol. 12)*, Emerald Group Publishing Limitedm Leeds, 2008, pp. 41–196.

_ **, Morten I Lau, and Elisabet Rutström**, "Identifying Time Preferences with Experiments: Comment," 2013.

Hausman, Jerry, "Contingent valuation: from dubious to hopeless," *Journal of Economic*

Perspectives, 2012, *26* (4), 43–56.

Hey, John D and Jinkwon Lee, "Do Subjects Separate (Or Are They Sophisticated)?," *Experimental Economics*, 2005, *8*, 233–265.

Hoffman, Elizabeth, Kevin McCabe, and Vernon L. Smith, "Social Distance and Other-Regarding Behavior in Dictator Games," *American Economic Review*, 1996, *86* (3), 653–660.

Holmén, Martin, Felix Holzmeister, Michael Kirchler, Matthias Stefan, and Erik Wengström, "Economic preferences and personality traits among finance professionals and the general population," *Economic Journal*, 2023, *133* (656), 2949–2977.

Holt, Charles A, "Preference Reversals and the Independence Axiom," *American Economic Review*, 1986, *76* (3), 508–515.

_ and Susan K Laury, "Risk aversion and incentive effects," *American Economic Review*, 2002, *92* (5), 1644–1655.

Holzmeister, Felix and Matthias Stefan, "The risk elicitation puzzle revisited: Across-methods (in) consistency?," *Experimental Economics*, 2021, *24*, 593–616.

Huber, Joel, Dan Ariely, and Gregory Fischer, "Expressing Preferences in a Principal-Agent Task: A Comparison of Choice, Rating, and Matching," *Organizational Behavior and Human Decision Processes*, 2002, *87* (1), 66–90.

Hunt, Melissa K, Derek R Hopko, Robert Bare, CW Lejuez, and EV Robinson, "Construct validity of the balloon analog risk task (BART) associations with psychopathy and impulsivity," *Assessment*, 2005, *12* (4), 416–428.

Imai, Taisuke and Colin F Camerer, "Estimating time preferences from budget set choices using optimal adaptive design," *Unpublished manuscript*, 2018.

_ and _ , "A large-scale, interdisciplinary meta-analysis on behavioral economics parameters," *The Economic Journal*, 2019, *131* (636), 1788–1814.

_ , Tom A Rutter, and Colin F Camerer, "Meta-Analysis of Present-Bias Estimation Using Convex Time Budgets," *Economic Journal*, 2021, *131* (636), 1788–1814.

Jagelka, Tomáš, "Are economists' preferences psychologists' personality traits? A structural approach," *Journal of Political Economy*, 2024, *132* (3), 910–970.

Johnson, Cathleen, Aurélien Baillon, Han Bleichrodt, Zhihua Li, Dennie Van Dolder, and Peter P Wakker, "Prince: An improved method for measuring incentivized preferences," *Journal of Risk and Uncertainty*, 2021, *62* (1), 1–28.

Ju, Uijong and Christian Wallraven, "Relationship between virtual reality balloon analogue risk task and risky decision-making," *Plos one*, 2023, *18* (2), e0282097.

Kahneman, Daniel and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 1979, *47* (2), 363–391.

_ , Jack L Knetsch, and Richard H Thaler, "Experimental Tests of the Endowment Effect and the Coase theorem," *Journal of Political Economy*, 1990, *98* (6), 1325–1348.

Kay, Aaron C and Lee Ross, "The perceptual push: The interplay of implicit cues and explicit situational construals on behavioral intentions in the Prisoner's Dilemma," *Journal of Experimental Social Psychology*, 2003, *39* (6), 634–643.

Kerschbamer, Rudolf and Daniel Müller, "Social preferences and political attitudes: An online experiment on a large heterogeneous sample," *Journal of Public Economics*, 2020, *182*, 104076.

Kessler, Judd B. and Stephan Meier, "Learning from (failed) replications: Cognitive load manipulations and charitable giving," *Journal of Economic Behavior & Organization*, 2014, *102*, 10–13.

Khaw, Mel Win, Ziang Li, and Michael Woodford, "Cognitive imprecision and small-stakes risk aversion," *The Review of Economic Studies*, 2021, *88* (4), 1979–2013.

Kirby, K. N. and N. N. Marakovic, "Delay-discounting probabilistic rewards: Rates decrease as amounts increase," *Psychonomic Bulletin & Review*, 1996, *3* (1), 100–104.

Kirby, Kris N., "One-year temporal stability of delay-discount rates," *Psychonomic Bulletin & Review*, 2009, *16* (3), 457–462.

Kittel, Bernhard, Georg Kanitsar, and Stefan Traub, "Knowledge, power, and self-interest," *Journal of Public Economics*, 2017, *150*, 39–52.

