UNIVERSITY OF CALIFORNIA SAN DIEGO

Essays in Experimental Economics: Social Choice

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Economics

by

Evgenii Baranov

Committee in charge:

Professor Joel Sobel, Chair
Professor Songzi Du
Professor Craig McKenzie
Professor Denis Shishkin
Professor Isabel Trevino

2024

The Dissertation of Evgenii Baranov is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

DEDICATION

To my dad, in loving memory.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Joel Sobel for his patience, advise, guidance, and unconditional support as the chair of my committee. Joel was incredibly supportive during the most difficult times of my life. I feel very lucky to have had a chance to work with him.

I am also indebted to my committee members Isabel Trevino, Songzi Du, Denis Shishkin, and Craig McKenzie for their encouragement and feedback. I also want to thank Emanuel Vespa, Joel Watson, and Mark Machina for many engaging discussions throughout the program.

I am grateful to my parents and my wife, Wendy, for making my dreams possible and always being there for me.

Chapter 1 is co-authored with Wolanski, Adrian and is currently being prepared for submission for publication. The dissertation author is one of primary authors of this chapter.

Chapter 2 is co-authored with Wolanski, Adrian and is currently being prepared for submission for publication. The dissertation author is one of primary authors of this chapter.

Chapter 3 is co-authored with Kisilev Dmitrii, Marchenkov Valerii, and Makarov Ilya, and is currently being prepared for submission for publication. The dissertation author is one of primary authors of this chapter.

All errors are my own.

VITA

2016	B.S in Economics, Higher School of Economics
2016	B.S in Economics, University of London
2018	M.A in Economics, University of California San Diego
2024	Ph.D in Economics, University of California San Diego

ABSTRACT OF THE DISSERTATION

Essays in Experimental Economics: Social Choice

by

Evgenii Baranov

Doctor of Philosophy in Economics

University of California San Diego, 2024

Professor Joel Sobel, Chair

This dissertation consists of three chapters that examine social preferences in various economic contexts.

In the first chapter, we provide a theoretical approach for investigating attitudes towards intertemporal inequality. We generalize the Pigou-Dalton principle to intertemporal settings by formulating several partial orders on the space of income streams. Three of the partial orders account for payments received over the lifespan of the stream and differ only by how intensely one stream dominates another. A fourth partial order only accounts for the level of inequality experienced in a specific period, rather than over the lifespan of the stream. We then perform a laboratory experiment to distinguish the empirical relevance of these different partial orders

and inequality rankings. We find that orders that rank whole streams accurately reflect how participants view streams for themselves, but these views do not translate into how they choose for others. Instead, many of our participants display inequality aversion based on period-wise outcomes.

In the second chapter, we study when it is possible to link a social planner's preferences across groups of different agents. We propose a preference consistency criterion that relates members of a family of social preferences across domains of different agents; this criterion requires preferences to be identical on domains differing only by adding agents with choice-independent payoffs. We derive additional domain changes for which consistent preferences are invariant, and test adherence to these predictions in an online laboratory experiment. While consistency rates are reasonably high, we document significant differences in consistency across the different types of domain changes. Additionally, we find that participants tend to choose options with higher inequality/lower inefficiency as domain size increases.

In the third chapter, we use large language models (LLMs) for representing human behavior, as a mean to bypass confidentiality constraints. We show that LLMs demonstrate behavior agreeing with real human subjects regarding social choice consistency. Moreover, we investigate the impact of environmental factors on social choice consistency, emphasizing the sensitivity of social choices to changes in the surrounding context, such as the presence of close friendships and shared characteristics among participants.

