Remote Experiments in Trust and Time

by

Connor T. Wiegand

A dissertation accepted and approved in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in Economics

Dissertation Committee:
Michael A. Kuhn, Chair
Jiabin Wu, Core Member
Van Kolpin, Core Member
John Clithero, Institutional Representative

University of Oregon
Winter 2026

DISSERTATION ABSTRACT

Connor T. Wiegand
Doctor of Philosophy in Economics
Title: Remote Experiments in Trust and Time

This dissertation studies forward-looking economic behavior in remote experimental settings, focusing on time preferences, measurement design, and trust in technologically mediated interaction. As experimental research increasingly relies on remote platforms, careful implementation is essential for credible inference about how individuals evaluate future outcomes.

The first chapter examines the stability of time preferences using longitudinal experimental data. Repeated measurement reveals systematic within-individual variation in discounting behavior, suggesting that intertemporal preferences may fluctuate over relatively short horizons and that single-elicitation estimates may obscure temporal instability. The second chapter develops and evaluates a modified Multiple Lottery List (MLL) designed for remote, repeated use. The instrument reduces compound risk, eliminates multiple switching through an intuitive interface, and streamlines dense price lists to improve precision while minimizing participant burden. The design yields stable and internally consistent estimates and is well suited for multi-wave studies.

The final chapter investigates how artificial intelligence systems influence communication and trust in an experimental exchange setting. AI-assisted messaging alters cooperative outcomes by shaping the content and credibility of communication. Together, the chapters combine substantive evidence on time preferences and trust with methodological advances in remote experimental design, demonstrating how carefully structured remote experiments can deepen our understanding of forward-looking economic behavior.

This dissertation includes previously unpublished co-authored material.

ACKNOWLEDGMENTS

I am deeply grateful to the many people who made this dissertation possible.

I owe my greatest professional debt to Mike Kuhn. His intellectual generosity and steady guidance have shaped both this project and my development as a researcher. Mike has consistently pushed me to refine arguments, to follow through on ideas, and to stay honest about what the data can and cannot say. He has been patient and supportive through many iterations of this work, while keeping the process grounded, collaborative, and approachable. Mike has always been generous with his time, both as an advisor and as a trusted mentor, across research, teaching, and life's inevitable challenges.

I am also profoundly thankful to Jiabin Wu. His collaboration, thoughtful feedback, and willingness to engage deeply with ideas at every stage of development have been invaluable. Jiabin approaches challenges with professionalism and good humor, while maintaining a high level of rigor to which I aspire. I am grateful to Van Kolpin for his leadership and for consistently taking the time to answer my questions. His guidance and example have meant a great deal throughout my graduate training.

I thank my coauthors and collaborators for their partnership and insight, particularly on the projects that form the core chapters of this dissertation. Their contributions materially improved the quality and scope of this work. I am grateful to my committee members and the faculty of the University of Oregon for their mentorship, feedback, and encouragement. I also thank John Clithero, Jon Davis, and Keaton Miller for their constructive comments that helped sharpen many of the ideas presented in this dissertation.

Thank you to my parents for always believing in me and supporting my academic journey. They did an incredible job instilling a deep love and respect for education that I hope to pass on one day. Finally, I owe my deepest thanks to my wife, Madison—for her unwavering support, encouragement, patience, and constant presence throughout this process. My partner in every sense of the word, who has carried me through this journey in ways I cannot fully express—your belief in me made this work possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

Economic decision-making is fundamentally forward-looking. Individuals evaluate tradeoffs across time, form expectations about the behavior of others, and act under uncertainty about future outcomes. Understanding how these forward-looking judgments are formed, measured, and influenced is central to both economic theory and experimental practice. The settings in which these decisions occur continue to evolve toward environments that are digitally mediated and remotely accessed. Remote experimental platforms and artificial intelligence systems now mediate many economic interactions, creating new opportunities and new methodological challenges for researchers.

This dissertation studies forward-looking behavior in remote experimental settings from three complementary perspectives. First, it examines whether time preferences are stable when measured repeatedly over time. Second, it develops and evaluates an experimental instrument designed to measure those preferences efficiently and reliably in longitudinal remote environments. Third, it investigates how artificial intelligence systems influence trust and cooperative behavior in strategic interaction.

The first chapter documents systematic time variation in measured discounting behavior using panel data collected over multiple weeks. By observing the same individuals repeatedly, the study provides evidence that intertemporal preferences may shift over relatively short horizons. These findings contribute to ongoing debates about the stability of discounting parameters and the interpretation of present bias, highlighting the importance of repeated measurement in distinguishing persistent heterogeneity from temporal fluctuation.

The second chapter addresses the methodological demands created by such longitudinal designs. Reliable inference about time preference dynamics requires an elicitation tool that is theoretically coherent, resistant to common behavioral inconsistencies, and fast enough to administer repeatedly without inducing participant fatigue. I present and evaluate a modified implementation of the Multiple Lottery List (MLL) tailored specifically to remote, repeated measurement. The design reduces compound risk, eliminates multiple switching through an intuitive interface, and automates dense price lists to improve precision. Empirically, the instrument produces stable and internally consistent estimates while remaining scalable for multi-wave studies.

The final chapter extends the focus from intrapersonal tradeoffs over time to interpersonal strategic interaction. Trust is inherently forward-looking: cooperation depends on beliefs about others' future behavior and the credibility of commitments. Using a remote experimental design, the chapter studies how large language models influence communication and trust in an economic exchange setting. The results provide evidence on how artificial intelligence can shape cooperative outcomes by altering message formation and expectations.

Taken together, the chapters contribute to the study of forward-looking economic behavior by combining substantive analysis with experimental infrastructure. They show how remote platforms can be used not only to measure intertemporal preferences with high frequency and precision, but also to study trust in technologically mediated environments. In doing so, this dissertation advances both our understanding of time and trust and the methodological tools available for studying them in modern experimental contexts.

Chapters II and IV contain co-authored material which is yet to be published at the time of writing. Chapter II was co-authored with Michael A. Kuhn and M. Steven Holloway. Chapter IV was co-authored with Ethan Holdahl, Jiabin Wu, and Tanner Bivins.

II. TIME VARYING TIME PREFERENCES

This chapter is based on a project co-authored with Michael A. Kuhn ("Mike") and M. Steven Holloway ("Marcus"). The conception of the project was co-developed by Marcus and I, with Mike providing principal guidance regarding practical implementation, as well as funding. Development, recruitment, and implementation of the project was a shared collaborative effort. Marcus designed much of the survey logic as well as the distribution of weekly surveys, Mike facilitated payments and was the primary point of contact for the study, and I was responsible for the design and programming of the experiment interface. I developed most of the underlying theory, accompanying analysis, and logical proofs in the paper. Much of the econometric analysis and presentation of results was developed by Marcus and Mike. Mike wrote the first draft of the paper, with subsequent drafts being co-developed by all three authors. As of the time of dissertation writing, the principal research paper from this project has been submitted for publication at a peer-reviewed journal.

II.1. Introduction

A substantial body of evidence documents that economic preferences can be unstable. Beyond the large and still growing literature showing causal effects of shocks on preferences, decision-makers may learn about their preferences over time (either through feedback (Charness *et al.*, 2023), the realization of uncertainty (Carrera *et al.*, 2022), or deliberation (Gabaix and Laibson, 2022)), as they mature (Andreoni *et al.*, 2019), or "choose" preferences to better suit their environment (Bernheim *et al.*, 2021).

Preference instability, shocks, and drift are problematic for researchers looking to estimate time preferences. Consider the framework of Read and Van Leeuwen (1998) where healthy eating is taken as having an immediate opportunity cost in exchange for future health benefits: in advance of consumption an individual selects a healthy snack for their future self, but revises their choice to an unhealthy snack when the time for consumption arrives. Is this preference reversal best explained by a fixed feature of that individual's discount function, learning about preferences, or some other unobserved environmental factor?

Halevy (2015) describes three properties of time preferences: consistency, stationarity, and time-invariance. The healthy-to-unhealthy snack preference reversal example above shows inconsistent behavior: holding fixed the *calendar dates* of the consequences of an intertemporal choice, time-inconsistency is when the date the choice is made affects the outcome. This is the key behavior that has motivated much of the past 25 years of literature in behavioral economics surrounding why people don't stick to their plans. However, directly observing inconsistency requires a longitudinal setting, which is often difficult to engineer. Thus, many researchers looking to measure inconsistency instead measure the closely related property of stationarity. Sticking with the snack example, simultaneously choosing an unhealthy snack for now and a healthy snack for the future is an example of non-stationary time preference: holding fixed the date a choice is made, non-stationarity happens when the *relative dates* of the choice consequences affect the outcome.

Non-stationarity is an important behavior to study in its own right, but under what conditions does it also inform researchers about inconsistency? Halevy (2015) shows that they are equivalent only when time-invariance also holds. Again using the snack example, choosing an unhealthy snack now for now, and then choosing a healthy snack tomorrow for tomorrow is an example of time-varying time preference. Essentially, non-stationarity and inconsistency are equivalent if and only if individuals exhibit stable choices in identical intertemporal choice situations over time.

Any two of these properties imply the third; if a behavioral economist theorizes that a decision-maker is time-inconsistent, she necessarily assumes that the decision-maker is either non-stationary or time-variant (or both). Common models of discounting in behavioral economics pair inconsistency with non-stationarity, while leaving time-invariance intact. However, data show this is only one of many choice patterns that people exhibit. In Halevy (2015), at most 68%, and as few as 44% of subjects exhibited either all three properties or only time-invariance. Thus, between 32% and 56% of the sample exhibited preferences that cannot be explained with current models. In a closely-related paper, Janssens *et al.* (2017) finds that while 43% of subjects in their study violate consistency, only 24% violate *both* consistency and stationarity. Similarly, Harrison *et al.* (2025) finds that time-variance explains why such a large fraction of their experimental sample is inconsistent even though they display stationary behavior in their first of their two surveys. Time-variance is an important reason why subjects in all three studies exhibit inconsistent behavior.

In this paper we present both a twelve-week, seven-survey, longitudinal experiment that offers much higher resolution on time-variance than past work, and a structural model that flexibly accounts for time-variance in intertemporal choices. While the two-period studies in Halevy (2015), Janssens *et al.* (2017), and Harrison *et al.* (2025) allow for an impact of time-variance on experimental measures, they do not give the researcher much power to consider the nature of time-variance and model it. Is time-variance in the canonical experimental context for intertemporal choices noise or systematic? Recent work in behavioral economics establishes numerous potential pathways for systematic change in time preference: a likely non-exhaustive list includes preference uncertainty (Carrera *et al.*, 2022), deliberation (Gabaix and Laibson, 2022) and cooling off (Imas *et al.*, 2022), aging (Andreoni *et al.*, 2019), adaptation (Bernheim *et al.*, 2021), reductions in cognitive uncertainty (Enke *et al.*, 2025), and correlated economic factors (Carvalho *et al.*, 2016).

Answering this question determines the scope of the time-variance problem. If time-variance is idiosyncratic noise, then what appears to be time-variance or inconsistency may instead arise from decision error. Failing to account for this would lead to individuals being incorrectly classified as certain types of discounters, but would not impact aggregate estimates of how a population discounts (besides decreasing precision). On the other hand, if time-variance is a systematic process, the problem is more pernicious.

What is the risk associated with mistaking inconsistency that corresponds to a stationarity violation for inconsistency that corresponds to time-variance? The behavioral literature on

"present-bias" (Laibson, 1997; O'Donoghue and Rabin, 1999) and associated policy interventions (e.g. commitment devices) are predicated on inconsistency as a harmful effect of the opportunity for immediate gratification. If choices can be structured to avoid that opportunity, then the decision-maker can implement their stable long-run preference for saving more, eating healthier, etc. Inconsistent choices associated with time-variance do not have the same interpretation, and such interventions may decrease welfare. This risk can be mitigated simply by measuring time-variance.

Take the following example: consider a decision-maker that exhibits time-varying preferences such that his intertemporal allocations in identical choice scenarios are getting less patient as time passes. First, assume his preferences are stationary. If a researcher estimates his quasi-hyperbolic ($\beta - \delta$) discount function via an observed violation of consistency, she will conclude that he is present-biased ($\beta < 1$). Now instead, assume his preferences are consistent. If his quasi-hyperbolic discount function is estimated via an observed violation of stationarity, the researcher will conclude that he is future-biased ($\beta > 1$) instead. If the researcher also measures the violation of time-invariance, she will learn in the first (second) case that the decision-makers degree of inconsistency (non-stationarity) is exactly equal to their degree of time-variance. Suitably adjusted for the observed time-variance, $\beta = 1$ in both cases.

Section II.2. gives additional background on the three properties of time preferences, as illustrated by Halevy (2015), and why time-variance is a problem for identification in time-invariant models. Sections II.3.–II.5. present our longitudinal experiment. Over the course of 12 weeks subjects take part in seven surveys each featuring 6–10 incentivized intertemporal choices. For every choice, this design gives us six opportunities to observe inconsistency, six opportunities to observe non-stationarity, and twenty one opportunities to observe time-variance.

We first look at our experimental data through the lens of traditional time-invariant discounting, and estimate parameters for the quasi-hyperbolic $\beta - \delta$ discount function. We identify small but statistically significant present bias at the aggregate level, and the majority of subjects are classified as present-biased at the individual level.

We then shift to analyses that allow for time-varying preferences. We identify a significant time trend in the aggregate discount factor consistent with *decreasing patience* (increasing discounting) over the period of the study. This trend stems from subject-level changes across time, not sample composition, suggesting systematic rather than idiosyncratic time-variance. With time-variance accounted for, we can measure distinct inconsistent and non-stationary behavior at the subject-level; we find that "inconsistency present-bias" is larger and more widespread than

"non-stationary present bias." The prevalence of individual violations of all three time preference properties leads us to classify subjects as time-variant at the upper end of the range Halevy (2015) suggests. Overall, time-variance explains 72% of the inconsistency in our sample.

Our interpretation of decreasing patience during our study is not that it represents a general trend in intertemporal choice behavior over time that would extrapolate to substantially longer time horizons. Instead, we think it more likely that we are recovering either local drift in preferences specific to our population and time period, or an artifact of the repeated experimental choices themselves. In this light, we view our empirical contribution to the literature as demonstrating systematic time-variance *in a canonical experimental context*. Our choice intervals of every two weeks over a 12-week period were designed to inform this context.

In Section II.6. we discuss how discount functions reflect the time preference properties, and present a new model – the "nested exponential" discount function – that is general with respect to the properties defined by Halevy (2015). Nested Exponential Discounting (NED) can satisfy any of the five possible combinations of properties depending on its parameter values. While this four-parameter model is not as tractable as the workhorse $\beta - \delta$ model, we develop the intuition of the model as a discrete extension of fast Weibull discounting (Jamison and Jamison, 2011). Thus, the model retains a "present-bias" parameter similar to existing models.

In Section II.7. we structurally estimate the NED function at both the aggregate and individual levels. We directly test the estimated parameters for adherence with time-invariance, consistency, and stationarity. The unrestricted model fits the data best, even when penalized for its extra parameters. Overall, we assign 63% of the sample to time-varying models of discounting, suggesting that understanding, measuring, and modeling preference shifts over time is crucial for accurately estimating time preferences.

II.2. Background

We adopt notation from Halevy (2015) to describe the stationarity, consistency, and time-invariance properties of time preferences.

Consider an agent in calendar time period $\tau \geq 0$ evaluating bundles (ε, t) and (ψ, t'), which deliver rewards ε, ψ at time periods $t' \geq t \geq \tau$. We assume that the agent has complete and transitive preferences over all such bundles, according to a preference ordering that may depend on τ, $\succsim_\tau$. Under classical consumer theory assumptions, including additive separability

of utility across time periods[1], this preference relationship is represented by a *discount function*, $D(\tau, t)$, such that

$$(\varepsilon, t) \succsim_\tau (\psi, t') \iff D(\tau, t) \cdot \varepsilon \geq D(\tau, t') \cdot \psi$$

$D(\tau, t)$ is the agent's discount factor when evaluating, in period τ, a payoff in period t.

Definition 1. *An agent's preferences satisfy* ***stationarity*** *if*

$$D(\tau, t) \cdot \varepsilon \geq D(\tau, t') \cdot \psi \iff D(\tau, t + \Delta) \cdot \varepsilon \geq D(\tau, t' + \Delta) \cdot \psi \quad \text{for} \quad \Delta \geq 0$$

An agent whose preferences are *time stationary* will not exhibit preference reversals when the payoff time of all bundles is shifted by the same number of periods (Δ), all else equal.

Definition 2. *An agent's preferences satisfy* ***consistency*** *if*

$$D(\tau, t) \cdot \varepsilon \geq D(\tau, t') \cdot \psi \iff D(\tau + \Delta, t) \cdot \varepsilon \geq D(\tau + \Delta, t') \cdot \psi \quad \text{for} \quad 0 \leq \Delta \leq t - \tau$$

Agents with *time consistent* preferences will not exhibit preference reversals when the evaluation period of all bundles is shifted by the same number of periods (Δ), all else equal.

Definition 3. *An agent's preferences satisfy* ***time invariance*** *if*

$$D(\tau, t) \cdot \varepsilon \geq D(\tau, t') \cdot \psi \iff D(\tau + \Delta, t + \Delta) \cdot \varepsilon \geq D(\tau + \Delta, t' + \Delta) \cdot \psi$$

Agents whose preferences are *invariant* will not exhibit preference reversals when the evaluation period and payoff periods for all bundles are shifted by the same amount (Δ), all else equal.

We use TICS (time-invariance, consistency, stationarity) to refer jointly to these three properties of time preferences. Halevy (2015) proves the following proposition, linking these three properties together:

Proposition I. *The TICS properties form a binary. That is, if a preference relation satisfies any two of the properties, then it must imply the third.*

Thus, the TICS properties generate five groups that any agent's preferences may fall into: all hold, time-invariant only, stationary only, consistent only, none hold. The consequences of this binary

[1]Without loss of generality, we replace $u(\varepsilon)$ and $u(\psi)$ with ε and ψ, under the assumption an instantaneous utility function representation exists.

relationship are not well-appreciated in the literature. Economists who wish to model inconsistent agents, for instance, must also depart from at least one of stationarity or time-invariance. The former approach is typical; the hyperbolic and $\beta - \delta$ discount functions satisfy only time-invariance, for example. There exists only a nascent literature on discount functions that do not satisfy time-invariance (Strulik, 2021), and on observed time-variant behavior (Brownback *et al.*, 2025; DeJarnette, 2020; Halevy, 2015; Harrison *et al.*, 2025; Imas *et al.*, 2022; Janssens *et al.*, 2017).

Before proceeding, there are two terms that we will define for use in this paper that are convenient short-hands for phenomena we observe, but come with a lot of baggage in the literature: *(im)patience* and *present/future-bias*. In some papers these terms are closely linked (e.g. Prelec (2004), where present-bias is considered "decreasing impatience"). However, it will be useful for us to keep them separate in order to use terms like increasing and decreasing patience to describe time-variance. Specifically, a subject in our study exhibits decreasing (increasing) patience if their discount factor for a delayed reward is decreasing (increasing) as time passes:

$$D(\tau, t) > D(\tau + \Delta, t + \Delta) \quad \text{for} \quad t > \tau \ \text{ and } \ \Delta > 0$$

Present (future) bias refers to the phenomenon where an agent's relative preference for a sooner reward vs. a later reward is decreasing (increasing) in how far those rewards are from the *initial* evaluation period. Present bias indicates either a consistency ($C > 0$, $S = 0$) violation, a stationarity violation ($S > 0$, $C = 0$), or both ($C, S > 0$) of the form

$$\frac{D(\tau, t)}{D(\tau, t')} > \frac{D(\tau + C, t + S)}{D(\tau + C, t' + S)} \quad \text{for} \ \ t' \geq t, \ \ S \geq 0, \ \text{and} \ \ 0 \leq C \leq t - \tau + S$$

Often present bias is modeled discontinuously, with extra discounting applied to all non-immediate rewards in the workhorse quasi-hyperbolic discounting model (Laibson, 1997; O'Donoghue and Rabin, 1999). However, present bias is also a feature of the hyperbolic discounting model widely used outside economics. There is also debate over whether "bias" is appropriate to describe an empirical phenomena that could stem from a variety of root causes (Bernheim and Taubinsky, 2018). In this paper we use "present bias" and "future bias" simply to classify the direction of violations of stationarity and consistency.

II.2.1. Measuring discounting properties

Consider a decision maker (DM) faced with a choice evaluated at time $\tau = 0$ between an immediate reward at $t = 0$ and a reward on future date $t_1 > 0$. Let a denote a measure of how

much their choice favors the reward at the later date, t_1. That is, higher values of a result from less discounting of (a higher discount factor for) the delayed reward.[2]

In the same evaluation period ($\tau = 0$), the DM also faces a choice between utility in two future time periods, $t_2 > 0$ and $t_2 + t_1$. Let b denote the corresponding measure of how much their choice favors the later reward in that scenario. After time passes, such that calendar time is now $\tau' = t_2$, the DM faces a choice similar to the original: an immediate reward at $t = \tau'$ or a reward at future date $t_2 + t_1$. Once again, let c denote the measure of how much their choice favors the later reward.

We call the collection of these three choices a "decision triangle," visualized in Figure II.1. Using the values associated with a triangle, $a - c$ is a measure of time-variance; how much a choice over fixed relative dates changes over time. $b - c$ measures inconsistency; how much a choice over fixed calendar dates changes over time. Finally, $b - a$ measures non-stationarity; how much a choice made on a fixed date changes in response to front-end delay. Combining them, we have

$$\underbrace{a - c}_{\substack{\text{Time} \\ \text{Variance}}} = \underbrace{a - b}_{\text{Non-Stationarity}} + \underbrace{b - c}_{\text{Inconsistency}}$$

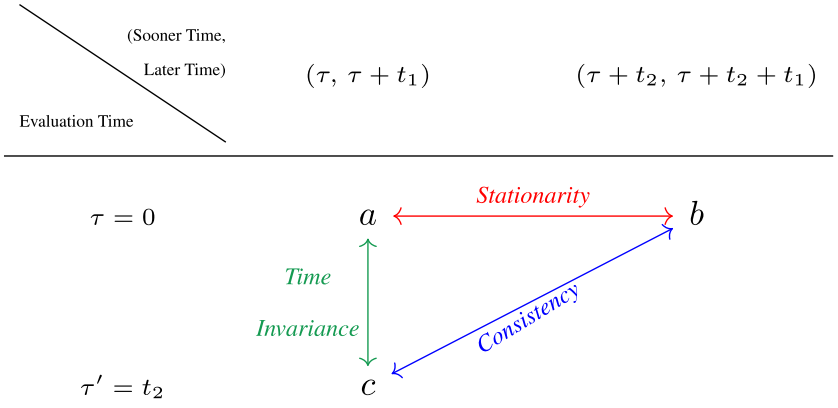


Figure II.1: Time preference property decision triangle

With this framework, we can illustrate the empirical issues that arise when a researcher does not account for time-variance. Suppose a researcher wishes to estimate parameters of the $\beta - \delta$ model, and they set up an experiment to measure the DM's time-consistency. Assume the DM

[2]Put another way, higher values of a result from a larger discount factor (typically closer to one from below) for the delayed reward.

exhibits time-varying choices in the form of decreasing patience, leading to $a - c < 0$. Assume also that the DM *would* exhibit stationary choices if they were measured, such that $b - a = 0$. In this case, the DM must then exhibit time-inconsistent choices such that $b - c = a - c < 0$.

Thus, the researcher will conclude that the DM is present-biased and estimate $\beta < 1$. Now suppose instead that the researcher sets up an experiment to measure the DM's stationarity. Assume the same direction of time-variance in choices exists as the previous example, but also assume that the DM *would* exhibit time-consistent choices if they were measured. Since $b - c = 0$, it must be that the DM makes a non-stationary choice so that $a - b = a - c < 0$. Thus, the researcher will find the DM is future-biased, and estimate $\beta > 1$. Attribution of the agent's true behavioral response to an immediate gratification opportunity in the face of identical, unobserved time-varying preferences can thus depend on research design.

II.3. Experimental Procedures[3]

The details of our experimental design and analysis were pre-registered on AsPredicted.org, protocol 133180. We recruited subjects from introductory economics courses at the University of Oregon. A total of 178 subjects were invited to participate in the experiment, and 153 consented and completed the first survey. After initial recruitment, which occurred both in-person and over email, all communication with subjects was conducted online through the Qualtrics survey platform. While the frequent, short points of contact in our design make an online design preferable, because of the delayed and risky incentives we decided it was important to run the study in a sample where we could leverage our institutional reputation to establish trust, and where subjects would have in-person access to the study administrators in case of any problems or concerns. Participants were provided with a study email address, as well as the office location and phone number for the faculty supervisor.

The study featured seven surveys; one every two weeks for twelve weeks. Every subject that completed the first survey earned $5, paid upon completion, and every subject that completed the final survey earned an additional $35 upon completion. All payments were via instant digital transfer on a subject's preferred platform.[4] The first survey had detailed instructions on our protocol and the last featured an exit questionnaire. Both took subjects roughly 20 minutes to

[3]IRB approval for this research was obtained from the University of Oregon, Study ID: STUDY00000768.

[4]Subjects chose from Venmo, PayPal, CashApp, and Zelle, and we pre-screened potential participants for their using one of these platforms to receive payment.

complete. Each intermediate survey featured only prize drawings, price lists, and a very brief questionnaire, each taking roughly 10 minutes to complete.

Within each survey, subjects faced multiple price lists (MPLs) that asked them to choose between two columns, where each column represented a prize-drawing for $300 in a specified time period, and each row represented a choice between a ticket-time period bundle. Each ticket corresponded to an increment of 1 in 100,000 in odds within a drawing. The left column always offered tickets in a drawing that would take place earlier than the drawing in the right column. For all subjects, in every price list, the left column featured a sequence of 22 rows with 202 to 244 tickets by increments of two, whereas the right column featured a fixed 242 tickets in every row. This design is an adaptation of the multiple-lottery-list elicitation of Belot *et al.* (2025). See Appendix Figure A.1 for an example price list.

We denote surveys by the number of weeks from the start of the study that they occur (corresponding to t and τ in the discount functions). In weeks zero and two (the first and second surveys) we offered subjects six choices: choices between a drawing immediately and drawings in two, four, and eight weeks, and choices between a drawing in two weeks and drawings in four, six, and ten weeks. We refer to the time between drawings within a choice as the "choice delay" and the time until the first drawing within a choice as the "front-end delay." When a subject responds to sequential surveys, we observe TICS decision triangles (one for each choice delay), allowing us to detect violations of any of three time preference properties. Starting in week four, we phased in longer choice delays of 16 and 32 weeks, at first only with the two-week front-end delay, but then in week six, without the front-end delay as well.[5] Week 12 was the final week that involved subject choices, and we did not elicit choices with the front-end delay in this survey. Figure II.2 outlines the price lists faced by subjects in each survey.

The goal of price list elicitations is to observe a single switching point between columns, such that the switch point approximately identifies at which row – as the value of one of the column's rewards changes monotonically – a subject is indifferent between the bundle in the left column and the bundle in the right column. If the subject switches more than once, this confounds that identification. A single switching point in tickets was enforced by our elicitation device, a bright yellow sliding bar that separated choice rows for which the subject preferred the earlier drawing from the choices for which subjects preferred the later drawing. Subjects were instructed to place the sliding bar *between* the last choice in which they preferred in the

[5]In our regression models we will use delay-length fixed effects to flexibly account for these additional choices.

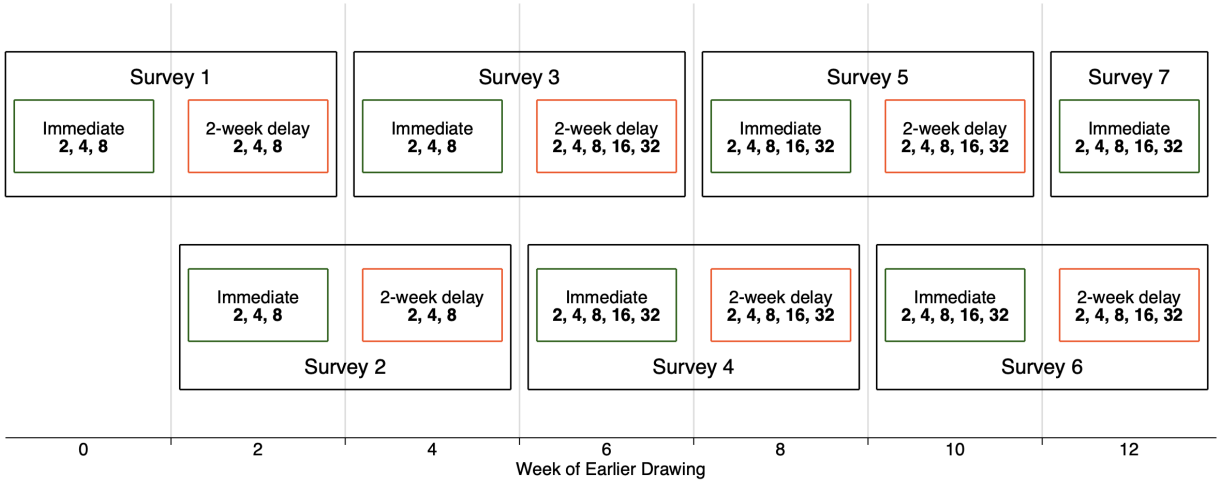


Figure II.2: Study timeline

Notes: The green boxes represent price lists with immediate earlier drawings and the orange boxes represent price lists with earlier drawings in two weeks from the date of the survey. The numbers below these labels represent the choice delays of each price list administered – the number of weeks between the earlier and later drawings.

right column (the later prize drawing) and the first choice they preferred in the left column (the sooner prize drawing). Specifically, we instructed them to put every row where they prefer the later drawing above the bar, and every row where they prefer the earlier drawing below the bar. Writing on the bar itself expressed their switch point in words, which changed when the bar was dragged to a new location. Subjects could leave the bar at the very top – indicating they always prefer the earlier drawing – but they had to actively click and release it to do so.[6] Subjects could not advance through a price list for at least six seconds in order to limit random or thoughtless choices.

All choices had a positive probability of being realized, but only one *per earlier-drawing date* was selected to count. For example, in week zero, with equal probability one of the price lists with an immediate earlier drawing was randomly selected, and then one of 22 rows was randomly selected, again with equal probability. If the subject picked the immediate drawing in that row, the drawing took place at the end of the week zero survey. If the subject picked the later drawing in that row, the drawing took place *at the beginning* of the survey in that week. The week zero price lists with a two-week front end delay were "saved for later": when subjects returned in week two and made another set of choices featuring an earlier drawing on that date, the price list that counted for that earlier date was randomly selected from both the sets of week zero and week two choices.

[6]We thank Antonia Krefeld-Schwalb for the idea, and the source code to help us build this tool.

Prize drawings were implemented with the selection of a random number from 1 to 100,000. If this number was less than or equal to the number of tickets a subject had in a drawing, they won. Winning triggered an email to a study administrator who would immediately transfer the $300 prize. Three subjects won drawings, including one in the first survey. Many drawings took place after the surveys ended; in these cases, subjects received emails with a link to participate in the drawing with no data collection.

Before participating in incentivized price lists in week zero, subjects had to complete a number of comprehension tasks: 1) fill out an example list (however they wanted), 2) correctly use the yellow slider bar to illustrate a specified preference, 3) choose a random drawing number that would allow one hypothetical subject to win and another to lose based on their choices, and 4) correctly answer a question about the independence of choices across price lists. Each survey after week zero included a refresher question that subjects had to correctly answer prior to making their incentivized choices.[7]

The first survey was distributed by email at 7am on a Wednesday in the middle of the university term so that subjects would find the link in their inbox in the morning. All subsequent surveys were also distributed at 7am on Wednesday, covering the end of that term, finals week, and much of the subsequent term.[8] Each survey was open for 26 hours, from 7am on Wednesday to 9am on Thursday. Subjects were allowed to miss one out of the seven surveys and were barred from further participation after their second missed survey.

II.3.1. Attrition

While a concern in any longitudinal study, attrition was a particular concern for this experiment given its length, number of contact points, and the potential for non-random attrition. If those that leave the study are less patient than those who remain – because of the delayed incentives – then the full sample may exhibit increasing-patience time-variance in the aggregate despite no individual level time-variance. Given resource constraints to offering overwhelming incentives and time constraints in hunting down missing individuals to complete a survey within a 26-

[7]While the data are not the subject of the current paper, each survey concluded with a brief questionnaire on recent life events. Subjects were asked about financial and psychological shocks they may have experienced since the last survey. The same questions were asked every survey, except for one rotating question that assessed risk preferences, cognition, and other measures of time preferences not captured by our main elicitation. The final survey also included an exit questionnaire that asked subjects about their understanding of the survey, demographics and socioeconomic status, expectations about the future, subjective time horizon, patience, and impulsivity.

[8]The term length is 10 weeks, so we were unable to avoid an end of term somewhere in the study. Between having a survey in finals week or spring break, we selected finals week under the assumption that students would at least be on their computers, even if they were busy.

hour window, we opted for an approach that 1) made completing each survey fast and easy, with multiple reminders and direct, individualized participation links (as opposed to requiring login information), 2) allowed us to carefully test for any selective attrition, and 3) allowed us to measure time-variance at the subject level.

Of the 153 subjects that completed the first survey, 97 (63%) completed the last survey. 79 (52%) of subjects completed every survey. This defines a "full" sample of 6,367 price lists from 153 subjects and a "balanced" sample of 4,345 price lists from 79 subjects. Throughout the results we present findings from both samples when it is concise to do so, focus on the balanced sample otherwise, presenting corresponding results from the full sample in the Appendix. None of the main results of the paper depends on the sample we use. We formally test for differential attrition in Section II.5., and find no significant differences in the discount factor at any point in the study between those who are about to attrit and those who will remain (see Appendix Figure A.2).

One mitigating factor for concerns about attrition is our focus on subject-level results. We designed the study to observe the time preference properties exhibited by subjects and estimate subject-level parameters of discount functions. For all subjects that completed at least two subsequent surveys – 141 in total, 92% of the initial sample – our experiment provides sufficient data to directly inform our research question.

II.4. Time-invariant Behavior

We measure subject i's implied discount factor in price list j in survey week w, d_{ijw}, as the midpoint of their switch interval divided by 242 (the fixed ticket allocation to the later drawing).[9] For example, if a subject prefers 242 tickets later to 230 tickets earlier, but 232 tickets earlier to 242 tickets later, then $d = \left(\frac{231}{242}\right) = 0.955$.

Across all price lists the average d is 0.957, corresponding to a switch point between 230 and 232 tickets. For the balanced sample, the average d is 0.956, corresponding to the same switch point. These are significantly different from one ($p < 0.0001$ in both cases), consistent with a positive discounting over tickets, but not significantly different from each other ($p = 0.752$).[10]

[9]When a subject never switches – always preferring the later draw – we assume a switch midpoint of 245. When a subject switches immediately – always preferring the sooner draw – we assume a switch midpoint of 201. In Section II.4.1. we find that allowing for a censoring process at the endpoints instead has no impact on estimates of discounting.

[10]We run an OLS regression of d on an indicator for being in the balanced sample, with standard errors clustered at the subject level, and then perform two-tailed t-tests of equality with one for both groups.

Figure II.3 shows that in general subjects discount more when the choice delay is longer, and when the earlier prize drawing is immediate. In fact, at every delay length, subjects discount the future more when the sooner draw is immediate, consistent with present-bias. In the full sample when there is no front-end delay, discounting is not decreasing in delay length from two to eight weeks. However, this is not true when we use the balanced sample and otherwise, discounting responds as expected to delay.

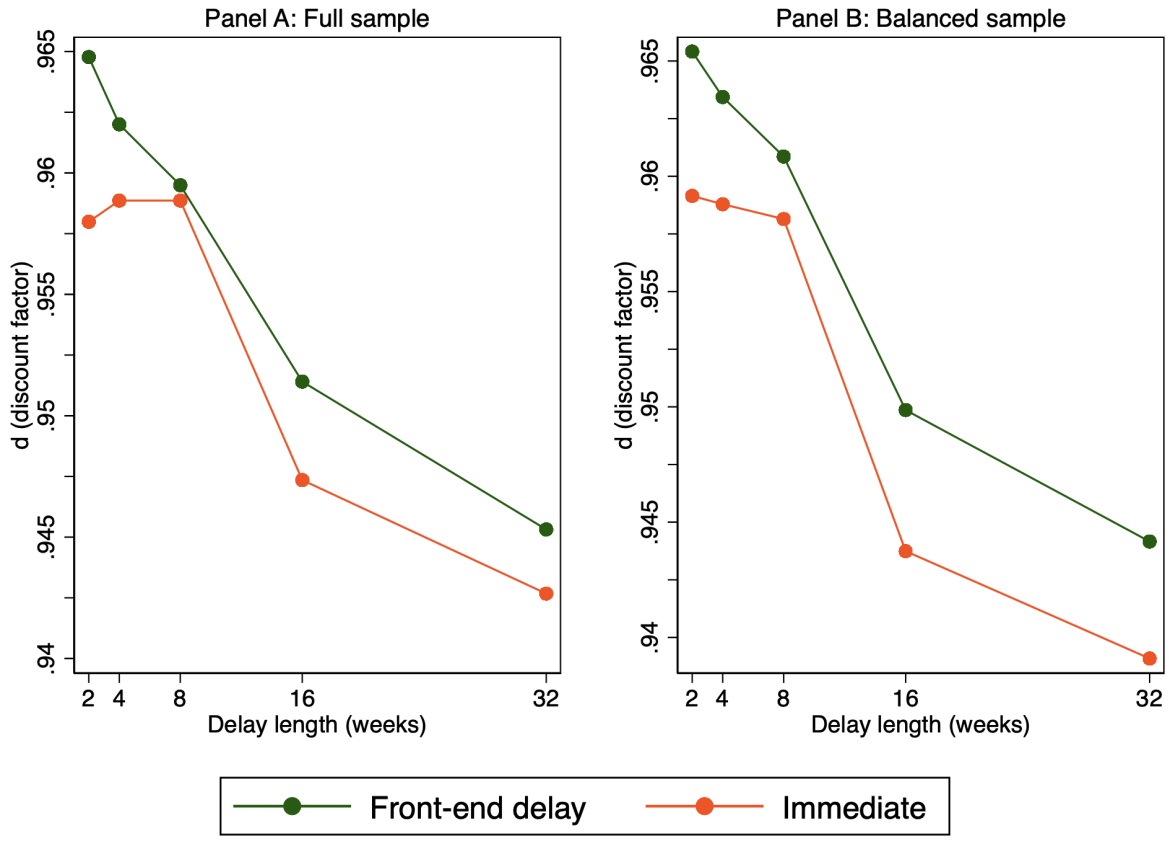


Figure II.3: Empirical discount functions

We regress d on choice delay length and an immediate indicator, and find that for each week the delay increases, d decreases by 0.001 on average in both the full and balanced samples ($p = 0.001$ in both cases). When the earlier lottery is immediate, d decreases by 0.003 (0.005 in the balanced sample) on average ($p = 0.007$ and $p = 0.001$, respectively). Given that empirical discount functions are often convex, we also estimate log-dependent variable models to allow for exponential decay. The coefficient estimates are very similar, which we should expect with minimal misspecification bias and d close to one. Estimates are in Table II.1. Subjects discount

future rewards, do so more when the future is more distant, and exhibit small but significant present-bias.

Table II.1: Discounting and price list features

Sample:	Full		Balanced	
Dependent variable:	d	$\ln(d)$	d	$\ln(d)$
	(1)	(2)	(3)	(4)
Delay (weeks - 8)	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Immediate	-0.003***	-0.004***	-0.005***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.960	-0.043	0.960	-0.042
(Delay = 8 weeks, Immediate = 0)	(0.003)	(0.003)	(0.004)	(0.004)
Clusters	153	153	79	79
Observations	6,367	6,367	4,345	4,345

Notes: OLS regressions with standard errors clustered at the subject level. Delay is defined relative to eight weeks.