Konovalov, Arkady and Ian Krajbich, "Revealed strength of preference: Inference from response times," *Judgment and Decision Making*, 2019, *14* (4), 381–394.

Kosfeld, Michael and Zahra Sharafi, "The Preference Survey Module: evidence on social preferences from Tehran," *Journal of the Economic Science Association*, 2024, *10* (1), 152–164.

Kosse, Fabian, Thomas Deckers, Pia Pinger, Hannah Schildberg-Hörisch, and Armin Falk, "The Formation of Prosociality: Causal Evidence on the Role of the Social Environment," *Journal of Political Economy*, 2020, *128*, 434–467.

Kőszegi, Botond and Matthew Rabin, "A Model of Reference-Dependent Preferences," *The Quarterly Journal of Economics*, 2006, *121* (4), 1133–1165.

_ and _ , "Reference-dependent Risk Attitudes," *American Economic Review*, 2007, *97* (4), 1047–1073.

Krawczyk, Michal and Fabrice Le Lec, "How to elicit distributional preferences: A stress-test of the equality equivalence test," *Journal of Economic Behavior & Organization*, 2021, *182*, 13–28.

Krosnick, Jon A. and Stanley Presser, "Question and Questionnaire Design," in Peter V Marsden and James D Wright, eds., *Handbook of Survey Research*, 2010, pp. 263–313.

Krupka, Erin and Roberto A Weber, "The focusing and informational effects of norms on pro-social behavior," *Journal of Economic Psychology*, 2009, *30* (3), 307–320.

Kulis, Brian and Michael I Jordan, "Revisiting k-means: new algorithms via Bayesian nonparametrics," in "Proceedings of the 29th International Coference on International Conference on Machine Learning" 2012, pp. 1131–1138.

Leboeuf, Robyn A., "Discount Rates for Time versus Dates: The Sensitivity of Discounting to Time-Interval Description," *Journal of Marketing Research*, 2006, *43* (1), 59–72.

Lejuez, Carl W, Jennifer P Read, Christopher W Kahler, Jerry B Richards, Susan E Ramsey, Gregory L Stuart, David R Strong, and Richard A Brown, "Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART).," *Journal of Experimental Psychology: Applied*, 2002, *8* (2), 75.

Lejuez, CW, Will M Aklin, Heather A Jones, Jerry B Richards, David R Strong, Christopher W Kahler, and Jennifer P Read, "The Balloon Analogue Risk Task (BART) differentiates smokers and nonsmokers.," *Experimental and Clinical Psychopharmacology*, 2003, *11* (1), 26.

Li, Ye, Antonia Krefeld-Schwalb, Daniel G. Wall, Eric J. Johnson, Olivier Toubia, and Daniel M. Bartels, "The More You Ask, the Less You Get: When Additional Questions Hurt External Validity," *Journal of Marketing Research*, 2022, *59* (5), 963–982.

Liberman, Varda, Steven M Samuels, and Lee Ross, "The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves," *Personality and social psychology bulletin*, 2004, *30* (9), 1175–1185.

Lichtenstein, Sarah and Paul Slovic, "Reversals of preference between bids and choices in gambling decisions.," *Journal of Experimental Psychology*, 1971, *89* (1), 46.

List, John A, "Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auctions for sportscards," *American Economic Review*, 2001, *91* (5), 1498–1507.

List, John and Todd Cherry, "Learning to Accept in Ultimatum Games: Evidence from an Experimental Design that Generates Low Offers," *Experimental Economics*, 2000, *3* (1), 11–29.

Luce, R Duncan, *Individual Choice Behavior: A Theoretic Analysis*, Vol. 4, Wiley New York, 1959.

Maniadis, Zacharias, Fabio Tufano, and John A List, "One swallow doesn't make a summer: New evidence on anchoring effects," *American Economic Review*, 2014, *104* (1), 277–290.

Mata, Rui, Renato Frey, David Richter, Jürgen Schupp, and Ralph Hertwig, "Risk preference: A view from psychology," *Journal of Economic Perspectives*, 2018, *32* (2), 155–172.

McFadden, D., "Conditional Logit Analysis of Qualitative Choice Behaviour," in P. Zarembka, ed., *Frontiers in Econometrics*, New York, NY, USA: Academic Press New York, 1973, pp. 105–142.

McFadden, Daniel, "Economic Choices," *American Economic Review*, 2001, *91* (3), 351–378.

Meier, Stephen and Charles D. Sprenger, "Temporal Stability of Time Preferences," *The Review of Economics and Statistics*, 2015, *97* (2), 273–286.

Mischel, W. and E.B. Ebbesen, "Attention in delay of gratification," *Journal of Personality and Social Psychology*, 1970, *16* (2), 329–337.

Morag, Dor and George F. Loewenstein, "Narratives and Valuations," *Available at SSRN: https://ssrn.com/abstract=3919471 or http://dx.doi.org/10.2139/ssrn.3919471*, 2023.