Chapter 1

Attitudes toward Intertemporal Inequality

1.1 Introduction

Issues of inequality and how to measure it have been studied in the economic literature for over a hundred years, with many researchers developing their own approaches to measuring inequality Gini (1921), Theil (1967), Atkinson (1970). These inequality measures offer slightly different interpretations of what contributes to higher levels of inequality. However, all of them agree on the Pigou-Dalton principle — a transfer from a poorer person to a richer person cannot decrease inequality in society. There have been approaches to generalize this principle from single-variable inequality to more general settings, starting with works of Kolm (1977), Maasoumi (1986), and Tsui (1999).

This paper seeks to evaluate the implications of the widely studied measures of inequality from one-period, one-good environments to a multi-period setting. In particular, we examine three different rankings for intertemporal income streams, that differ by the access to saving and borrowing technologies: one assumes only direct hand-to-mouth consumption, another allows saving money under the mattress, and the third one allows borrowing or lending at a given interest rate. We use these income rankings as a base for establishing inequality-reducing transfers in line with the Pigou-Dalton principle and also offer the fourth alternative – measuring inequality levels at each period separately, rather than over the entire stream.

In a laboratory experiment, we test the proposed income rankings for personal decisions and find that the subjects exhibit strict preferences for receiving weakly more money, but when receiving the same total amount spread over the periods, their choices are indistinguishable from random. This suggests the lack of the use of borrowing/lending resources by student subjects in our given timeframe.

In general, we consider the planner's problem of ranking social allocation streams of income. Every agent receives an income stream paid out across 3 time periods. Besides that, every agent is assumed to be identical. We study social preferences of the planner in this setting. The problem is identical to studying multidimensional inequality, but with additional structure imposed since each dimension represents the same good (money) at a different point in time. It is not obvious how the bundles (10 apples, 2 bananas) and (5 apples, 3 bananas) should be compared by an agent, but if instead those bundles were to be ($10 today, $2 tomorrow) and ($5 today, $3 tomorrow), then it is reasonable to claim that the former is better. Our proposed income rankings can be viewed as a generalization of this simplistic example.

Pigou-Dalton transfer principle is a basis for measuring inequality, and there are multiple reasonable generalizations of it from the single good single period case. With our application to intertemporal income streams, there are multiple ways to define what it means to have a regressive transfer. Our proposed notions of stream dominance are designed to address that part of the generalization. We then provide generalizations of the Pigou-Dalton principle to multi-period settings for two types of dominance rankings: period-wise and whole-stream-wise. Period-wise dominance rankings only look at the income received up to the point where the decision is made and do not account for the future, not-yet-received, income. Our whole-stream dominance ranking instead aggregates all of the payments coming from an income stream, including past, present, and future not-yet-received payments. Of particular interest is the fact that these notions need not always agree—one stream might have lower period-wise inequality but higher whole-stream inequality than another.

Our weakest notion of stream dominance, vector dominance, is explored in Lasso de la Vega et al. (2010), Basili et al. (2017), which use standard vector rankings and majorization criteria for generalizing Pigou-Dalton principle to multidimensional inequality. One income stream would vector dominate another one if it pays at least as much money. Our findings confirm that it is indeed the most natural notion for ranking individual streams, commonly used by agents.

Bosmans et al. (2009) also proposes a multidimensional generalization of Pigou-Dalton, that uses the planner's trade-off rates between different attributes to define regressivity. In that regard, our work could be viewed as a special case of multidimensional generalization of Bosmans et al. (2009), but we have additional structure imposed from the fact that the agents are consuming the same good at every period. This allows us to propose more sensible dominance rankings that are unlikely to arise from individual trade-off rates between different goods. Additionally, individual trade-off rates between goods could differ drastically from one agent to the other, since it is unclear how to rank access to healthcare against a college degree. It is much clearer, however, how to rank money today versus money tomorrow. For the income streams, we find that a notion of cumulative dominance—comparing streams on the basis of accumulated sums at every period—is commonly spread.