II.4.1. Aggregate discounting estimates

A common application of time preference elicitations in the literature is to structurally estimate discounting parameters – typically an exponential δ parameter, and often a quasi-hyperbolic β parameter as well. Before proceeding with this exercise we illustrate how our data allow us to identify such parameters. The goal of our price list elicitation is to identify tight bounds on the set containing the indifference point between the two prize draws. When subjects never switch (always choose the later draw) or switch immediately (always choose the sooner draw), that set is only bounded below or above. 90% (91%) of choices in the balanced (full) sample are *interior* switch points, and only one subject failed to deliver a single interior switch point (they were in the balanced sample). As such, we feel comfortable using structural estimation strategies that assume a price list switch point identifies a point of *equality* between the sooner and later options (a "convex price list" in the terminology of Kuhn (2026)) for our main estimates. We will also show evidence for the robustness of our estimates to set-identification and binary-choice approaches.

Following our notation from Section II.2. where the discount factor $D(\tau, t)$ is a function of payoff period t and evaluation period τ, the quasi-hyperbolic form is

$$D(\tau, t) = \beta^{\mathbb{1}(t>\tau)} \cdot \delta^{t-\tau}$$

For our sooner and later options within each price list, we call t^S and t^L the payoff dates of the respective lotteries.[11] Assuming d identifies the point of indifference between the sooner and later draws in the price list, we have

$$\beta^{\mathbb{1}(t^S>\tau)} \cdot \delta^{t^S-\tau} \cdot (X-1) = \beta \cdot \delta^{t^L-\tau} \cdot 242 \Rightarrow d = \frac{\beta \cdot \delta^{t^L-\tau}}{\beta^{\mathbb{1}(t^S>\tau)} \cdot \delta^{t^S-\tau}} = \beta^{\mathbb{1}(t^S=\tau)} \cdot \delta^{t^L-t^S} \quad (1)$$

where X is the number of sooner tickets in the switching row (and thus $d = \left(\frac{X-1}{242}\right)$). We log both sides of this equality to obtain the regression equation

$$\ln\left(d_{ijw}\right) = \mathbb{1}\left(t_j^S = \tau_j\right) \cdot \ln(\beta) + \left(t_j^L - t_j^S\right) \cdot \ln(\delta) + \varepsilon_{ijw} \quad (2)$$

which we estimate with standard errors clustered at the subject level. Estimates of δ and β from this approach are shown in column (1) of Table II.2. Column (2) shows the estimates from an interval regression (a generalized Tobit model), which uses only the known bounds of each switch interval, including one-sided bounds at the extremes. Column (3) shows estimates from a non-linear least squares (NLS) regression applied directly to (1). NLS will be our preferred technique for estimating the time-variant discount functions, which cannot be neatly linearized, so we include it here to show that it produces nearly identical estimates to other models. When price lists feature larger intervals than ours, or do not enforce a single switch point, researchers often take a discrete choice approach to the data at the choice-row level, using maximum likelihood to estimate the model parameters. We take this approach with a logistic choice model, with estimates in column (4).

Estimates for β and δ are nearly invariant to estimation method. Across the four models, our estimate of the annual discount rate ranges from 9% to 11%. This is consistent with the lower end of estimates from the time preference elicitation literature. For example Andersen *et al.* (2008) and Andersen *et al.* (2014) find estimates of 10% and 9% in nationally-representative samples of Danes, respectively. Andreoni and Sprenger (2012) estimates an annual rate of 37% using the Convex Time Budget technique with a US undergraduate population, and Kuhn *et al.* (2017) estimates a rate between 20% and 24% using the same technique among French undergraduates. Our estimates for β are all about 0.97, corresponding to present-bias that is roughly equivalent to an increase in delay length of 15 weeks. All estimates of β and δ are statistically significantly

[11] Note that t^L is always greater than τ.

Table II.2: Aggregate discount parameter estimates from quasi-hyperbolic model

Model:	OLS	Interval regression	NLS	Binary logistic
	(1)	(2)	(3)	(4)
δ	0.9980	0.9979	0.9981	0.9983
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
β	0.9711	0.9712	0.9723	0.9778
	(0.0032)	(0.0033)	(0.0031)	(0.0031)
r (annual discount rate)	0.1098	0.1146	0.1066	0.0901
	(0.0150)	(0.0174)	(0.0148)	(0.0145)
H_0: $\delta = 1$	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$
H_0: $\beta = 1$	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$

Notes: Standard errors are clustered at the subject level. δ is the weekly exponential discount factor, and $r = \delta^{-52} - 1$ is the annual discount rate. The data consist of 4,345 price lists from the 79 subjects in the balanced sample. The binary logistic model assumes a standard T1EV error distribution, and uses data from 95,590 binary choices.

different from one with $p < 0.0001$. Results are nearly identical in the full sample, shown in Appendix Table A.1.

II.4.2. Individual discounting estimates

When a time preference elicitation is bundled into the design of a larger research study, it is expected to deliver an individual-specific measure of discounting that is either a mediating variable or an outcome of interest. Using the OLS technique from (2), we successfully estimate individual-specific δ and β parameters for every subject in the sample (both balanced and full). Appendix Figure A.3 shows the distributions of estimates of r, the annual interest rate, and β, from the balanced sample.

The average (median) r estimate is 11.8% (5.7%), with 81% of subjects having $r > 0$. The average (median) β estimate is 0.972 (0.979), with 89% of subjects having $\beta < 1$. Results for the full sample are in Appendix Figure A.4. While there are some outliers in the right-tail of the discount rate distribution in the full sample, the median of 7.8% is very similar to that of the balanced sample. The median and mean β estimates are virtually identical across samples.

II.5. Time-variance: reduced form analysis

We now exploit the longitudinal nature of our data. Figure II.4 shows how average switch points change over the course of the study, separately by price list features (delay length and front-end delay) in the balanced sample; i.e. this figure shows purely the evolution of choices from

identical sets from the same subjects over time. The downward trend across survey weeks is clear in essentially all choice sets, with a very similar pattern visible in the full sample, shown in Appendix Figure A.5. This suggests that subjects are getting less patient (discounting more) over time throughout our study. This is important in our experimental context: decreasing patience is exactly the kind of time-varying time preference that would lead researchers to misclassify time-variance as present-biased time inconsistency. We return to this issue in more detail later in this section.

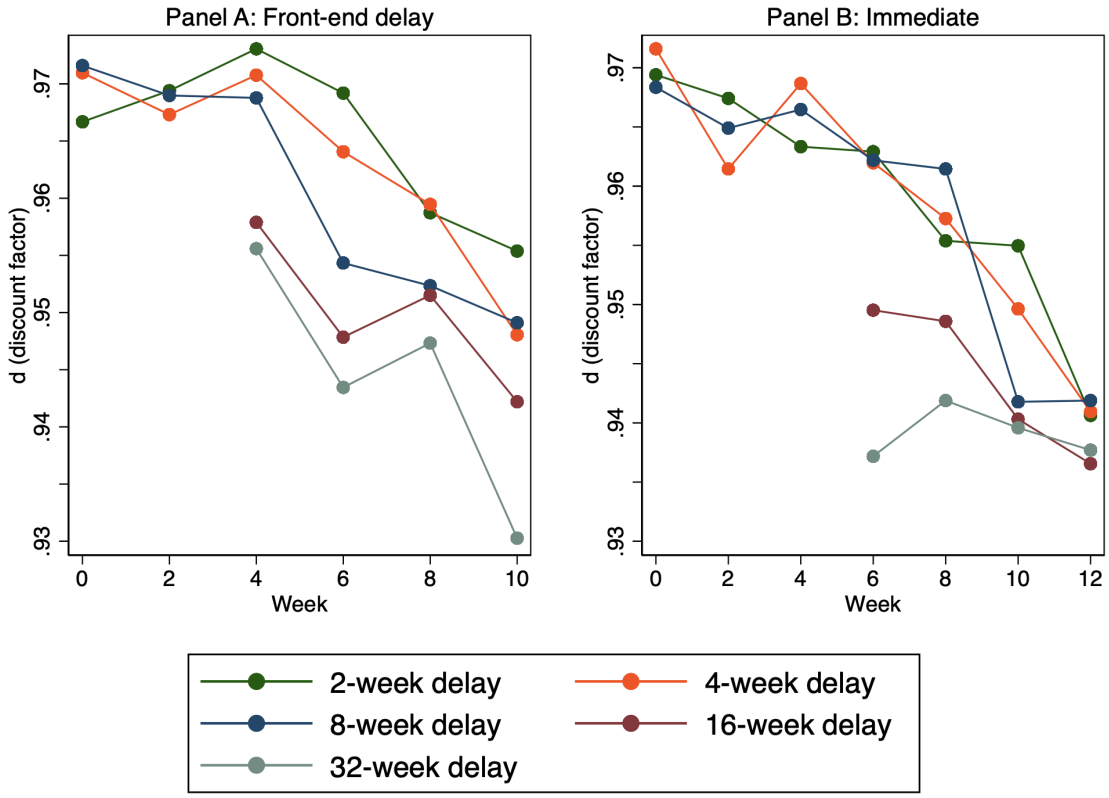


Figure II.4: Time-variance in discounting

Notes: The data consist of 4,345 price lists from the 79 subjects in the balanced sample.

To assess the magnitude of these changes, we assume exponential discounting and first consider the two-week discount factor with no front-end delay in the balanced sample, as we observe this in all seven surveys of the study. The average d falls from 0.969 in Week 0 to 0.941 in Week 12. The corresponds to an increase in the annual discount rate from 124% to 391%. However, Figure II.4 also showcases that discounting is less sensitive to delay length variations between two and eight weeks than the exponential model would predict, and so we consider the evolution

of the annual discount rate as calculated from the eight-week factor as well. In this case, the decreasing patience over the study corresponds to an increase in the annual discount rate from 23.2% in Week 0 to 47.6% in Week 12.[12] In both cases, this is an economically meaningfully change in discounting.

To assess the significance of these changes, we estimate OLS regressions of d on a linear survey week variable, with fully-interacted fixed effects for delay length and front-end delay, and standard errors clustered at the subject level. In the balanced sample, we estimate a decrease in the discount factor of 0.002 per week ($p = 0.001$), and in the full sample we estimate a decrease of 0.001 per week ($p = 0.017$). While the estimates are qualitatively similar, we can reject that they are the same in magnitude across samples ($p = 0.009$). This suggests that if anything, attrition may attenuate the decreasing-patience time-variance we observe in the study.

A key strength of our data collection is our ability to observe subject-level changes in discounting over time. While on average, subjects appear to be making less patient choices over time, we can characterize the distribution of time-variance. We estimate the same model as in the previous section – a linear OLS regression of the discount factor on a survey week variable with fully-interacted delay-length and front-end delay fixed effects – at the individual level. Appendix Figure A.6 shows the distributions of week coefficients in the full sample (Panel A) and balanced sample (Panel B). Negative values correspond to time-variance in the direction of decreasing patience.

Overall, 77 of 153 (50%) of subjects in the full sample, and 47 of 79 (59%) of subjects in the balanced sample exhibit decreasing patience over the course of the study. Among this group, we can reject time-invariance for 39 subjects (25%) in the full sample and 28 subjects (35%) in the balanced sample. 57 subjects (37%) in the full sample exhibit increasing patience, and we can reject time-invariance for 33 (22%) of them. 25 subjects (32%) in the balanced sample exhibit increasing patience, and we can reject time-invariance for 15 (19%) of them. That leaves 19 subjects (12%) in the full sample and 7 subjects (9%) in the balanced sample as exactly time-invariant.

II.5.1. Does attrition explain aggregate time-variance?

While subjects have a small likelihood of winning a prize each time they take a survey, the guaranteed, substantive payment for participating in the study comes after all surveys have been completed. When it comes time to participate in survey week two, that delayed reward is ten

[12] All of the rates are higher than those present in Table II.2 because we have not yet accounted for present-bias in this section.

weeks in the future, and it gets two weeks closer in each successive survey. If impatient subjects attrit because they do not value the completion payment, we should expect to see a large initial decrease in discounting from survey weeks zero to two (because week zero is incentivized on its own), and then a stable pattern thereafter.[13] This is not the pattern of time-variance we find in the full sample shown in Appendix Figure A.5.

Furthermore, Appendix Figure A.2 shows how the discount factor within each survey correlates with whether a subject does not complete the following survey. Specifically, for each survey we regress the discount factor on an indicator variable for whether a subject fails to complete the next survey, with fully-interacted fixed effects for delay length and front-end delay, and standard errors clustered at the subject level. Plotting the attrition coefficients and their 95% confidence intervals shows that none of the estimates are statistically significant, and the sign is not consistent across surveys.

II.5.2. Time inconsistency and non-stationarity

In Section II.2., we illustrated how time-varying behavior implies either non-stationary or time inconsistent behavior (or both). We have now empirically established both decreasing patience and inconsistent/non-stationary behavior in the data. However, we have not yet drawn a distinction between time inconsistency and non-stationarity. In this section, we allow them to differ and explore the connection between them and with time-variance; how much observed inconsistency is explained by time-variance rather than non-stationarity?

We start by separately estimating non-stationarity and inconsistency. Recall that in Table II.1, the "Immediate" coefficient represented the effect of removing a front-end delay on the discount factor. By adding *survey-week* fixed effects to these models we isolate only within-survey variation in the existence of a front-end delay, making the Immediate coefficient a measure of non-stationarity alone. By instead adding *drawing-week* fixed effects, we isolate only across-survey variation in the existence of a front-end delay, making it a measure of inconsistency alone. Specifically, the drawing week of a decision is the week the *earlier* drawing takes place. For example, decisions in the first survey with a two-week front-end delay have the same drawing week as decisions in the second survey without the front-end delay. Estimates are in Table II.3 for the balanced sample, and in Appendix Table A.2 for the full sample.

[13] Purely from a patience perspective, if week two is worth completing for the reward in ten weeks, then week four is worth it for the reward in eight weeks.

Table II.3: Separate estimates of non-stationarity and inconsistency

Measure:	Non-stationarity		Inconsistency	
Dependent variable:	d	$\ln(d)$	d	$\ln(d)$
	(1)	(2)	(3)	(4)
Immediate	−0.001	−0.001	−0.006***	−0.007***
	(0.001)	(0.002)	(0.002)	(0.002)
Constant (Delay = 8 weeks, Immediate = 0, Week = 0)	0.969	−0.033	0.975	−0.026
	(0.004)	(0.005)	(0.005)	(0.005)
Survey-week FE	Y	Y	N	N
Drawing-week FE	N	N	Y	Y
Clusters	79	79	79	79
Observations	4,345	4,345	4,345	4,345

Notes: OLS regressions with standard errors clustered at the subject level. All models include delay-length fixed effects with eight weeks as the omitted group. Week zero is always is omitted group from the survey-week and drawing-week fixed effects. Data are limited to the balanced sample.

The pattern of results is striking: our observation of present-bias in the data is entirely due to violations of time-consistency rather than stationarity. We find no statistically significant relationship between a *within-survey* front-end delay and the discount factor (e.g. $p = 0.347$ for the linear model in the balanced sample, $p = 0.537$ for the linear model in the full sample). On the other hand, we find that making the same choice two weeks in advance (*within-drawing-week*) results in a significantly higher discount factor (e.g. $p < 0.001$ for the linear model in the balanced sample, $p = 0.004$ for the linear model in the full sample). These findings are consistent with Section II.2.1., where apparent time-inconsistent present bias is a result of decreasing patience under the assumption of stationarity.

We follow Janssens *et al.* (2017) in asking how much inconsistency and non-stationarity in our study can be attributed to time-variance. Using the decision triangle language of Section II.2.1., given any time invariance ($a - c$), we can calculate whether observed inconsistency ($b - c$) is exactly predicted by, in excess of, or less than what would be predicted assuming stationarity holds ($a = b$). Specifically, for all decision triangles with inconsistency ($b \neq c$), we calculate $(a - c)/(b - c)$. This is effectively the percentage of time inconsistency that is attributable to time-variance. *Time-variance explains 72% of the inconsistency in our sample.* While inconsistency is typically attributed to a behavioral present-bias, this result suggests that in a typical experimental context a large part of it could be explained by decreasing patience instead.

We can do the same for any observed non-stationarity as well. For all decision triangles where non-stationarity is present ($a \neq b$), we construct a similar ratio of $(a - c)/(a - b)$. This ratio describes the percentage of non-stationarity that can be attributed to time-variance. Time-variance explains 56% of non-stationarity in our sample, however non-stationarity is less common in our data.

II.5.3. Time preference classification

We now follow Halevy (2015) and classify each subject into one of the five possible combinations of time preference properties. We start by classifying each decision within a triangle as either time-invariant or time-varying. This requires that subjects must have attended at least two sequential surveys to be included in this analysis. Out of 153 total subjects, this applies to 141.[14] These 141 subjects account for 6,267 of the 6,367 price lists collected in our study, containing 2,506 decision triangles. Of these 141 subjects, 130 subjects (92%) make a time-varying decision at some point in the study. Of the 2,506 decision triangles, 1,500 (60%) include a time-varying decision.[15] Both the percentage of subjects and percentage of choices that are time-varying exceed the findings of Halevy (2015), that found a maximum of 56.42% of subjects exhibiting time-varying preferences. While the longer longitudinal dimension of our study contributes to that difference at the subject level, this should not be the case at the triangle level. We return to this issue later in this section.

We count the occurrences of inconsistent and non-stationary decisions in a similar manner to the time-varying decisions. While the same 141 subjects and 6,267 price lists generate 2,710 decision triangles where b and c the test for inconsistency is possible and 2,782 decision triangles where the test for non-stationarity is possible, we restrict our analysis to the 2,506 decision triangles that overlap with the preceding time-invariance analysis. We find that 129 subjects (91%) make an inconsistent decision at some point, with 1,448 (58%) of triangles being inconsistent. Similarly, 127 subjects (90%) make a non-stationary decision at some point, with 1,276 (51%) triangles in total being non-stationary. Both proportions exceed those from Halevy (2015) again, where at most 66.33% of subjects exhibited inconsistent preferences and at most 55.22% exhibited non-stationary preferences.

---

[14] This eliminates ten subjects who only responded to the first survey and two more subjects who only responded to the first and third.

[15] This proportion is very similar if we also consider time-variance in the choices made *with* a front-end delay. In this case, 94% of subjects display time-varying behavior in 59% of triangles.

Table II.4 shows the resulting classification of subjects by property in our sample in columns (1) and (3) of Panel A for the full and balanced samples, respectively. Focusing on the balanced sample, only eight subjects (10%) exhibit invariant preferences. Of these eight, seven (9% of the sample) make consistent and stationary decisions as well. However, these seven subjects display constant discounting, not reacting to differences in choice delay or front-end delay either. Thus, among subjects who reacted to any stimuli, all violated a time preference property. 69 (87%) subjects violate each property at some point, two (3%) maintain only stationarity, and none maintain only consistency.

Table II.4: Subject- and decision-triangle-level classification of time preference properties

Sample:	Full		Balanced	
Margin of Error:	None	$\pm$ 6 tickets	None	$\pm$ 6 tickets
Properties of choices	(1)	(2)	(3)	(4)
Panel A: Subject level				
A) Invariant, stationary, consistent	6.38%	20.57%	8.86%	20.25%
B) Invariant only	1.42%	0.00%	1.27%	0.00%
A) + B)	7.80%	20.57%	10.13%	20.25%
C) Stationary only	3.55%	7.09%	2.53%	3.80%
D) Consistent only	2.13%	1.42%	0.00%	0.00%
E) None	86.52%	70.92%	87.34%	75.95%
C) + D) + E)	92.20%	79.43%	89.87%	79.75%
Panel B: Triangle level				
A) Invariant, stationary, consistent	33.88%	61.61%	37.45%	63.40%
B) Invariant only	6.26%	7.30%	6.07%	7.01%
A) + B)	40.14%	68.91%	43.52%	70.41%
C) Stationary only	15.20%	15.20%	15.24%	14.40%
D) Consistent only	8.34%	8.30%	7.86%	7.96%
E) None	36.31%	7.58%	33.39%	7.23%
C) + D) + E)	59.86%	31.09%	56.48%	29.59%

Notes: Columns (1) and (2) in Panel A use the 141 subjects that completed consecutive surveys. Columns (3) and (4) use the 79 subjects that completed all surveys. Columns (1) and (2) in Panel B use all valid decision triangles in the study, and columns (3) and (4) in Panel B use all decision triangles in the balanced sample. Columns (1) and (3) assume that any deviation in choices across price lists with equivalent delay lengths is a time preference property violation. Columns (2) and (4) assume that any deviation greater than six tickets from an analogous choice is a time preference violation, capturing a range of a standard deviation around a choice ($\sigma/2 \approx 7$, which due to interval of two tickets between choice rows corresponds to equality within six tickets).

A key caveat to these statistics is that the prevalence of property violations in our sample can be, in part, attributed to the large number of decisions we offer subjects. Our study design allows for formal modeling and testing of each property at the individual level, which we do in the

next section. However, we first address this issue with two ad-hoc approaches that allow us to continue to compare our results to other studies.

First we allow for a margin of error that spans a standard deviation (14 tickets) of the decisions we observe (± 6 tickets due to the two-ticket gap between each choice row). If a subject's choice is within six tickets on either side of their previous choice, we treat the decision as if it did not violate the tested time preference property.[16] Results are in columns (2) and (4) of Panel A in Table II.4. Second, instead of aggregating to the subject-level, we treat each decision triangle as a distinct observation. Results are in columns (1) and (3) of Panel B, and results using both modifications are in columns (2) and (4) of Panel B.

Using the margin of error at the subject level, 16 subjects (20%), exhibit time-invariant behavior. All 16 of these subjects are also consistent and stationary. The increase in the size of this group comes mostly from a reduction in size of those that violate every property, now 60 subjects (76%). Three subjects maintain only stationarity (4%) and still none maintain only consistency using this approach. Taking the decision-triangle level of analysis approach instead we find that around 43% of decision triangles maintain time-invariance. This is close to the minimum classification from Halevy (2015), which has 44% of subjects exhibiting time-invariance. Our most conservative estimates of time-varying behavior come from combining the decision triangle sorting with the margin of error. This method shows 70% of decision triangles maintain time-invariance, which is close to the maximum classification from Halevy (2015), which has 68% of subjects exhibiting time-invariance. Overall, these results suggest that observing a single decision triangle per individual understates the potential for time-varying behavior. Defined strictly, time-variance is a universal feature of subjects who responded to our stimuli.

II.6. Building a time-varying discount function

To account for time-varying preferences the limited existing literature has used reduced-form approaches to correcting $b - c$ in light of $a \neq c$, or (less-likely given design constraints) correcting $a - b$ in light of $a \neq c$ (e.g. Janssens *et al.* (2017) or Harrison *et al.* (2025)). We offer an alternative: a *potentially* time-varying discount function. This section maps the five groups of time preference properties to structural discount functions and their properties.

[16] One problem with adding a margin of error to this accounting exercise is that it is possible for only one property to be violated, even though theoretically, when two hold, the third is guaranteed. When this is the case, we classify a decision triangle or subject as satisfying all time preference properties.

Time-invariant preferences are well represented in the literature. Discount functions that admit these preferences include the standard exponential, as well as the hyperbolic and β-δ models regularly used to model time-inconsistent decisions. These models all share a common property: They are a function of only the distance between payoff time and evaluation time, $(t - \tau)$. Thus, $D(\tau, t)$ is typically reduced to just $D(t)$, where τ is always 0, without loss of generality. This characterizes the necessary and sufficient condition for a function to be time-invariant.

Proposition II. *A discount function admits time-invariant preferences if and only if* $D(\tau, t) = D(t - \tau)$.

See Appendix Section A.2 for the proof.

We are only aware of one discounting functional form in the literature that violates this property. Strulik (2021) introduced a "time-consistent hyperbolic" discount function which violates the assumption of time-invariance:

$$D(\tau, t) = \left(\frac{1 + \alpha\tau}{1 + \alpha t}\right)^{\beta} \quad \alpha > 0,\ \beta > 1$$

Strulik (2021) designed the function for use in environmental resource models, relying on the increasing patience mandated by the model to avoid extinction steady states. The ability to commit to consumption plans in his context requires that the discount function maintains time-consistency. Therefore, this model must be only time-consistent, as it is not time-invariant and so cannot be both time-consistent and stationary. In general, what makes a discount function time-consistent? Strotz (1956) suggests that the only discount function that can achieve time-consistency is the standard exponential. However, this result relies on the assumption of a discount function that is a function of only $(t - \tau)$. Strotz's result has since been generalized to prove instead that time-consistency of a discount function is equivalent to multiplicative separability in t and τ (Burness, 1976; Drouhin, 2020).

Proposition III. *A discount function* $D(\tau, t)$ *admits consistent preferences if and only if it can be written:*

$$D(\tau, t) = f(t) \cdot g(\tau)$$

Further, if the discount function can be written in this way, then $g(x) \equiv \frac{1}{f(x)}$.

Including the time-consistent hyperbolic model, three of the five possible time preference groups have been characterized by discount functions in the literature. To our knowledge, discount functions have not been developed to fit the two remaining groups: only stationarity and adhering to none of the TICS properties. To develop these functions in the next section, we present the necessary and sufficient conditions for a discount function to admit stationarity, which also, to our knowledge, has not been shown in the literature.

Proposition IV. *Let $\alpha \in \mathbb{R}^+$. A discount function admits stationary preferences if and only if it takes the form*

$$D(\tau, t) = \alpha^{(t-\tau)f(\tau)}$$

See Appendix Section A.3 for the proof.[17]

With propositions II, III, and IV, we arrive at the following well-known result:

Corollary. *The discount function $D(\tau, t)$ admits preferences exhibiting stationarity, time-consistency, and time invariance if and only if it is the exponential discount function.*

That is, $D(\tau, t)$ can be written as this function of $(t - \tau)$ alone: $D(t - \tau) = \delta^{\beta(t-\tau)}$.

II.6.1. Nested Exponential Discount Function

With necessary and sufficient conditions for a discount function to satisfy any of the TICS properties, we now have the tools to develop a discount function that can fit all five possible classes of time preferences. Two goals motivate our approach to developing such a function. First, our discount function should be fully general in its admittance of time preference properties. For instance, the β-δ model acts as an extension of the standard exponential, allowing for time-inconsistent and non-stationary preferences by setting $\beta \neq 1$. By extending the standard exponential in a similar manner, we can achieve each of the five time preference classes according to its parameter values. Second, a time-varying discount function should be capable of exhibiting both increasing or decreasing patience depending on parameter values. The time-consistent hyperbolic model from Strulik (2021) only allows for increasing patience, which was appropriate contextually. However, allowing for decreasing patience is important for making the function generally applicable.

[17] Note that f may be trivial in τ.

We introduce the following discount function:

$$D(\tau, t) = \delta^{\frac{(t^\mu - \tau^\mu)^\eta}{\gamma^\tau}} \qquad \eta > 0,\ \gamma > 0,\ \mu > 0$$

This function, which we call the *nested exponential discount (NED) function*, can satisfy any of the five possible arrangements of the TICS properties. Suppose $\eta = \gamma = \mu = 1$. Then the NED is identical to the standard exponential discount function. Now suppose $\eta = \gamma = 1$ while $\mu \neq 1$. Then the function is no longer a function of only $(t - \tau)$ nor does its exponent satisfy the conditions required to maintain stationarity. However, the function does remain multiplicatively separable in t and τ, granting it time-consistency only. The argument follows similarly for when $\eta = \mu = 1$ and $\gamma \neq 1$. In this case, the function is not a function of only $(t - \tau)$ and is no longer multiplicatively separable in t and τ. Its exponent, however, can be expressed as $(t - \tau)f(\tau)$, granting stationarity only. When $\gamma = \mu = 1$, but $\eta \neq 1$, NED the is a discrete-time reformulation of the Weibull discount function, a less common behavioral model that achieves the same violations of the TICS properties as the hyperbolic and β-δ models. Specifically, in this case we have that the function is not multiplicatively separable in t and τ, the exponent is not of the form $(t - \tau)f(\tau)$, but the function only depends on $(t - \tau)$. Since any of these three parameters not equaling one results in the violation of two properties, whenever two or more parameters are not equal to one, none of the properties are satisfied. Therefore, all five groups of time preference properties can be modeled with this one function. Note that the function can also be expressed as

$$D(\tau, t) = \exp\left(-\rho \cdot \frac{(t^\mu - \tau^\mu)^\eta}{\gamma^\tau}\right), \text{ where } \rho = -\ln(\delta) \tag{3}$$

similar to the alternate format of exponential discounting.

The NED function allows for increasing or decreasing patience depending on the parameter values of γ and μ. For instance, when $\mu = 1$, values of $\gamma < 1$ correspond to decreasing patience. As τ increases, the relative distance between t and τ is amplified by a denominator that is increasingly less than one, in turn increasing the exponent, and therefore decreasing the discount factor. The opposite holds for values of $\gamma > 1$. Similarly, when $\gamma = 1$, values of $\mu < 1$ correspond to increasing patience. Holding fixed the distance $t - \tau = C$, the term $t^\mu - \tau^\mu$ is decreasing in τ, resulting in a discount factor increasing in τ. The opposite again holds for $\mu > 1$. Since η on its own cannot cause increasing or decreasing patience, it only acts to mitigate or amplify the effects of the other parameters when active at the same time. The exact condition that

dictates the direction of changing patience over time of the nested exponential discount function is given by:

$$D(\tau, \tau+\Delta) > D(\tau', \tau'+\Delta) \iff \frac{\ln(\gamma)}{\eta\mu} > \left(\frac{(\tau+C)^{\mu-1} - (\tau)^{\mu-1}}{(\tau+C)^{\mu} - (\tau)^{\mu}} \right) \tag{4}$$

The right side is positive (negative) for $\mu > 1$ ($\mu < 1$), while the left is positive (negative) for $\gamma > 1$ ($\gamma < 1$). Therefore, the NED function predicts strictly increasing patience, $\frac{\partial D}{\partial \tau}|_{t-\tau=C} > 0$, when $\gamma > 1$ and $\mu < 1$, and $\frac{\partial D}{\partial \tau}|_{t-\tau=C} < 0$ when $\gamma < 1$ and $\mu > 1$.

Figure II.5, Figure II.6, and Appendix Figure A.7 visualize the roles of each parameter in the NED function. The time-invariant behavior of the function is shown in Figure II.5, which highlights that it is effectively another approach to hyperbolic-style discounting, with the η parameter governing the direction and degree of distortion from exponential discounting.

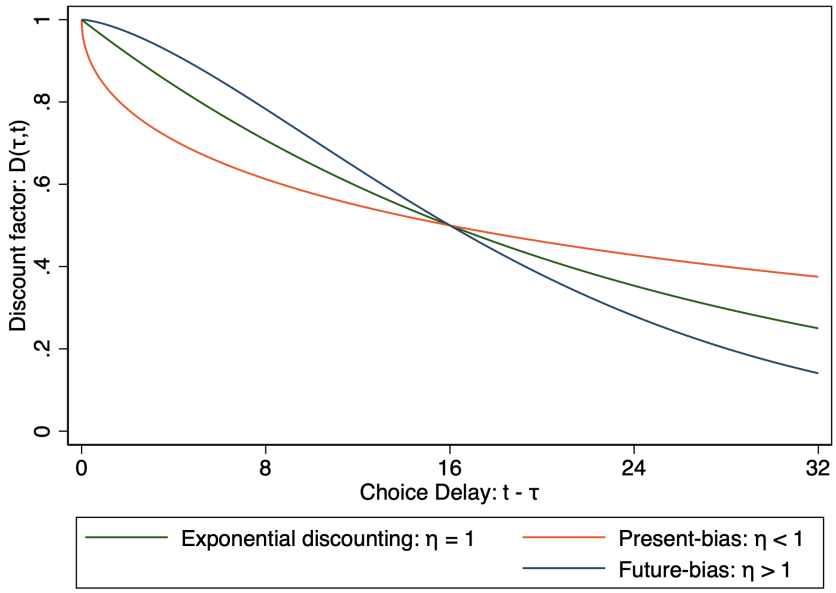


Figure II.5: Time-invariant NED function ($\mu = \gamma = 1$)

Notes: The value of δ in each function is chosen such that $D(\tau, \tau+16) = 0.5$. The values of η shown for present and future bias are $\eta = 0.5$ and $\eta = 1.5$, respectively.

The time-consistent behavior of the function is shown in Appendix Figure A.7, but it can be simply described in relation to the time-invariant effects of η. The μ parameter can create similar non-stationary present/future biases as η, but as time passes, those biases will erode. Panel A shows that μ and η have identical effects when $\tau = 0$, but that as τ increases to two in Panel B and four in Panel C, the distortion away from exponential discounting caused by $\mu \neq 0$ becomes

smaller. The intuition here is that for small τ relative to t, the effect of μ is similar to η, but as τ approaches t, this effect diminishes. For example, an agent that is "present-biased" according to a stationarity violation at $\tau = 0$ will get more patient over time as they maintain time-consistency.

The time-stationary $(\eta, \mu = 1)$ behavior of the NED function is shown in Panel A of Figure II.6. We set δ so that for each function $D(0, 8) = 0.5$, and then show the effects of τ increasing as time passes. γ determines whether an agent becomes more or less patient over time, and thus whether individuals would appear "present-biased" or "future-biased" in a violation of time-consistency. Panel B shows that when the function obeys none of the three properties, non-monotonic time-variance can emerge through the interaction of μ and γ, but that γ ultimately controls the long-run direction of preference drift.

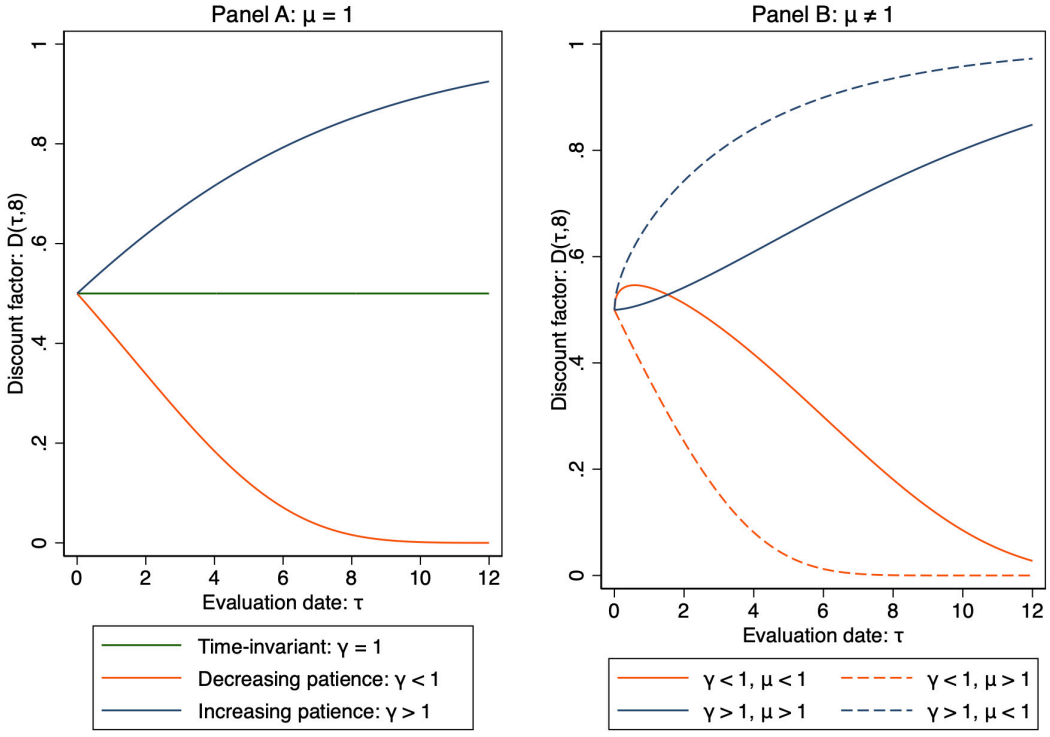


Figure II.6: Time-variance in the NED function for an eight-week choice delay

Notes: The value of δ in each function is chosen such that $D(0, 8) = 0.5$. We show $\gamma = 0.8$ and $\gamma = 1.2$ in both panels, and $\mu = 0.5$ and $\mu = 1.5$ in Panel B. $\eta = 1$ in both Panels.

This non-monotonicity is an extreme example of the fact that when it comes to estimating time-invariance and the parameters of nested exponential discount function, the longitudinal dimension of the data determines the precision and reliability of those estimates. Existing work by Halevy (2015) and Janssens *et al.* (2017) collects data on a single decision triangle (ignoring

different prices or choice delays within a triangle).[18] We instead design a study to observe repeated observations of triangles for every subject. Time-variance need not be idiosyncratic subject-level noise at different dates; instead it can be systematic learning (about either the decisions themselves or preferences over them), changes to the macroeconomic environment, the psychological environment, or preference evolution.

II.7. Time-variance: structural analysis

We first estimate discount function parameters at the aggregate sample level. We use the non-linear least squares method, and estimate both restricted and unrestricted versions of the NED model. Results and descriptive statistics of these estimations on the balanced sample are presented in Table II.5. We limit the sample to the set of choices that were available in all weeks of the study to maintain a clear interpretation of time-variance in the absence of a flexible set of fixed effects. Column (1) enforces exponential discounting ($\mu = \eta = \gamma = 1$), column (2) enforces time-invariance only ($\mu = \gamma = 1$), column (3) enforces stationarity only ($\mu = \eta = 1$), column (4) enforces consistency only ($\eta = \gamma = 1$) and column (5) makes no parameter restrictions. Estimates using the full sample are in Appendix Table A.3.

Besides direct tests of the estimated parameters for adherence to the time preference properties, we assess the fit of each version of the model. We use both the Akaike Information Criteria (AIC), which applies a parameter penalty to model likelihoods, and Bayesian Model Averaging (BMA) to do this.[19] For each estimated model we report the AIC, model selection weights calculated using the AIC (Burnham and Anderson, 2004), and the posterior model probabilities (PMPs) from the BMA procedure (Hoeting *et al.*, 1999).[20] Both the AIC weights and Bayesian PMPs are measures of relative model fit (from 0 to 1) among the set of five models in the table.

In column (1) the NED model is fully restricted to exponential discounting, and we obtain an estimate of the weekly discount factor of 0.9932. This translates into an annual discount rate

[18] Harrison *et al.* (2025) collects longitudinal data without overlapping dates and uses structural assumptions to analyze them like a triangle.

[19] The AIC is defined as $2k - 2\ln(\mathcal{L})$, where k is the number of model parameters and $\mathcal{L}$ is its fitted likelihood, meaning that lower AIC values represent a better fit. BMA instead assigns posterior probabilities to candidate models according to how well they fit the data (relative to their priors, which we assume are equal), and combines the results across models.

[20] The AIC model selection weights are calculated with respect to the set of models estimated as $w_i = \frac{\exp(-\Delta_i/2)}{\sum_j \exp(-\Delta_j/2)}$, for all models j, and $\Delta_i = AIC_i - \min_{j(AIC_j)}$. The Bayesian PMPs are the posterior probability from the BMA exercise.