Nielsen, Kirby and John Rehbeck, "When choices are mistakes," *American Economic Review*, 2022, *112* (7), 2237–2268.

O'Donoghue, Ted and Charles Sprenger, "Reference-Dependent Preferences," in Bernheim B. Douglas, DellaVigna Stefano, and Laibson David, eds., *Handbook of Behavioral Economics: Foundations and Applications*, Vol. 1, Elsevier, 2018.

Oprea, Ryan, "What makes a rule complex?," *American Economic Review*, 2020, *110* (12), 3913–3951.

_ , "Decisions Under Risk are Decisions Under Complexity," *American Economic Review*, 2024, *114* (12), 3789–3811.

Payne, John W., James R. Bettman, and Eric J. Johnson, *The adaptive decision maker*, New York: Cambridge University Press, 1993.

Payzan-LeNestour, Elise and James Doran, "Craving money? Evidence from the laboratory and the field," *Science Advances*, 2024, *10* (2), eadi5034.

Perez, Fabien, Guillaume Hollard, and Radu Vranceanu, "How serious is the measurement-error problem in risk-aversion tasks?," *Journal of Risk and Uncertainty*, 2021, *63* (3), 319–342.

Peterson, Joshua C, David D Bourgin, Mayank Agrawal, Daniel Reichman, and Thomas L Griffiths, "Using large-scale experiments and machine learning to discover theories of human decision-making," *Science*, 2021, *372* (6547), 1209–1214.

Peysakhovich, Alexander and Jeffrey Naecker, "Using methods from machine learning to evaluate behavioral models of choice under risk and ambiguity," *Journal of Economic Behavior & Organization*, 2017, *133*, 373–384.

Platt, Michael L. and Scott A. Huettel, "Risky Business: The Neuroeconomics of Decision Making Under Uncertainty," *Nature Neuroscience*, 2008, *11*, 398–403.

Plott, Charles R and Kathryn Zeiler, "The willingness to pay–willingness to accept gap, the "endowment effect," subject misconceptions, and experimental procedures for eliciting valuations," *American Economic Review*, 2005, *95* (3), 530–545.

Preuss, Malte, "Intra-individual stability of two survey measures on forward-looking attitude," *Journal of Economic Behavior & Organization*, 2021, *190*, 201–227.

Quidt, Jonathan De, Lise Vesterlund, and Alistair J Wilson, "Experimenter demand effects," in "Handbook of research methods and applications in experimental economics," Edward Elgar Publishing, 2019, pp. 384–400.

Ranganathan, Kavitha and Tomás Lejarraga, "Elicitation of risk preferences through satisficing," *Journal of Behavioral and Experimental Finance*, 2021, *32*, 100570.

Rao, Hengyi, Marc Korczykowski, John Pluta, Angela Hoang, and John A Detre, "Neural correlates of voluntary and involuntary risk taking in the human brain: an fMRI Study of the Balloon Analog Risk Task (BART)," *Neuroimage*, 2008, *42* (2), 902–910.

Ray, Debajyoti, Daniel Golovin, Andreas Krause, and Colin Camerer, "Bayesian Rapid Optimal Adaptive Design (BROAD): Method and Application Distinguishing Models of Risky Choice," 2012. Caltech, *mimeo.*

Read, Daniel, Shane Frederick, Burcu Orsel, and Juwaria Rahman, "Four Score and Seven Years from Now: The Date/Delay Effect in Temporal Discounting," *Management Science*, 2005, *51* (9), 1326–1335.

Rieger, Marc Oliver, Mei Wang, and Thorsten Hens, "Estimating cumulative prospect theory parameters from an international survey," *Theory and Decision*, 2017, *82*, 567–596.

Roth, Alvin E, "The Early History of Experimental Economics," *Journal of the History of Economic Thought*, 1993, *15* (2), 184–209.

Rustagi, Devesh, Stefanie Engel, and Michael Kosfeld, "Conditional cooperation and costly monitoring explain success in forest commons management," *Science*, 2010, *330* (6006), 961–965.

Ryan, Elizabeth G. Christopher C. Drovandi, James M. McGree, and Anthony N. Pettitt, "A Review of Modern Computational Algorithms for Bayesian Optimal Design," *International Statistical Review*, 2016, *84* (1), 128–154.

Schneider, Florian, Fanny Brun, and Roberto A Weber, "Sorting and wage premiums in immoral work," *Review of Economics and Statistics*, forthcoming. rest_a_01536.

Schotter, Andrew and Isabel Trevino, "Is Response Time Predictive of Choice? An Experimental Study of Threshold Strategies," *Experimental Economics*, 2021, *24*, 87—-117.

Seidl, Christian, "Preference reversal," *Journal of Economic Surveys*, 2002, *16* (5), 621–655.