While there are many possible rankings of streams and these rankings often produce agreeing definitions of regressive transfers, this is not always the case. It is possible for one transfer to be regressive under one ranking but progressive under another. As an example, consider two agents, Ann and Bob, with respective consumption streams $C_a = (6,1,6)$ and $C_b = (1,11,1)$ that denote \$ amounts consumed in each of the 3 periods. Our third notion of whole-stream inequality comes from a ranking using exponential discounting. Note that regardless of the discount factor, C_a always yields a higher discounted value than C_b, since: $6+\delta+6\delta^2 > 1+11\delta+\delta^2 \to 5-10\delta+5\delta^2 > 0 \to 5(1-\delta)^2 > 0$, which is true for $0 < \delta < 1$. An example of an inequality-increasing transfer would take \$2 from Bob in period 2 and give it to Ann, and then take \$2 from Ann in period 3 and give it to Bob. This results in new outcomes $\hat{C}_a = (6,3,4)$ and $\hat{C}_b = (1,9,3)$, which is regressive under exponential discounting. However, note that in period 2, Ann has accumulated \$7 while Bob has \$12, while the new outcome gives

Ann an accumulated income of \$9 in period 2 and Bob an accumulated income of \$10 in period 2, without changing the accumulated income in any other period. Such a transfer is then progressive in terms of reducing differences in period-wise accumulated income.

As this example illustrates, the same transfer may be viewed as both regressive and progressive depending on whether the decision-maker ranks streams using period-wise or whole-stream income ranking. This paper argues that it is possible to have different agents express aversion to inequality, but still disagree on whether a particular transfer is regressive or progressive depending on the ranking they use. We show experimentally that one of these rankings, the period-wise ranking of cumulative payments, is empirically supported by far more participants than any whole-stream rankings of payments (even the most stringent ranking of vector dominance used frequently in the literature). To summarize, our contribution comes in three parts: we note the tension between whole-stream and period-wise dominance rankings used to generalize regressivity of the transfers, and thus leading to contradictory predictions for changes in inequality; we then take it to the laboratory experiment and show that there are indeed agents of different types: the majority cares about period-wise inequality, but some do care about life-time comparisons; finally, we observe individual decisions that are inconsistent with the commonly used exponential discounting.

Overall, we recruited 118 participants for a laboratory experiment in which each subject acts as a social planner selecting payment streams for a pair of other participants. In this environment, we find evidence that about 22% of the participants do not have any altruistic motives. Among the others, however, a much larger share of subjects' behavior is explained by period-wise inequality aversion rather than whole-stream inequality aversion. We also offer participants choices over streams for themselves that differ by various whole-stream rankings, and find that a large share of participants chooses dominated options (especially for our weakest ranking based on exponential discounting).

We also offer some choices that are slight variations of each other, where we introduce a very small inefficiency into the transfer. This allows us to see if some planners would be willing to forgo some efficiency to reduce inequality. While this is related to the literature on altruistic behavior and warm-glow Andreoni (1990), the main difference is that planners are giving up other people's payments (rather than their own) to reduce inequality. For an extensive overview of experimental work on inequality aversion, we refer the reader to Clark and d'Ambrosio (2015).

Zuber (2011) and Jackson and Yariv (2015) point out that if both individual and social preferences are represented by Exponential Discounted Utility with different discount factors, then a social criterion satisfying stationarity, time consistency, and the Pareto principle is dictatorial. There have been many different approaches to weakening some of these axioms to avoid the impossibility result: Billot and Qu (2022) and Feng and Ke (2018) weaken the Pareto Condition, Hayashi (2016) and Millner and Heal (2018) argue against time invariance, while Miyagishima (2022) weakens axioms on inequality aversion. Our paper provides experimental evidence in favor of the latter two approaches; our participants often choose exponentially discounting dominanted options over dominant options, and display aversion to period-wise rather than whole-stream inequality.

The rest of the paper is structured as follows. Section 2 formally defines stream dominance relationships, that are later used to provide whole-stream and period-wise generalizations of the Pigou-Dalton principle. Section 3 provides details on the experimental design. Section 4 shares our main findings from the experiment and section 5 concludes the paper.