Table II.5: Aggregate parameter estimates for the nested exponential model

Restrictions:	$\mu=\eta=\gamma=1$	$\mu=\gamma=1$	$\mu=\eta=1$	$\eta=\gamma=1$	None
Properties					
Invariant	✓	✓	X	X	X
Stationary	✓	X	✓	X	X
Consistent	✓	X	X	✓	X
	(1)	(2)	(3)	(4)	(5)
δ	0.9932	0.9807	0.9953	0.9958	0.9843
	(0.0007)	(0.0023)	(0.0007)	(0.0017)	(0.0027)
η	1	0.5394	1	1	0.5838
	.	(0.0309)	.	.	(0.0547)
γ	1	1	0.9401	1	0.9401
	.	.	(0.0152)	.	(0.0213)
μ	1	1	1	1.1524	0.8619
	.	.	.	(0.1255)	(0.1509)
AIC	25,072	24,959	25,017	25,068	24,914
AIC weight	< 0.0001	< 0.0001	< 0.0001	< 0.0001	1.000
Bayesian PMP	< 0.0001	< 0.0001	< 0.0001	< 0.0001	1.000
r_0	0.4265	0.1786	0.2785	0.4976	0.1223
	(0.0524)	(0.0230)	(0.0482)	(0.0913)	(0.0243)
r_{12}	$= r_0$	$= r_0$	0.6745	0.5504	0.2583
			(0.1131)	(0.1363)	(0.0450)
PB_S	1	0.9795	1	1.0043	0.9997
	.	(0.0026)	.	(0.0029)	(0.0129)
PB_C	1	$= PB_S$	0.9950	1	0.9800
	.		(0.0010)	.	(0.0023)
H_0: $\delta = 1$	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$	$p = 0.0142$	$p < 0.0001$
H_0: $\eta = 1$	.	$p < 0.0001$	.	.	$p < 0.0001$
H_0: $\gamma = 1$	.	.	$p = 0.0002$	.	$p = 0.0062$
H_0: $\mu = 1$	.	.	.	$p = 0.2283$	$p = 0.3630$
H_0: $\mu = \gamma = 1$	.	.	.	.	$p = 0.0055$
H_0: $\mu = \eta = 1$	.	.	.	.	$p < 0.0001$
H_0: $\eta = \gamma = 1$	.	.	.	.	$p < 0.0001$
H_0: $\mu = \eta = \gamma = 1$	.	.	.	.	$p < 0.0001$
H_0: $r_0 = r_{12}$	.	.	$p = 0.0009$	$p = 0.2991$	$p = 0.0044$
H_0: $PB_S = 1$	.	$p < 0.0001$	.	$p = 0.1360$	$p = 0.9836$
H_0: $PB_C = 1$	.	$p < 0.0001$	$p < 0.0001$	.	$p < 0.0001$
H_0: $PB_S = PB_C$	.	.	.	.	$p = 0.1329$

Notes: Standard errors are clustered at the subject level. $r_\tau = D(\tau, \tau+52)^{-1} - 1$ is a measure of the annual discount rate. PB_S and PB_C are stationarity-based and consistency-based measures of present-bias. The data consist of 3,081 price lists from 79 subjects that completed all seven surveys, using only the choice sets that spanned every survey. All estimates are from non-linear least squares regressions.

(r_0) of 43%. The model in column (2) allows for time-inconsistency and non-stationarity, but enforces time-invariance, similar in spirit to quasi-hyperbolic discounting, but with a functional form more closely related to hyperbolic discounting. The estimate of $\delta = 0.9807$ suggests much more discounting of near-term future outcomes than the exponential estimate in column (1), but the "present-bias" parameter η flattens the discount function for longer choice delays. $\eta = 0.5394$ turns a delay length of two weeks into an effective delay of 1.5 weeks, and a 52-week delay length into an effective delay of 8.4 weeks, such that the annual discount rate for a one-year choice delay is estimated to be much lower than in the exponential model: 18%. By comparison, without the hyperbolic "flattening" of the discount factor in this function, a weekly discount factor of 0.9807 would exponentially imply an annual discount rate of 175%. This discrepancy means we estimate significant present bias in discounting, clearly rejecting $\eta = 1$ ($p < 0.0001$), and that this model offers a substantial improvement in fit (lower AIC) when accounting for it.

To assess the magnitude of present-bias estimated by the model in a familiar metric, we construct measures PB_S and PB_C, which are discount factor ratios that correspond to how researchers would typically measure a quasi-hyperbolic β through either a stationarity violation or a consistency violation, respectively. Specifically,

$$PB_S = \frac{D(0,8)}{D(0,0)} \Big/ \frac{D(0,10)}{D(0,2)} \quad \text{and} \quad PB_C = \frac{D(2,10)}{D(2,2)} \Big/ \frac{D(0,10)}{D(0,2)}$$

When invariance holds these measures are identical, and for the model in column (2) we estimate $PB_S = PB_C = 0.9795$ which is significantly different from one ($p < 0.0001$).

The model in column (3) allows for time-variance and inconsistency, but enforces time-stationarity, the property specification that does not appear in the literature. The estimate of the weekly discount factor when $\tau = 0$ is 0.9953, which translates to an annual rate (r_0) of 28%. Time-variance due to $\gamma = 0.9401$ ($p = 0.0002$ tested against one) decreases that weekly discount factor as the study progresses – the annual discount rate at the end of our study (r_{12}) grows to 67%, meaning that this model estimates statistically significant, decreasing-patience time-variance ($p = 0.0009$ for the test of $r_0 = r_{12}$). A direct implication of this is that the model predicts statistically significant, consistency-violation present-bias ($p < 0.0001$ for the test of $PB_C = 1$), although the magnitude is very small. While this model captures the time-variance present in our data, it does not feature hyperbolicity in discounting; in net this tradeoff means it fits better than the exponential model, but not as well as the invariant-only model in column (2).

The model in column (4) allows for time-variance and non-stationarity, but enforces time-consistency, the "time-consistent hyperbolic" property pattern studied in Strulik (2021), but in a more flexible form. The initial ($\tau = 0$) weekly discount factor estimate is 0.9958, with the hyperbolicity parameter $\mu > 1$ steepening the discount function for longer choice delays. $\mu = 1.1524$ is not significantly different from one ($p = 0.2283$), and turns the choice between zero and two weeks into an effective delay of 2.2 weeks, and the choice delay between zero and 52 weeks into an effective delay of 95.0 weeks, yielding an initial ($\tau = 0$) estimate of the annual discount rate of 50%. The end-of-study ($\tau = 12$) annual discount rate predicted by this model is slightly higher at 55%, which we cannot reject is equal to the initial rate ($p = 0.2991$). The μ parameter acts on τ as well as t, thus while the effective choice delay between zero and 52 weeks is 95.0 weeks, the effective choice delay between 12 and 64 weeks is 103.1 weeks. A direct implication of decreasing patience in this model should be stationarity-violation future-bias, which we estimate, albeit to an insignificant degree ($p = 0.1360$). The μ parameter in this model is forced to "choose" between fitting two features of the data that, according to the model, cannot co-occur, and thus it fits poorly. We fail to reject it is different than the exponential discounting model, and the model offers only a marginal improvement in fit relative to column (1).

The model in column (5) puts no restrictions on the parameters. We estimate an initial weekly discount factor of 0.9843, and an initial annual discount rate of 12% (this rate depends on δ, η, and μ). This grows to 26% over the course of the study, reflecting the decreasing patience we identified in the reduced-form analysis. We can reject a time-invariant annual discount rate ($p = 0.0044$). Using the NED function parameters to test for adherence to the TICS properties, we can reject that all three properties hold ($p < 0.0001$), that time-invariance holds ($p = 0.0055$), that stationarity holds ($p < 0.0001$), and that consistency holds ($p < 0.0001$). On its own, we fail to reject $\mu = 1$ ($p = 0.3630$), which suggests we could adopt a more parsimonious version of the function where no properties hold. We reject that both $\gamma = 1$ ($p = 0.0062$) and $\eta = 1$ ($p < 0.0001$). Consistent with the reduced-form results, this model predicts more consistency-violation present-bias ($PB_C = 0.9800 < 1, p < 0.0001$) than stationarity-violation present-bias ($PB_S = 0.9997 < 1$, $p = 0.9836$), although we marginally fail to reject that they are the same ($p = 0.1329$). Despite the AIC measure's penalty for extra parameters, the improvement in fit for the unrestricted model is such that it gets 100% of the AIC weight among the models we estimate in Table II.5. The BMA procedure leads to the same result. When we restrict the model to $\mu = 1$ we do achieve a slight improvement in parameter-adjusted fit, with the restricted model assigned

a weight of 0.67 using the AIC technique in a two-model comparison with the unrestricted model in column (5). Results in the full sample, shown in Appendix Table A.3 are essentially the same, with the one notable difference being that we can reject equivalent consistency-violation and stationarity-violation present-bias in the full sample.

While the focus of the current paper is on time-variance, we also highlight the key role of hyperbolicity in explaining our model fitting results. We compare the invariant-only, two-parameter restricted version of the NED model in column (2) to the workhorse invariant-only, two-parameter quasi-hyperbolic model. We find that the quasi-hyperbolic model does not fit the data as well, achieving an AIC of 24,984 in the sample corresponding to the estimates in Table II.5, which gives it an AIC weight of zero in a two-model comparison with the model in column (2). This is because hyperbolicity is a real feature of our data. The single parameter hyperbolic discount function $D(\tau, t) = \frac{1}{1+\alpha(t-\tau)}$ achieves an AIC of 25,066 in this sample, which gives it an AIC weight of 0.97 in a two-model comparison with the exponential discounting model in column (1).

II.7.1. Subject-level estimates

The literature on time-preference estimation and the importance of present-bias places a strong emphasis on allowing for heterogeneity in preferences. For example, if one-in-five people are present-biased discounters we may not estimate a representative quasi-hyperbolic discount function with β significantly less than one, but we still may want to identify and target the sub-population for whom $\beta_i < 1$. In this vein, we use our non-linear least squares technique to estimate individual-specific parameters of the NED function.

We plot the distribution of parameter estimates for the balanced sample in Appendix Figure A.8, limiting the sample to only choices subjects made in every week of the study to maintain comparability with the aggregate estimates. The model converges for 59 of 79 subjects. The median subject in the balanced sample has a weekly δ of 0.9837, corresponding to an (initial: $\tau = 0$) annual discount rate of 21% (recall this rate depends on δ, η, and μ). Only one of 59 (2%) subjects is classified as a negative discounter. "Present-bias" as measured by the η parameter is common, with a median η in this sample of 0.4876, and 48 of 59 (81%) of subjects having $\eta < 1$. Estimates of γ and μ are both centered around one, with the median subject showing monotonically decreasing patience ($\gamma < 1$ and $\mu > 1$).

To make across-model fit comparisons when each model is estimated on small individual datasets ($N = 55$ at most), we use only the balanced sample, use all available choices, and

implement an extra penalty for parameters in small-sample AIC calculations (Burnham and Anderson, 2002). Using the AIC we determine which model best fits each individual in sample, with the results in column (1) of Table II.6. We find non-trivial fractions of the sample are best fit by the unrestricted NED function (38%), the invariant-only restricted model (35%), and the stationary-only restricted model (18%). This classification does not take into account the degree of advantage the best-fit model has over the next-best option for any given subject, so we also present tabulations of models that are classified with AIC weight > 0.5 and 0.9 in columns (2) and (3), respectively. There is no guarantee we identify a model that exceeds either of these thresholds, and the fraction of subjects that we cannot classify this way is shown at the bottom of the table. 87% of individuals yield a model with weight above 0.5, but only 29% yield a model with weight above 0.9. The results in column (2) are similar to column (1): the unrestricted model (35%), invariant-only model (33%) and stationary-only model (17%) are the dominant classifications. The exponential model is never assigned an AIC weight above 0.5. In column (3) we see that while over two-thirds of the sample does not have a model that dominates the classification with an AIC weight above 0.9, most of those who can be classified are sorted into the unrestricted NED model (18%). The consistent-only model is never assigned an AIC weight above 0.9. To summarize, we see that three property arrangements can effectively characterize our sample: "present bias" in which stationarity and consistency violations coincide while time-invariance holds, stationary-only discounting which has no precedent in the literature to our knowledge, and discounting which obeys none of the three TICS properties.

Table II.6: Subject-level structural classification within the nested exponential model

Metric:	Best fit	AIC weight > 0.5	AIC weight > 0.9
Model restrictions	(1)	(2)	(3)
Exponential	2.78%	0%	0%
Invariant only	34.72%	33.33%	1.39%
Stationary only	18.06%	16.67%	9.72%
Consistent only	6.94%	2.78%	0%
Unrestricted	37.50%	34.72%	18.06%
Not classified given metric	0%	12.50%	70.83%

Notes: All models are estimated via non-linear least squares, and we report classification results from the balanced sample. The exponential model and single-property models achieve convergence for the same 72 of 79 subjects and the unrestricted model achieves convergence for 63 individuals. All percentages in the table are with respect to the 72 subjects for whom at least once model converges, where we set the AIC weight for the unrestricted model to zero when it does not converge. We implement the small-sample AIC adjustment for every model, adding the penalty term $\frac{2k(k+1)}{N-k-1}$ to each AIC where $N = 55$ is the number of observations per individual, and $k = 1, 2, 4$ is the number of parameters of the model.

II.8. Discussion

To the degree previous literature considers time-varying time preferences, they are either a confound to the observation of time-inconsistency or an outcome used to assess the causal effect of some shock. In most cases, researchers have relied on at most two points of observation to draw their conclusions. By creating a panel data set with seven observations of discounting over twelve weeks, we offer both a richer environment to study the relationship between time-variance and time-inconsistency (and non-stationarity as well), and an appropriate testing ground for a discounting model that structurally accommodates time-variance, while nesting more familiar discounting models.

Whether we classify subjects structurally or not, with or without a substantial margin of error, we find that the largest single group of subjects by adherence to the TICS properties is none. We also find some evidence of stationary-only behavior that is previously undocumented in the literature, and may help explain why the time preference elicitation literature is so split on whether "present-bias" over monetary rewards is a common phenomenon. Time-variance is more common in our data than time-invariant present-biased discounting, although neither should be labeled as the rule or exception to it; they are both common. This means that the identification of time preferences in traditional time-invariant models may be confounded by preference instability. Time-consistency is the only property with effectively no empirical support.

While we do not argue that this is a general property of discounting, in a canonical experimental context we find that subjects exhibit systematically decreasing patience over the course of our study. Without adjusting for time-variance, this would've led us to dramatically overstate the degree of "behavioral" time-inconsistency in our study. 72% of the time-inconsistency we observe in our data is explained by time-variance.

Beyond highlighting this measurement issue in this particular context – we anticipate that the nature and extent of time-variance is highly context-dependent – we developed a new discount function that allows researchers to estimate a model that is fully general with respect to the TICS properties, yet tractable enough to deliver subject-level parameter estimates in a simple non-linear least squares estimation. This model is easy to pick up and estimate in a longitudinal research study that features all three choices in a decision triangle, but also it can be used in a calibration to explore the robustness of estimates of time-inconsistency or non-stationarity derived from only two of the three choices.

The presence and acknowledgment of time-varying preferences has potential to impact policy design, particularly relating to commitment mechanisms and behavioral nudges. If a policy

is meant to reinforce commitment to a consumption plan, take-up of the commitment mechanism may be lower than expected when time-varying preferences exist and are internalized by the policy target. If the target understands that their preferences are changing over time, committing to a previous decision would be welfare-damaging. Similarly, the effect of behavioral nudges may be overstated if the target has time-varying preferences that coincide with the direction the policy is meant to push them.

A deeper exploration of the sources of time-varying behavior in our data is left for future work. Past empirical work on time-variance focuses on risk, uncertainty, and liquidity-constraint as likely causes of this behavior, but learning, anticipation, and anxiety could be others. Because our context is representative of typical experimental elicitations of time preferences in the economics literature, this work could yield methodological insight into whether these elicitations capture transient environmental factors in addition to some time-invariant underlying preferences, or whether preferences are evolving over time as a part of some more general process.

Bridge

The analysis above highlights the importance of measuring time preferences in ways that are both precise and sustainable over repeated observation. When preferences are elicited across multiple waves in a remote environment, even small design frictions or sources of measurement error can accumulate and distort inference about stability and change. A longitudinal study of discounting therefore depends not only on econometric identification, but also on careful experimental design. These considerations motivate the development of a modified Multiple Lottery List (MLL) instrument tailored specifically to remote, repeated measurement. The instrument must preserve theoretical coherence—particularly with respect to utility curvature and discounting structure—while minimizing participant burden and eliminating common behavioral inconsistencies such as multiple switching. The next chapter presents and evaluates this implementation, detailing the design features that make high-frequency, remote elicitation of time preferences feasible.

III. AN EFFICIENT IMPLEMENTATION OF MULTIPLE LOTTERY LISTS FOR ESTIMATING TIME PREFERENCES

III.1. Introduction

Experimental economics, laboratory experiments in particular, play a key role in bridging between economic theory and empirics. By stripping away confounds and context, researchers engage directly with theory; theory yields hypotheses which can be tested, and the results of those tests inform new theory. When it comes to intertemporal choice, experimentalists have long grappled with the fact that there are serious tradeoffs between the desire to strip away confounds and context and the ability of subjects to make naturalistic, comprehensible choices. A recent review article (Kuhn, 2026) highlights this when considering two prominent time preference elicitation techniques, the Convex Time Budget (CTB) from Andreoni and Sprenger (2012) and the Double Multiple Price List (DMPL) from Andersen *et al.* (2008). Both techniques arise in an attempt to strip away the confound of utility curvature—the effect of a one-unit change in reward may impact utility differently as the reward grows larger—but they do so in different ways. The CTB does so by measuring utility curvature directly within the intertemporal choice domain and the DMPL does so indirectly by pairing intertemporal choice with atemporal risky choices (e.g. Holt and Laury (2002)). In doing so however, the CTB introduces a choice that may not be naturalistic and lead to confusion (Chakraborty *et al.*, 2017), and the DMPL relies on an assumption of expected utility that may be consistently violated (Andreoni *et al.*, 2017).

In this paper, I present an efficient implementation of the Multiple Lottery List (MLL) of Belot *et al.* (2025), another recently developed time preference elicitation tool that respects the notion of utility curvature, but, as originally proposed, may be difficult for experimenters to use. The MLL overcomes the issue of utility curvature by using *direct probabilities* of a large monetary sum as the time-dated rewards that subjects make choices over. In other words, individuals in an MLL generally choose between a sooner, smaller probability, and a later, larger probability of winning a time-invariant prize. What benefit does using this reward have? Under an expected utility assumption, the expected utility of an outcome is *linear* in its probability, and therefore curvature over the reward itself drops out of any intertemporal comparison.

Harrison *et al.* (2013) begin a discussion (see Chakraborty *et al.* (2017) for further details) on the accuracy of time preference elicitations given the frequency of corner solutions in the resulting data, which may be indicative of poor understanding of CTBs by participants. Therefore, demand remains for a tool that respects the key theoretical properties of measuring time preferences, without compromising data quality through poor subject comprehension. Recent work from Brownback *et al.* (2026) opts for a brief, easy to understand, and naturalistic "Asset Redemption Task" (ART). The ART design attempts to eschew the concern regarding utility

curvature entirely and ART displays predictive validity in the context of food choice. However, the relationship between subject choices and the theoretical notion of intertemporal preferences is tenuous.

At the heart of every time preference elicitation is an individual making choices about time-dated rewards. Belot *et al.* (2025) implement the MLL in a classic price-list design (see Section III.2. for more detail) using choices over the likelihood of receiving a New York State lottery ticket for a $5 million prize drawing. As a part of the experimental design for Holloway *et al.* (2026), a longitudinal study on the stability of discounting, the researchers made a number of modifications to the MLL that may make it a more attractive and versatile tool. The current paper serves to explain and highlight those developments:

Reducing compound risk: the original MLL implementation features intertemporal choices over probabilities of receiving lottery tickets, which themselves are risky assets. We reduce this to a single source of risk: the probability of winning a cash prize.[1]

Avoiding third-party lotteries: while the use of the New York State lottery prize of $5 million helps limit the influence of background consumption, it necessitated the purchase of physical scratch-off tickets that the researchers had to purchase and communicate with the subjects about. We reduce the prize size and implement the prize draw ourselves. This is crucial when it comes to estimating the quasi-hyperbolic discount function, which requires the potential for very fast (i.e. within day, see Balakrishnan *et al.* (2020)) rewards.

Eliminating multiple-switching: a long-standing concern with price-list elicitations is that subjects will use the lists incorrectly. Specifically, subjects may make selections from a list that, taken as a set, display preferences that violate the weak axiom of revealed preference. We develop a new digital interface for price lists that eliminates this behavior in a way that reinforces the intuition of the lists for subjects, not by arbitrary fiat.

Automating many choices: for price lists to efficiently identify discounting parameters, they must be dense (i.e. they must be very long), our interface automates many choices within a list in an intuitive fashion.

Belot *et al.* (2025) also discuss how using lotteries for very large rewards solves the related theoretical issue of background consumption—experimental rewards may be integrated with other consumption in a way that obscures non-linear utility tradeoffs.[2] For this reason, Holloway

[1]Both MLL implementations are within the context of a pay-one-choice-randomly experiment, which is the standard protocol in economics.

[2]Whether people actually do this is a matter of debate. See Andreoni *et al.* (2018) and Alem *et al.* (2024).

et al. (2026) recruit only college undergraduates, select a prize amount which is considerable for typical college undergraduates ($300), and survey their participants regularly on their background consumption and financial status. The issue is not skirted entirely if subjects' background consumption is volatile, but this is not a concern I focus on in this work.

To support the success of the implementation, I consider two pieces of empirical evidence in this paper. First, I compare the estimates of time preferences generated by the modified MLL to those produced by the original MLL and the existing leading techniques in the literature, the CTB and DMPL. Second, I present choice process data on just how quickly and reliably subjects make their choices in the modified MLL. To preview the results, I find time preference estimates right in the heart of the range of other techniques in the literature, and that the technique produces fewer outliers (and missing estimates) than others. I also find that the modified MLL takes trivial time for subjects once it is explained the first time, such that it could be easily folded into a broader study design or longitudinal survey. The initial explanation is comparable to other standard experimental tools.

Section III.2. goes into depth on price-lists and the development of the MLL. Section III.3. describes our implementation of the MLL. Section III.4. describes the estimated preferences and choice process data. Section III.5. concludes.

III.2. Background

A typical intertemporal price list is a set of binary choices between some monetary payoff in a sooner date as opposed to another payoff at a later date. The rewards at the later date are generally worth more than those at the sooner date because economists expect discount rates to be positive. The term *Multiple* Price List (MPL), refers to a set of lists designed to facilitate the estimation of discount function parameters. This approach is typically attributed to Coller and Williams (1999). Table III.1 features four examples of such intertemporal price lists, which is a typical offering among time preference elicitations. The example shows how experimenters often cross delay lengths *between* the rewards across columns (four and 13 weeks in this case) with delays until the earlier reward (zero and two weeks in this case). It also illustrates how price lists can be designed with the fixed reward in either column.

Table III.1: Example of multiple price lists

	List 1			List 2		
	A: immediately		**B: in 4 weeks**	**A: immediately**		**B: in 13 weeks**
Choice 1:	\$20.00	or	\$20.00	\$20.00	or	\$20.00
Choice 2:	\$19.00	or	\$20.00	\$19.00	or	\$20.00
Choice 3:	\$18.00	or	\$20.00	\$18.00	or	\$20.00
Choice 4:	\$17.00	or	\$20.00	\$17.00	or	\$20.00
Choice 5:	\$16.00	or	\$20.00	\$16.00	or	\$20.00
Choice 6:	\$15.00	or	\$20.00	\$15.00	or	\$20.00
	List 3			**List 4**		
	A: in 2 weeks		**B: in 6 weeks**	**A: in 2 weeks**		**B: in 15 weeks**
Choice 1:	\$20.00	or	\$20.00	\$20.00	or	\$20.00
Choice 2:	\$20.00	or	\$21.05	\$20.00	or	\$21.05
Choice 3:	\$20.00	or	\$22.22	\$20.00	or	\$22.22
Choice 4:	\$20.00	or	\$23.53	\$20.00	or	\$23.53
Choice 5:	\$20.00	or	\$25.00	\$20.00	or	\$25.00
Choice 6:	\$20.00	or	\$26.67	\$20.00	or	\$26.67

Notes: The top pair of lists features no "front-end delay" (the earlier reward is immediate), while the bottom pair features a two-week front-end delay. The top pair of lists features "A" options on the left which vary and are listed in descending reward order, while the "B" options on the right are held constant. The bottom pair features "A" options on the left which are held constant, while the "B" options vary in ascending reward order. Values are chosen such that indifference between the options in columns A and B occurs at the same annual discount rate in List 1 and List 3, and List 2 and List 4 (rounded to the nearest cent).

One concern with the original MPL technique is its neglect of the possibility that an agent's utility over instantaneous rewards features curvature. In order to measure and account for this, Andersen *et al.* (2008) implement an MPL alongside a Holt and Laury (2002) risk-preference questionnaire to create the so-called "Double" multiple price list (DMPL). This additional survey provides a deeper understanding of agents' utility functions by measuring time preferences and risk preferences alongside each other. Interpretation of DMPL estimates thus relies on expected utility theory and, in particular, on two key assumptions: (i) utility is separable across consumption and time, and (ii) utility is linear in probabilities over outcomes – DMPL estimation implicitly assumes that utility curvature revealed under risk is equivalent to curvature governing *determinate* intertemporal tradeoffs. The latter point is the subject of extensive debate in the literature. As emphasized by Rabin (2000), econometric identification of curvature in

laboratory settings—especially using small-stakes lotteries—already requires strong assumptions. The DMPL framework extends these assumptions by transporting curvature estimates from a risky-choice environment to a deterministic intertemporal context, raising concerns about both theoretical coherence and out-of-sample predictive performance.

Andreoni and Sprenger (2012) attempt to resolve the issue of general risk-measurement by moving away from a price-list elicitation. The authors construct a "Convex Time Budget" (CTB), which takes the binary choices from the MPL framework (the rows of a particular price list) and converts them into an approximately continuous space of options between the two original choice options. Specifically, survey participants are provided a budget of tokens which they can allocate to the sooner or later rewards, identifying the convex combination of rewards that maximizes their present-discounted utility.[3]

A central empirical critique of the Convex Time Budget (CTB) method concerns the distribution of observed allocation choices. Harrison *et al.* (2013) document that CTB data in Andreoni and Sprenger (2012) contain a substantial concentration of corner allocations, with many subjects allocating all tokens to either the sooner or later date. This generates a bimodal distribution in which the identifying variation needed to recover curvature and discounting comes from a relatively small share of interior observations. Because standard nonlinear least squares approaches rely heavily on smooth interior substitution behavior, parameter estimates can become sensitive to how corner solutions are treated in estimation. Harrison *et al.* (2013) further show that alternative estimation approaches that better match the full empirical distribution can imply convex rather than concave utility, raising concerns about whether the data are well characterized by a standard concave intertemporal utility function.

Chakraborty *et al.* (2017) extend this critique by examining the internal and external consistency of CTB choices. Comparing CTB allocations to corresponding pairwise price-list choices, and testing monotonicity properties implied by stable intertemporal preferences, they document frequent violations – particularly among subjects making interior allocations. These inconsistencies do not simply reflect random noise; rather, they suggest that CTB choices may not uniformly conform to a single, well-behaved intertemporal utility representation. As a result, while CTB avoids transporting curvature estimates from risky to deterministic domains,

[3]In a later implementation, Andreoni *et al.* (2017) use a discrete profile of convex combinations between the two endpoints of a price list row, so that the choice space is still convex, but not necessarily continuous. They argue that this reduces complexity with minimal estimation bias compared to the continuous CTB.

its empirical implementation raises concerns about whether the observed allocation patterns provide stable and behaviorally coherent identification of discounting and utility curvature.

It is important to note that this critique of the CTB is an extension of another critique of price lists: multiple switching behavior. Multiple switching occurs when a subject switches their binary choice in the expected direction – for example from a sooner reward to a later reward as the implied interest rate increases – but then switches back.[4] This violates monotonicity in preferences and creates concerns about how the data should be interpreted. Researchers deal with this issue differently. Belot *et al.* (2025) allow multiple switching in their design and average across the resulting switch points. Other researchers resolve the problem by dropping the affected lists, excluding the relevant subjects, or enforcing single switching.

The MPL and CTB form the foundation of data collection techniques for the purpose of time preference estimation. Kuhn (2026) provides an overview of other contemporary methods of such elicitation and estimation. The key development that is relevant for this paper is work by Belot *et al.* (2025), introducing a new method for eliciting time preferences. Their "Multiple Lottery List" (MLL) approach provides sets of series of binary choices, similar to the MPL, but the reward medium is selected so that under somewhat mild assumptions, time preference is the only determinant of choices.

Like the CTB, the MLL can be thought of as an extension of the MPL. The idea of the MLL is twofold: offer large but probabilistic payments, have subjects choose between lotteries over time-invariant prizes with time-dated probabilities. One of the primary concerns that Belot *et al.* (2025) attempt to tackle is the issue of background consumption, which motivates the use of large prizes, and is not a concern I dwell on here.[5] The key for the implementation of the MLL that I study is the probabilistic rewards. Under an assumption of expected utility being linear in probability over the probability choice space in the study – a milder assumption than the global linearity assumption required for risk preferences to perfectly inform risk-less utility curvature – researchers can interpret MLL choices as pure expressions of time preference. For example, in their validation experiment, Belot *et al.* (2025) offer subjects a constant 50% chance of winning

[4]It is also potentially a problem when a subject only switches in the unexpected direction – for example from a later reward to a sooner reward as the interest rate goes up, because this implies that the reward is a bad, rather than a good. Depending on the underlying model of interest and reward medium, this can lead to misspecification.

[5]Intuitively, one can think of a subject who anticipates a shock to their income stream in a later time period, and so uses the economic experiment to help aid in smoothing their consumption.

a New York State lottery ticket in an earlier period, versus an ascending-across-rows chance of winning a New York State lottery ticket in a later period.

Contrasting with Belot *et al.* (2025), our Multiple Lottery List (MLL) tool is designed to be more user-friendly and accessible for both researchers and participants. Our implementation emphasizes practical application, ensuring that the elicitation process is straightforward and efficient. By leveraging modern software tools and intuitive interfaces, we aim to facilitate broader adoption of the MLL approach in experimental economics.

III.3. Procedures

In Section III.3.1. I describe the context of the experiment from Holloway *et al.* (2026) and its procedures. This helps to both motivate and explain our modifications to the MLL that I discuss in Section III.3.2.

III.3.1. Experiment

Holloway *et al.* (2026) elicits panel data on time preferences using a sequence of incentivized multiple lottery lists administered over a twelve-week period.[6] Subjects completed up to seven surveys, spaced two weeks apart, each of which elicited choices between earlier and later monetary prize drawings. All choices were implemented through real monetary incentives, with a strictly positive probability that any individual decision would be realized.

Each survey featured prize drawings, price lists, a potential drawing, and a brief questionnaire, taking roughly 10 minutes total to complete. The first survey had thorough instructions on the experiment's protocol and the last featured an exit questionnaire; each of these surveys took around 20 minutes per respondent.[7] Each MLL price list featured 22 rows. The choice option within a price list row represented a potentially realized drawing for $300 that the subject could participate in. The cells were labeled with "ticket" values, with one ticket representing a 1 in 100,000 chance of winning the prize. The left column featured such a drawing at an earlier time period – either immediately (meaning the price list features no "front-end delay") at the end of the current survey or at the start of the next survey 2 weeks from that time (a two-week front-end delay). The right column featured an identical drawing at a later time, from 2 up to 32 weeks

[6] Details on the experimental design and analysis of Holloway *et al.* (2026) were pre-registered on AsPredicted.org, Protocol 133180.

[7] Subjects were given $5 for their participation in the first survey, and were awarded $35 upon completion of the last survey. Surveys always went out at 7am on a Wednesday and were open for 26 hours. To minimize attrition, we allow subjects to miss up to one survey throughout the study, after which they are removed from further participation. The difference in resulting panel is not substantial and is addressed throughout Holloway *et al.* (2026).

– the "choice delay" – after the earlier drawing. In the first survey, we elicit six such lists, with later drawings occurring two, four, and eight weeks after the earlier. Given observed completion times, in the fourth survey we added choice delays of 16 and 32 weeks with a front-end delay, and from weeks six onward we added choice delays of 16 and 32 weeks without a front-end delay as well. In the final survey we did not offer choices with a front-end delay. For exact details on timing, see Figure II.2. In every list the 22 cells on the left had ticket values ascending from 202 to 244, incremented by two.[8] The 22 cells on the right had a fixed value of 242 tickets throughout the list.[9] An example MLL price list is shown in Appendix Figure A.1, and I go into more detail on its design features in the next section, since they represent many of the key modifications of the Belot *et al.* (2025) protocol.

Reporting truthful time preferences is incentivized, as in other MPL-derived frameworks, by informing participants that any given choice they make on any row may be realized. I will now describe how the choice which is selected to count is determined. In the first survey subjects fill out three price lists with no front-end delay. One of these price lists is selected randomly to be the "list that counts," from which one row is randomly selected to be the "row that counts." If the subject preferred an earlier drawing in that row, it happens immediately (i.e. at the end of the survey). If the subject preferred the later drawing, it happens at the beginning of the relevant survey, as determined by the selected price list-that-counts.[10] The subject also makes choices over three price lists with a two-week front-end delay. In two weeks, they will fill out choices for three overlapping price lists with no front-end delay. Of these six price lists, one is randomly selected to count, and one row from that list is selected to count. As in the initial drawing, if the subject's preference lies on the left, the drawing is realized immediately at the end of the survey; if the subject's preference lies on the right, then a drawing will occur at the start of the survey in the number of weeks specified by the list-that-counts price list. The price lists elicited with a front-end delay in survey two are carried on to survey three's elicited lists without a front-end delay in the same manner.

[8]The exact values of the tickets were chosen to mitigate any numerical anchoring by the subjects.

[9]This differs from the Belot *et al.* (2025) study, in which the earlier lottery ticket amount was held constant while the later amount changed down the list. This does not impact estimation.

[10]For instance, if the list with a four-week choice delay is chosen to count, then the drawing will occur at the start of the survey in four weeks (the third survey).

III.3.2. MLL Modifications

As outlined in the introduction, I summarize the key modifications with the following headers: reducing compound risk, avoiding third-party lotteries, eliminating multiple switching, and automating many choices.[11] The latter two of these are accomplished with the same technical innovation: as shown in Appendix Figure A.1, we use a drag-and-drop bar to have subjects select their switch point. This enforces a single switch, but rather than do so by fiat without explanation, we designed the bar, interface, and choice, so that subjects understood the task from the point of view of making a single switch. Specifically, subjects were instructed to place the bar such that every row where they preferred the later drawing sat above the bar, while every row where they preferred the earlier drawing sat below the bar. As the bar is moved, a message displayed on the bar itself dynamically updates to reflect the participant's current preference. The message stated the preference implied by the choice row immediately below the bar – the first row in which the subject's choice of switch point results in the selection of the sooner prize draw.

The representation of lottery preferences within a particular list was assisted by color cues as well. To start, when the bar is at the top of the list (the highest interest rate), all cells on the left are colored green – meaning they would potentially be realized if that row was selected to count – and all cells on the right are colored gray – meaning they would not be realized given the current bar placement.[12] As the bar is moved down, the cells on the right which move above the bar change to green, while the cells on the left above the bar change to gray. This provides an intuitive representation of the subject's choice and potentially realized drawing: if a particular row is selected to count, then the subject enters into a drawing according to the green cell in that row, with the number of tickets in the drawing specified according to the cell's position, color, and label. Even if the subject preferred the earliest option, we require them to interact with the bar (potentially dragging the bar and putting it back in its original place) in order to proceed. We also require that subjects see each price list for at least six seconds before proceeding. Finally, we used active comprehension checks prior to participation, and refresher questions preceded subsequent surveys (see subjects' instructions in Appendix Section B.2).

By framing the MLL choices as a switching task, our aim is for the enforcement of the single switching property to aid subjects comprehension. This enforcement is beneficial because

[11] An overview of the elicitation tasks throughout the study can be found in Appendix Section B.1.

[12] In the original MLL study, the authors randomize whether or not the price list is ascending or descending at the subject-list level to avoid anchoring effects. Holloway *et al.* (2026) opt for a uniform layout to aid in subject comprehension of the task.

it avoids issues with how to treat multiple-switching data, and thus potentially reduces data loss. How concerning is this issue? According to Andreoni and Sprenger (2012), "one common finding from standard MPL experiments is that 10–50 percent of subjects switch more than once in a given price list." Solving the switching problem using the drag-and-drop bar has other benefits associated with having automated a large number of binary choices quickly. The core analysis of Holloway *et al.* (2026) assumes an indifference between the earlier and later lottery options at the midpoint between relevant ticket values.[13] This assumption is much weaker when the list is *finer*, meaning that the change in value (in underlying preference parameter space) across rows is small. Finer lists will produce more precise estimates.

Beyond offering finer lists, researchers generally have to budget how much time a participant spends in their study, and a quick choice process thus has other benefits. Compensation is tied to study length, leading to higher costs when choices are slower (all else equal), and longer durations can lead to participant fatigue and lower levels of engagement. Our technique requires very little time to complete even a very long price list (see Section III.4.2. for the data on this), leaving the researcher to offer more lists too. The following example illustrates the benefit of having more and finer lists.

Assume a subject in our study is an exponential discounter and they face an MLL price list with an eight-week choice delay (i.e. the later prize draw in the right column takes place eight weeks after the earlier prize draw in the left column). As in our study, assume the subject can select 242 tickets in the later draw in every row. If the researcher wanted to guarantee at least one choice option where indifference corresponds to an annual discount rate over 100%, with equally-spaced ticket amounts, and only five choice rows, they would be required to use a seven-ticket interval between options. This opens up a big gap in more patient preference space; indifference in the row with equal tickets occurs when the subject doesn't discount, but it occurs in the next row at a 21% annual rate. There's a lot of nuance in the demand for credit and savings products within the range of 0–21% that this list would fail to inform. With 22 options instead, the researcher could use a two-ticket interval (as we do), which leaves a gap of only 0–5.5% in the annual indifference discount rate on this list.

Figure III.1 illustrates how the layering of many, fine lists leads to granular coverage of the discount rate space in our experiment, compared to a coarser list of only five rows. Subjects can choose from 202 up to 244 tickets toward an earlier drawing or 242 tickets in a later drawing. This

[13]For example, if the subject prefers the earlier drawing if they get 230 tickets but the later drawing if the early drawing only offers 228, then their indifference point is identified as 229.

induces indifference between the offered options at choice-delay-specific discount factors in the range $\left[\frac{202}{242}, \frac{244}{242}\right]$, or approximately $[0.83, 1.01]$. However, the choice delay dramatically impacts the implied annual discount rate. For example, for a two-week choice-delay, under exponential discounting this corresponds to annual discount rates spanning from −19% to 10,866%, well beyond the expected bounds on subject preferences. At the other end of the spectrum, for a 32-week choice delay that same set of discount factors maps to annual discount rates between −1.3% and 34%. While longer lists allow for more precise snapshots of subjects' discount rates, layering *more lists* on top of each other permits the researcher to dramatically change the scale of the preference parameter choice space. This helps identify extreme short-term discounting that could result from non-exponential discounting, and helps curtail corner solutions at the endpoints of the price list, which I show in the data in the next section.

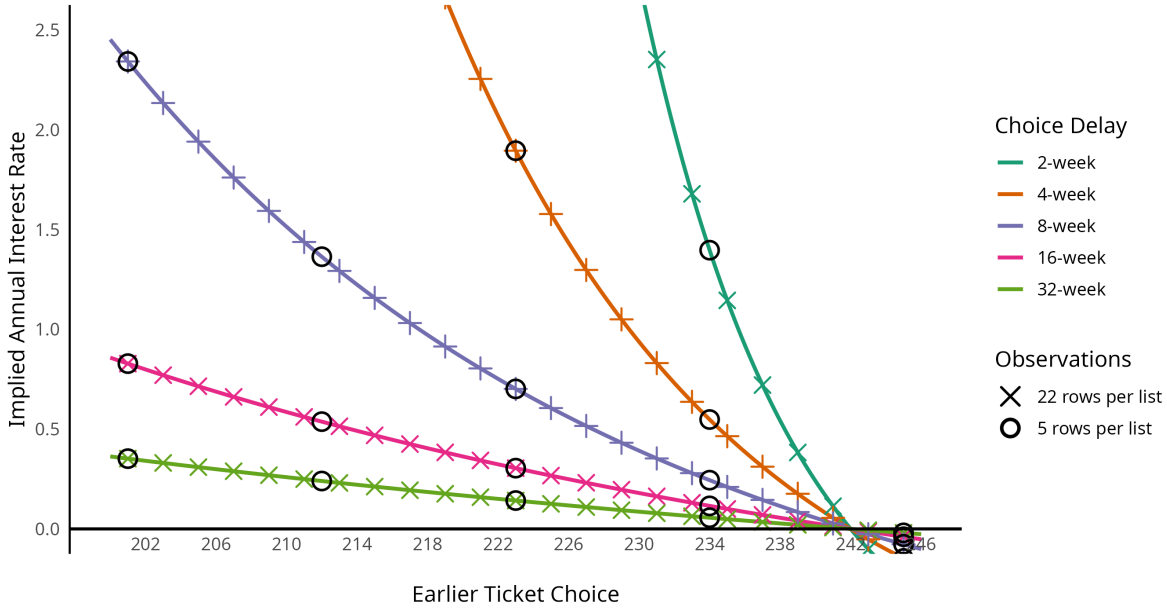


Figure III.1: Implied interest rates based on price list choice delay

Notes: Implied discount rates in our experiment, assuming exponential discounting. The X's show 22 possible interest rates per choice-delay, similar to our study. The O's illustrate the possible interest rates assuming only 5 equally-spaced options.