Shu, Suzanne B. and John W. Payne, "Life Expectation Judgments, Fairness, and Loss Aversion in the Psychology of Social Security Claiming Decision," *Working Paper RRC NB13-05, National Bureau for Economic Research*, 2016.

Slovic, Paul, "Risk-taking in children: Age and sex differences," *Child Development*, 1966, pp. 169–176.

_ , "The Construction of Preference," *American Psychologist*, 1995, *50*, 364–371.

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

Sokol-Hessner, Peter, Ming Hsu, Nina G. Curley, Mauricio R. Delgado, Colin F Camerer, and Elizabeth A. Phelps, "Thinking like a trader selectively reduces individuals' loss aversion," *PNAS*, 2009, *106* (13), 5035–5040.

Sprenger, Charles, "An endowment effect for risk: Experimental tests of stochastic reference points," *Journal of Political Economy*, 2015, *123* (6), 1456–1499.

Stango, Victor and Jonathan Zinman, "Behavioral biases are temporally stable," Technical Report, National Bureau of Economic Research 2020.

_ **and** _ , "We Are All Behavioural, More, or Less: A Taxonomy of Consumer Decision-Making," *The Review of Economic Studies*, 2023, *90* (3), 1470–1498.

Stantcheva, Stefanie, "How to run surveys: A guide to creating your own identifying variation and revealing the invisible," *Annual Review of Economics*, 2023, *15*, 205–234.

Sunde, Uwe, Thomas Dohmen, Benjamin Enke, Armin Falk, David Huffman, and Gerrit Meyerheim, "Patience and comparative development," *The Review of Economic Studies*, 2022, *89* (5), 2806–2840.

Tanaka, Tomomi, Colin F Camerer, and Quang Nguyen, "Risk and time preferences: Linking experimental and household survey data from Vietnam," *American Economic Review*, 2010, *100* (1), 557–571.

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

Tertilt, Michael and Peter Scholz, "To advise, or not to advise—how robo-advisors evaluate the risk preferences of private investors," *The Journal of Wealth Management*, 2018, *21* (2), 70–84.

Thaler, Richard, "Some empirical evidence on dynamic inconsistency," *Economics Letters*, 1981, *8* (3), 201–207.

Thaler, Richard H and Eric J Johnson, "Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice," *Management Science*, 1990, *36* (6), 643–660.

Thurstone, Louis L, "The Indifference Function," *The Journal of Social Psychology*, 1931, *2* (2), 139–167.

Toubia, Olivier, Eric Johnson, Theodoros Evgeniou, and Philippe Delquié, "Dynamic Experiments for Estimating Preferences: An Adaptive Method of Eliciting Time and Risk Parameters," *Management Science*, 2013, *59* (3), 613–640.

Trautmann, Stefan T and Gijs van de Kuilen, "Higher order risk attitudes: A review of experimental evidence," *European Economic Review*, 2018, *103*, 108–124.

van Dolder, Dennie and Jürgen Vandenbroucke, "Behavioral Risk Profiling: Measuring Loss Aversion of Individual Investors," *Available at SSRN 4199169*, 2022.

Vieider, Ferdinand M, Mathieu Lefebvre, Ranoua Bouchouicha, Thorsten Chmura, Rustamdjan Hakimov, Michal Krawczyk, and Peter Martinsson, "Common Components of Risk and Uncertainty Attitudes across Contexts and Domains: Evidence from 30 Countries," *Journal of the European Economic Association*, 2015, *13* (3), 421–452.

Wallis, W Allen and Milton Friedman, "The Empirical Derivation of Indifference Functions," 1942.

Wang, Joseph Tao-Yi, Michael Spezio, and Colin F. Camerer, "Pinocchio's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth Telling and Deception in Sender-Receiver Games," *American Economic Review*, 2010, *100*, 984—1007.

Wang, Xinghua and Daniel Navarro-Martinez, "Increasing the external validity of social

preference games by reducing measurement error," *Games and Economic Behavior*, 2023, *141*, 261–285.
Weber, Elke U, Ann-Renee Blais, and Nancy E Betz, "A Domain-Specific Risk-Attitude Scale: Measuring Risk Perceptions and Risk Behaviors," *Journal of Behavioral Decision Making*, 2002, *15* (4), 263–290.
Yang, Liu, Olivier Toubia, and Martijn G de Jong, "Attention, information processing, and choice in incentive-aligned choice experiments," *Journal of Marketing Research*, 2018, *55* (6), 783–800.
Yu, Chi Wai, Y Jane Zhang, and Sharon Xuejing Zuo, "Multiple switching and data quality in the multiple price list," *Review of Economics and Statistics*, 2021, *103* (1), 136–150.
Zhou, Wenting and John Hey, "Context Matters," *Experimental Economics*, 2018, *21*, 723–756.