1.2 Theoretical Framework

1.2.1 Setup

A society consists of $n = 2$ agents [1], that live for $T \in \mathbb{N}$ periods. Let $I^j \in \mathbb{R}^T$ be the stream of income[2] received by agent j, with $I_t^j \in \mathbb{R}$ being the income received by agent j in period t.

[1] It is straightforward to generalize our structure to settings with more than two agents; we consider this case only for expositional simplicity.

[2] We focus on income rather than consumption; as such, we will be discussing indirect utility functions over income rather than utility function over consumption.

1.2.2 Intertemporal choice problems

Each agent in society faces an intertemporal optimization problem of the form:

$$\max_{(c_1,\ldots,c_T)} U(c_1,\ldots,c_T)$$

$$\text{Subject to: } \sum_{t=1}^{T} c_t p_t \leq \sum_{t=1}^{T} I_t p_t$$

Where p_t measures relative prices of consumption in period t, given existing saving/borrowing techniques available to the consumers. To account for the value of flexibility, we will assume that $p_t \geq p_{t+1}$ for all t. There are many additional restrictions we can impose on the budget set. One example is when the agent does not have access to any saving or borrowing technology. In that case, the agent's constraint would be

Setup 1.

$$\forall t \leq T;\ c_t \leq I_t$$

Such a consumer would always be hand-to-mouth regardless of the level of income they receive.

Another example is when the agent has access to a savings technology without interest or depreciation, but cannot borrow.

Setup 2.

$$\forall \tau \leq T;\ \sum_{t=1}^{\tau} c_t \leq \sum_{t=1}^{\tau} I_t$$

An additional constraint possibility is that the agent can save or borrow between any periods at a constant interest rate of r. Such a constraint would be

Setup 3.

$$\sum_{t=1}^{T} \frac{c_t}{(1+r)^{t-1}} \leq \sum_{t=1}^{T} \frac{I_t}{(1+r)^{t-1}}$$

It is important to note that we have assumed nothing about the agents preferences here; we have only provided a description of the intertemporal budget constraints the agent faces.

1.2.3 Partial Orders on Streams

The difficulty of discussing inequality in intertemporal settings is that it is not always obvious when one person's income stream is better than another's, since streams (vectors) of income are harder to rank than scalars. To this end, we define three variations of *stream dominance rankings*, which are partial orders over the space of income streams $I = (I_1, I_2, \ldots, I_T)$. The first of these is the standard partial order over $\mathbb{R}^T$, in which one stream is better than another if it pays a weakly larger amount in each period.

Definition 1. The stream $I^A \in X$ is called *vector dominant* over the stream $I^B \in X$ if $\forall t \;\; I^A_t \geq I^B_t$, and we denote it as $I^A R^V I^B$.

Proposition 1. *The indirect utility for an agent facing any of the three setups in section 2.2 is always weakly higher from stream I^A than from stream I^B whenever $I^A R^V I^B$.*

The second partial order we define makes use of the natural relationship between periods in an intertemporal budget: moving income from a later period to an earlier period provides greater flexibility

Definition 2. The stream $I^A \in X$ is called *cumulatively dominant* over the stream $I^B \in X$ if $\forall \tau \;\; \sum_1^\tau I^A_t \geq \sum_1^\tau I^B_t$, and we denote it as $I^A R^C I^B$.

Cumulatively dominant streams have sufficiently more consumption earlier in time such that there is a way to rearrange consumption only through savings that results in the previously cumulatively dominant stream now being vector dominant.[3] An example of a cumulatively dominant stream relationship is the example provided in section 1 of $I^A = (10, 2)$ and $I^B = (5, 3)$. While I^A does not vector dominate I^B, it does cumulatively dominate it since $I^A_1 = 10 > I^B_1 = 5$ and $I^A_1 + I^A_2 = 10 + 2 > I^B_1 + I^B_2 = 8$.