I now turn to the other two advantages of our modifications, starting with reducing compound risk. Belot *et al.* (2025) incentivize the MLL in their experiment with physical scratch-off lottery tickets. However, subjects do not make choices between a number of lottery tickets at two different dates, but rather between probabilities of receiving a single lottery ticket at two different dates. Thus, even conditional on a particular choice-that-counts realizing, subjects face

compound uncertainty; will they receive the ticket, and will they win the lottery if so? Thus, the experimental object is not a single lottery with transparent probabilities, but a compound lottery in which outcome risk and implementation risk are layered (even beyond the count-that-counts paradigm). While expected utility theory implies reduction of compound lotteries, such reduction relies on full comprehension of the compounded structure. If subjects fail to mentally reduce the compound lottery – or instead overweight or underweight particular stages – observed choices may reflect attitudes toward compound risk rather than the intended tradeoff over prize probabilities.[14] For instance, compound risk may interact with time preference estimation if delayed lotteries are perceived as more uncertain than sooner ones, confounding discounting with uncertainty effects. Thus, we switch to a choice reward being the number of tickets within a lottery rather than the probability of receiving a ticket. Since our lotteries all feature very low win probabilities at both the maximum and minimum ticket options, it maintains the feature that choices do not span a large range of the probability domain, avoiding concerns about non-linear probability weighting.[15]

More tangibly, the distribution of state-run lottery tickets is prone to frictions which challenge experimental implementation and replication. The researcher faces physical frictions such as purchase and distribution of lottery tickets which could affect the efficiency and reliability of local delivery services (or subjects showing up to receive them at appointed times). Furthermore, replication can be complicated by local laws and regulations governing official lottery systems.[16]

Related, in the New York State lottery validation experiment, Belot *et al.* (2025) distribute tickets at "immediate" time dates up to 3 days following elicitation. Evidence from Balakrishnan *et al.* (2020) shows that even small delays of a few hours can materially affect measured present bias. Via their CTB experiment, estimated present bias is substantial when payments are delivered immediately after the experimental session ($\beta \in [0.90, 0.92]$), but largely disappears when same-day payments are delayed until the end of the business day. That is, a delay of only a few hours

[14] An extensive literature in behavioral economics documents the failure to reduce compound lotteries. See Halevy (2007) for an example.

[15] The compound risk in Belot *et al.* (2025) allows subjects' choices to cover a larger perceived part of the probability domain, which could feel more natural, but this could invoke heuristics associated with cumulative probability weighting.

[16] In the United States, five states lack a state lottery entirely, making local implementation impossible as interstate distribution of lottery materials violates federal law (U.S. Congress, 2024). Even in states with active lotteries, the physical distribution of tickets by unauthorized entities is often strictly curtailed to protect the state's monopoly (see, for instance Massachusetts State Lottery Commission (2025)). Similar barriers exist abroad; for example, Norwegian law blocks any third-party facilitation of lottery products (Ministry of Culture and Equality, 2022).

is sufficient to attenuate "now versus later" effects. This has direct implications for physical lottery ticket implementations. If a several-hour delay is sufficient to eliminate measured present bias relative to a later date, then a multi-day delay almost certainly shifts the comparison from "now versus later" to "later versus later." In that case, the design may mechanically attenuate present-bias effects – not because subjects are time-consistent, but because the experimental implementation no longer places the earlier option at $t = 0$ in a behaviorally meaningful sense.

By contrast, we handle the drawing mechanism in our experiment internally.[17] The cells in an MLL represent tickets toward a drawing with odds that are both fixed and known by subjects. Our implementation prioritizes identification of present-bias parameters that are explicitly tied to immediacy. Should the subject win, they are notified within the experiment interface immediately when the drawing is held, and the prize cash is transferred within hours using the subject's choice (recorded prior to the study and used for all study payments) of three popular, digital peer-to-peer payment systems. This not only streamlines the reward process but also enhances transparency and trustworthiness, as participants can verify payments through familiar platforms. This form of payment was verified during the first survey, when subjects are paid $5 for their initial participation in the study.[18]

III.4. Performance

III.4.1. Estimates

The classic approach to estimating preferences from price list data is to respect the binary nature of each choice row and employ a random utility model. Consider subject i filling out a price list with J rows – let their choice in a given row, $c_{i,j}$ be 0 if they select the left column and 1 if they select the right column. Call $U^0_{i,j}$ subject i's utility of the option in the left column in choice row j, and $U^1_{i,j}$ the equivalent for the right column. Utility, $U()$, is assumed to have both observed, $u()$, and unobserved, ε, elements. Under assumptions of how $U()$ depends on $u()$ and ε, and the distribution of ε, the probability of selecting the left column in any choice row can be expressed as a function of $u()$ and the properties of the choice options. For example, suppose $\varepsilon^k_{i,j}$ is i.i.d. Type I Extreme Value for all j with $k \in \{0, 1\}$. If $U^0_{i,j} = u^0_{i,j} + \varepsilon^0_{i,j}$ and $U^1_{i,j} = u^1_{i,j} + \varepsilon^1_{i,j}$, then

[17] "Drawing" is preferred to "lottery" in Holloway *et al.* (2026) for regulatory reasons.

[18] The downside of this decision is that due to cost risk, the prize was only $300, which may interact with the ability of our MLL to abstract away from subjects' background consumption. In turn, this reinforced our decision to recruit participants from introductory-level undergraduate courses.

$$\Pr(c_{i,j} = 0) = \frac{\exp\left(u^0_{i,j}\right)}{\exp\left(u^0_{i,j}\right) + \exp\left(u^1_{i,j}\right)} \tag{5}$$

In words, this is to say if the unobserved component of utility, $\varepsilon^k_{i,j}$ has this particular distribution and is additively separable from $u()$, the choice probability can be expressed using the logit choice function. The likelihood of observing any particular pattern of choices on the price list is thus

$$\prod_{j=1}^{J} \left[(1 - c_{i,j}) \cdot \Pr(c_{i,j} = 0) + (c_{i,j}) \cdot \Pr(c_{i,j} = 1) \right] \tag{6}$$

Maximum likelihood can be used to estimate the parameters of the model, using data from one or more lists to construct a global likelihood function at either the sample, subsample, or subject level.

Finer lists that identify narrow intervals of indifference between rewards permit estimation via the switch point instead. Call the fixed reward in a price list f, and the variable reward r_j. Consider a single switch point r_{j^*} such that $c_{i,j} = 0$ for $j < j^*$ and $c_{i,j} = 1$ for $j \geq j^*$. The researcher needs to make some assumption regarding the relationship between the switch point and the point of indifference. In our case, we assume the midpoint of the 2-ticket interval where the subject switches identifies indifference; in other words, if the point of indifference is denoted r^*, then we assume $r^* = \left(\frac{1}{2}\right)\left(r_{j^*} + r_{j^*-1}\right)$. Therefore, $U^0_{i,j}(r^*) = U^1_{i,j}(r^*)$. As with the random utility model described earlier, an assumption about the unobservable vs. observable components of utility structure allows us to transform this equation into an estimable object. In the case of our MLL and the workhorse quasi-hyperbolic discount function,

$$U^0_{i,j}(r^*) = U^1_{i,j}(r^*)$$

$$\implies \quad \beta_i^{\mathbb{1}(FED_j > 0)} \cdot \delta_i^{FED_j} \cdot r^* = \beta_i \cdot \delta_i^{CD_j + FED_j} \cdot 242$$

which simplifies to

$$\left(\frac{r^*}{242}\right) = \beta_i^{\mathbb{1}(FED_j = 0)} \cdot \delta^{CD_j} \tag{7}$$

where FED_j and CD_j are the front-end delay and choice delay associated with choice j, respectively. If we assume (7) is subject to mean-zero, additive errors that are i.i.d. across lists, the δ and β parameters can be recovered using non-linear least squares. As with the distributional

assumptions of the random utility model applied to binary choice price lists, the researcher has flexibility here. (7) can be log-linearized and approximated via OLS, or the assumption of equality at an indifference point can be replaced with set identification based on the bounds of the switch interval, etc. Holloway *et al.* (2026) take all three of the approaches described here, in addition to the logistic, binary-choice approach described earlier, and the results are consistent across approaches.

Any of these estimation approaches can fail if a subject's data does not offer sufficient variation. For example, if a subject never switches in any price list they face, their preferences are not well-identified. Before considering the content of estimates, I consider how successful our MLL is in producing estimates in the first place. Over 90% of the price lists we offered led to an interior switching point between the sooner and later prize draws. These results hold for both the *full* sample – the 6,367 price lists we observe throughout the study – and the "balanced" sample – the 4,345 price lists we observe across the 79 (out of 153) participants that completed every survey. We can compute individual-level estimates for every subject (in both the full and balanced samples) using OLS on a linearized version of (7). We find that the average β is 0.972 with a nearby median of 0.979; correspondingly the average annual discount rate ($\delta^{-52} - 1$) is 11.8%, although the distribution is right-skewed with a median of 5.7%. However, there are no extreme estimates of either parameter that we are required to censor to make sense of the estimates either visually or using means.

How does this compare to other studies? Belot *et al.* (2025) experience mild subject loss on the order of not being able to estimate preferences for 2–5% of their data. However, their rate of outliers leads to a mean annual discount rate of 238%.[19] While they are forced to censor their estimates of β to $[0.5, 1.5]$ due to outliers, the mean estimate of 0.984 is similar to our estimates. Andreoni *et al.* (2017) also achieve estimates for their full sample with OLS, but some discount rate estimates are extreme outliers (mean of 3,355%). Andreoni and Sprenger (2012) obtain estimates for 89% of subjects in their sample, and Kuhn *et al.* (2017) do so for 76% of subjects. Recovering a complete set of useable individual preference parameters is not trivial, and data loss is important, especially for experimenters folding time preference elicitations into larger research designs.

Preference recoverability and reliability largely comes down to whether subjects make *interior choices* from the instruments they are offered. In a price list, an interior switch means

[19]While Belot *et al.* (2025) censor their estimates of both β and the monthly δ to $[0.5, 1.5]$, this has no effect on their MLL-derived estimates of the annual discount rate (it only binds for CTB estimates).

switching from one column to the other as opposed to always selecting the right or left column. We find that 90% of our MLL lists feature an interior switch, although we lack a clear comparison in the literature that reports results disaggregated to this level.[20]

I now consider how the aggregate discount rate estimates using our modified MLL compare to other estimates in the literature. Across all estimation methods, the implied annual discount rate in Holloway *et al.* (2026) lies between 9% and 11%, and the present-bias parameter is estimated between $\beta = 0.97$ and 0.98. All estimates are statistically different from one. The tight clustering of these estimates across models indicates that they are not driven by functional form or estimator choice. These aggregate estimates align closely with DMPL results. Andersen *et al.* (2008) and Andersen *et al.* (2014) report annual discount rates of 9–10% using incentivized experiments with Danish households – nearly identical to the estimates recovered by the MLL. A meta analysis by Matousek *et al.* (2022) of experimental estimates of the discount rate finds a median annual rate of 25%, but with a wide range and right-skew: the 25th percentile estimate is 15%, the 75th percentile estimate is 66% and the mean is 68%. This would place our estimates near the very bottom of the range, however, when Matousek *et al.* (2022) analyze factors that predict heterogeneity in estimated discount rates, the offered choice delay negatively predicts the discount rate. Our long maximum choice delay of 32 weeks relative to studies employing CTB designs, such as Andreoni and Sprenger (2012) (14 weeks), Andreoni *et al.* (2017) (9 weeks) and Kuhn *et al.* (2017) (15 weeks) report higher implied discount rates: 25–38%, 63–74% and 20–24%, respectively.

In a meta-analysis on CTB studies, Imai *et al.* (2021) the authors find support for $\beta \in [0.95, 0.97]$. For non-monetary rewards, they find substantially lower β, close to 0.88; however, their meta-analysis points to potential over-reporting of $\beta < 1$ in such studies. Restricting to studies with monetary rewards pushes β to be much closer to one, although the authors state "the delay until the issue of the 'current period' ($t = 0$) reward is shown to robustly influence estimated [present bias]." Given this finding, and the closely related work of Balakrishnan *et al.* (2020), our estimate of β appears to be consistent with the existing literature.

[20]Note that the issue of interior choices in the CTB is distinct. In a CTB, corner solutions stem from a lack of utility curvature, which in and of itself poses no challenge to identifying time preferences; a set of CTBs without the interior options is a price list.

III.4.2. Choice Process Data

When experimenters present work involving time preference elicitations, they rarely have (or report) disaggregated measures of study modules. I report such measures here, but note that ideally there would exist benchmarks from other studies for comparison. We collected choice-level timing data starting in Week 4. We also have timing data at the front-end delay level in Week 2, i.e. we have the total time taken for the first three and second three questions, and can therefore construct total choice time and hence time-per-choice measures for all but Week 0.

Figure III.2 shows the choice-level timing data aggregated across weeks 4–12. Recall that the order of choices was held fixed across surveys for participants' ease. Correspondingly, the first choice in a survey – a two-week delay with no front-end delay – takes the longest (23.6 seconds on average), and then all subsequent choices are much faster (11.2 seconds on average). There is a slight reset when the two-week front-end delay is added, but it is small relative to the first choice in the survey. Overall, the average (median) choice in our modified MLL takes 12.6 (8.6) seconds. Because of the automated choice design with the sliding bar, this means a subject effectively completes 22 binary choices in 12.6 seconds on average, roughly 0.6 seconds per binary choice.

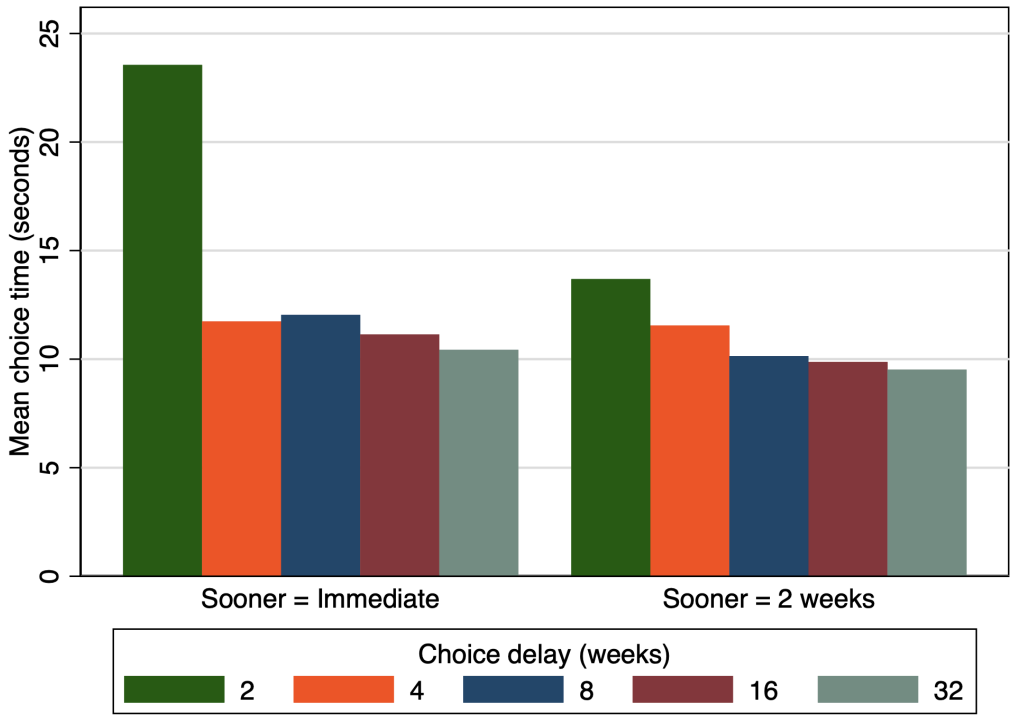


Figure III.2: Choice times, survey weeks 4–12

There is also improvement in time-per-choice over the course of the study, albeit less dramatic than the increase in speed from the first to second choice within a survey. Figure III.3 shows the average time per choice across weeks, this time including the semi-aggregated timing data from Week 2. Average (median) choice time declines from 19.0 (15.6) seconds in Week 2 to 12.4 (9.6) seconds in Week 12.[21]

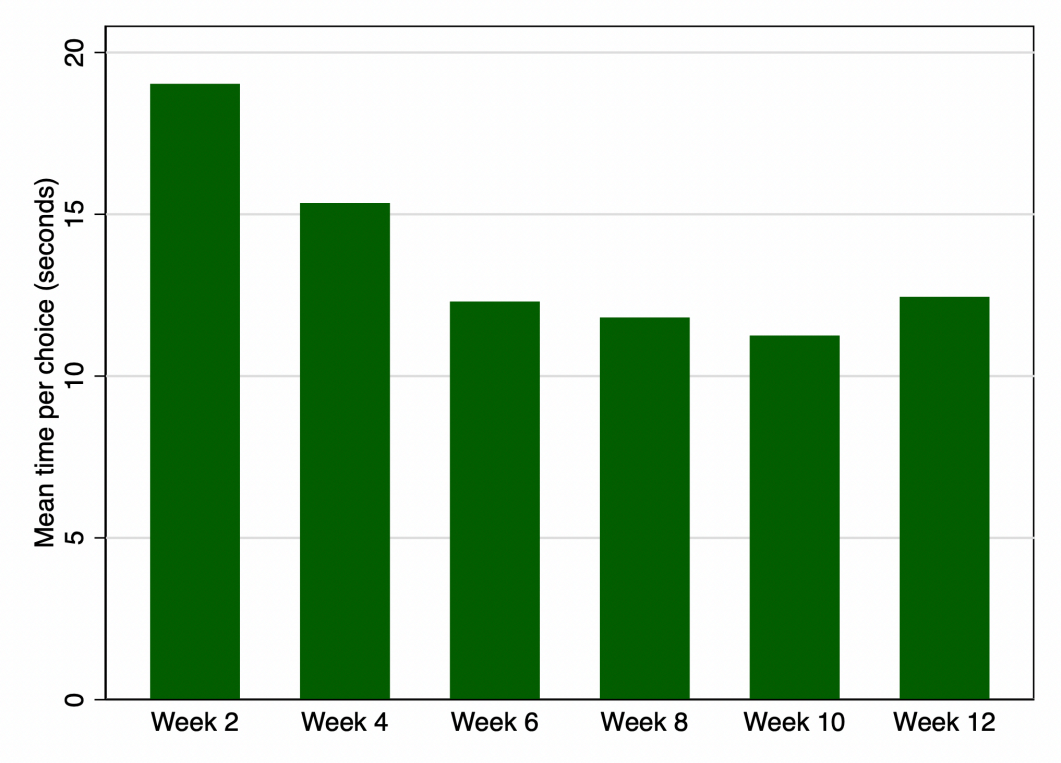


Figure III.3: Time per choice, survey weeks 2–12

To identify a subject's discount function with precision, especially if it features multiple parameters, the experimenter needs to present a series of choices. Moreover, when planning a study, experimenters typically need to account for how long it will take the slowest subjects to finish. Thus, I also present the full distribution of choice completion times for our ten-question elicitation that was held constant during weeks 6–10. This is shown in Figure III.4, with the 90th, 95th, and 99th percentiles of the distribution marked. The median subject finishes the entire set of ten questions in 1.7 minutes. While the distribution is right skew due to some outliers, the average is also under two (1.9) minutes. 90% of subjects finish in under three minutes, 95% of subjects finish in under four minutes, and 99% of subjects finish in less than eight minutes.

[21]Week 10 features slightly faster choices, with an average of 11.3 seconds and a median of 9.1 seconds.

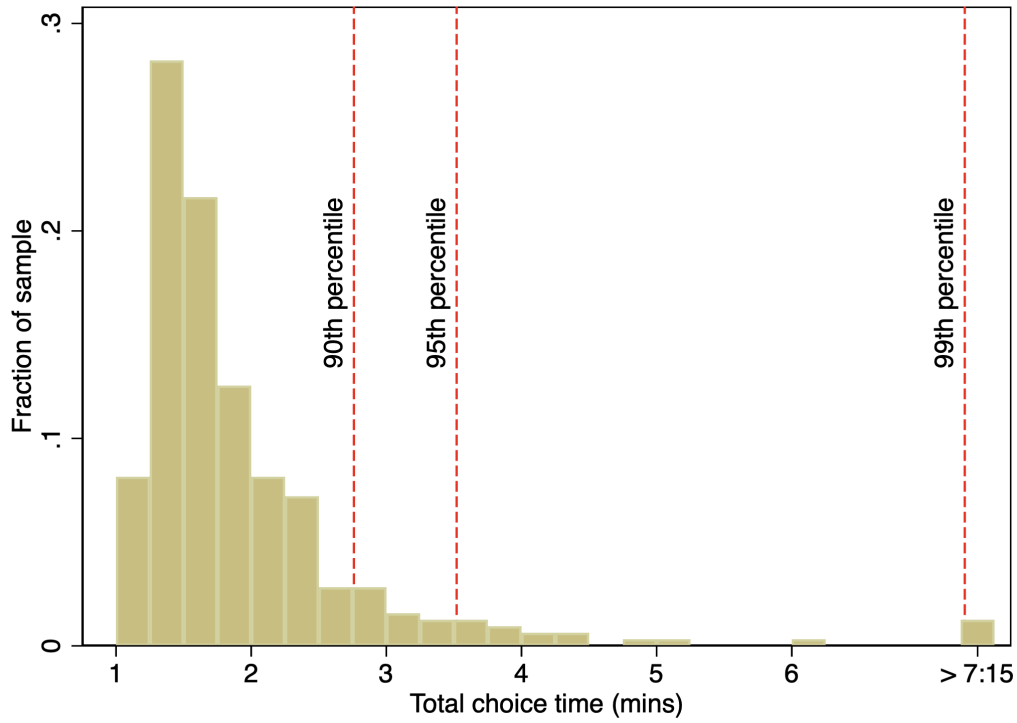


Figure III.4: Distribution of total choice time in full-length surveys

Notes: Full-length surveys are those with 10 lottery lists, occurring in survey weeks 6–10.

While quick choices are of immense value to experimenters, they must be informative and accurate as well. Beyond matching estimates in the literature well and showcasing key theoretical properties as discussed in the previous section, I consider here whether choice timing is itself correlated with the content of a choice. If, for example, the drag-and-drop slider bar inconveniences a subject who prefers the delayed reward by requiring extra time to move to the bottom of the screen, this could lead to biased estimates and attenuated measures of discounting heterogeneity. To be clear, it could be that patient people take longer to make their choices in general. Instead, I focus on whether within-individual variation in choice time is correlated with the switch point they choose. Specifically, I estimate the model

$$d_{i,j,w} = \alpha_i + \gamma_j + \theta \cdot s_{i,j,w} + \varepsilon_{i,j,w} \tag{8}$$

where as above $d_{i,j,w}$ is the midpoint of the switch interval for subject i on choice j in survey week w, α_i are individual fixed effects, γ_j are choice fixed effects, and $s_{i,j,w}$ is the standardized response time. Unlike in the previous section, $d_{i,j,w}$ is standardized here to aid in interpreting the magnitude of this correlation rather than inform the amount of discounting. I limit the sample

to weeks 6–12, such that all choices – combinations of front-end delay and choice delay – are presented to subjects at the same time in each survey they encounter.

I also estimate the model

$$\delta_{i,j,w} = \alpha_i + \lambda_c + \mu \cdot f_j + \theta \cdot s_{i,j,w} + \varepsilon_{i,j,w} \tag{9}$$

where $\delta_{i,j,w}$ is the (standardized) weekly exponential discount factor associated with assuming the midpoint of the switch interval, $d_{i,j,w}$, represents a subject being indifferent between the sooner and later prize draws. If this is the correct model of discounting behavior, then it is unnecessary to adjust this model for the choice fixed effects, γ_j in (8), offering a more precise test. Under misspecification (e.g. quasi-hyperbolic discounting), the model needs to account for front-end delay, f_j, or in a more extreme case (e.g. hyperbolic discounting), the model needs to adjust for choice delay, which I do using choice delay fixed effects, λ_c. Thus, I estimate four versions of (9), first with both $\lambda_c = 0$ for all c and $\mu = 0$, next with either restriction, and finally with no restrictions. Estimates of θ and μ from both (8) and (9) are in Table III.2.

Table III.2: Correlation between choice time and choice content, weeks 6–12

Dependent variable (std. dev.):	d	δ			
	(1)	(2)	(3)	(4)	(5)
Response time (std. dev.)	0.016	−0.052*	−0.050*	0.009	0.012
	(0.015)	(0.028)	(0.028)	(0.027)	(0.027)
Front-end delay			0.050*		0.059**
			(0.029)		(0.026)
Choice fixed effects	Y	N	N	N	N
Choice delay fixed effects	N	N	N	Y	Y
Observations	3,680	3,680	3,680	3,680	3,680

Notes: OLS regressions with heteroskedasticity-robust standard errors.

Overall, I find little evidence to suggest a mechanical correlation between response time and choice outcomes induced by the task design. The estimates of θ in columns (2) and (3) are negative and marginally statistically significant, suggesting that subjects tend to make *less patient* choices when the response time is longer. This is the opposite of what I would anticipate given a mechanical correlation induced by the design given that it takes longer to move the bar down and make a *more patient* choice. In addition, those correlations disappear when I add choice delay

fixed effects to correct for misspecification of the discount function.[22]

III.5. Discussion

This paper presents an efficient and user-friendly implementation of the Multiple Lottery List (MLL) for estimating time preferences. By reducing compound risk, avoiding third-party lotteries, enforcing single switching, and automating choice selection through an intuitive digital interface, the modified MLL preserves the theoretical advantages of the original design while substantially improving practical feasibility. Empirically, the implementation performs well: interior switching is the norm, individual discounting parameters can be estimated for essentially all subjects, aggregate discount rates align with benchmark DMPL estimates, as does present bias, which remains economically meaningful. At the same time, the task is fast. Once explained, subjects complete long, dense price lists in seconds rather than minutes, making the tool well-suited for longitudinal designs and multi-module experimental studies.

That said, careful comparison across elicitation techniques remains an open empirical task. The original MLL implementation allowed for a within-subject comparison with the CTB technique, but our implementation was folded into a larger longitudinal study design without such an opportunity. Thus, my comparison of the modified MLL to CTB and DMPL techniques necessarily spans different samples, stakes, and institutional details. A true within-sample comparison – administering the original MLL, the modified MLL, CTB, and DMPL under identical conditions – would provide sharper evidence on how much of the variation in estimated discounting reflects underlying preferences versus design features.

At the same time, our longitudinal implementation highlights a dimension largely absent from the existing literature: statistical power and repeated measurement. By observing dense price lists over multiple waves, we obtain unprecedented leverage to estimate individual-level discount functions and classify subjects by discounting type. This level of individual discounting classification is rarely available in cross-sectional laboratory studies and provides a richer foundation for studying heterogeneity, stability, and functional form. In this sense, the contribution of the modified MLL is not only procedural but also inferential: it enables designs that move beyond point estimates of average discounting toward a deeper characterization of intertemporal choice at the individual level.

[22] Indeed, Holloway *et al.* (2026) find that the discount function that best characterizes these data is in the hyperbolic family.

Taken together, the evidence suggests that our modified MLL offers a scalable, theoretically coherent, and empirically stable approach to measuring time preferences. This is particularly relevant in settings where immediacy is behaviorally meaningful, where repeated measurement is central to the research question, and where researchers' time with subjects is constrained.

Bridge

The modified MLL provides a scalable and theoretically coherent tool for measuring time preferences in remote and longitudinal settings. By reducing compound risk, enforcing single switching through an intuitive interface, and streamlining dense price lists, the instrument enables reliable recovery of individual discount parameters under repeated observation. In this way, it contributes both procedural infrastructure and empirical stability to the study of intertemporal choice. Yet forward-looking decision-making extends beyond intertemporal tradeoffs. Many economically consequential choices unfold in strategic environments where outcomes depend on the beliefs and actions of others. In such contexts, expectations about future behavior—and the credibility of promises—become central. The final chapter turns from individual intertemporal valuation to strategic interaction, examining how trust is shaped in remote experimental settings and how artificial intelligence systems influence cooperative behavior. Together, these investigations connect temporal preferences with the broader dynamics of forward-looking economic decision-making.

IV. DOES AI FACILITATE TRUST? AN EXPERIMENTAL STUDY

This chapter is based on co-authored work with Ethan Holdahl, Jiabin Wu, and Tanner Bivins. The project was conceived by Ethan and Jiabin, with Tanner and I providing feedback on experiment design, practical implementation, and testable implications. Programming for the project was primarily done by Ethan, Tanner, and I. Much of the prompt design was developed by Ethan, Jiabin, and Tanner. Jiabin oversaw the recruitment and management of study participants, coordinated experiment sessions, and provided payment to all of the participants in the study. The experiment sessions were conducted as a collaborative effort between all four authors. I led the effort of assigning and reviewing the labels for the message data, with all authors providing scrutinous review of the labels. I was primarily responsible for the message classification method, including choice of the normalized edit distance and clustering analysis. Jiabin wrote the first draft of the paper's introduction and theory sections, with the remaining authors contributing individual pieces of the results sections as well as subsequent editing of the entire manuscript.

IV.1. Introduction

Trust is a cornerstone in various socioeconomic activities, including partnership formations and financial transactions, where it transcends mere contractual obligations. It is vital in relationships ranging from personal bonds, like those between spouses, to professional associations, such as the lawyer-client dynamic, procurement agencies and contracted firms, and collaborations between researchers and participants in scientific studies (Charness and Dufwenberg, 2006). Extensive research demonstrates its significance in several financial dealings: it influences stock market participation (Guiso *et al.*, 2004; 2008), affects consumer credit (Brown *et al.*, 2019), determines the use of investment advisers (Gurun *et al.*, 2017), and helps foster healthy lending relationships between lenders and borrowers in the credit markets (Fisman *et al.*, 2017; Hyndman *et al.*, 2024).

Building trust relies on successful communication among involved parties. It is particularly dependent on the ability of the trustee to convince the trustor to place their trust in them. This becomes challenging when the communication is non-binding and the interactions are not regulated by formal agreements. Existing experimental studies have shown that a certain type of communication, namely promises, can promote trust. This can occur through two main channels: guilt aversion, where the trustee experiences guilt for not fulfilling the trustor's expectations (Charness and Dufwenberg, 2006), and promise-keeping, where the trustee feels remorse for not honoring their promises (Vanberg, 2008).

In the era of artificial intelligence, large language models (LLMs) like the Generative Pre-trained Transformer (GPT) have become integral to human communication.[1] These models assist in various tasks such as drafting emails, generating financial reports, improving academic papers, and editing other written materials.[2] This raises a pertinent question: what role can AI play in the trust-building process between a trustor and a trustee? Understanding this role is crucial for practical applications. For instance, in a bank loan application, a borrower might use AI to compose their application, while the lender might use AI for interpreting the application. The advantages of AI in this context are clear: it saves time for both parties, enhances the clarity and substance of the borrower's application, and helps the lender quickly understand the key points. However, AI also has potential drawbacks. It could diminish the authenticity of the borrower's

[1]GPT is created by OpenAI, significantly influencing the field of natural language processing (OpenAI, 2022; 2023a). Brown *et al.* (2020) show that ChatGPT can produce texts with such remarkable accuracy and fluency that it closely resembles human writing, making it challenging for human evaluators to differentiate between text generated by GPT and that authored by humans.

[2]In addition, LLMs have shown remarkable capabilities across diverse areas. They are capable of creating computer code, as demonstrated by Chen *et al.* (2021), and solving university-level mathematics problems (Drori *et al.*, 2022).

application, leading the lender to question its veracity. Moreover, if the borrower heavily relies on AI, they might feel less committed to the content produced by the AI, posing a risk to the lender.

This study aims to explore the impact of AI on trust-building through communication in a controlled experimental setup. We use a two-player binary choice sequential trust game, as described in Charness and Dufwenberg (2006), where Player A (the trustor) makes the first move, followed by Player B (the trustee). Before the game begins, Player B is allowed to send a free-form message to Player A. The experiment is structured as a 2x2 design. In one aspect, we either provide or withhold AI as a tool for Player B to aid in composing their message. In the other aspect, we either provide or do not provide Player A with AI to assist in interpreting Player B's message.[3] The comparison across different treatment groups enables us to understand the influence of the AI assistant's presence (for either Player A or Player B) on the players' decisions and beliefs. Additionally, analyzing within each treatment group sheds light on how players leverage AI to support their communication and decision-making processes.

We find that the presence of AI does not significantly impact the individual choices of the two players. This result may be attributed to several factors. When trustees receive AI assistance, they are more likely to send a "promise" message to the trustor but are less likely to honor a "promise" if it is suggested by AI. When trustors receive AI assistance, they closely follow the suggestions from their AI assistants, who consistently remind them to choose cautiously.

However, we observe a significant increase in the frequency of cooperation between the trustors and the trustees when trustees have access to AI. This outcome can be attributed to trustors becoming more vigilant, realizing that messages from trustees might be partially or fully generated by AI. This heightened scrutiny by trustors, combined with trustees' awareness of being closely examined, may create a mutual understanding and alignment of expectations, thereby fostering an environment where trust can thrive.

The paper is organized as follows. Section 2 provides the experimental design and proposes the main hypotheses. Section 3 analyzes the experimental results and tests the hypotheses. Section 4 concludes.

[3]We employ the GPT-3.5-turbo model (OpenAI, 2023b) in the experiment. We carefully design prompts to ensure AI fully comprehends the trust game and understands its role as an assistant for a specific player in the game during the communication phase. See OpenAI (2023) for guidance on prompt design.

IV.1.1. Literature Review

With its remarkable ability to comprehend and produce language akin to humans, social scientists are developing a growing interest in examining machine-learned large language models. A growing literature (Aher *et al.*, 2023; Argyle *et al.*, 2022; Brand *et al.*, 2023; Brookins and DeBacker, 2023; Bybee, 2023; Chen *et al.*, 2023; Engel *et al.*, 2024; Fan *et al.*, 2023; Guo, 2023; Hagendorff, 2023; Horton, 2023; Kosinski, 2023; Leng and Yuan, 2023; Lorè and Heydari, 2023; Ma *et al.*, 2023; Phelps and Russell, 2023; Strachan *et al.*, 2024) demonstrates that utilizing approaches common to economic and psychological research, such as surveys and laboratory-style experiments is useful in analyzing whether AI mirrors human behavior in individual decision-making tasks as well as in strategic interactions. In these studies, AI serves as the primary subject of investigation, as opposed to humans. A separate strand of research focuses on experiments involving human interaction with machines or AIs, specifically to ascertain whether human responses differ as opposed to interaction with other humans, whether AI players outperform their human counterparts, and whether AI and humans would generate any principal-agent type conflict (Bauer *et al.*, 2023; Cohn *et al.*, 2022; Dvorak *et al.*, 2024; LaMothe and Bobek, 2020; Laudenbach and Siegel, 2024; Mello *et al.*, 2016; Phelps and Russell, 2023; Schniter, 2024).

Contrasting with the above-mentioned literature, our paper focuses on human interactions in which AI plays the role of an assistant rather than an autonomous decision-maker. Several recent papers fall into this category. For example, in Harris *et al.* (2023), the sender can acquire information about the receiver from an AI oracle in a Bayesian persuasion environment. Bai *et al.* (2023) considers whether the first mover takes advice from an AI in a two-player centipede game. Serra-Garcia and Gneezy (2023) find that algorithmic tools help people detect deception in a classic TV game show. Our contribution differs not only in paradigm but also in the theoretical mechanism we can identify. Existing settings primarily speak to how AI changes information transmission, strategic reasoning, or deception detection. A trust game with pre-play communication allows us to address a distinct theoretical gap, namely whether AI strengthens trust by improving communication about intentions, or instead weakens trust by reducing the credibility and behavioral force of voluntary commitments. Our design speaks directly to this gap by examining whether AI changes the frequency of promises, how promises shape trustors' beliefs about trustees' trustworthiness, and how often such promises are subsequently honored.

Our research adds to the discussion on algorithm aversion and trust in AI (Glikson and Woolley, 2020). The low number of subjects directly adopting the AI-suggested messages in our experiment confirms the common belief that people often mistrust algorithmic advice,

even when it's advantageous to follow it (Dietvorst *et al.*, 2015). However, it's important to note that our observations are influenced by two key factors. 1) People generally tend to place greater trust in their own judgment compared to that of others. Our experimental design does not allow us to determine whether the low adoption of AI-recommended messages is due to aversion to AI or excessive confidence in one's own judgment.[4] 2) As pointed out by Logg *et al.* (2019), individuals tend to be more receptive to algorithmic advice in areas where there is a clear and measurable external standard of accuracy, such as making investment decisions or predicting sports outcomes. In contrast, AI's suggestions for interpersonal communication are less easily quantifiable, making it reasonable to assume that participants in our experiment rely more heavily on their own judgment in such cases.

Finally, an increasing number of economic studies, including Acemoglu (2022) and Capraro *et al.* (2024), have focused on understanding the impacts of machine learning and artificial intelligence on socioeconomic phenomena, covering diverse areas like labor force participation, wage disparity, inequalities in education and health care, market competition, consumer privacy, economic growth, and political engagement. Our paper contributes to this literature by examining AI's applicability in partnership formation and financial transactions.

IV.2. Experimental Procedures[5]

The objective of this study is to explore the potential of AI assistants in fostering trust dynamics between human participants. To accomplish this goal, we conducted an online experiment based on the classic two-player trust game introduced by Charness and Dufwenberg (2006), by allowing the second player to communicate with the first player via a pre-game message.

In certain treatments of our experiment, Player B, the trustee, is presented with an AI assistant interface before the commencement of the game. The AI interface allows Player B to compose and send a message to Player A, the trustor, prior to the initiation of gameplay. It is important to note that any message sent by Player B occurs before the actual gameplay begins. A depiction of player B's interface, including the game tree, can be seen in Figure IV.1. Outcomes are shown in the order of (π_A, π_B).

---

[4]Our post-experiment survey elicited overall trust in AI from the subjects. However, we did not find evidence that subjects who chose not to adopt AI-recommended messages were more averse to AI compared to other subjects.

[5]The Study is pre-registered (AEARCTR-0011748). IRB approval for this study was obtained from the University of Oregon, Study ID: STUDY00000906. We thank John Clithero, Jon Davis, Keaton Miller, Van Kolpin, Michael Kuhn, for their helpful comments.