[3] We refer to this statement as proposition 0, which is stated formally and proved in the appendix.

Proposition 2. *The indirect utility for an agent facing setup 2 or 3 in section 2.2 is always weakly higher from stream I^A than from stream I^B whenever $I^A R^C I^B$.*

Definition 3. The stream $I^A \in X$ is called *discounting dominant*[4] over the stream $I^B \in X$ if $\forall \delta \in (0,1); \sum_1^T \delta^t I_t^A \geq \sum_1^T \delta^t I_t^B$, and we denote it as $I^A R^D I^B$.

A discounting dominant stream has sufficiently more income earlier in time such that there is a way to rearrange both streams through saving and borrowing at an interest rate of $1-\delta$ per period that results in a vector dominant stream regardless of the value of δ.[5].

Proposition 3. *The indirect utility for an agent facing setup 3 in section 2.2 is always weakly higher from stream I^A than from stream I^B whenever $I^A R^D I^B$.*

While these rankings[6] capture many ways in which one stream can be better than another, they all describe the totality of the streams without describing what happens at various points in time. It may be reasonable to instead care about which stream is better *at some point in time* rather than in total. We propose the following:

Definition 4. The *cumulative income* of stream I at time t is $w_t^I = \sum_{\tau=1}^t I_\tau$. We say that stream I^A is *richer or wealthier* than stream I^B at time t if $w_t^{I^A} \geq w_t^{I^B}$.

Note that if one stream is cumulatively dominant over another, the cumulative consumption of the dominant stream is higher than that of the dominated stream at each period. In general, however, the rankings of cumulative income can change over time. This definition allows us to compare any two streams period-by-period without requiring a relationship across the whole stream.

[4]This definition and its characterization is due Chambers and Echenique (2018)

[5]This is merely a reinterpretation of the definition of discounting dominance, where the streams are rearranged to have positive income in the first period only and this consumption level is the present value of the stream with discount factor δ

[6]Note that these dominance rankings are nested: vector dominance implies cumulative dominance, and cumulative dominance implies discounting dominance.

1.2.4 Inequality Aversion

Proposed dominance notions serve to identify who is richer and who is poorer in an intertemporal setting. We use these to predict how a planner who dislikes inequality might rank pairs of streams. We construct generalized variants of the Pigou-Dalton principle[7], which utilize the different stream dominance rankings discussed above. We have:

Definition 5. The *generalized Pigou-Dalton principle with stream dominance ranking R*: given (I^A, I^B) and $(I^A - \varepsilon, I^B + \varepsilon)$, for all $\varepsilon \in \mathbb{R}^T$ such[8] that $I^A R (I^A - \varepsilon) R (I^B + \varepsilon) R I^B$ then $(I^A - \varepsilon, I^B + \varepsilon) \succeq (I^A, I^B)$.

That is, a transfer from the dominant to the dominated stream results in an allocation the planner prefers (provided that the transfer is sufficiently small to preserve the dominance ranking between the new streams). Proposition 4 states that such transfers can always be found for our dominance notions.

Proposition 4. *For $R = R^V, R^C, R^D$, if $I^A R I^B$ and $I^A \neq I^B$, there exists $\varepsilon \neq 0$ such that $I^A R (I^A - \varepsilon) R (I^B + \varepsilon) R I^B$.*

We then provide a definition of the generalized Pigou-Dalton principle and an analog of proposition 2 for period-wise inequality rankings.

Definition 6. The *generalized Pigou-Dalton principle for period-wise inequality*: if $w_\tau^{I^A} \geq w_\tau^{I^A - \varepsilon} \geq w_\tau^{I^B + \varepsilon} \geq w_\tau^{I^B}$, or $w_\tau^{I^B} \geq w_\tau^{I^B + \varepsilon} \geq w_\tau^{I^A - \varepsilon} \geq w_\tau^{I^A}$ for all τ, then we have $(I^A - \varepsilon, I^B + \varepsilon) \succeq (I^A, I^B)$.