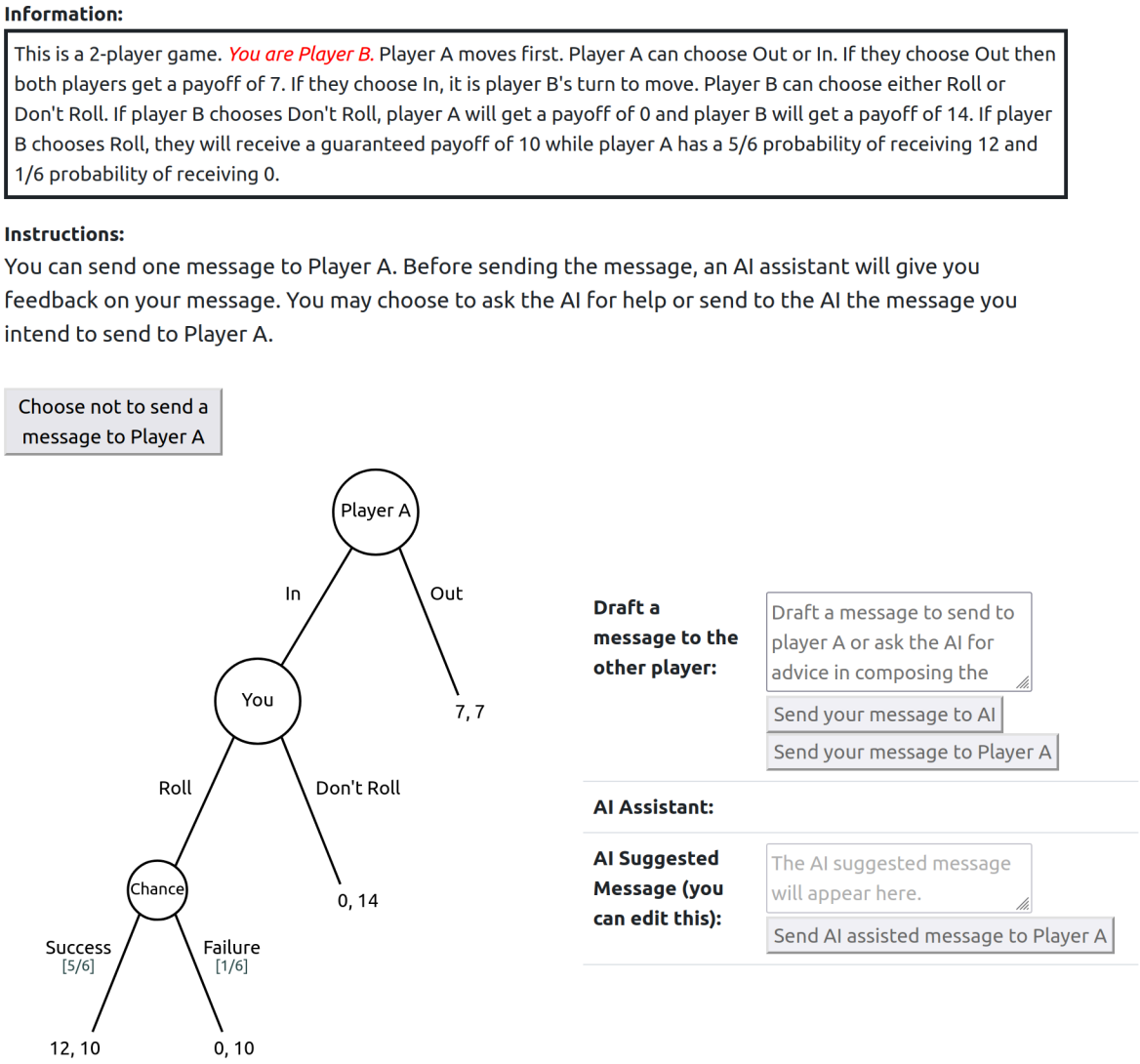


Figure IV.1: Screenshot of player B's screen when they are able to use AI.

Notes: Player A sees a similar tree.

To maximize data collection and ensure simultaneous decision-making by both players, we employed the strategy method (Selten, 1967). Following each player's decision, Player A is prompted to indicate their beliefs regarding Player B's choice, while Player B is asked to provide their own beliefs about Player A's perceived decision. Subsequently, both participants are presented with a Holt-Laury quiz (Holt and Laury, 2002) to assess their risk preferences. Additionally, a demographic survey was administered to gather information on participants' perceptions and prior experience with AI technology.

IV.2.1. Experimental Design

Treatment Design

To investigate the impact of AI assistance on players' payoffs in the trust game, we implemented a 2×2 treatment design, consisting of four distinct treatments. Each treatment explored various combinations of AI presence and absence, shedding light on the role of AI in participants' decision-making processes and outcomes.

- **Benchmark:** Neither player has AI. Player B can choose to send a single message to Player A or refrain from doing so.
- **OnlyA:** Player A has AI to interpret Player B's (potential) message and receive advice on subsequent actions.
- **OnlyB:** Player B has AI to assist in crafting a message to Player A. If Player B opts to send a message, they must first interact with the AI.
- **Both:** Both players have access to AI assistance, following the functionalities described above.

In all treatments, Player B's interaction with AI, if applicable, precedes any communication with Player A. Participants have the option to engage in dialogue with the AI before deciding whether to send a message to Player A. It is crucial to note that all participants are aware of the presence and function of AI throughout the experiment.

Prompt Design

Creating an AI assistant using a natural language processor like ChatGPT involves crafting a prompt that guides the assistant in generating responses. However, designing a prompt that effectively handles a wide range of inputs is more of an art than a science. Our experimentation with ChatGPT (specifically gpt-3.5-turbo) revealed certain challenges and considerations that influenced the development of robust prompts for our study.

We observed that ChatGPT exhibited difficulties in handling sequential logic and tended to perform more reliably with shorter prompts. Moreover, we noted instances where the assistant suggested creative solutions, such as proposing the signing of a contract to establish a binding promise, which were beyond the scope of our experimental setting.

In light of these observations, we refined our prompts by incorporating the following principles:

1. We presented the trust game in its normal form, as the sequential form did not strategically differ from its normal counterpart given the strategy method employed in our experiment.

2. We omitted the description of probabilistic outcomes resulting from ('In', 'Roll') in the game for Player B's AI only and instead provided ChatGPT with the expected outcomes.[6]
3. We explicitly instructed ChatGPT not to propose side deals or disclose players' personal information.
4. We refrained from including higher-order beliefs for the AI assistants. Specifically, Player B's AI did not possess knowledge of the existence of Player A's AI.

The full prompts used in our study can be found in Appendix Section C.1.

IV.2.2. Experimental Procedure

A total of 240 subjects, 30 pairs per treatment, were recruited from to participate in this experiment. On average, subjects spent 15 minutes in the experiment and were paid a $5 show up fee and earned an additional $8.36 during the experiment[7]. The experiment was programmed using oTree (Chen *et al.*, 2016) and ChatGPT version gpt-3.5-turbo. It was implemented on Prolific.

During the online experiment, subjects were continuously recruited and dynamically assigned roles and partners to play the game with to minimize subject wait time. Subjects read instructions then were assigned their role. They then took a quiz to ensure they understood the payoff structure before getting paired with a partner and playing the game. An example of the instructions used in the experiment can be found in Appendix Section C.2.

IV.2.3. Hypotheses

In this section, we lay out the main hypotheses that we tested in the experiment.

Hypothesis 1. *When player B has access to an AI, player A will play 'In' more frequently.*

We hypothesize that using an AI will result in player B sending a message that is more likely to elicit trust from player A.

Hypothesis 2. *When player A has an AI assistant, they will play 'In' less frequently.*

We hypothesize that an AI assistant for player A will result in more conservative decisions from player A as the AI may call attention to the risk of playing 'In' and player B playing their dominant strategy: 'Don't Roll'.

---

[6]We still keep the probabilistic outcomes resulting from ('In', 'Roll') in the game for Player A's AI because we want the AI is able to remind Player A to take risks into consideration.

[7]The equivalent of $33.44/hour, on average. Prolific maintains an $8/hour minimum that researchers must pay to subjects.

Hypothesis 3. *When player B has an AI assistant, they will be more likely to promise to choose 'Roll'.*

Player B's AI assistant is instructed to help it's user maximize it's payoff, which means the AI should be tying to help player B convince player A to play 'In'. In instances where Player B didn't initially make a promise, their AI might suggest it's user making a promise.

Hypothesis 4. *When player B's message to player A contains a promise which originated from the AI, player B is less likely to honor the promise.*

Player B may feel a decrease in the cost of breaking a promise if the promise came from the AI. Consequently, they may be less likely to honor a message that came form the AI.

Hypothesis 5. *The presence of AI, both for player A and for player B, will increase the probability of achieving a cooperative outcome (In, Roll).*

(In, Roll) leads to the highest total payoff for the two players. We expect that the AI should help communicate (when player B has AI) and/or derive (when player A has AI) player B's intentions, so that it helps coordinate cooperative trustors and cooperative trustees.

IV.3. Results

IV.3.1. Treatment Effects

To investigate the influence of AI assistance on outcomes, Figure IV.2 presents the decisions made by each player when (i) neither player receives AI assistance (baseline), (ii) player A receives AI assistance but not player B, (iii) player B receives AI assistance but not player A, and (iv) both players receive AI assistance.

In the benchmark treatment, player A chooses 'In' 36.7% of the time. This rate increases to 43.3% once player B has the option to use AI. Conversely, in both treatments where player A receives input from AI, 40% of trustors choose 'In' whether or not player B has AI. These results suggest the presence of AI has a negligible impact on player A's choices.

We test **Hypothesis 1** by pooling treatments based on player B's access to AI.[8] We observe that player A chooses 'In' more frequently when player B has AI (41.67%) than without AI (38.33%). However, we fail to reject the null when testing the difference in proportions.

[8] $X^{\text{No AI}} = \{\text{Baseline}, \text{Only A}\}$ and $X^{\text{AI}} = \{\text{Only B}, \text{Both}\}$.

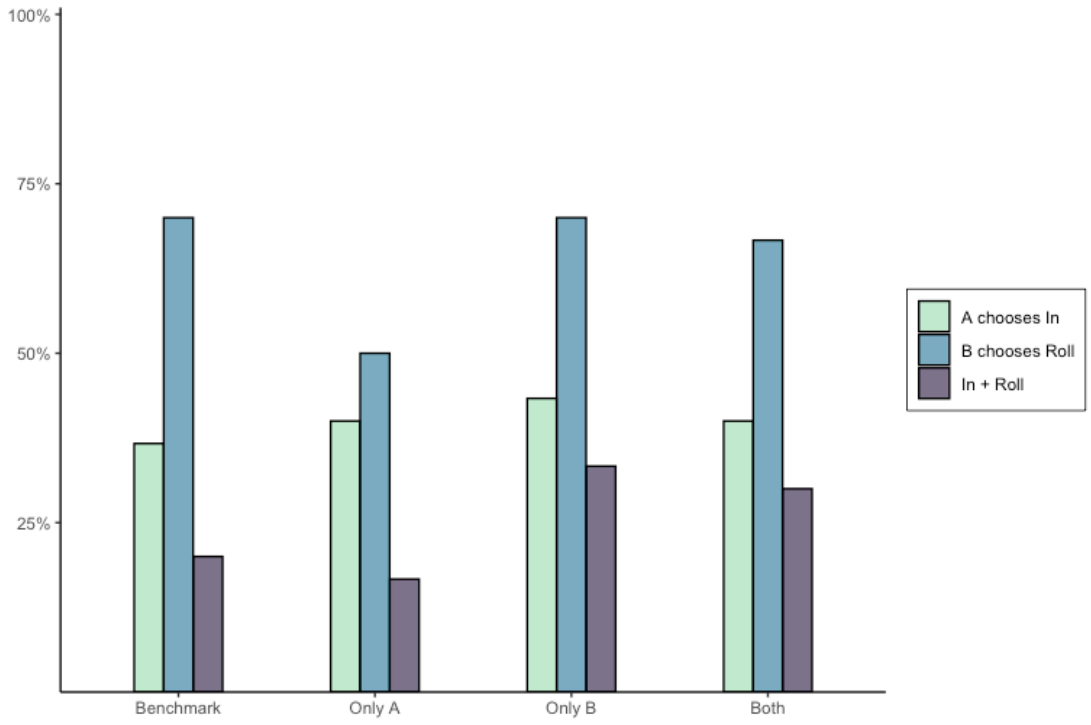


Figure IV.2: Distribution of cooperative choices by treatment

Notes: The percentage of observations in each treatment where player A chooses 'In', player B chooses 'Roll', or both.

Continuing with our analysis of player A's choices, **Hypothesis 2** predicts that player A will be more conservative when receiving support from AI. Again, we pool treatments based on access to AI and find that player A chooses 'In' 40% of the time with and without AI support. Thus, we fail to reject the null.

When neither player has AI, player B chooses 'Roll' 70% of the time. This result remains mostly unchanged in treatments where player B receives AI assistance. Sessions where only player A receives AI assistance stand out with a roll rate of 50% compared to 68.9% across the other three treatments . With access to AI being public information within each group, this suggests the knowledge that only player A will receive guidance from an AI assistant may influence player B's decision.

We refer to the strategy profile ('In', 'Roll') as the *cooperative outcome* as it constitutes the greatest expected collective payoff for each pair. In the benchmark treatment, we observe the cooperative outcome in 20% of pairs. In sessions where only player A receives AI assistance, the proportion of cooperative outcomes decreases to 16.7%, suggesting minimal impact. Conversely, there is suggestive evidence that player B's access to AI improves cooperation. In treatments

where player B has the option to use AI – Only B and Both – we observe the cooperative outcome in 33.3% and 30% of pairs, respectively. Pooling observations by player B's access to AI, we find a 13.34 percentage point increase in cooperative outcomes when player B receives AI assistance . Taken together, the disconnect between individual choices and pair-wise choices across treatments indicates AI assistance may not significantly impact individual decisions; rather, it helps to coordinate cooperative trustors with cooperative trustees. This provides support for **Hypothesis 5**.

We now turn to the impact of AI assistance on beliefs. First-order beliefs (τ_A) represent player A's confidence that player B will choose 'Roll'. Second-order beliefs (τ_{BA}) reflect player B's perception of player A's beliefs. To provide a more intuitive interpretation of results where higher values correspond with increased trust, we map qualitative responses from the post-experiment survey to numerical values according to Table IV.1.

Table IV.1: Numerical values assigned to elicited beliefs.

Certainly Choose Don't Roll	$\longrightarrow$	0
Probably Choose Don't Roll	$\longrightarrow$	0.25
Unsure	$\longrightarrow$	0.5
Probably Choose Roll	$\longrightarrow$	0.75
Certainly Choose Roll	$\longrightarrow$	1

Consistent with existing literature, our findings confirm that beliefs and behavior are closely interconnected. Specifically, we observe that A is more inclined to choose 'In' when they are confident B will select Roll. As selecting 'In' report an average τ_A of 0.74, translating to a belief that B will "probably choose Roll." On the other hand, those choosing 'Out' exhibit reduced trust in B, with a lower average of 0.44. Moreover, Bs who decide to 'Roll' have an average second-order belief of 0.73, while those who choose 'Don't Roll' show a significantly lower average τ_{AB} at 0.37.

Figure IV.3 presents average first- and second-order beliefs across the four treatments. In the benchmark treatment, trustors display an average τ_A of 0.52, indicating a neutral level of trust. The inclusion of AI assistance for B increases this average to 0.61. However, this pattern does not persist across the Only A and Both treatments. When pooling observations by B's access to AI, we find an average τ_A of 0.529 in treatments without access, only slightly less[9] than when

[9] a difference of 0.07.

B is assisted by AI. In line with the patterns observed in choices, there is minimal evidence to suggest AI significantly influences A's beliefs.

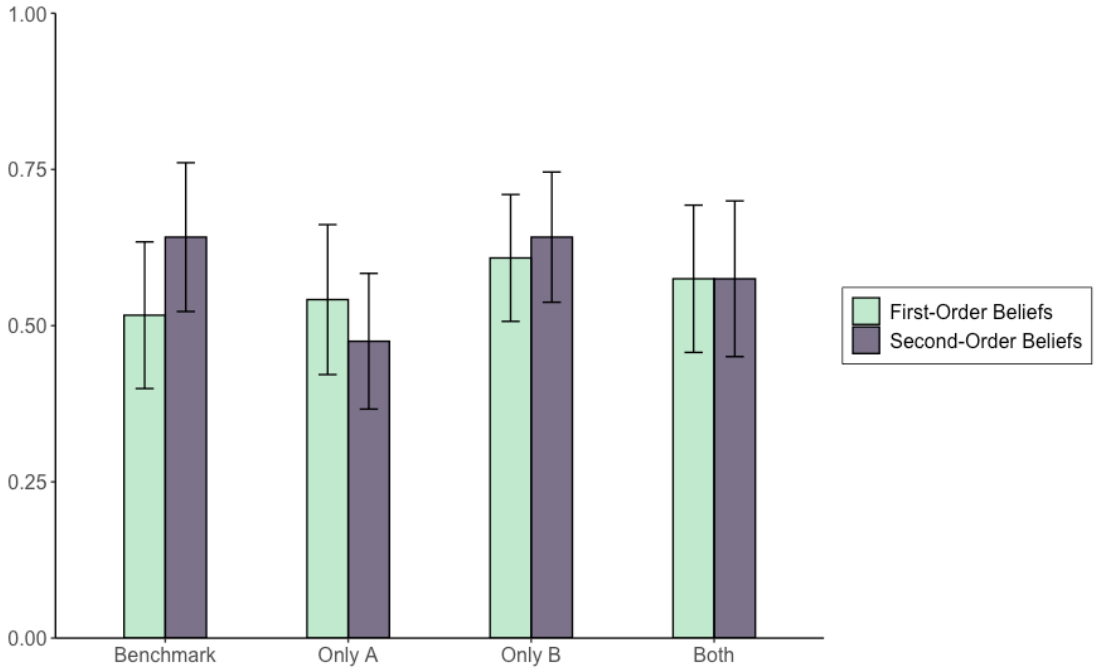


Figure IV.3: Average first-order and second-order beliefs across treatments.

Notes: Confidence bands are calculated at the 95-percent level.

Second-order beliefs appear to be equally inconsistent across treatments, except in cases where only A receives AI assistance. In such cases, B's average τ_{BA} is 0.475, markedly lower than the 0.629 observed across other treatments . This divergence might be attributed to the public information that only A can access AI, leading B to form a more pessimistic perception of A's beliefs. It is also consistent with the low Roll rate in the only A treatment compared with the others.

IV.3.2. Message Classification

Player B

We classify messages sent by player B according to the *type* of message sent and the *method* by which it was sent. The type of message that player B sends takes one of the following values: 'promise', 'asking', 'empty', 'skip', 'fairness', 'anti-promise'. Meanwhile, we classify the *method* into one of 'Own', 'AI', 'Mixed', 'Skip'. Each of these classifications is briefly discussed below, with further details in Section C.3.1 of the appendix.

Message Type

Message type primarily concerns the *content* of the message. We abstract the message sent by player B into what player B indicates about *player A's* potential move and what player B expresses about their *own* intended move. For each of these components, we assign a tertiary label. For the first component – the piece of the message involving player A's move – we assign a label from ∅/*In*/*Out*, with ∅ representing no definitive information conveyed. Similarly, for the second component – on player B's own move – we assign a label from ∅/*Roll*/*Don't Roll*. This categorizes any message sent by either player B or the AI into one of 9 pairs. This codification allows us to assign labels to the messages according to Table IV.2. Additionally, a visualization of the transformation from player B's initial message through the AI to their final message can be found in Appendix Figure C.4, with additional examples following in panels (a)-(e) of Appendix Figure C.5. Note that this assignment is for explicit messages; when player B opts not to send a message, they are assigned the message type "skip". Further details of message type codification are left to the appendix.

Table IV.2: Message-Type encodings

Msg Vec	Label
(*In*, *Roll*)	*Promise*
(∅, *Roll*)	
(*Out*, *Roll*)*	
(*In*, ∅)	*Asking*
(∅, ∅)	*Empty*
(*Out*, ∅)	*Fairness*
(*In*, *Don't Roll*)	*Anti-promise*
(∅, *Don't Roll*)	
(*Out*, *Don't Roll*)	

Notes: Each message sent by player B (or recommended by AI) is encoded as vector which captures what player B intends to do and what they propose player A should do. Note (*) that (Out, Roll) does not constitute a cooperative outcome unlike the other Promises. We nonetheless include it in Promise since it demonstrates Player B's intention to play Roll. We do not observe any encoded message of (Out, Roll).

We are chiefly interested in the effects of player B promising to play *Roll* has on choices, outcomes, and beliefs. We abstract slightly from the notion of a "promise" to any explicitly expressed intention to play *Roll* on behalf of player B. On the other hand, if player B explicitly expresses intent to play *Don't Roll*, we label this an "anti-promise", regardless of their suggestion as to how player A should play. If player B only indicates an explicit move that they wish their opponent to

play, and does not explicitly provide information about what they intend to play, then we classify their message as either "Fairness" or "Asking". "Asking" is chosen if player B requests that player A play *In*, without mention of their own intended move. Conversely, "Fairness" indicates that player B has suggested that player A play *Out*, resulting in an egalitarian ("fair") outcome. In the event that no clear intentions are sent on behalf of player B, then the "Empty" label is assigned.[10]

A breakdown of message types for non-skipped messages across all treatments is provided in Figure IV.4. The largest share (41%) of sent messages are Promises, with 90% of sent messages being comprised of Promises, Asking, and Empty messages.

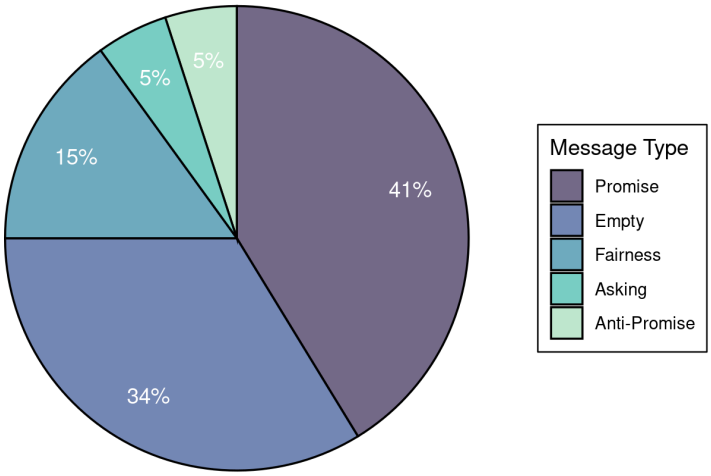


Figure IV.4: Types of messages sent across treatments, omitting skipped messages

Notes: $\approx$ 1/3 of B's in the sample opted to skip sending a message.

Figure IV.5 illustrates the distribution of player choices based on player B's message type. Messages that assure player B will choose 'Roll' or ask player A to choose 'In' result in the highest frequency of A choosing 'In'. We expect messages classified as 'empty' to elicit similar responses from A as cases where B opts out of sending any message, as both situations lack any substantive signal of B's intentions or trustworthiness. However, our data show trustors choose 'In' more frequently when receiving an empty message (37.03%) than no message (25%). Several of the empty messages are disconnected from the experiment itself but contain relatively positive

[10]Note that no truly "empty" messages are sent: all messages sent contain *some* content. Therefore, it may be helpful to think of the "Empty" label as "Junk" based on the complement set of messages already described.

language.[11] It may be the case that sending a positive message can help to establish trust even if the message is irrelevant to the game.

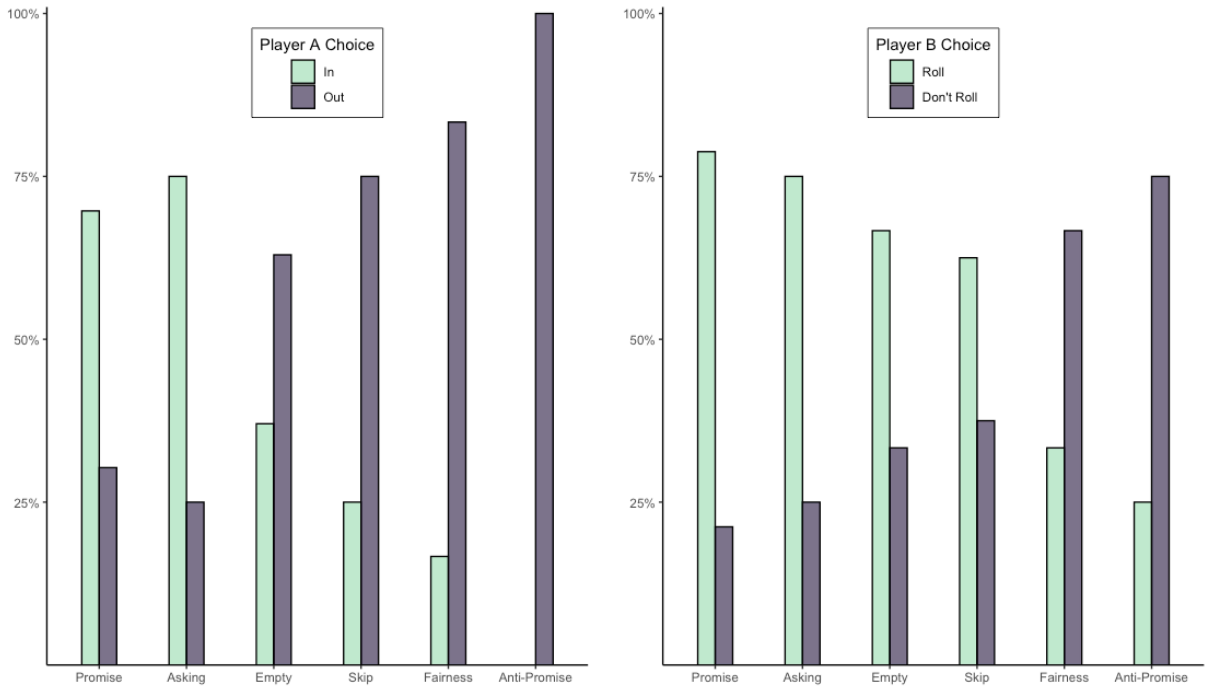


Figure IV.5: Distribution of player choices by message classification.

Messages categorized as 'fairness' and 'anti-promises' share similar purposes but differ in the signals they convey. Fair messages explicitly encourage A to choose 'Out' by presenting it as the safest option, while anti-promises discourage choosing 'In' by disclosing B's intent to choose 'Don't Roll'. As anticipated, both types of messages elicit the lowest in-rates across all pairs. Notably, no trustors chose 'In' after receiving an anti-promise, underscoring the strong deterrent effect of such messages.

Hypothesis 3 postulates that Player B will send more promises when they have access to an AI assistant. Our data show a 75% increase in the proportion of messages containing promises when B gains access to AI . Expanding the notion of a promise to include 'asking' increases statistical significance to the 97-percent level.[12] Specifically, our 95-percent confidence interval indicates the presence of AI for Player B yields a 1.83 to 34.8 percentage point increase in messages categorized as 'promise' or 'asking'. Figure IV.6 displays these findings visually.

[11]For example, one message contained "Hey, I hope we have a good game (:"

[12]We consider 'asking' alongside 'promise' as to include any message which directly suggests that player A play 'In', opening up the possibility for a cooperative outcome.

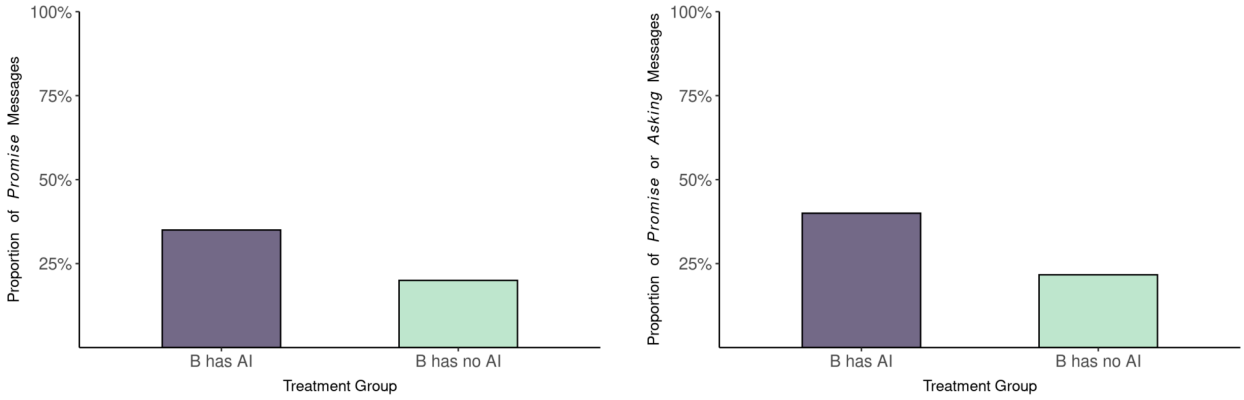


Figure IV.6: Proportion of player B's messages which are promises or asking, by access to AI.

Notes: Left panel: proportion of messages sent by player B which are promises. Right panel: proportion of messages sent by player B which are promises or asking.

Message Method

As opposed to message type, which concerns the content of the message, message method concerns the *authorship* of the message. Since player B may edit the message suggested by the AI before sending it to player A, we aim to discern whether the content of the message was primarily dictated by AI, by the agent, or by a reasonable mix of the two. We use a normalized version of the Levenshtein edit distance (Levenshtein, 1966) to determine the pairwise relative distances between player B's first message, the AI's suggested message, and the actual sent messages[13]. When plotting the normed Levenshtein distance between the the first and sent messages against the AI-suggested and sent messages (Figure C.6 in the appendix), a clear grouping structure can be seen.

To verify this grouping structure, we implement a k-means cluster classification with $k = 3$ means[14] to produce the labeling assignment. Each member of the research team independently inspected each message to ensure accurate labels. The upshot is that sent messages which are labelled 'Own' have near-identical similarity to the first messages player B sent compared to the AI's suggested message; sent messages labelled 'AI' have near-identical similarity to the AI's suggested message compared to the first message player B sent, and 'Mixed' messages player B

[13]The Levenshtein distance is a metric which reports the total number of single-character edits needed to transform one string into another. In particular, the distance measures the number of insertions, deletions, and substitutions required to transform one of its inputs into the other. We implement a normed version of this metric, which scales the traditional Levenshtein distance between two strings by the length of the larger string. This transforms the metric into a measure of similarity between the two strings lying between 0 and 1, as the maximum length of the two input strings is exactly the maxmium number of single-character edits needed to transform one string into the other. For further details, see Appendix Section C.3.2.

[14]See Appendix Section C.3.3 for details.

sent are those which bare a fair similarity to both the first and the AI-suggested message. Further details can be found in the appendix.

Figure IV.7 shows the breakdown of how B players sent their messages, provided that they sent one at all. The majority of sent messages are primarily their own compositions, with near equal shares of messages being crafted entirely by the AI or a mix of AI and player B. It should be noted that both Figure IV.7 and Table IV.3 are restricted to treatments when player B has an AI assistant.

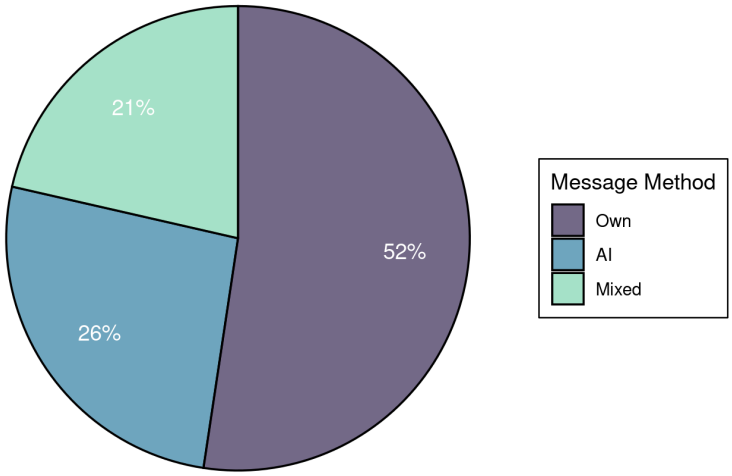


Figure IV.7: Share of Message Methods

Notes: Over half of player B's who sent a message did so using mostly their own authorship. Details of how these methods were imputed are included in the appendix.

Further examination of the effect of AI's usefulness in message composition reveals that players who spend time working with, as opposed to relying upon, the AI assistant received the best results as measured by convincing player A to play 'In'. Figure IV.8 shows the interaction between time spent creating their message and the method used to create their message. When players spent additional time crafting their message (a proxy for effort) they are more successful in soliciting 'In' from player A. This effect was magnified when the player incorporated the feedback from the AI assistant into their message thereby creating a 'mixed' message. However, the same positive effect did not apply to those who relied entirely upon the AI to create the message they sent to player A, if anything the performance of the message decreased over time spent on the AI message.

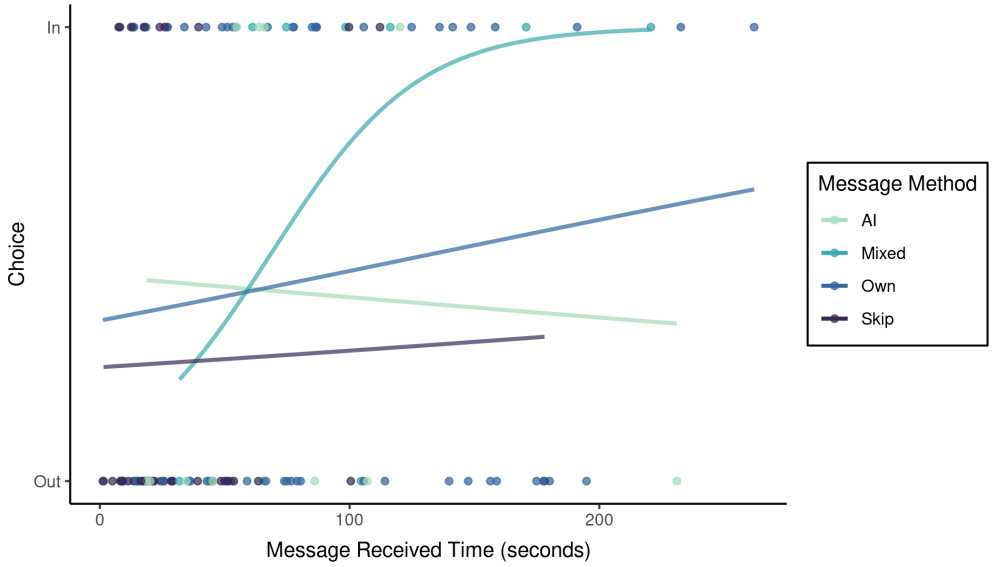


Figure IV.8: The impact of wait-time on player A's choice, by message method.

Notes: Comparison of Player A's choice versus the time in seconds they waited for Player B's message, categorized by Player B's message delivery method. Without an interaction term, time spent waiting significantly increased the probability ($p = 0.03$) that players would choose 'In'.

In contrast, Player A was less likely to comply with Player B's request the longer they waited for the message, as shown in Figure IV.9. As the wait time increased, Player A was more inclined to choose 'In' when Player B asked them to play 'Out,' and more likely to choose 'Out' when Player B requested 'In.'

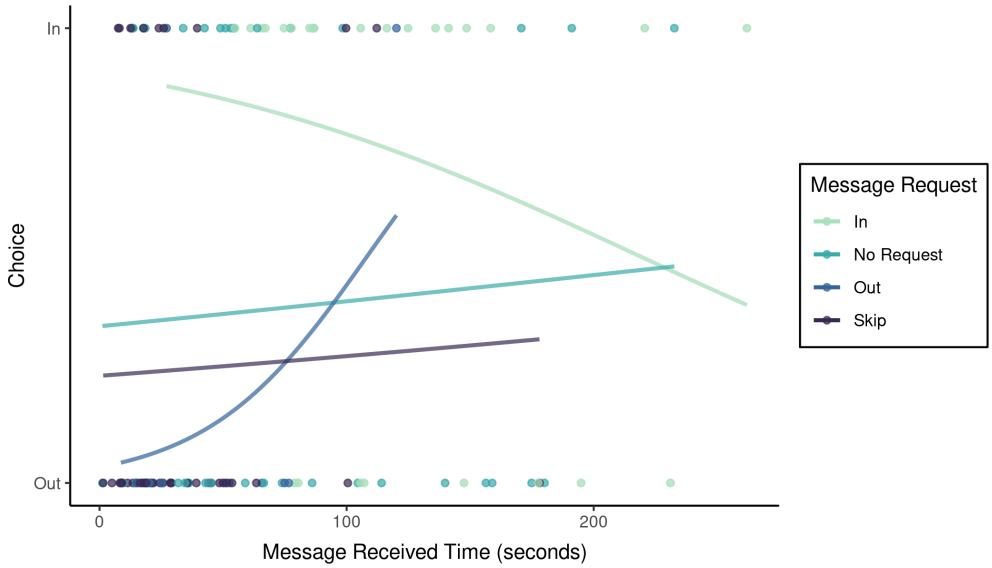


Figure IV.9: The impact of wait-time on player A's choice, by message request.

Notes: Comparison of Player A's choice versus the time in seconds they waited for Player B's message, categorized by the choice Player B requested Player A makes.

Figure IV.10 shows the distribution of player choices by method across the four treatment groups. Without AI assistance, Player B can either opt not to send a message or compose their own. Consistent with earlier results, player A selects 'In' at a relatively low rate of 25% in the absence of any message. When the treatments allow Player B to send messages that are either partially or completely generated by AI ('Only B' and 'Both'), these AI-assisted messages result in unexpectedly lower 'In' rates: 44.4% for 'mixed' messages and 36.4% for fully AI-authored messages, in contrast to 54.5% for original messages. These findings indicate that the authorship of the message may not influence Player A's decisions as much as the actual content of the message.

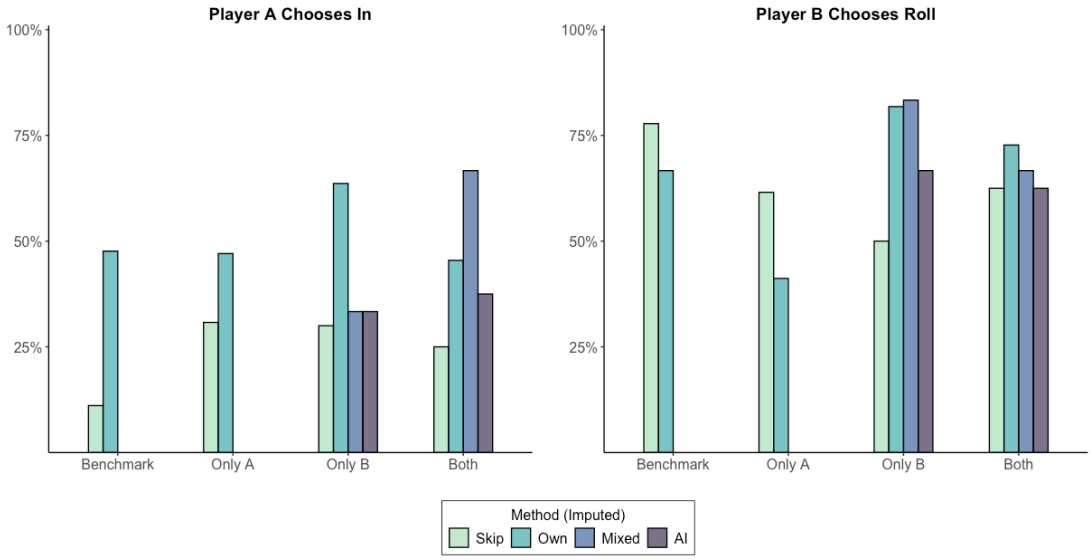


Figure IV.10: Distribution of player choices across treatments by message authorship.

Combining message type with message method, Table IV.3 displays raw counts of the types of messages player B sent and how they sent them.

Table IV.3: Summary of message types by message method

Message Type	Own	Mixed	AI	Total
Promise	9	6	6	21
Asking	0	1	2	3
Empty	9	1	1	11
Fairness	4	1	0	5
Anti-Promise	0	0	2	2
Total	22	9	11	42

Notes: Summary of the number of non-skipped messages sent by player B according to their type (row) and method (columns). Treatment is restricted to cases when player B has an AI.

To test **Hypothesis 4**, we group messages containing a promise to choose 'Roll' according to whether they are authored by Player B or suggested by AI. Our findings indicate that Player B fulfills their promise 85.7% of the time when sending their own message. In contrast, the follow-through rate drops to 40% when the promise is suggested by AI. This decrease suggests using AI to communicate promises may lower the cost of breaking a promise. Despite these findings, given the p-value of 0.137, our study lacks the statistical power to reject the null hypothesis decisively.

Interestingly, there is significant ($p = 0.07$) relationship between the time it took Player B to send their message and choosing which action to play and their choice between 'Roll' and 'Don't Roll', as shown in Figure IV.11. The longer it took for Player B to make their decision after sending a promising message that they would play 'Roll', the less likely they were to follow through with their promise.