That is, a transfer that preserves who is richer at each period but (weakly) decreases the difference in cumulative income in each period generates a preferred allocation. Such transfers

[7]The standard (single period single good) Pigou-Dalton principle says that if $I^A > I^A - \varepsilon > I^B + \varepsilon > I^B$, then $(I^A - \varepsilon, I^B + \varepsilon) \succ (I^A, I^B)$. All of our results reduce to the standard Pigou-Dalton principle when $T = 1$.

[8]The components of ε can be positive or negative; the only restriction on the transfer stream is that $I^A R (I^A - \varepsilon) R (I^B + \varepsilon) R I^B$ then $(I^A - \varepsilon, I^B + \varepsilon) \succeq (I^A, I^B)$.

are always guaranteed to exist for any pair of (different) streams. We provide the proof in the appendix[9].

Proposition 5. *For any streams I^A and I^B where $I^A \neq I^B$, there exists an ε such that a planner satisfying the generalized Pigou-Dalton principle for period-wise inequality prefers $(I^A - \varepsilon, I^B + \varepsilon)$ to (I^A, I^B).*

Both generalized Pigou-Dalton principles are consistent with the idea that moving income from richer individuals to poorer individuals reduces inequality, and that lower inequality is preferred to higher inequality. These versions differ only by how they determine who is richer and who is poorer. Both principles also agree that providing each agent with the same stream minimizes the level of inequality and yields the most preferred outcome.

Proposition 6. *If a planner satisfies either the generalized Pigou-Dalton principle for whole-stream or for period-wise inequality, then $\forall I$, there does not exist $\varepsilon \neq 0$ such that $(I - \varepsilon, I + \varepsilon) \succeq (I, I)$.*

Transfers that are preferred by planners satisfying one generalized Pigou-Dalton principle may not always be preferred by planners satisfying the other. That is, there are pairs of streams and transfers from a discounting dominant stream to a discounting dominated stream that preserve the dominance ranking, but also weakly increase the difference between cumulative consumption in all periods. Consider the following example from section 1 between $I_a = (6,1,6)$ and $I_b = (1,11,1)$, and $\hat{I}_a = (6,3,4)$ and $\hat{I}_b = (1,9,3)$. Note that $\hat{I}_A R^D I_A R^D I_B R^D \hat{I}_B$, so a planner who is whole-stream inequality averse with discounting dominance would prefer (I_A, I_B) to $(\hat{I}_A, \hat{I}_b)$. However, note that the cumulative consumption vectors of the streams are $w_{I_A} = (6,7,13)$, $w_{I_B} = (1,12,13)$, $w_{\hat{I}_A} = (6,9,13)$ and $w_{\hat{I}_B} = (1,10,13)$. The pair $(\hat{I}_A, \hat{I}_b)$ has a smaller difference in cumulative consumption between Ann and Bob in period 2 than (I_A, I_B) does, with identical

[9] The idea behind the proof of both of these propositions is largely the same; that the partial orders we are using in each case are convex binary relations

differences in cumulative consumption in the other periods. A planner who is period-wise inequality averse would therefore prefer $(\hat{I}_A, \hat{I}_b)$ to (I_A, I_B).

We provide this example to demonstrate that it is reasonable for people to be inequality-averse but disagree over the inequality effects of a particular policy due to using different definitions of inequality. This feature is unique to the intertemporal setting, as neither the single period nor the general multidimensional setting admits multiple sensible ways of discussing inequality reduction. This then leads us to an empirical exercise to determine which, if any, of our notions of intertemporal inequality aversion describe empirical behavior.