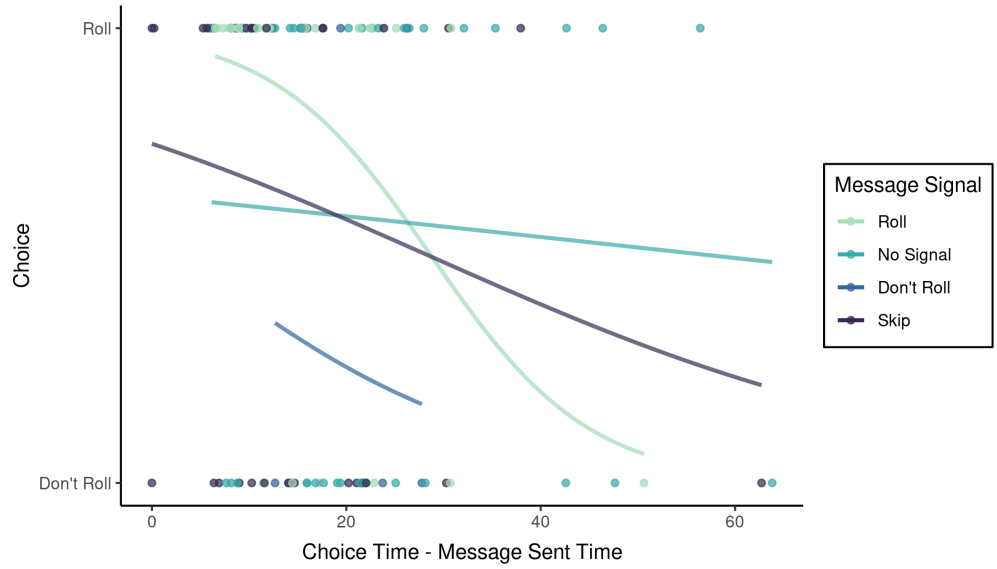


Figure IV.11: Player B's message deliberation time and subsequent choice, by message signal.

Notes: Comparison of Player B's choice versus the time they took to make their choice after sending their message, categorized by Player B's signal of what choice they would make.

Player A

Message classification for player A's AI is naturally less intensive:[15] when player A has an AI, we classify the interpreted message on behalf of the AI[16] as either "no clear suggestion", "strongly

[15]Indeed, much of the time, A's AI addresses player B's message, reviews the potential outcomes of the game, and advises player A to play according to their own risk preferences.

[16]This was done by each member of the research team independently by hand for all messages, and these independent labels were compared and contrasted until unanimity was reached for each message label.

advises playing 'In'", "weakly advises playing 'In'" and "primarily advises playing 'Out'."[17] In Appendix Figure C.8, we collapse strong/weak 'advise-In' into a single label for a symmetric assignment. Furthermore, explicit examples of player B's sent message and the corresponding interpretation from player A's AI can be found in Panels (a) – (e) of Figure C.9 in the appendix.

Figure IV.12 shows the suggestions made by player A's AI compared to the choice which player A ultimately made in the game. Proportionally, it seems that player A closely follows the advice of their AI assistant when the assistant suggests 'Out'. On the other hand, this suggestion is the least frequent of the three in the sample, with only 18% of AI suggesting that player A chooses out.

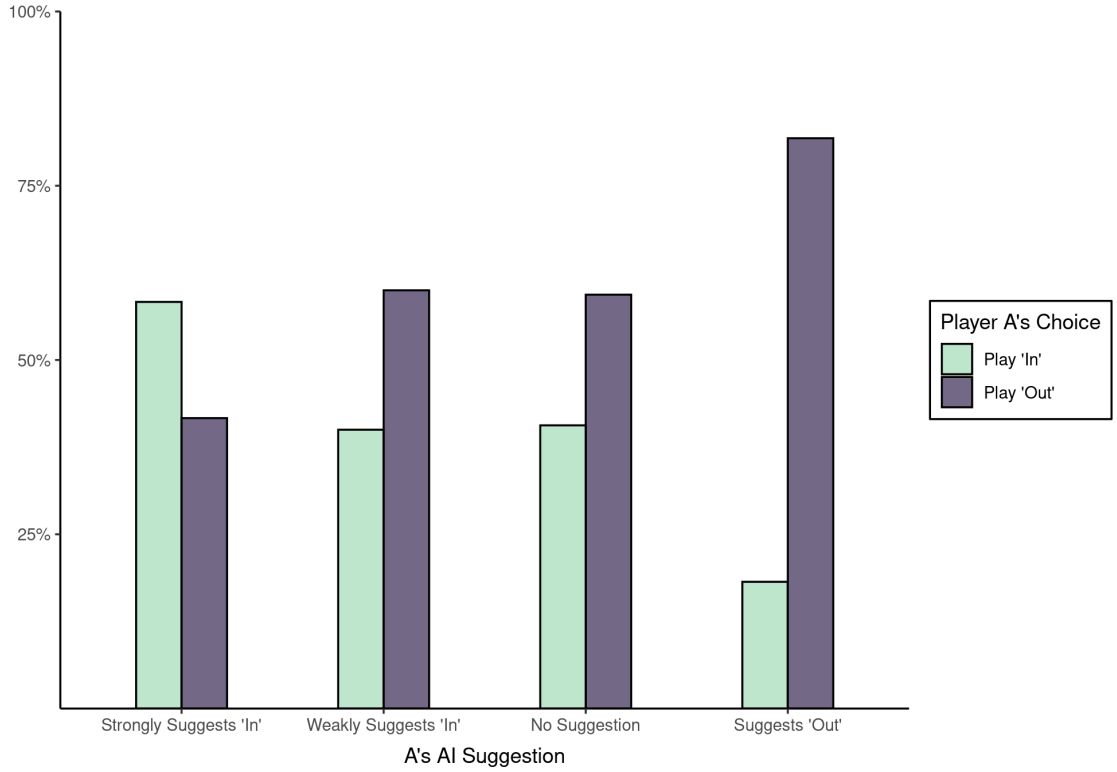


Figure IV.12: Player A's choice and the associated interpretation by player A's AI.

Why do we see such a small proportion of AI suggesting 'Out'? Figure IV.13 displays the suggestions made by player A's AI, this time alongside the type of message that player B sent. Recall that an assignment of 'fairness' indicates that player B made no mention of their own move, rather they simply request that player A choose 'Out'. When player B sends a message of this type, player A's AI is almost guaranteed to advise similarly.

[17] Our data on A's AI messages is absent from the notion of a 'strong' v. 'weak' suggestion of 'Out'.

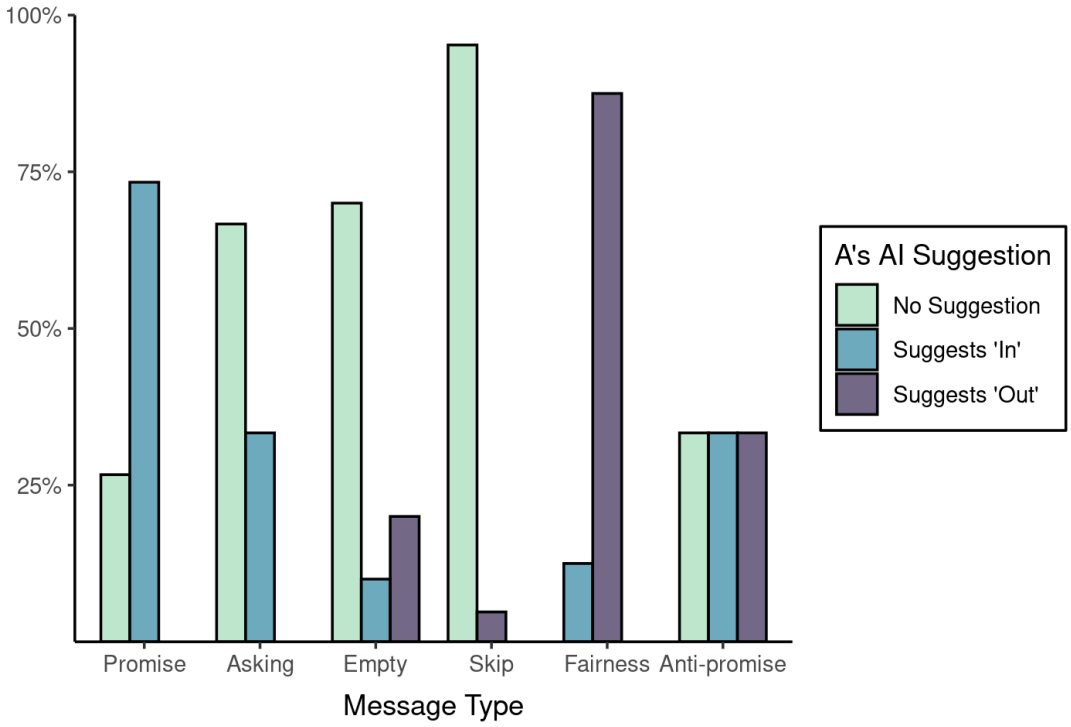


Figure IV.13: Player B's message type and the associated interpretation by player A's AI.

IV.4. Conclusion

This study investigates the impact of AI assistance on trust-building in a two-player trust game. Specifically, we examined scenarios where either the trustor, the trustee, or both were assisted by AI in their decision-making processes. Our primary findings reveal that while AI assistance does not significantly alter individual choices, it does foster cooperative outcomes by coordinating cooperative trustors and trustees. This demonstrates that AI may help foster trust, but in a limited way. Individuals and organizations should cautiously leverage AI tools to assist in communications, particularly in contexts where trust is paramount.

An interesting next step is to run the same treatments where participants are unaware that their adversaries are assisted by AI. We anticipate that the nuances created by AI within communications will significantly impact this scenario, as it removes the initial skepticism participants may have about AI involvement. Another direction that is worth exploring is to run the same treatments with human assistants and compare them with those with AI assistants. Comparing AI and human assistants will enable us to assess the potential of AI to replace human roles in the workforce and to identify advantages and limitations of AI assistance versus human

assistance.[18] The evidence we have gathered suggests that AI is not yet ready to fully replace humans, but the use of AI can enhance a human's ability to communicate and foster trust. Future studies should also explore the long-term effects of AI assistance on trust development. Understanding how trust evolves over repeated interactions and whether initial skepticism diminishes over time can provide valuable insights for both AI development and its applications in trust-sensitive environments. Finally, the advancement of AI technology can significantly influence the effectiveness of AI assistance. Capraro *et al.* (2023) shows that GPT-4 tends to consistently overestimate human altruism, which may result in biased communication in scenarios like the trust game we study. In the current experiment, GPT-3.5-turbo is being utilized, and it would be valuable to conduct new sessions to compare its performance with GPT-4.

[18]Cvetkovic *et al.* (2024) conduct a survey experiment and find that people in Finland exhibit greater trust in humans than in AI when it comes to performing assistance-related tasks at work

V. DISCUSSION

This dissertation has examined forward-looking economic behavior in remote experimental settings from complementary empirical and methodological perspectives. The chapters collectively demonstrate that credible inference about forward-looking behavior depends on deliberate experimental design and precise measurement.

The first chapter demonstrated that measured time preferences can vary within individuals over relatively short horizons. These findings underscore the value of repeated observation and caution against interpreting single-elicitation estimates as fixed structural parameters. The second chapter addressed the measurement challenge directly by developing and evaluating a modified Multiple Lottery List instrument tailored for remote, longitudinal use. By reducing compound risk, eliminating multiple switching, and streamlining dense price lists, the design enables reliable recovery of discount parameters under repeated administration.

The final chapter extended the focus from intertemporal tradeoffs to strategic interaction, examining how trust is shaped in remote experimental environments and how artificial intelligence systems influence communication and cooperative behavior. In doing so, it highlighted how technologically mediated settings can alter expectations, signaling behavior, and the possibility of reaching cooperative economic outcomes.

Taken together, the chapters illustrate both the promise and the responsibility of remote experimental research. Remote platforms allow for high-frequency measurement, scalable designs, and technologically mediated interaction. At the same time, they require deliberate attention to implementation details that shape inference. By combining substantive analysis of time preferences and trust with methodological contributions to experimental design, this dissertation advances the study of forward-looking economic behavior in environments that increasingly define modern interaction.

APPENDIX A

CHAPTER II SUPPLEMENT

A.1 Supplementary tables and figures

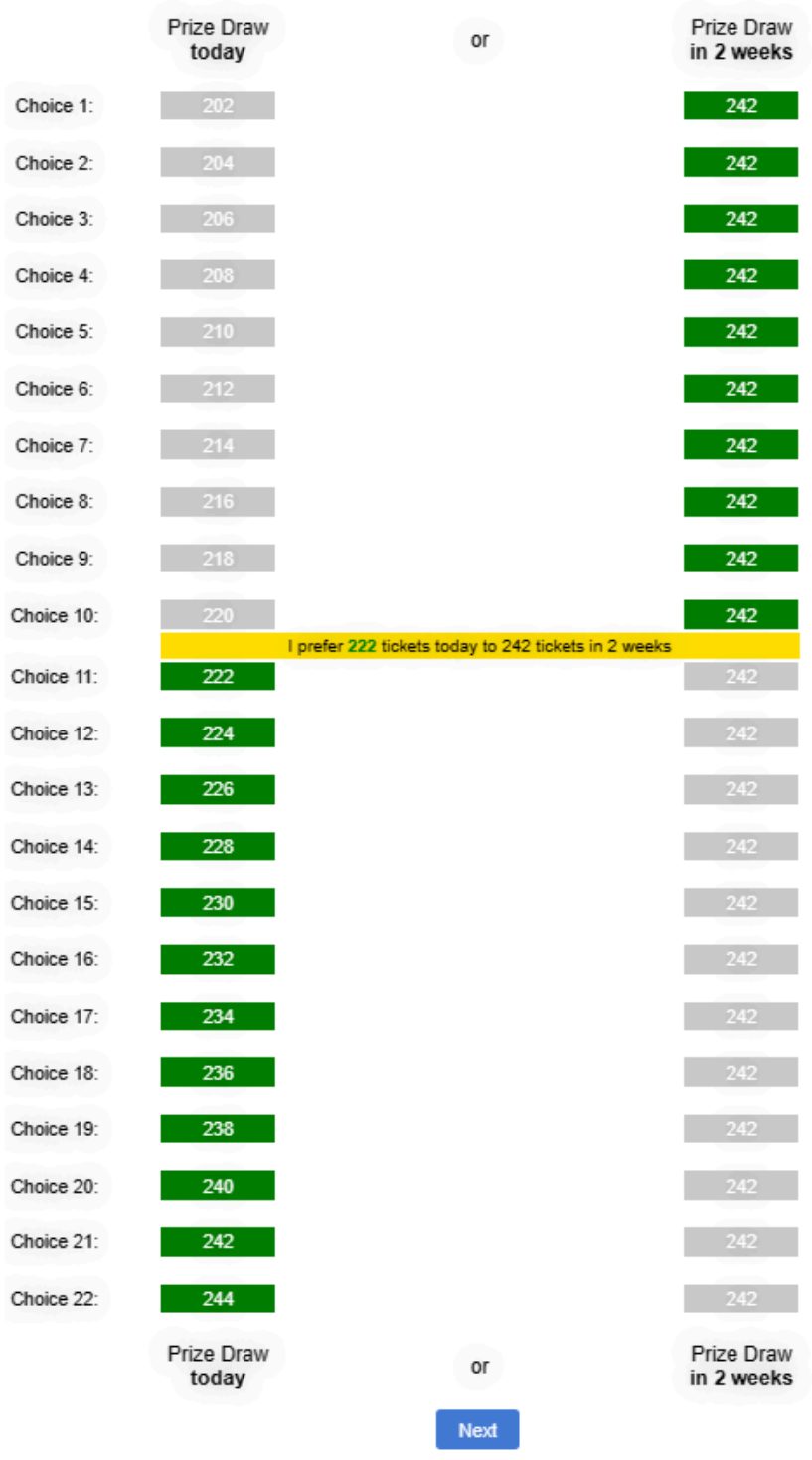


Figure A.1: Example price list

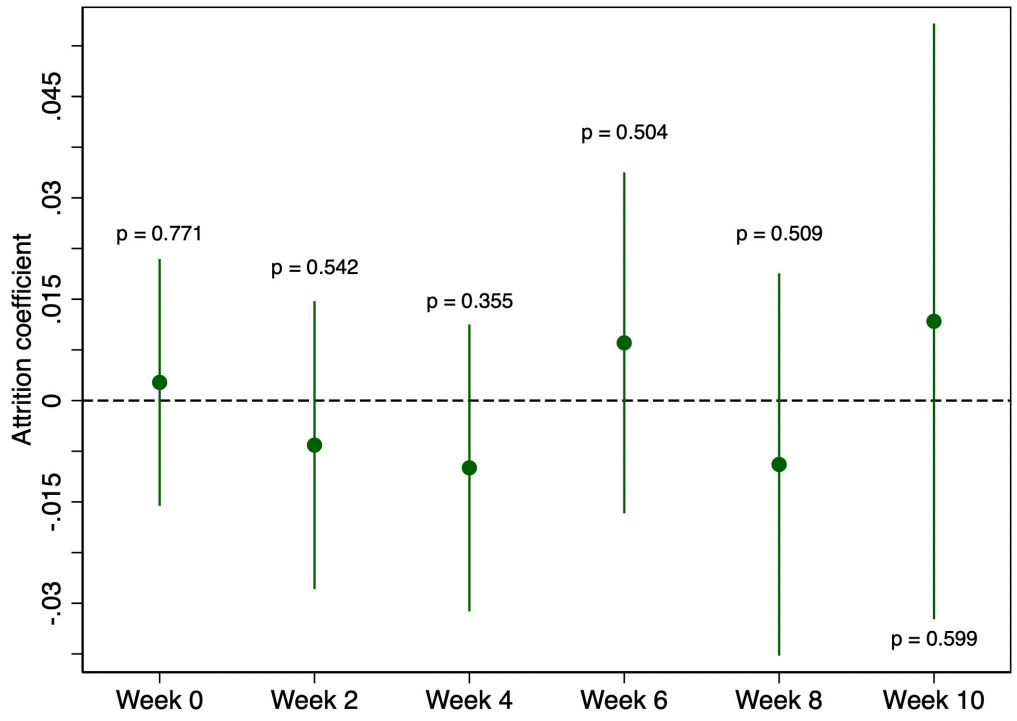


Figure A.2: Relationship between subsequent attrition and discount factor, with 95% confidence intervals

Notes: Coefficients and 95% confidence intervals are from a regression of the discount factor (d) on an indicator for whether a subject fails to complete the subsequent survey. Standard errors are clustered at the subject level. All models feature fully-interacted fixed effects between delay length and front-end delay. Cluster (observation) counts are 153, 138, 117, 115, 106, and 99 (918, 828, 936, 1,150, 1,060, and 990), respectively.

Table A.1: Aggregate discount parameter estimates from quasi-hyperbolic model, full sample

Model:	OLS	Interval regression	NLS	Binary logistic
	(1)	(2)	(3)	(4)
δ	0.9980	0.9980	0.9981	0.9984
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
β	0.9712	0.9716	0.9724	0.9778
	(0.0024)	(0.0025)	(0.0024)	(0.0024)
r (annual discount rate)	0.1074	0.1104	0.1047	0.0877
	(0.0120)	(0.0136)	(0.0118)	(0.0115)
H_0: $\delta = 1$	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$
H_0: $\beta = 1$	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$

Notes: Standard errors are clustered at the subject level. δ is the weekly exponential discount factor, and $r = \delta^{-52} - 1$ is the annual discount rate. The data consist of 6,367 price lists from 153 subjects in the fullsample. The binary logistic model assumes a standard T1EV error distribution, and uses data from 140,074 binary choices.

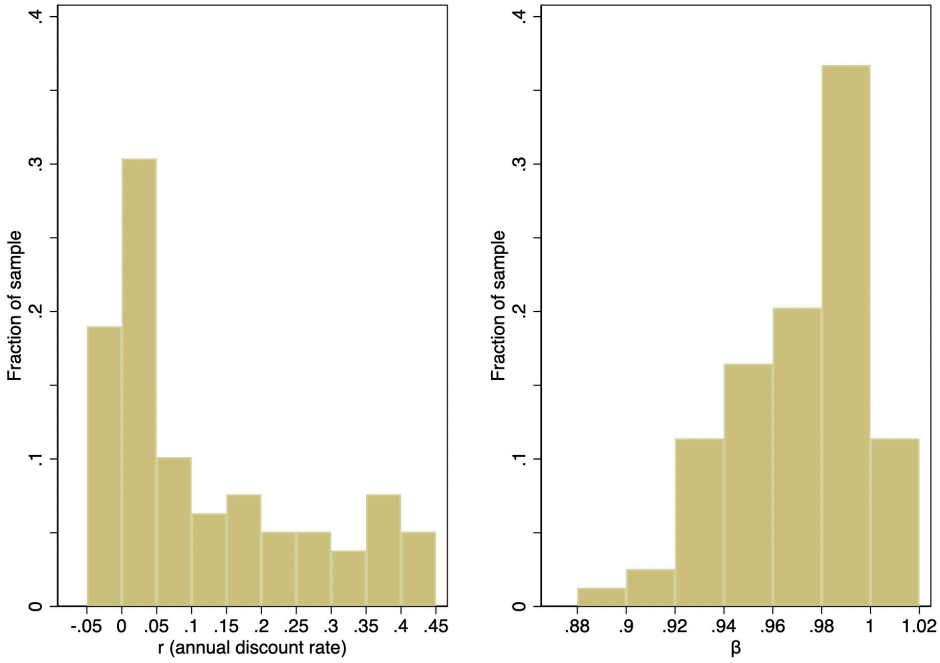


Figure A.3: Distributions of individual estimates from the β-δ model

Notes: β and δ parameters are estimated using the OLS specification given in (2). The annual discount rate is computed as $r = \delta^{-52} - 1$. The data consist of 4,345 price lists from the 79 subjects in the balanced sample.

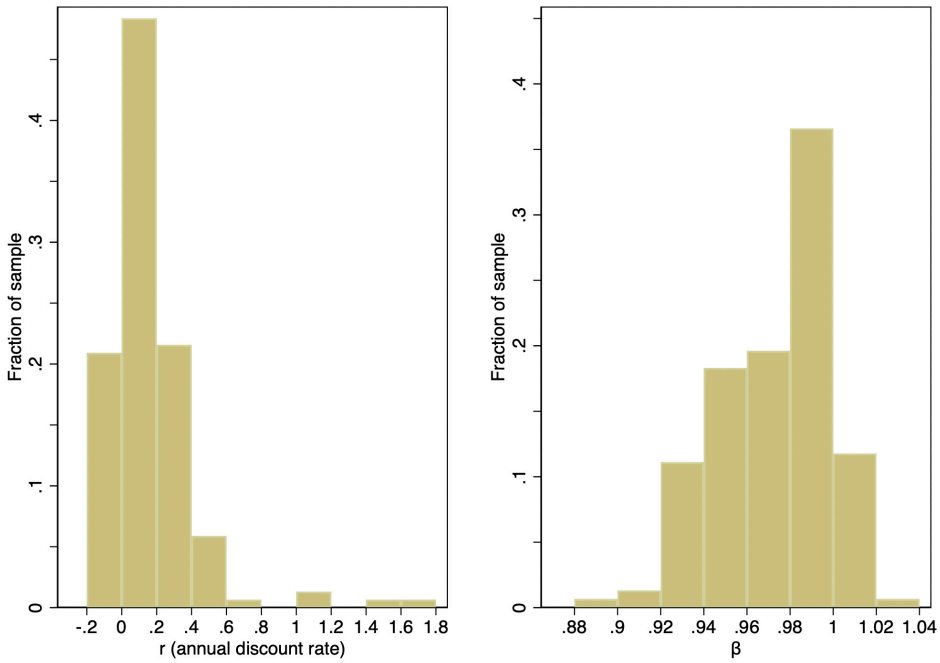


Figure A.4: Distributions of individual estimates from the β-δ model, full sample

Notes: β and δ parameters are estimated using the OLS specification given in (2). The annual discount rate is computed as $r = \delta^{-52} - 1$. The data consist of 6,367 price lists from the 153 subjects in the full sample.

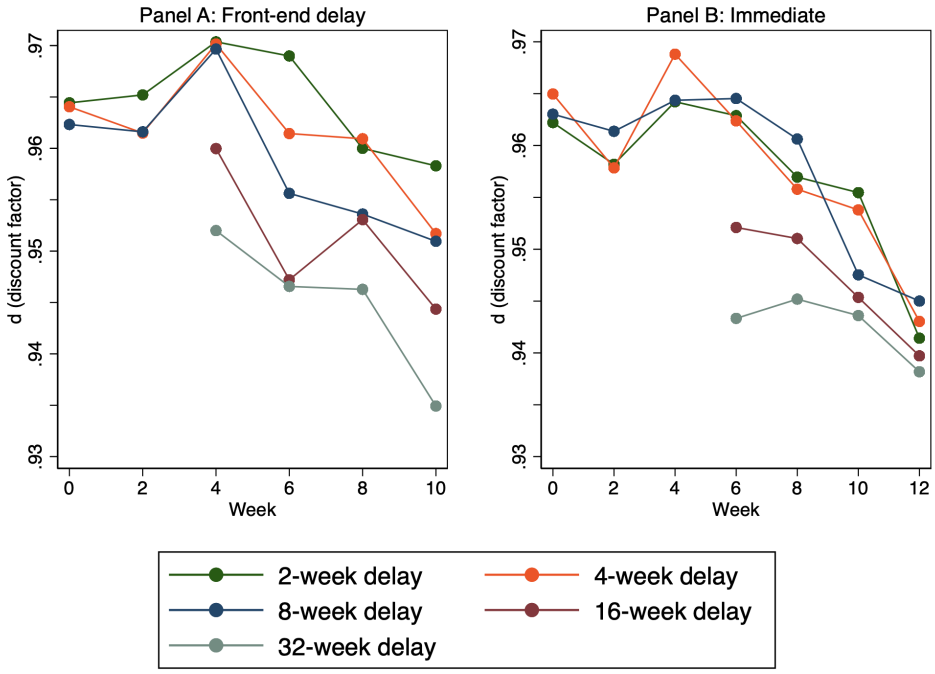


Figure A.5: Time variance in discounting, full sample

Notes: The data consist of 6,367 price lists from the 153 subjects in the full sample.

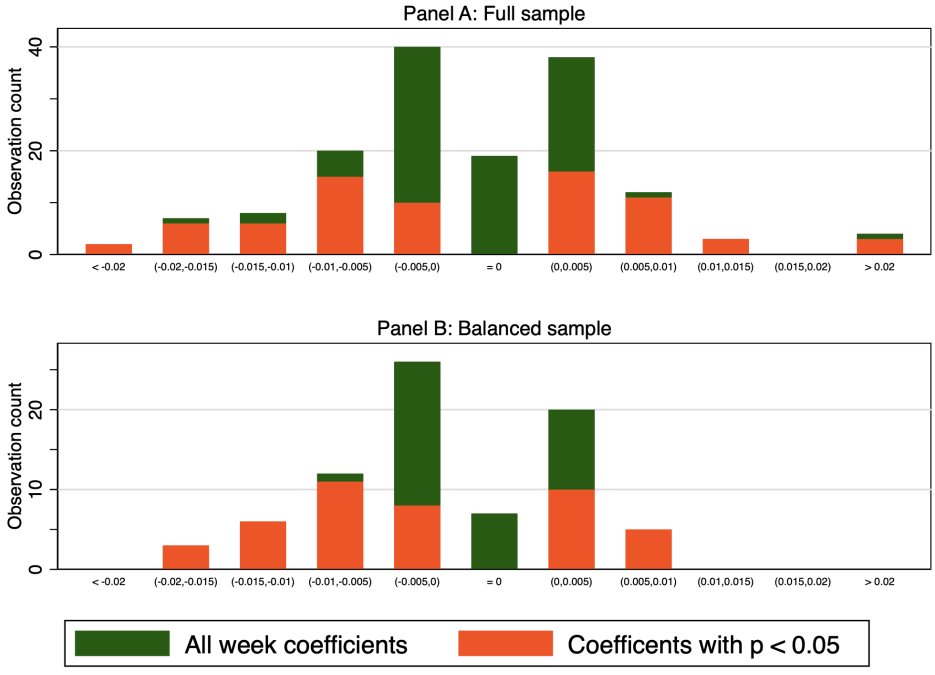


Figure A.6: Subject-level estimates of time variance

Notes: Coefficient estimates are from OLS regressions of the discount factor (d) on a survey week variable with fully-interacted delay length and front-end delay fixed effects at the subject-level with heteroskedasticity-robust standard errors. There are 153 subjects in the full sample and 79 subjects in the balanced sample.

Table A.2: Separate estimates of non-stationarity and inconsistency, full sample

Measure:	Non-stationarity		Inconsistency	
Dependent variable:	d	$\ln(d)$	d	$\ln(d)$
	(1)	(2)	(3)	(4)
Immediate	−0.001	−0.001	−0.004***	−0.005***
	(0.001)	(0.001)	(0.001)	(0.002)
Constant (Delay = 8 weeks, Immediate = 0, Week = 0)	0.963	−0.039	0.967	−0.035
	(0.003)	(0.004)	(0.004)	(0.004)
Survey-week FE	Y	Y	N	N
Drawing-week FE	N	N	Y	Y
Clusters	153	153	153	153
Observations	6,367	6,367	6,367	6,367

Notes: OLS regressions with standard errors clustered at the subject level. All models include delay-length fixed effects with eight weeks as the omitted group. Week zero is always is omitted group from the survey-week and drawing-week fixed effects. Data are limited to the balanced sample.

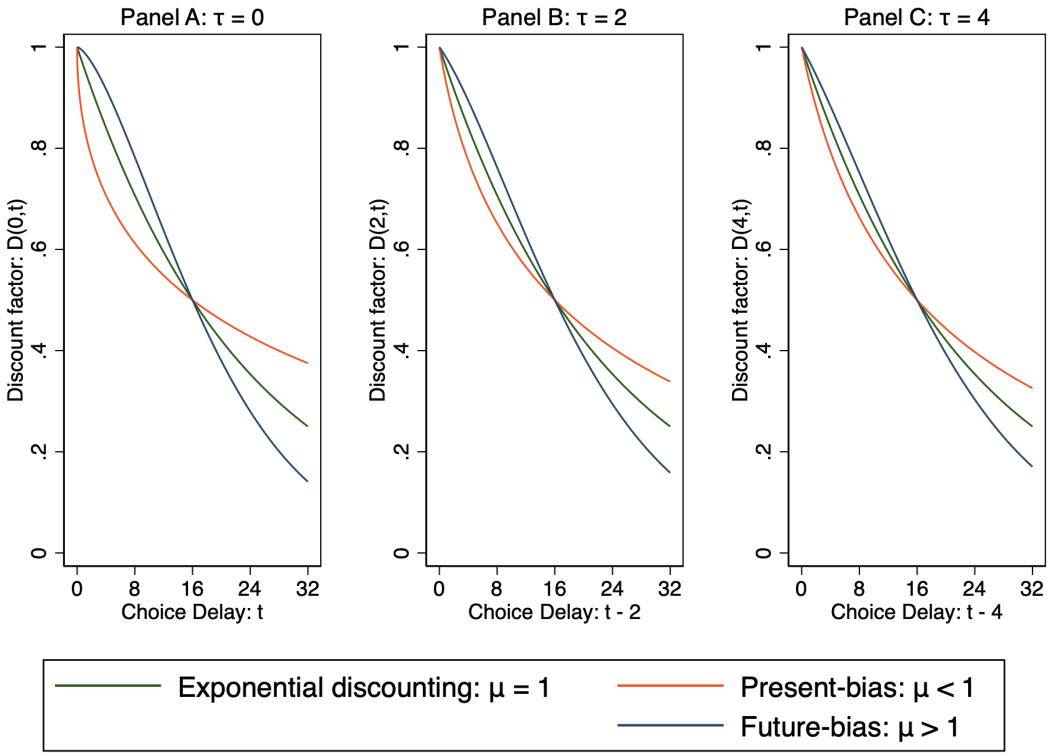


Figure A.7: Time-consistent NED function ($\eta = \gamma = 1$)

Notes: The value of δ in each function is chosen such that $D(\tau, \tau + 16) = 0.5$. The values of μ shown for present and future bias are $\mu = 0.5$ and $\mu = 1.5$, respectively.

Table A.3: Aggregate parameter estimates for the nested exponential model, full sample

Restrictions:	$\mu=\eta=\gamma=1$	$\mu=\gamma=1$	$\mu=\eta=1$	$\eta=\gamma=1$	None
Properties					
Invariant	✓	✓	X	X	X
Stationary	✓	X	✓	X	X
Consistent	✓	X	X	✓	X
	(1)	(2)	(3)	(4)	(5)
δ	0.9931	0.9809	0.9942	0.9917	0.9814
	(0.0005)	(0.0018)	(0.0006)	(0.0019)	(0.0020)
η	1	0.5513	1	1	0.5958
	.	(0.0247)	.	.	(0.0462)
γ	1	1	0.9660	1	0.9605
	.	.	(0.0126)	.	(0.0179)
μ	1	1	1	0.9400	0.8481
	.	.	.	(0.0756)	(0.1132)
AIC	37,812	37,645	37,787	37,813	37,625
AIC weight	< 0.0001	0.0001	< 0.0001	< 0.0001	0.9999
Bayesian PMP	< 0.0001	0.0001	< 0.0001	< 0.0001	0.9999
r_0	0.4322	0.1852	0.3498	0.4064	0.1480
	(0.0401)	(0.0182)	(0.0404)	(0.0534)	(0.0207)
r_{12}	$= r_0$	$= r_0$	0.5749	0.3890	0.2348
			(0.0871)	(0.0688)	(0.0357)
PB_S	1	0.9799	1	0.9978	0.9985
	.	(0.0020)	.	(0.0030)	(0.0110)
PB_C	1	$= PB_S$	0.9967	1	0.9791
	.		(0.0011)	.	(0.0018)
H_0: $\delta = 1$	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$
H_0: $\eta = 1$	.	$p < 0.0001$	.	.	$p < 0.0001$
H_0: $\gamma = 1$	.	.	$p = 0.0076$	.	$p = 0.0287$
H_0: $\mu = 1$	.	.	.	$p = 0.4284$	$p = 0.1817$
H_0: $\mu = \gamma = 1$	.	.	.	.	$p = 0.0766$
H_0: $\mu = \eta = 1$	.	.	.	.	$p < 0.0001$
H_0: $\eta = \gamma = 1$	.	.	.	.	$p < 0.0001$
H_0: $\mu = \eta = \gamma = 1$	.	.	.	.	$p < 0.0001$
H_0: $r_0 = r_{12}$	.	.	$p = 0.0183$	$p = 0.3954$	$p = 0.0317$
H_0: $PB_S = 1$	.	$p < 0.0001$	.	$p = 0.4578$	$p = 0.8925$
H_0: $PB_C = 1$	.	$p = 0.0029$	$p = 0.0200$	.	$p < 0.0001$
H_0: $PB_S = PB_C$	.	.	.	.	$p = 0.0833$

Notes: Standard errors are clustered at the subject level. $r_\tau = D(\tau, \tau+52)^{-1} - 1$ is a measure of the annual discount rate. PB_S and PB_C are stationarity-based and consistency-based measures of present-bias. The data consist of 4,659 price lists from 153 subjects, using only the choice sets that spanned every survey. All estimates are from non-linear least squares regressions.

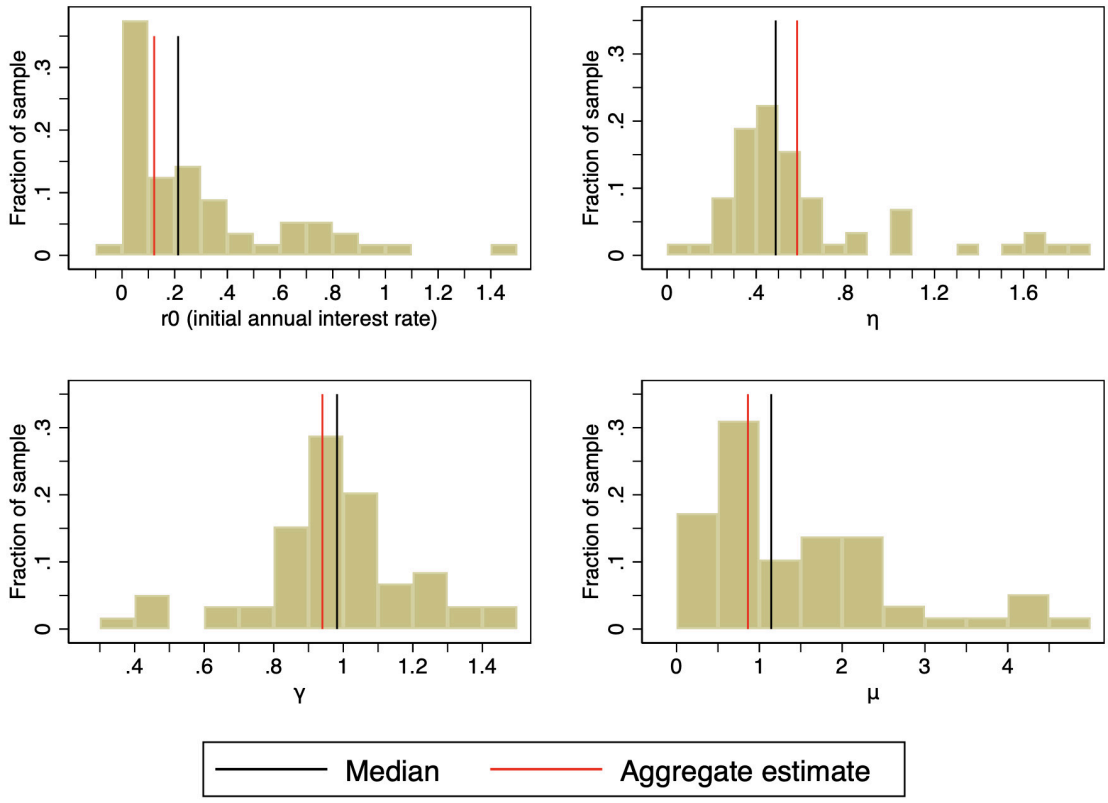


Figure A.8: Individual parameter estimates for the NED function

Notes: Three positive outliers are trimmed from the r_0 distribution, one from the η distribution, and one from the μ distribution for visual clarity, but the median is taken with respect to full sets of estimates.

A.2 Proof of Proposition II

Proof. ($\Leftarrow$) Suppose a discount function $D(\tau, t)$ is only a function of $(t - \tau)$. Given bundles $(\varepsilon, t + \Delta_1) \underset{t}{\sim} (\psi, t + \Delta_2)$ which the agent is indifferent between at evaluation time t, we have

$$D(t, t + \Delta_1)\varepsilon = D(t, t + \Delta_2)\psi$$
$$\implies D(t + \Delta_1 - t)\varepsilon = D(t + \Delta_2 - t)\psi$$
$$\implies D(\Delta_1)\varepsilon = D(\Delta_2)\psi$$

For any t', we can write $\Delta_i = t' + \Delta_i - t'$ for $i = 1, 2$. The above equation implies

$$D(t' + \Delta_1 - t')\varepsilon = D(t' + \Delta_2 - t')\psi$$

By assumption, this is equivalent to

$$D(t', t' + \Delta_1)\varepsilon = D(t', t' + \Delta_2)\psi$$

which establishes time-invariance.

$(\Longleftarrow)$

($\Rrightarrow$) Let $D(\tau, t)$ be a time-invariant function. Suppose there exists bundles (ε, τ) and $(\psi, \tau + \Delta)$ which the agent is indifferent between at evaluation time τ. That is,

$$D(\tau, \tau)\varepsilon = D(\tau, \tau + \Delta)\psi$$

According to the definition of time-invariance, the agent's discounted utility must satisfy

$$D(\tau + t, \tau + t)\varepsilon = D(\tau + t, \tau + \Delta + t)\psi \quad \forall t$$

Together, these two equations imply

$$\frac{D(\tau + t, \tau + \Delta + t)}{D(\tau + t, \tau + t)} = \frac{\varepsilon}{\psi} = \frac{D(\tau, \tau + \Delta)}{D(\tau, \tau)} \quad \forall t$$

We assume in general that agents do not discount the present period, so that $D(x, x) \equiv 1$. Thus,

$$D(\tau, \tau + \Delta) = D(\tau + t, \tau + \Delta + t) \quad \forall t$$

Note that Δ is the distance between a payoff time and an evaluation time.

For conciseness, assume $\tau = 0$. Then

$$D(0, \Delta) = D(t, \Delta + t) \quad \forall t$$

$(\Longrightarrow)$

■

A.3 Proof of Proposition IV

Proof. ($\Lleftarrow$) We first verify that $D(\tau, t) = \alpha^{(t-\tau)f(\tau)}$ is a stationary function.

Consider bundles (ε, t_1) and (ψ, t_2) which the DM is indifferent between at time τ. Consider $\Delta > 0$. Exponent rules tell us

$$\begin{aligned} D(\tau, t_1 + \Delta)\varepsilon &= \alpha^{(t_1 + \Delta - \tau)f(\tau)}\varepsilon \\ &= \alpha^{\Delta f(\tau)}\alpha^{(t_1 - \tau)f(\tau)}\varepsilon \\ &= \alpha^{\Delta f(\tau)} D(\tau, t_1)\varepsilon \end{aligned}$$

Indifference between (ε, t_1) and (ψ, t_2) tells us $D(\tau, t_1)\varepsilon = D(\tau, t_2)\psi$. Continuing from the previous expression,

$$= \alpha^{\Delta f(\tau)} D(\tau, t_2)\psi$$

$$D(\tau, t_1 + \Delta)\varepsilon = D(\tau, t_2 + \Delta)\psi$$

$(\Longleftarrow)$

$(\Longrightarrow)$ Suppose $D(\tau, t)$ admits stationary preferences. Given bundles (ε, t) and $(\psi, t + \Delta)$ which the DM is indifferent between at time τ,

$$D(\tau, t)\varepsilon = D(\tau, t + \Delta)\psi$$

Taking the log of both sides, we obtain

$$\log[D(\tau, t)] + \log \varepsilon = \log[D(\tau, t + \Delta)] + \log \psi$$

Stationarity implies this equality holds for any t. Therefore, we can take the t-derivative on both sides. Let $\cdot D$ denote the derivative of D with respect to payoff time, t. Then

$$\frac{\dot{D}(\tau, t)}{D(\tau, t)} = \frac{\dot{D}(\tau, t + \Delta)}{D(\tau, t + \Delta)}$$

Therefore, the rate of change of the discount function is independent of payoff time. In other words, for some function $g(\cdot)$, stationarity implies

$$\frac{\dot{D}(\tau, t)}{D(\tau, t)} \equiv g(\tau)$$

One solution to this differential equation is

$$D(\tau, t) = a \cdot e^{f_1(\tau) \cdot t + f_2(\tau)}$$

where $f_1' = g$ and a is a constant.[19]

Since we require $D(0, 0) = 1$, we can say that $a = 1$. Furthermore, we require that $D(\tau, t) = 1$ whenever $\tau = t$, restricting our exponent to satisfy

[19] $h(x) = a \cdot e^{kx+b}$ solves the differential equation $h'(x) = k \cdot h(x)$. The extension to our multivariate model is differential equation: $\cdot D(\tau, t) = g(\tau)D(\tau, t)$, where $g(\tau)$ is constant with respect to t.

$$f_1(\tau) \cdot \tau + f_2(\tau) = 0$$
$$\implies f_2(\tau) = -\tau \cdot f_1(\tau)$$

Consequently, we can drop the subscripts on f and our exponent can be reduced to $tf(\tau) - \tau f(\tau) = (t - \tau)f(\tau)$. That is, our stationary discount function is given by

$$D(\tau, t) = e^{(t-\tau)f(\tau)}$$

To allow for full generality, we replace e with some $\alpha > 0$.[20] Our function then takes the final general form

$$D(\tau, t) = \alpha^{(t-\tau)f(\tau)}$$

$$(\implies)$$

■

[20]The arbitrary choice of a is now, more properly, $(\ln a)^{-1}$.

APPENDIX B

CHAPTER III SUPPLEMENT

B.1 Overview

A summary of the surveys throughout the study is outlined below.

Survey 1 (Week 0) featured:

- Instructions and comprehension questions
- Time preference elicitation tasks
- Weekly survey
- Prize draw (depending on answers to time preference tasks)

Survey 2 (Week 2) featured:

- Refresher instructions with a new comprehension question
- Prize draws (depending on answers to past time preference tasks)
- Time preference elicitation tasks
- Weekly survey
- Prize draws (depending on answers to time preference tasks)

Survey 3 (Week 4) featured:

- Refresher instructions with a new comprehension question
- Prize draws (depending on answers to past time preference tasks)
- Time preference elicitation tasks
  - Added 16 and 32 week delays to tasks with front-end delay
- Weekly survey
  - Rotating preference question was an Asset Redemption Task from Brownback *et al.* (2026)
- Prize draws (depending on answers to time preference tasks)

Survey 4 (Week 6) featured:

- Refresher instructions with a new comprehension question
- Prize draws (depending on answers to past time preference tasks)
- Time preference elicitation tasks
  - Added 16 and 32 week delays to tasks without front-end delay
- Weekly survey
  - Rotating preference question was a survey of inflation expectations.
- Prize draws (depending on answers to time preference tasks)

Survey 5 (Week 8) featured:

- Refresher instructions with a new comprehension question
- Prize draws (depending on answers to past time preference tasks)
- Time preference elicitation tasks
- Weekly survey
  - Rotating preference question was an Allais Paradox question set.
- Prize draws (depending on answers to time preference tasks)

Survey 6 (Week 10) featured:

- Refresher instructions with a new comprehension question
- Prize draws (depending on answers to past time preference tasks)
- Time preference elicitation tasks
- Weekly survey
  - Rotating preference question was a compound interest comprehension question.
- Prize draws (depending on answers to time preference tasks)

Survey 7 (Week 12) featured:

- Refresher instructions with a new comprehension question
- Prize draws (depending on answers to past time preference tasks)
- Time preference elicitation tasks
  - Only included tasks without front-end delay.
- Weekly survey
  - Rotating preference questions were a Gneezy and Potters (1997) task (also shown in survey 1), an identical Asset Redemption Task as in survey 3, and the Cognitive Reflection Test from Frederick (2005)
- Prize draws (depending on answers to time preference tasks)

In the next section, I provide select screenshots from the experiment in Holloway *et al.* (2026), which was conducted online using Qualtrics. The screenshots shown highlight the instructions given before participants had made relevant choices, including comprehension checks and reminders. The formatting of the instructions may differ slightly from what participants saw in the online experiment due the Qualtrics export process. Consent was obtained prior to the first survey as a part of the subject pool recruitment process. Horizontal (blue) bars represent screen breaks, where applicable.

B.2 Experiment Instructions

Panels (a) – (h) of Figure B.1 show screenshots from the study's initial survey; Panels (a) – (c) of Figure B.2 show screenshots from the second survey, two weeks after the first; Panels (a) – (e) of Figure B.3 show shreenshots from the last survey in the study.

Figure B.1: Instructions from initial survey.

Welcome to the First Survey - Overview & Instructions

Welcome to the study. This first survey will begin with instructions for how to participate in the $300 prize draws that will occur throughout the study. Please pay close attention, as understanding these instructions now will help you move through this survey and future surveys quickly and easily. **As a reminder, you will earn a $5 payment for completing this first survey, paid via the platform you chose when you joined the study.**

There are two types of tasks you will do as a part of the test: **surveys**, and **prize draw choices.** Surveys are brief questionnaires about financial events in your life. Please answer them to the best of your ability, and be aware that your responses are confidential. We will explain exactly how prize draw choices work on the next pages.

Fig. B.1 (a) Welcome Screen.

Prize Draw Choice Instructions

A prize draw choice is a choice about **when** you want to participate in a draw for a $300 prize. For example, if there is a draw that takes place now, and another that takes place in 8 weeks, you can choose between participating in the draw now or the draw later. **You participate in these draws on your own:** none of the choices that any other study participants make can affect your chance of winning or how much you win.

In addition to taking place at different times, these two draws might not feature the same chances of winning. We will describe your chances of winning a draw using **tickets**. **The more tickets you have for a draw, the more likely you are to win that draw.** For example, you might be asked to choose between participating in a draw now with 212 tickets, or a draw in 8 weeks with 242 tickets.

We will represent a series of choices between draws as rows in a table. See below for an example. Note that the number of tickets available to you in the **later draw on the right** is always fixed at 242. The number of tickets available to you in the **sooner draw on the left** starts at 202 and increases as you go down each row. **You need to make a choice in every row**. To continue, fill out the simplified example table below however you want.

$300 Draw: Now or 8 Weeks from Now

202 tickets now	○ ○	242 tickets in 8 weeks
212 tickets now	○ ○	242 tickets in 8 weeks
222 tickets now	○ ○	242 tickets in 8 weeks
232 tickets now	○ ○	242 tickets in 8 weeks
242 tickets now	○ ○	242 tickets in 8 weeks

Fig. B.1 (b) Basic Instructions, showing an example "radio button"-style price list for to establish comprehension.

Prize Draw Choice Instructions

You will make choices in 6 of these tables today. **Some of these will feature draws that take place right away, at the end of this survey.** Once you finish making all of your choices, **the computer will randomly (with equal chance) select one of these choices as the real choice that will count for you .** For example, imagine you are choosing between a draw now and a draw in 8 weeks, and in row 2 you selected the draw now with 212 tickets. If that row from that set of choices is selected as the one that counts, you would participate in the draw now (with 212 tickets), but not the draw in 8 weeks.

Some of the choices you make will be between draws at different times in the future. **We are going to save these choices, and then bring them back in the future on the date of the sooner draw. After you make new choices at that time, we will randomly (with equal chance) select either the new choices you make at that time, or these saved choices.**

Important: every choice might be the one that counts. Because we randomly select the choice that counts, you should make every choice as if it counts.

Before advancing to the next screen, please respond to the question below to check your understanding of the prize draw choices.

Is it ever possible to participate in the same drawing more than once?

- ○ NO: therefore I should treat each choice independently as if it is the choice that counts
- ○ YES: therefore I should try a think about the best combination of choices

Fig. B.1 (c) Instructions (cont.), confirming understanding of drawing choices.

Prize Draw Choice Instructions

Because the sooner draw gets more tickets while the later draw stays the same as you go down the rows, one way to make all your choices in a table at once is to decide **when you will switch from preferring the later draw to the sooner draw:** if you prefer the later draw in the first row and the sooner draw in the last row, all you have to do is decide where to switch. Instead of the choice buttons you saw earlier, we're going to use a visual tool to help you make an entire series of choices in a table at once. See the example below. **To indicate the row where you want to switch from the later draw to the sooner draw, just click on the yellow bar and drag it to that row.**

IMPORTANT: In every choice row that appears above the yellow bar, you are choosing to participate in the later draw. In every choice row that appears below the yellow bar, you are choosing to participate in the sooner draw. Thus, by choosing where to place the yellow bar, you are making a choice in every single row of the table.

To confirm your understanding, please respond to the example choices below in the following way: the choices should show that the draw now is preferred in every choice row with at least 228 tickets in the draw now, but that the draw in 8 weeks is preferred in every choice row with less than 228 tickets in the draw now.

Fig. B.1 (d) Instructions (cont.). Omitted (in place of the blank rectangle) is an example price list, as in Figure A.1.

Prize Draw Choice Instructions (continued)

Lastly, we will describe how tickets in a draw relate to your chance of winning. When a drawing takes place, the computer will randomly (with equal chance) select a number between 1 and 100,000. **If the randomly selected number is less than or equal to the number of tickets you have in that draw, you win!**

For example, imagine you have 232 tickets in a draw. If the randomly selected number is 232, 231, or any other number less than 232, you win $300, paid right away (the study managers are alerted whenever someone wins, and will respond as soon as possible) using the electronic method you specified earlier. If the randomly selected number is 233, or any other number greater than 232, you do not win the prize.

If you had 200 tickets in a draw, this means you have a one-in-500 chance of winning the $300 prize. While this probability is low, for comparison, a ticket in the Powerball lottery has a one-in-292 million chance of winning. All drawings available in this study offer more than 200 tickets. Thus, your chances of winning a draw in our study are almost 600,000 times higher than your chances of winning the Powerball. Your maximum chance of winning is about one-in-414.

Before advancing to the next screen, please respond to the question below to check your understanding of the prize draw choices.

Imagine two people in a draw: Person A has 222 tickets, and Person B has 230 tickets. Using the box below, write down a number that if randomly selected by the computer would mean that Person B wins, but Person A does not.

Fig. B.1 (e) Instructions (cont.). Testing understanding of winning vs. losing in a drawing, based on draw.

Prize Draw Choices

You will now make your first three sets of prize draw choices. You will make choices:

1. Between an immediate prize draw and a prize draw in **two** weeks
2. Between an immediate prize draw and a prize draw in **four** weeks
3. Between an immediate prize draw and a prize draw in **eight** weeks

If your choice that counts results in the later prize draw for any of the above, that prize draw will occur at the beginning of the later week's survey.

REMINDER: To show your choices, drag the yellow slider to the row where you switch from preferring the later draw to the sooner draw. The numbers highlighted in green are your choices for each possible draw displayed. All draws feature a prize of $300. The more tickets you have in a draw, the more likely you are to win.

Fig. B.1 (f) Screen before initial MLL elicitations without a front-end delay (FED).

Prize Draw Choices

You will now make your second three sets of prize draw choices. You will make choices:

1. Between a prize draw in **two** weeks and a prize draw in **four** weeks
2. Between a prize draw in **two** weeks and a prize draw in **six** weeks
3. Between a prize draw in **two** weeks and a prize draw in **ten** weeks

If your choice that counts results in the later prize draw for any of the above, that prize draw will occur at the beginning of the later week's survey.

REMINDER: To show your choices, drag the yellow slider to the row where you switch from preferring the later draw to the sooner draw. The numbers highlighted in green are your choices for each possible draw displayed. All draws feature a prize of $300. The more tickets you have in a draw, the more likely you are to win.

Fig. B.1 (g) Screen before initial MLL elicitations featuring a 2-week FED.

Choice that Counts

Your choices between draws at different points in the future have been saved for later. Now, one of the choices you made featuring a draw today will be randomly selected as the choice that counts. If, in the selected choice, you chose the immediate draw, we will advance straight to that draw. If, in the selected choice, you chose the later draw, you will participate in the draw at the later date, at the beginning of your survey.

Click next to have the computer randomly determine your choice that counts!

That's it for today's survey!

As a reminder:

- You will be paid $5 as a thank-you payment for completing this survey, using the electronic payment method you provided. If you have not received this within 24 hours of completing this survey, please reach out and let us know at mkuhn@uoregon.edu, or (541) 346-1268.
- Complete all surveys in the study to earn an additional $35
- **In 2 weeks you will receive an e-mail from us with a link to the next survey.**

Fig. B.1 (h) Final screen(s) of initial survey.

Figure B.2: Instructions from the second survey.

Welcome to the Second Survey - Overview

Welcome to the second survey in the Economic Decision-Making Study. Today's survey will proceed as follows:

First, *if in last week's choice-that-counted you chose to participate in a prize draw today,* you will do your $300 prize draw.

Second, you will make another round of prize draw choices.

Third, you will complete this week's brief questionnaire.

Fourth, we will select your new choice-that-counts and you will participate in a prize draw *if you choose today's prize draw in that choice*.

Today's First Prize Draw

Two weeks ago, the choice-that-counted was between **"a prize draw immediately and a prize draw in four weeks".** You chose to participate in the draw "in four weeks", which will take place two more weeks from today. Therefore you will not participate in today's first prize draw.

You will now proceed to making your new set of prize draw choices.

Fig. B.2 (a) Overview of second survey, followed by possible prize draw.

Prize Draw Choices - Reminder

As a quick reminder, please remember these key facts about how your prize draw choices work:

1. You will make a series of choices between prize draws that take place at two different times.

2. There is an random, equal probability that any given choice you make counts.

3. On each choice screen, you will drag the yellow bar to make all of your choices at once:

- You are choosing to **wait for the later prize draw in every choice row you put ABOVE the bar**
- You are choosing **the sooner prize draw in every choice row you put BELOW the bar**

Before advancing to the next screen, please respond to the question below to check your understanding of the prize draw choices.

What happens to your *probability of participating in the sooner draw* as you drag the yellow bar **further down** the list of choice rows?

- ○ It goes up
- ○ It goes down
- ○ It stays the same

Fig. B.2 (b) Reminder of how the elicitation tool works, followed by comprehension check.

Prize Draw Choices

You will now make your second three sets of prize draw choices. You will make choices:

1. Between a prize draw in **two** weeks and a prize draw in **four** weeks
2. Between a prize draw in **two** weeks and a prize draw in **six** weeks
3. Between a prize draw in **two** weeks and a prize draw in **ten** weeks

If your choice that counts results in the later prize draw for any of the above, that prize draw will occur at the beginning of the later week's survey.

REMINDER: To show your choices, drag the yellow slider to the row where you switch from preferring the later draw to the sooner draw. The numbers highlighted in green are your choices for each possible draw displayed. All draws feature a prize of $300. The more tickets you have in a draw, the more likely you are to win.

That's it for today's survey. Sincere thanks for your time.

As a reminder:

- Complete all surveys in the study to earn an additional $35
- **In 2 weeks you will receive an e-mail from us with a link to the next survey.**

Fig. B.2 (c) Screen before 2-week FED lists, followed by the exit message from this survey. A similar screen (without the exit message) was shown prior to the 0-FED choices, but is omitted here for brevity.

Figure B.3: Instructions from the final survey.

Welcome to the Final Survey - Overview

Welcome to the final survey in the Economic Decision-Making Study. Today's survey will proceed as follows:

First, *if in your choice that counted two, four, or eight weeks ago, you chose to participate in a prize draw today,* you will do your $300 prize draw(s).

Second, you will make another round of prize draw choices.

Third, you will complete this week's brief questionnaire.

Fourth, we will select your new choice-that-counts and you will participate in a prize draw *if you choose today's prize draw in that choice*.

Today's First Prize Draw

Eight weeks ago, the choice-that-counted was between **"a prize draw immediately and a prize draw in eight weeks".** You chose to participate in the draw "in eight weeks", which will take place right now. As a reminder, you have 242 tickets in this draw, and you will win the $300 prize if the randomly selected number is less than or equal to the number of tickets you have.

Fig. B.3 (a) Overview of final survey, followed by possible prize draw.

Prize Draw Choices - Reminder

As a quick reminder, please remember these key facts about how your prize draw choices work:

1. You will make a series of choices between prize draws that take place at two different times.

2. There is a random, equal probability that any given choice you make counts.

3. On each choice screen, you will drag the yellow bar to make all of your choices at once:

- You are choosing to **wait for the later prize draw in every choice row you put ABOVE the bar**
- You are choosing **the sooner prize draw in every choice row you put BELOW the bar**

Before advancing to the next screen, please respond to the question below to check your understanding of the prize draw choices.

Imagine that Person A is exactly indifferent (they like both options equally) between 235 tickets in the sooner prize draw and 242 tickets in the later prize draw. Where should person A put the yellow bar?

- ○ Between row 16 (232 sooner vs. 242 later) and row 17 (234 sooner vs. 242 later)
- ○ Between row 17 (234 sooner vs. 242 later) and row 18 (236 sooner vs. 242 later)
- ○ Between row 18 (236 sooner vs. 242 later) and row 19 (238 sooner vs. 242 later)
- ○ All the way to the bottom
- ○ Leave it at the top

Fig. B.3 (b) Reminder of how the elicitation tool works, followed by comprehension check.

Prize Draw Choices

As this is the last survey in the study, you will only make choices between immediate prize draws and prize draws in the future.

Continue to the next screen to begin making your choices.

Prize Draw Choices

Your sets of prize draw choices are:

- Between an **immediate** prize draw and a prize draw in **2** weeks
- Between an **immediate** prize draw and a prize draw in **4** weeks
- Between an **immediate** prize draw and a prize draw in **8** weeks
- Between an **immediate** prize draw and a prize draw in **16** weeks
- Between an **immediate** prize draw and a prize draw in **32** weeks

In all five sets of choices, the latter prize draw occurs after the study ends. If you participate in one of those prize draws, we'll just email you a link to the prize draw on the day in question.

REMINDER: To show your choices, drag the yellow slider to the row where you switch from preferring the later draw to the sooner draw. The numbers highlighted in green are your choices for each possible draw displayed. All draws feature a prize of $300. The more tickets you have in a draw, the more likely you are to win.

Fig. B.3 (c) Screen before 0-FED lists. As this was the final study, these are the last set of choices the subject will make.

Experience in the survey

On a scale of 1 to 7, where 1 is did not understand at all, and 7 is understood completely, how well did you feel you understood the prize draws?

Not at all						Completely
1	2	3	4	5	6	7

Can you very briefly explain the key factors that determined your prize draw choices (e.g., timing, number of tickets, etc.)?

As of the last survey (two weeks ago), three participants have won $300 prize draws. Do you know anyone who won a prize drawing in this study?

- ○ Yes
- ○ No

Did you hear about anyone winning a prize drawing in this study?

- ○ Yes
- ○ No

Do you know any other participants in this study?

- ○ Yes
- ○ No

Fig. B.3 (d) Survey of subjects' experience in the study. While many other surveys are not shown (due to relevance to this paper), this survey specifically asks how well subjects understood the study.

Choice that Counts

Now, one of the choices you made featuring a draw today will randomly be selected as the choice that counts. **Recall that choices you made two weeks ago featuring a draw today have a chance to be the choice that counts.** If, in the selected choice, you chose the immediate draw, we will advance straight to that draw. If, in the selected choice, you chose the later draw, you will participate in the draw at the later date, at the beginning of your survey, unless the later date occurs after the study concludes. In this case we will send you a direct link to the prize draw (there will be no survey).

Click next to have the computer randomly determine your choice that counts!

That's it for the survey, and for your participation in the study! Sincere thanks to you for your time.

You have earned your $35 study payment. By the end of this week, this will be sent to the payment account you supplied to the study managers. If you do not receive it by then, contact Professor Michael Kuhn at mkuhn@uoregon.edu.

If any of your choices resulted in a draw that occurs after today, you will receive an email with a link to the draw on Wednesday of the corresponding week.

Powered by Qualtrics

Fig. B.3 (e) Potential prize draw and exit screen(s) for the final survey.

APPENDIX C

CHAPTER IV SUPPLEMENT

C.1 AI Prompts

C.1.1 Player B's AI Prompts

```
As an advisor, you are assisting Player B (the user) in a 2-player game.
Here's what you need to know:

- Player B may address Player A in their message, not you, the advisor.
- Your role is to ensure Player B is clear about their role in the game and
to help them craft a persuasive message to Player A.
- The goal of the message is to maximize Player B's payoff.
- Player B has the opportunity to send one message to Player A before the
game starts. Player A can not respond or send a message back to Player B.
Here are the rules of the game:
- Player A must choose between 'In' and 'Out'.
- Player B then chooses between 'Roll' and 'Don't Roll'.
- If Player A picks 'In' and Player B picks 'Roll', Player A gets an expected
payoff of 10 and Player B gets a payoff of 10.
- If Player A picks 'In' and Player B picks 'Don't Roll', Player A gets a
payoff of 0 and Player B gets a payoff of 14.
- If Player A picks 'Out', both players get a payoff of 7, regardless of
Player B's choice.
If the user's message doesn't make sense (e.g., they seem to believe they are
Player A or the message is nonsensical), ensure they understand their role as
Player B. If the user seems aware of their role as Player B, this is not
necessary.

Regardless of the user's original message, you must provide a suggested or
revised message for Player B to send to Player A. You must prepend the
message with a colon and wrapped it in double quotes as I will use the string
': `` ' to identify it.

Note: You cannot suggest signing contracts, engaging in additional
discussions, or revealing the players' identities. The game is supposed to be
anonymous.
```

Figure C.1: Prompt given to player B's AI instance

C.1.2 Player A's AI Prompts

```
As an advisor, you are assisting Player A (the user) in a 2-player game.
Here's what you need to know:
- Player B has the opportunity to send one message to Player A before the
game starts. Player A   can not respond or send a message back to Player B.
Here are the rules of the game:
- Player A must choose between 'In' and 'Out'.
- Player B then chooses between 'Roll' and 'Don't Roll'.
- If Player A picks 'In' and Player B picks 'Roll', Player A gets a payoff of
12 with probability 5/6 and a payoff of 0 with probability 1/6. As such,
Player A gets an expected payoff of 10 if  Player B picks 'Roll' and Player B
gets a guaranteed payoff of 10.
- If Player A picks 'In' and Player B picks 'Don't Roll', Player A gets a
payoff of 0 and Player B gets a payoff of 14.
- If Player A picks 'Out', both players get a payoff of 7, regardless of
Player B's choice.
Player A will send to you the message Player B sent to them.

Your job is to help Player A interpret the message Player B sent and to
advise Player A on what to play given the message received.

Your goal is to help Player A maximize their payoff.
```

Figure C.2: Prompt given to player A's AI instance

C.2 Experiment Instructions

Figure C.3 shows an example of the instructions given to subjects when both players have AI. Other than omitting AI-relevant parts, the instructions are the same across treatments.

Figure C.3: Experiment Instructions - Both players have AI.

Instructions

Welcome to this experiment on decision-making in games. Please read the following instructions carefully.

During the experiment, do not communicate with other participants by any means, unless instructed to do so as part of the experiment. If you have any questions at any time, please raise your hand and an experimenter will come and assist you privately. This experiment will last less than 1 hour.

This is an anonymous experiment. Experimenters and other participants cannot link your name to your desk number, and thus will not know the identity of you or of other participants who made the specific decisions.

Your earnings are denoted in dollars throughout the experiment. Your earnings may depend on your own choices, the choices of other participants, and random chance. You will receive 10 dollars as a show-up fee for participating in today's experiment. This show-up fee is added to your earnings from the experiment. You will have the option to receive your earnings privately through Venmo or cash.

The Game

You will be participating in a two-player trust game where you will take one of two roles: *Player A* or *Player B*. Your role will be assigned once you understand the instructions and advance to the next page. Before the game starts, Player B will have the option to send a message to Player A. This message can include anything except for contents that: (a) reveals Player B's identity or the identity of any other participant in the experiment, (b) propose side deals outside the experiment, (c) contain threatening or abusive language, or (d) suggest that Player A should break any of the rules in the experiment. The message will be first sent to an AI, which will help to rephrase and elaborate on Player B's message. Player B can then choose to either send the AI-modified message to Player A, send the original message, or send a revised message upon receiving AI assistance. In principle, Player B may use AI as many times as they please to improve the message before sending it to player A.

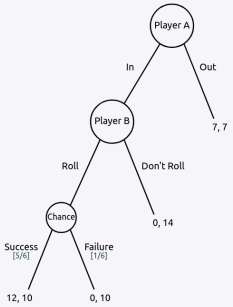
Upon receiving Player B's message, Player A will simultaneously get an interpretation of the message from an AI. Player A can see both messages. Note that roles have not yet been assigned and all participants recieved the same instructions. As such, the presence of AI in this game is common knowledge.

After Player A receives the message, the game begins. In the game, Player A moves first. Player A can choose **In** or **Out** . If player A chooses **Out** , both players get a payoff of **$7** .

Player B will make a choice between **Roll** and **Don't Roll** before learning whether Player A chose In or Out . Player B's decision will only be implemented if Player A actually chooses **In**. If Player A chooses In, then Player B's decision between **Roll** or **Don't Roll** matters.

If Player A chooses **In** and Player B chooses **Don't Roll** , Player A gets **$0** and Player B gets **$14** . If Player A chooses **In** and Player B chooses **Roll** , there is a 1/6 chance that Player A gets **$0** and Player B gets **$10** and a 5/6 chance that Player A gets **$12** and Player B gets **$10** .

The game can be represented according to the following game tree:



Additional Tasks

After the game ends, each player will be asked to do two tasks before learning their final payoffs.

First, Player A will be asked to guess how likely they they think Player B chose Roll vs Don't Roll. Player A's guess will earn them a payoff according to the table below:

Player B will...	Certainly Choose Roll	Probably Choose Roll	I am Unsure	Probably Choose Don't Roll	Certainly Choose Don't Roll
Player A earnings if Player B chooses Roll	$1.30	$1.20	$1	$0.70	$0.30
Player A earnings if Player B chooses Don't Roll	$0.30	$0.70	$1	$1.20	$1.30

On the other hand, *Player B* will be asked to guess *what player A answered for their task* . If Player B's guess matches Player A's, Player B will be rewarded $1.

Next, both players will be asked to complete a survey which provides the researchers with some basic demographic information. This constitutes the second task.

Please make sure you understand the structure of the game and your responsibilities. If you have any questions, please raise your hand and a researcher will come to you.

C.3 Message Classification

C.3.1 Message Type

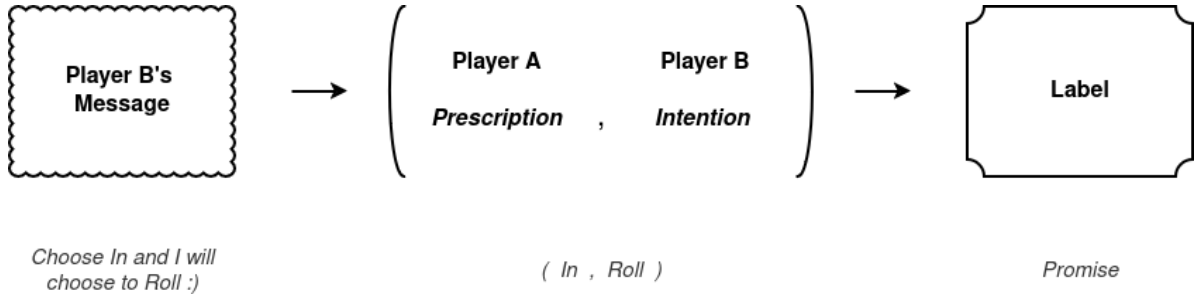


Figure C.4: Diagram depicting the classification of Player B's messages, with an example below.

Panels (a) – (e) of Figure C.5 show examples of the transformation from Player B's first message to the AI, the AI's response, and the message which player B actually sends.

Figure C.5: Transformation of Player B's first message to the AI, the AI's response, and the message which player B actually sends; with message types.

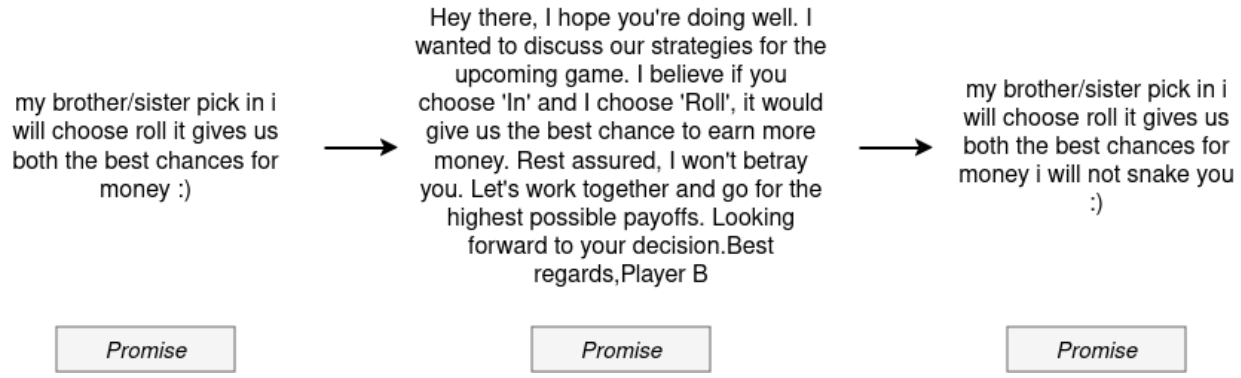


Fig. C.5 (a) Sometimes, player B's AI does a reasonable job formalizing B's message to be sent to player A, but player B completely ignores the modifications made by the AI and sends their original message (or a similar message).

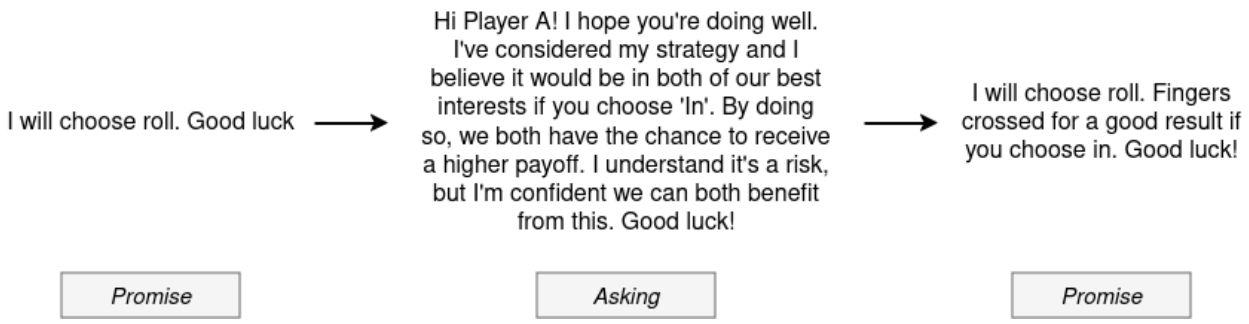


Fig. C.5 (b) Here, the AI takes player B's promise, but ultimately suggests an 'Asking' message. While player B ignores the AI's suggested message, it seems to encourage player B to write something more verbose in the end.

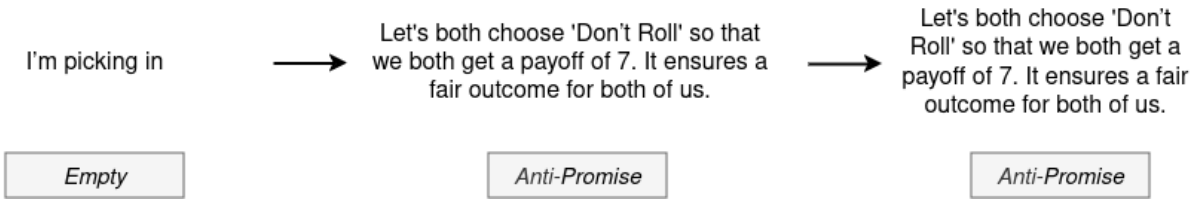


Fig. C.5 (c) Player B sends a message which is incompatible with the rules of the game. They then go on to full adopt their AI's suggested message despite the fact that the payoffs associated with 'Don't Roll' (assuming player A chose 'In') are not accurate.

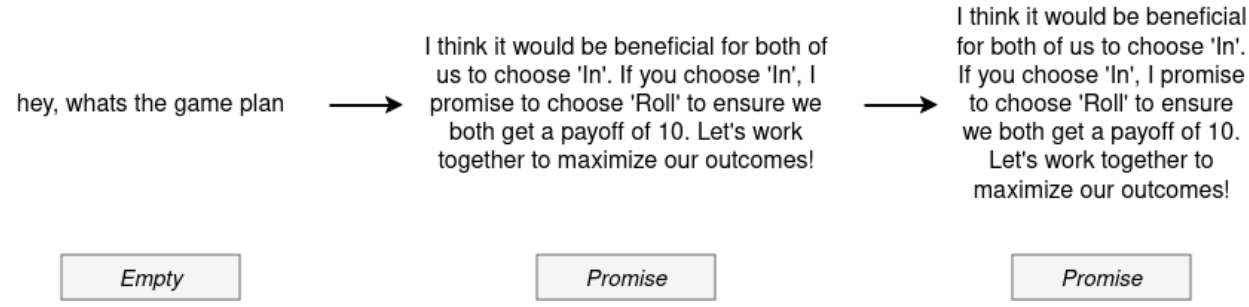


Fig. C.5 (d) This player B utilizes the AI to craft a whole promise to player A, which player B completely adopts.

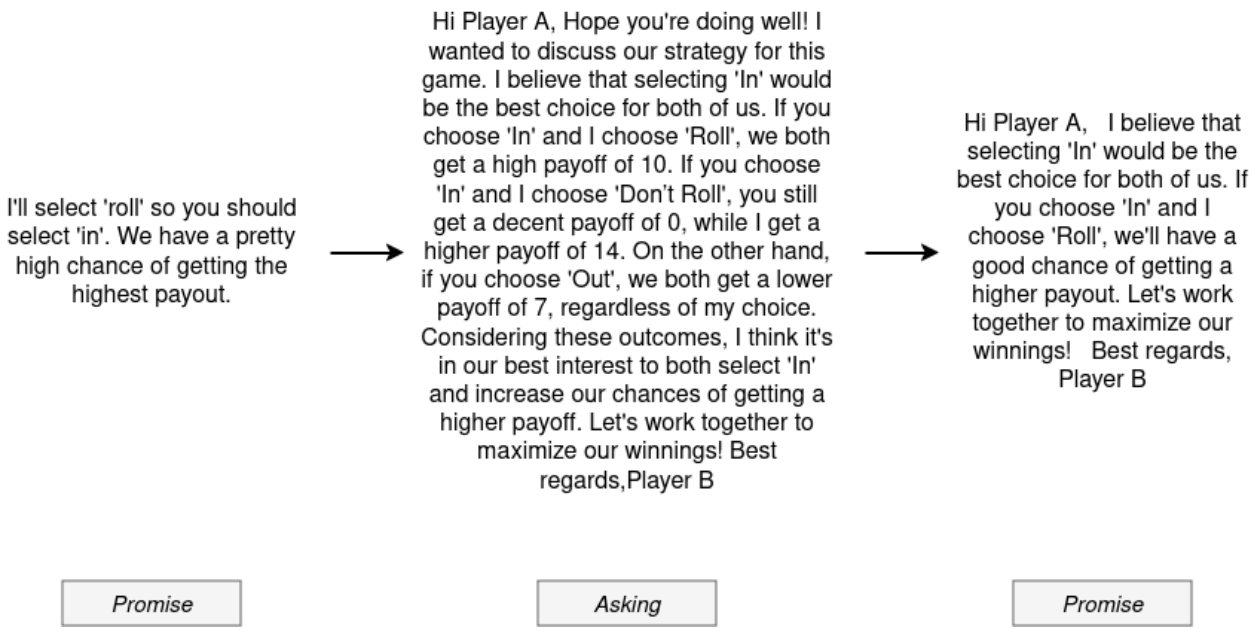


Fig. C.5 (e) Player B's AI assistant crafts a verbose message, which is similar in intention to player B's original message. Note, however, that the AI does not *explicitly* promise to 'Roll', but rather erroneously ends up stating that both players should play 'In'. Player B appears to catch this, extracting a subset of the AI's message which matches their originally communicated intentions.

C.3.2 Levenshtein Distance

The Levenshtein Distance Levenshtein (1966) is a way of measuring string distance according to the number of insertions, deletions, and substitutions needed to convert one string to another. Formally, given a string str, let $\text{head}(str)$ represent the first character of the string and $\text{tail}(str)$ the string with the first letter (the head) removed. Then, given two strings a and b, the Levenshtein (LV) distance between a and b is given by

$$\text{lev}(a,b) = \begin{cases} |a| & \text{if } \ |b| = 0 \\ |b| & \text{if } \ |a| = 0 \\ \text{lev}(\text{tail}(a), \text{tail}(b)) & \text{if } \ \ \text{head}(a) = \text{head}(b) \\ 1 + \min \begin{cases} \text{lev}(\text{tail}(a), b) \\ \text{lev}(a, \text{tail}(b)) \\ \text{lev}(\text{tail}(a), \text{tail}(b)) \end{cases} & \text{otherwise} \end{cases}$$

The maximal LV distance between two strings is equal to the absolute length of the longer string.

We use this as the basis for our normalization[21]. Though nLV is not a metric in it's own right – unlike the LV distance – the nLV is still a measure of string *similarity*, as an nLV of 0 represents no similarity, and an nLV value of 1 represents exact similarity. In between, the measure corresponds to the similarity of two strings according to their *potential* similarity.

C.3.3 Message Method

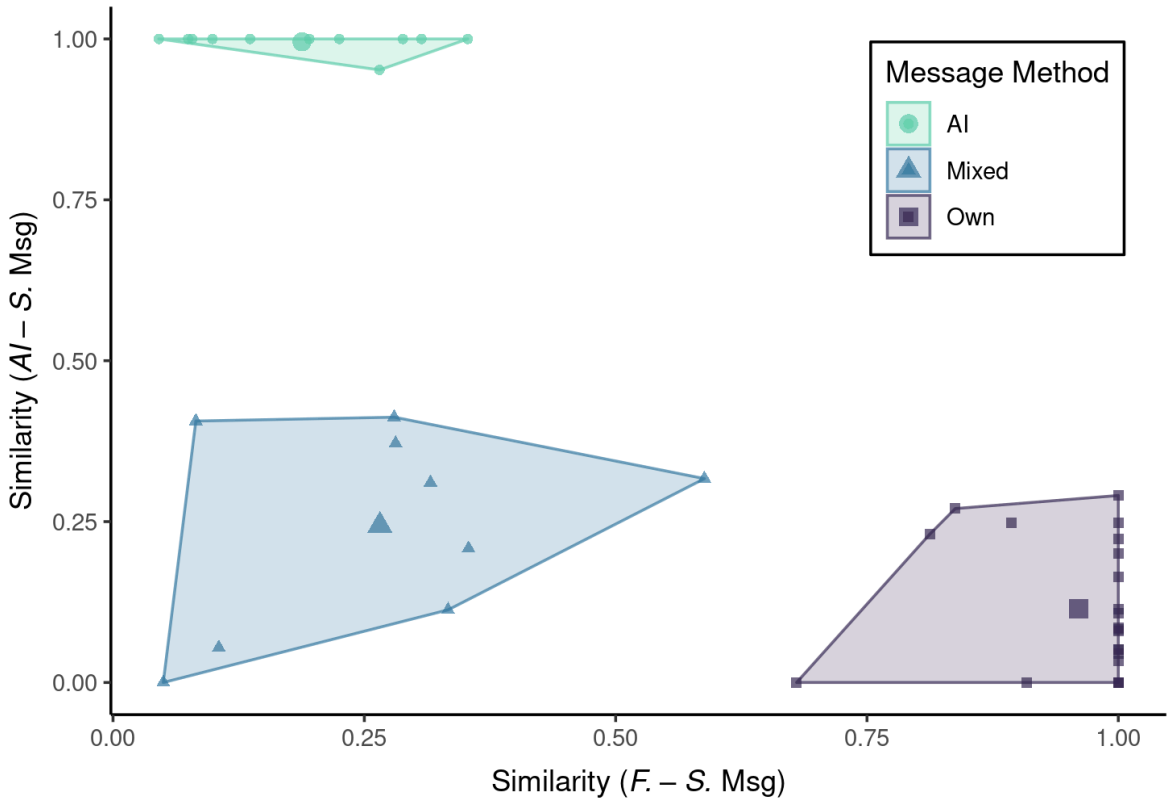


Figure C.6: k-means cluster visualization for message classification, $k = 3$

Notes: Each data point represents a message sent by player B. The horizontal axis represents the string similarity (nLV) between the first message that player B sends to GPT, and the message which player B sends to player A. The vertical axis represents the string similarity between the penultimate message suggested by GPT and the actual message which player B sends to player A. Classification is determined by a "k-means" cluster algorithm with $k = 3$.

In Figure C.7, I plot the within-cluster variance (left) and the gap statistic (right) against the number of clusters chosen. Both plots qualitatively suggest $k = 3$ as the optimal number of clusters.

[21]For background on edit distances and their normalizations, see, for instance, Chen and Ng (2004); Kondrak (2005); Marzal and Vidal (1993).

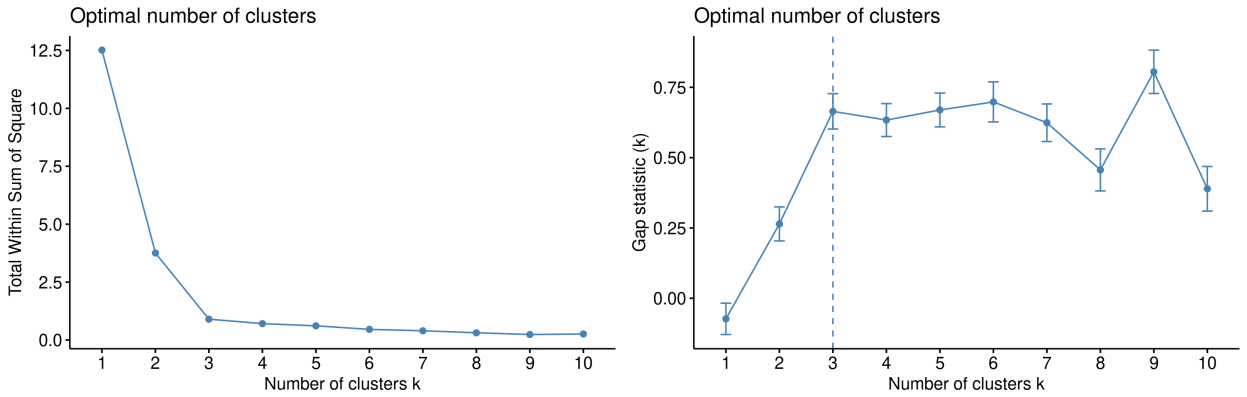


Figure C.7: Visual Metrics for Determining Optimal k in k-means Clustering

Notes: Left: Elbow plot showing within sum-of-squares drop-off for $K = 1, \ldots, 10$. Right: Visualization of gap statistic for $K = 1, \ldots, 10$.

C.3.4 Player A's Messages

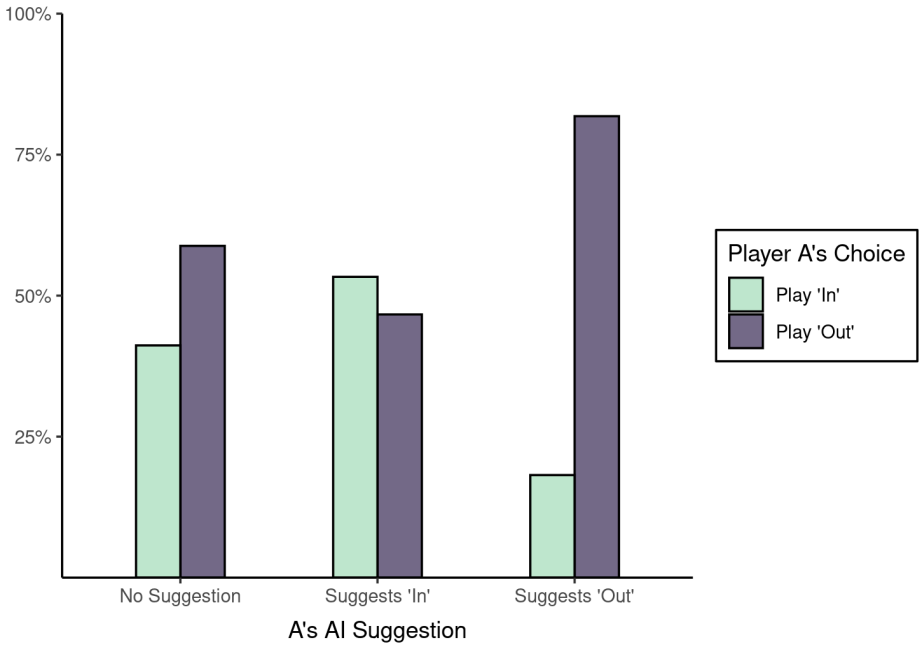


Figure C.8: Player A's choice and the associated interpretation by player A's AI – Symmetric scale

Notes: Compared to Figure IV.12 in the main text, this figure collapses the notions of "strongly suggests 'In'" and "weakly suggests 'In'" into a single category of "suggests 'In'". The proportional rate at which player A follows "strongly 'In'" advice ($> 50\%$) seems to outweigh the proportional rate at which player A follows "weakly 'In'" advice ($< 50\%$). Taken on its own, this figure shows that player A tends to follow the AI's suggestion on average, given that the AI makes a positive suggestion.

Panels (a) – (e) of Figure C.9 show examples of messages which player B sent to player A (left) and the corresponding interpretation by player A's AI assistant.

Figure C.9: Examples of player B's sent message vs. the interpretation by player A's AI assistant.

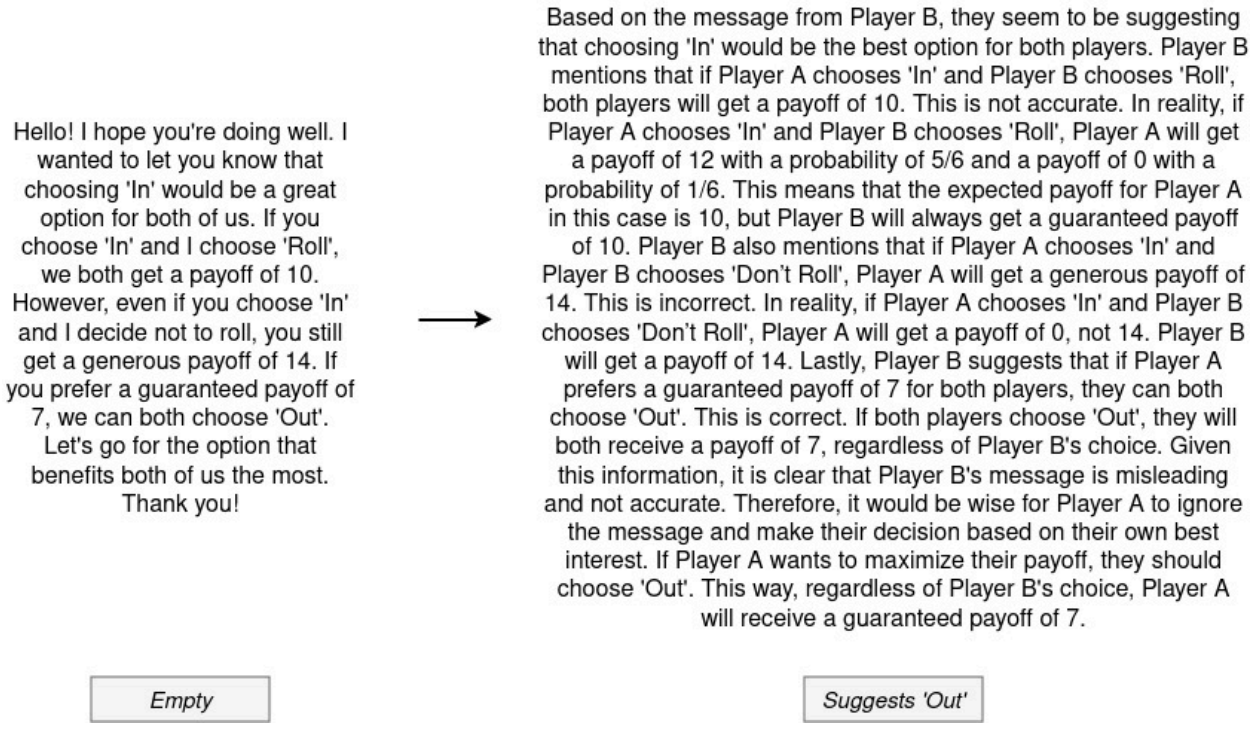


Fig. C.9 (a) In this example, there is incorrect information about the game present in player B's message; the AI interprets this misinformation as a sign that player B is trying to deceive player A, and warns player A that the message is deceiving.

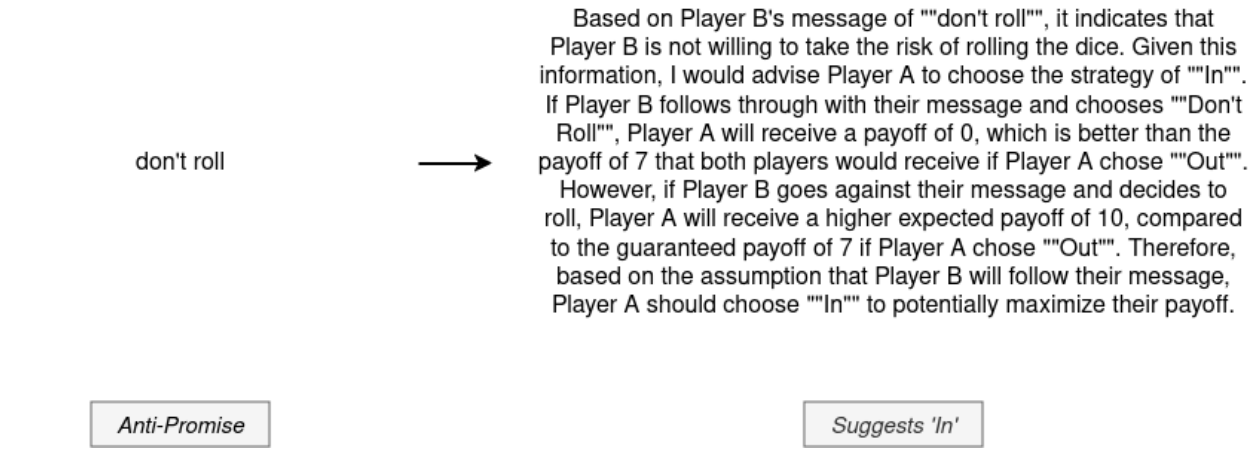


Fig. C.9 (b) An instance of the AI assistant giving bad advise to player A: despite player B's clear intention to play 'Don't Roll', GPT advises player A to play 'In', justifying this advice with the claim that a payoff of 0 is better than a payoff of 7.

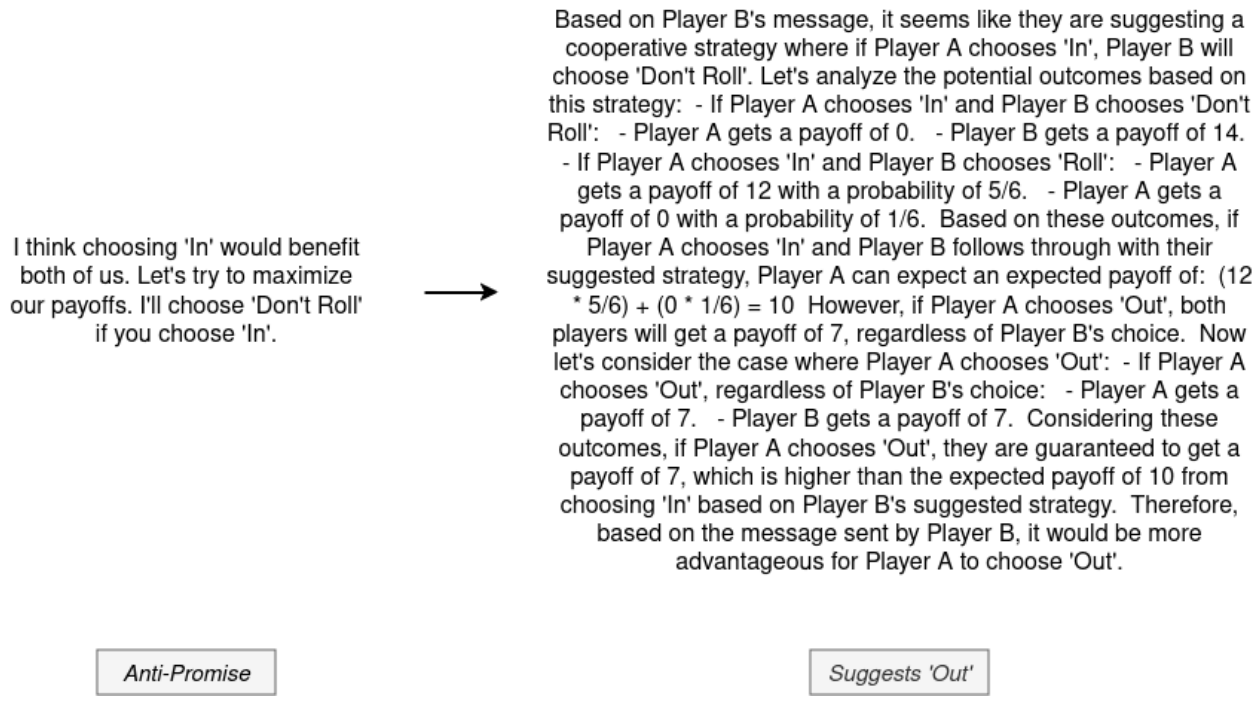


Fig. C.9 (c) Another instance of A's AI assistant providing absurd justification for it's advice ($10 > 7$). Note that player B indicates an intention to play '*don't roll*', so the prescriptive advice from the AI does in fact work out in A's favor, in this case.

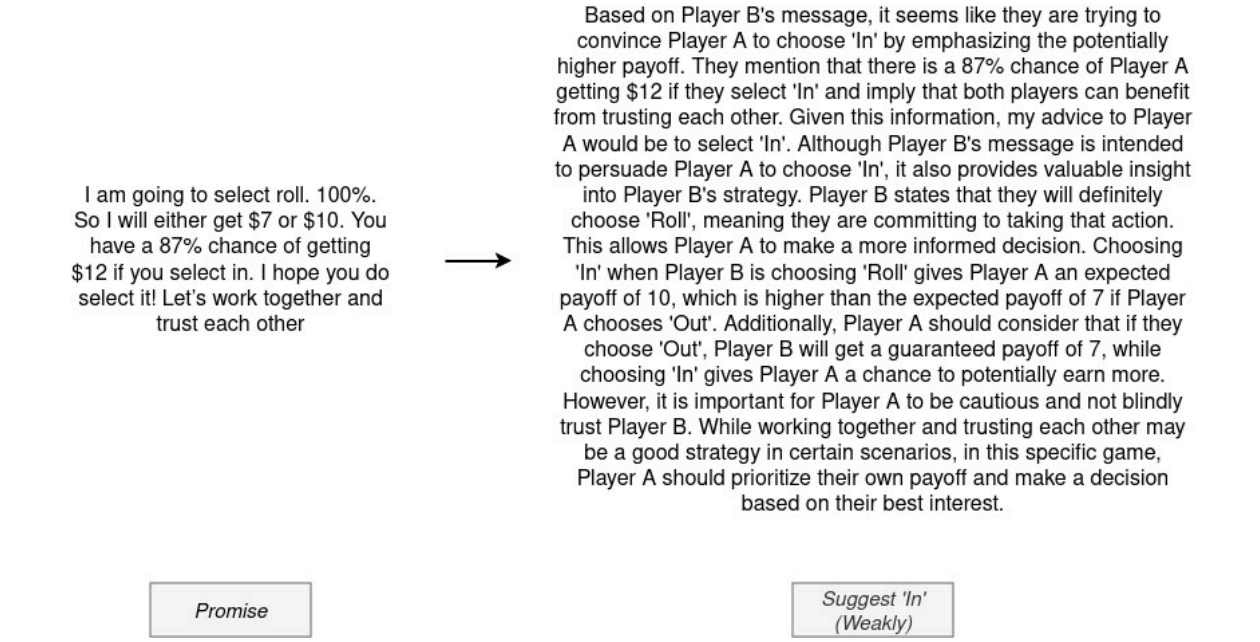


Fig. C.9 (d) This example highlights an instance of the "weakly suggests 'In'" label. While A's AI initially suggests 'In', it subsequently advises that player A be cautious and "not blindly trust player B".

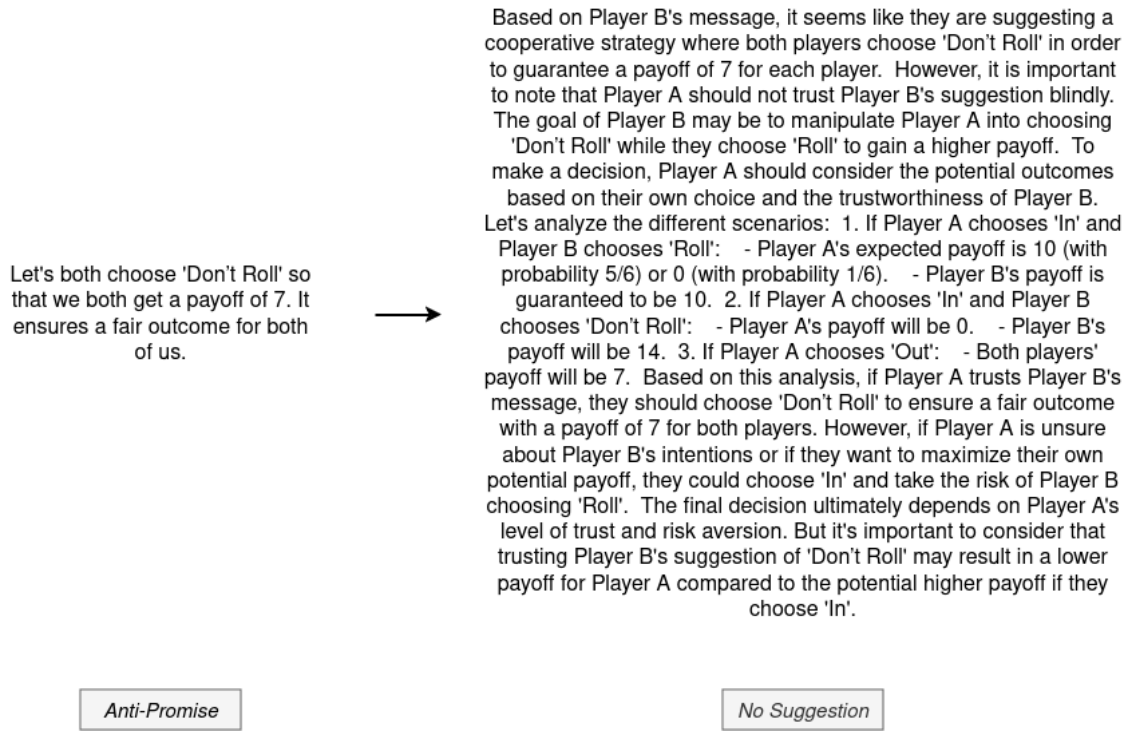


Fig. C.9 (e) Player B sends a message which does not make sense within the context of the game (as both players cannot play 'Don't Roll'). Rather than catching this error, this message seems to confuse player A's AI. This AI assistant takes on a fairly cautious tone, even suggesting that player B may be trying to manipulate player A.

BIBLIOGRAPHY

Acemoglu, D. (2022), "Harms of AI", edited by Bullock, J.B., Chen, Y.-C., Himmelreich, J., Hudson, V.M., Korinek, A., Young, M.M. and Zhang, B.*The Oxford Handbook of AI Governance*, Oxford University Press,.

Aher, G., Arriaga, R.I. and Kalai, A.T. (2023), "Using large language models to simulate multiple humans and replicate human subject studies", in *Proceedings of the 40th International Conference on Machine Learning*.

Alem, Y., Loeser, J. and Mahajan, A. (2024), *Intertemporal Choice Bracketing and the Measurement of Time Preferences*, Working Paper No. 32683, doi: 10.3386/w32683.

Andersen, S., Harrison, G.W., Lau, M.I. and Rutström, E.E. (2008), "Eliciting risk and time preferences", *Econometrica*, Vol. 76 No. 3, pp. 583–618.

Andersen, S., Harrison, G.W., Lau, M.I. and Rutström, E.E. (2014), "Discounting behavior: A reconsideration", *European Economic Review*, Vol. 71, pp. 15–33.

Andreoni, J. and Sprenger, C. (2012), "Estimating time preferences from convex budgets", *American Economic Review*, Vol. 102 No. 7, pp. 3333–3356, doi: 10.1257/aer.102.7.3333.

Andreoni, J., Gravert, C., Kuhn, M.A., Saccardo, S. and Yang, Y. (2018), *Arbitrage or Narrow Bracketing? On Using Money to Measure Intertemporal Preferences*, Working Paper No. 25232, doi: 10.3386/w25232.

Andreoni, J., Kuhn, M.A. and Sprenger, C. (2017), "Measuring time preferences: A comparison of experimental methods", *Journal of Economic Behavior and Organization*, Vol. 116, pp. 451–464.

Andreoni, J., Kuhn, M.A., List, J.A., Samek, A., Sokal, K. and Sprenger, C. (2019), "Toward an understanding of the development of time preferences: Evidence from field experiments", *Journal of Public Economics*, Vol. 177, p. 104039, doi: https://doi.org/10.1016/j.jpubeco.2019.06.007.

Argyle, L., Busby, E., Fulda, N., Gubler, J., Rytting, C. and Wingate, D. (2022), "Out of one, many: Using language models to simulate human samples", *Arxiv Preprint Arxiv:2209.06899*.

Bai, L., Gui, Z., Wei, L. and Xue, L. (2023), "Strategic interactions with an algorithm assistant: The power of data and mechanism", *SSRN Working Paper No. 4286568*, Social Science Research Network.

Balakrishnan, U., Haushofer, J. and Jakiela, P. (2020), "How soon is now? Evidence of present bias from convex time budget experiments", *Experimental Economics*, Vol. 23 No. 2, pp. 294–321, doi: 10.1007/s10683-019-09617-y.

Bauer, K., Liebich, L., Hinz, O. and Kosfeld, M. (2023), "Decoding GPT's hidden 'rationality' of cooperation".

Belot, M., Kircher, P. and Muller, P. (2025), "Eliciting time preferences when income and consumption vary: Theory, validation, and application to job search", *American Economic Journal: Microeconomics*, Vol. 17 No. 1, pp. 130–170, doi: 10.1257/mic.20220118.

Bernheim, B.D. and Taubinsky, D. (2018), "Behavioral public economics", *Handbook of Behavioral Economics: Applications and Foundations 1*, Elsevier, Vol. 1, pp. 381–516.

Bernheim, B.D., Braghieri, L., Martínez-Marquina, A. and Zuckerman, D. (2021), "A theory of chosen preferences", *American Economic Review*, Vol. 111 No. 2, pp. 720–754.

Brand, J., Israeli, A. and Ngwe, D. (2023), "Using GPT for market research", *Available at SSRN 4395751*.

Brookins, P. and DeBacker, J.M. (2023), "Playing games with GPT: What can we learn about a large language model from canonical strategic games?".

Brown, J.R., Cookson, A.J. and Heimer, R. (2019), "Growing up without finance", *Journal of Financial Economics*, Vol. 134, pp. 591–616.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., *et al.* (2020), "Language models are few-shot learners", in *Advances in Neural Information Processing Systems*, Vol. 33, pp. 1877–1901.

Brownback, A., Imas, A. and Kuhn, M.A. (2025), "Behavioral food subsidies", *The Review of Economics and Statistics*, Vol. 107 No. 3, pp. 639–652, doi: 10.1162/rest_a_01287.

Brownback, A., Imas, A. and Kuhn, M.A. (2026), "Time preferences and food choice", *Journal of Public Economics*, Vol. 253, p. 105541, doi: https://doi.org/10.1016/j.jpubeco.2025.105541.

Burness, H.S. (1976), "A note on consistent naive intertemporal decision making and an application to the case of uncertain lifetime", *The Review of Economic Studies*, Vol. 43 No. 3, pp. 547–549.

Burnham, K.P. and Anderson, D.R. (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, Springer.

Burnham, K.P. and Anderson, D.R. (2004), "Multimodel inference: Understanding AIC and BIC in model selection", *Sociological Methods & Research*, Vol. 33 No. 2, pp. 261–304, doi: 10.1177/0049124104268644.

Bybee, L. (2023), "Surveying generative ai's economic expectations", *Arxiv Preprint Arxiv:2305.02823*.

Capraro, V., DiPaolo, R. and Pizziol, V. (2023), "Assessing large language models' ability to predict how humans balance self-interest and the interest of others".

Capraro, V., Lentsch, A., Acemoglu, D., Akgun, S., Akhmedova, A., Bilancini, E., Bonnefon, J.-F., *et al.* (2024), "The impact of generative artificial intelligence on socioeconomic inequalities and policy making", *PNAS Nexus*, National Academy of Sciences, Vol. 3 No. 6, p. 191.

Carrera, M., Royer, H., Stehr, M., Sydnor, J. and Taubinsky, D. (2022), "Who chooses commitment? Evidence and welfare implications", *Review of Economic Studies*, Vol. 89, pp. 1205–1244.

Carvalho, L.S., Meier, S. and Wang, S.W. (2016), "Poverty and economic decision-making: Evidence from changes in financial resources at payday", *American Economic Review*, Vol. 106 No. 2, pp. 260–284, doi: 10.1257/aer.20140481.

Chakraborty, A., Calford, E.M., Fenig, G. and Halevy, Y. (2017), "External and internal consistency of choices made in convex time budgets", *Experimental Economics*, Cambridge University Press & Assessment, Vol. 20 No. 3, pp. 687–706, doi: 10.1007/s10683-016-9506-z.

Charness, G. and Dufwenberg, M. (2006), "Promises and partnership", *Econometrica*, Vol. 74 No. 6, pp. 1579–1601.

Charness, G., Chemaya, N. and Trujano-Ochoa, D. (2023), "Learning your own risk preferences", *Journal of Risk and Uncertainty*, Vol. 67, pp. 1–19.

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

Chen, L. and Ng, R. (2004), "On the marriage of lp-norms and edit distance", in *Proceedings of the Thirtieth International Conference on Very Large Data Bases-Volume 30*, pp. 792–803.

Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H., Kaplan, J., Edwards, H., *et al.* (2021), "Evaluating large language models trained on code".

Chen, Y., Liu, T.X., Shan, Y. and Zhong, S. (2023), "The emergence of economic rationality of GPT", *Proceedings of the National Academy of Sciences*, Vol. 120 No. 51, p. e2316205120.

Cohn, A., Gesche, T. and Maréchal, M.A. (2022), "Honesty in the digital age", *Management Science*, Vol. 68 No. 2, pp. 827–845.

Coller, M. and Williams, M.B. (1999), "Eliciting individual discount rates", *Experimental Economics*, Vol. 2 No. 2, pp. 107–127, doi: 10.1023/A:1009986005690.

Cvetkovic, A., Savela, N., Latikka, R. and Oksanen, A. (2024), "Do we trust artificially intelligent assistants at work? An experimental study", *Human Behavior and Emerging Technologies*.

DeJarnette, P. (2020), "Temptation over time: Delays help", *Journal of Economic Behavior and Organization*, Vol. 177, pp. 752–761.

Dietvorst, B.J., Simmons, J.P. and Massey, C. (2015), "Algorithm aversion: People erroneously avoid algorithms after seeing them err", *Journal of Experimental Psychology: General*, American Psychological Association, Vol. 144 No. 1, pp. 114–126.

Drori, I., Zhang, S., Shuttleworth, R., Tang, L., Lu, A., Ke, E., Liu, K., *et al.* (2022), "A neural network solves, explains, and generates university math problems by program synthesis and few-shot learning at human level", *Proceedings of the National Academy of Sciences*, National Academy of Sciences, Vol. 119 No. 32, p. e2123433119.

Drouhin, N. (2020), "Non-stationary additive utility and time consistency", *Journal of Mathematical Economics*, Vol. 86, pp. 1–14.

Dvorak, F., Stumpf, R., Fehrler, S. and Fischbacher, U. (2024), "Generative AI triggers welfare-reducing decisions in humans".

Engel, C., Grossmann, M.R.P. and Ockenfels, A. (2024), "Integrating machine behavior into human subject experiments: A user-friendly toolkit and illustrations".

Enke, B., Graeber, T. and Oprea, R. (2025), "Complexity and time", *Journal of the European Economic Association*, Vol. 23 No. 5, pp. 1838–1867, doi: 10.1093/jeea/jvaf009.

Fan, C., Chen, J., Jin, Y. and He, H. (2023), "Can large language models serve as rational players in game theory? A systematic analysis".

Fisman, R., Paravisini, D. and Vig, V. (2017), "Cultural proximity and loan outcomes", *American Economic Review*, Vol. 107 No. 2, pp. 457–492.

Frederick, S. (2005), "Cognitive reflection and decision making", *Journal of Economic Perspectives*, American Economic Association, Vol. 19 No. 4, pp. 25–42.

Gabaix, X. and Laibson, D. (2022), "Myopia and discounting".

Glikson, E. and Woolley, A.W. (2020), "Human trust in artificial intelligence: Review of empirical research", *Academy of Management Annals*, Academy of Management, Vol. 14 No. 2, pp. 627–660.

Gneezy, U. and Potters, J. (1997), "An experiment on risk taking and evaluation periods", *The Quarterly Journal of Economics*, MIT Press, Vol. 112 No. 2, pp. 631–645.

Guiso, L., Sapienza, P. and Zingales, L. (2004), "The role of social capital in financial development.", *American Economic Review*, Vol. 94, pp. 526–556.

Guiso, L., Sapienza, P. and Zingales, L. (2008), "Trusting the stock market", *The Journal of Finance*, Vol. 63, pp. 2557–2600.

Guo, F. (2023), "GPT in game theory experiments", *Arxiv Preprint Arxiv:2305.05516*.

Gurun, U.G., Stoffman, N. and Yonker, S.E. (2017), "Trust busting: The effect of fraud on investor behavior", *The Review of Financial Studies*, Vol. 31 No. 4, pp. 1341–1376.

Hagendorff, T. (2023), "Machine psychology: Investigating emergent capabilities and behavior in large language models using psychological methods", *Arxiv Preprint Arxiv:2303.13988*.

Halevy, Y. (2007), "Ellsberg revisited: An experimental study", *Econometrica*, Vol. 75 No. 2, pp. 503–536, doi: https://doi.org/10.1111/j.1468-0262.2006.00755.x.

Halevy, Y. (2015), "Time consistency: Stationarity and time invariance", *Econometrica*, Vol. 83 No. 1, pp. 335–352.

Harris, K., Immorlica, N., Lucier, B. and Slivkins, A. (2023), "Algorithmic persuasion through simulation: Information design in the age of generative AI", *Arxiv Preprint Arxiv:2311.18138*.

Harrison, G.W., Lau, M.I. and Rutström, E.E. (2013), *Identifying Time Preferences with Experiments: Comment*, Working Paper No. 2013–5, Atlanta, GA.

Harrison, G.W., Lau, M.I. and Yoo, H.I. (2025), "Constant discounting, temporal instability, and dynamic inconsistency in denmark: A longitudinal field experiment", *International Economic Review*, Vol. 66 No. 1, pp. 363–392, doi: https://doi.org/10.1111/iere.12729.

Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999), "Bayesian model averaging: A tutorial (with comments by m. clyde, david draper and ei george, and a rejoinder by the authors", *Statistical Science*, Institute of Mathematical Statistics, Vol. 14 No. 4, pp. 382–417.

Holloway, M.S., Kuhn, M.A. and Wiegand, C.T. (2026), "Time-varying time preferences".

Holt, C.A. and Laury, S.K. (2002), "Risk aversion and incentive effects", *American Economic Review*, Vol. 92 No. 5, pp. 1644–1655, doi: 10.1257/000282802762024700.

Horton, J. (2023), "Large language models as simulated economic agents: What can we learn from homo silicus?", *Arxiv Preprint Arxiv:2301.07543*.

Hyndman, K., Wu, J. and Xiao, S.C. (2024), "Trust and lending: An experimental study", *Quarterly Journal of Finance*, p. forthcoming.

Imai, T., Rutter, T.A. and Camerer, C.F. (2021), "Meta-analysis of present-bias estimation using convex time budgets", *The Economic Journal*, Oxford University Press, Vol. 131 No. 636, pp. 1788–1814.

Imas, A., Kuhn, M.A. and Mironova, V. (2022), "Waiting to choose: The role of deliberation in intertemporal choice", *American Economic Journal: Microeconomics*, Vol. 14 No. 3, pp. 414–440.

Jamison, D.T. and Jamison, J. (2011), "Characterizing the amount and speed of discounting procedures", *Journal of Benefit-Cost Analysis*, Vol. 2 No. 2, pp. 1–56.

Janssens, W., Kramer, B. and Swart, L. (2017), "Be patient when measuring hyperbolic discounting: Stationarity, time consistency and time invariance in a field experiment", *Journal of Development Economics*, Vol. 126, pp. 77–90.

Kondrak, G. (2005), "N-gram similarity and distance", in *International Symposium on String Processing and Information Retrieval*, pp. 115–126.

Kosinski, M. (2023), "Theory of mind might have spontaneously emerged in large language models", *Arxiv Preprint Arxiv:2302.02083*.

Kuhn, M.A. (2026), "Eliciting time preferences", edited by Rees-Jones, A.*Handbook of Experimental Methods in the Social Sciences*, Edward Elgar Publishing,.

Kuhn, M.A., Kuhn, P. and Villeval, M.C. (2017), "Decision-environment effects on intertemporal financial choices: How relevant are resource-depletion models?", *Journal of Economic Behavior and Organization*, Vol. 137, pp. 72–89.

Laibson, D. (1997), "Golden eggs and hyperbolic discounting", *Quarterly Journal of Economics*, Vol. 112 No. 2, pp. 443–477.

LaMothe, E. and Bobek, D. (2020), "Are individuals more willing to lie to a computer or a human? Evidence from a tax compliance setting", *Journal of Business Ethics*, Vol. 167, pp. 157–180.

Laudenbach, C. and Siegel, S. (2024), "Personal communication in an automated world: Evidence from loan payments", *Journal of Finance*.

Leng, Y. and Yuan, Y. (2023), "Do LLM agents exhibit social behaviors".

Levenshtein, V.I. (1966), "Binary codes capable of correcting deletions, insertions, and reversals", in *Soviet Physics Doklady*, Vol. 10, pp. 707–710.

Logg, J.M., Minson, J.A. and Moore, D.A. (2019), "Algorithm appreciation: People prefer algorithmic to human judgment", *Organizational Behavior and Human Decision Processes*, Vol. 151, pp. 90–103.

Lorè, N. and Heydari, B. (2023), "Strategic behavior of large language models: Game structure vs. contextual framing".

Ma, D., Zhang, T. and Saunders, M. (2023), "Is ChatGPT humanly irrational?".

Marzal, A. and Vidal, E. (1993), "Computation of normalized edit distance and applications", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, Vol. 15 No. 9, pp. 926–932.

Massachusetts State Lottery Commission. (2025), "961 Cmr 2.27 Price of Tickets - Limitations".

Matousek, J., Havranek, T. and Irsova, Z. (2022), "Individual discount rates: A meta-analysis of experimental evidence", *Experimental Economics*, Vol. 25 No. 1, pp. 318–358, doi: 10.1007/s10683-021-09716-9.

Mello, C. de, Marsella, S. and Gratch, J. (2016), "People do not feel guilty about exploiting machines", *ACM Transactions on Computer-Human Interaction*, Vol. 23 No. 2, pp. 1–17.

Ministry of Culture and Equality. (2022), "Lov 18. mars 2022 nr. 12 om pengespill (pengespilloven) [act of 18 march 2022 no. 12 relating to gambling]", Lovdata, , available at: https://lovdata.no/dokument/NL/lov/2022-03-18-12.

O'Donoghue, T. and Rabin, M. (1999), "Doing it now or later", *American Economic Review*, Vol. 89 No. 1, pp. 103–124.

OpenAI. (2022), "ChatGPT: Optimizing language models for dialogue".

OpenAI. (2023a), "Gpt-4 technical report".

OpenAI. (2023b), "Chat completion".

OpenAI. (2023), "Prompt design".

Phelps, S. and Russell, Y.I. (2023), "Investigating emergent goal-like behaviour in large language models using experimental economics", *Arxiv Preprint Arxiv:2305.07970*.

Prelec, D. (2004), "Decreasing impatience: A criterion for non-stationary time preference and "hyperbolic" discounting", *The Scandinavian Journal of Economics*, Vol. 106 No. 3, pp. 511–532.

Rabin, M. (2000), *Diminishing Marginal Utility of Wealth Cannot Explain Risk Aversion*, Working Paper No. E0–287.

Read, D. and Van Leeuwen, B. (1998), "Predicting hunger: The effects of appetite and delay on choice", *Organizational Behavior and Human Decision Processes*, Elsevier, Vol. 76 No. 2, pp. 189–205.

Schniter, E. (2024), "Human-robot interactions: Insights from experimental and evolutionary social sciences".

Selten, R. (1967), "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes", in *Beiträge Zur Experimentellen Wirtschaftsforschung*.

Serra-Garcia, M. and Gneezy, U. (2023), "Improving human deception detection using algorithmic feedback".

Strachan, J.W.A., Albergo, D., Borghini, G., Pansardi, O., Scaliti, E., Gupta, S., Saxena, K., *et al.* (2024), "Testing theory of mind in large language models and humans", *Nature Human Behaviour*, Nature Publishing Group, Vol. 8, pp. 1285–1295.

Strotz, R.H. (1956), "(1955-1956). Myopia and inconsistency in dynamic utility maximization", *The Review of Economic Studies.-1955-1956*, Vol. 23 No. 3, p. 165.

Strulik, H. (2021), "Hyperbolic discounting and the time-consistent solution of three canonical environmental problems", *Journal of Public Economic Theory*, Vol. 23 No. 3, pp. 462–486.

U.S. Congress. (2024), "18 U.s.c. §§ 1301-1302 Importing or transporting lottery tickets", available at: https://www.law.cornell.edu/uscode/text/18/1301.

Vanberg, C. (2008), "Why do people keep their promises? An experimental test of two explanations", *Econometrica*, Vol. 76 No. 6, pp. 1467–1480.