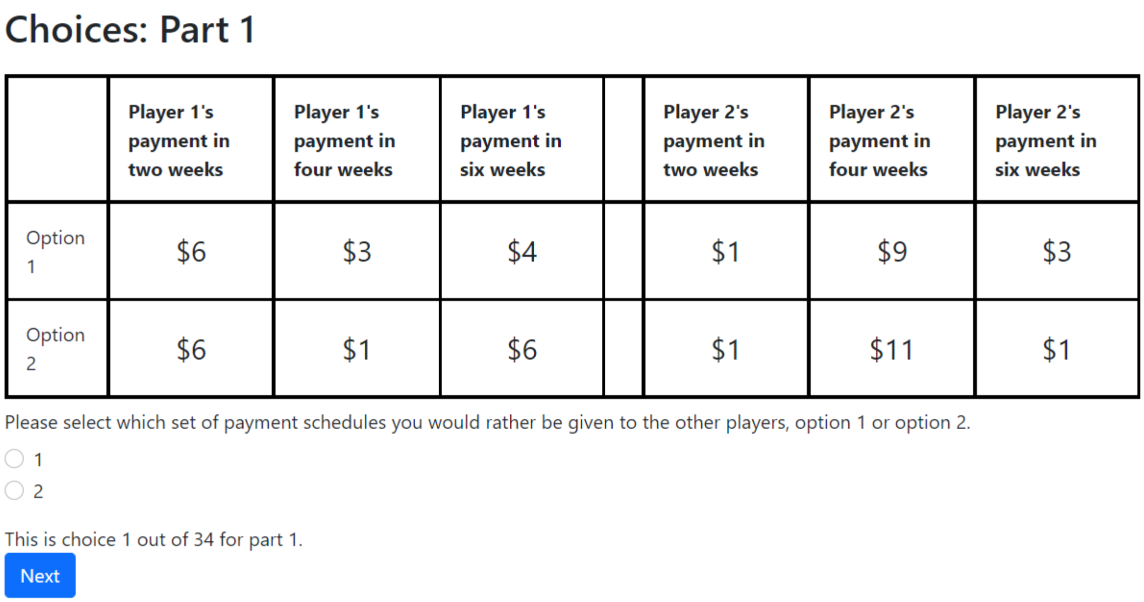
1.3 Experimental Design

There are two goals of our laboratory experiments. The first is investigating if participants find our intertemporal inequality definitions sufficiently descriptive. The second purpose is attempting to disentangle which, if any, stream rankings participants find more important.

We recruited 118 UCSD students using the UCSD Economics Laboratory subject pool. Participants make two sets of decisions[10], one set of choices as social planners for pairs of other participants, and a second set of decisions only for themselves. All choices were binary over pairs of streams for the relevant agents. When choosing for others, the pair of participants remained the same across all decisions but the identity of these other participants is unknown to the planner. They were simply referred to as player 1 and player 2, with the planners do knowing which participant is assigned to which role. An example of the decision interface the participants saw is included as Figure 1.

These social planner choices are designed to measure consistency with our various generalized Pigou-Dalton principles. The pairs of streams in many of these choices are related by an inequality-reducing transfer (according to one of our generalized Pigou-Dalton principles). Eleven choices have a transfer that reduces both whole-stream (2 for vector dominance, 5 for cumulative dominance, and 4 for discounting dominance) and period-wise inequality. Another

[10]The experiment is conducted using the oTree platform created by Chen et al. (2016)



Choices: Part 1

	Player 1's payment in two weeks	Player 1's payment in four weeks	Player 1's payment in six weeks		Player 2's payment in two weeks	Player 2's payment in four weeks	Player 2's payment in six weeks
Option 1	$6	$3	$4		$1	$9	$3
Option 2	$6	$1	$6		$1	$11	$1

Please select which set of payment schedules you would rather be given to the other players, option 1 or option 2.

- 1
- 2

This is choice 1 out of 34 for part 1.

Next

Figure 1.1. Example of decision interface

5 choices involve a transfer that reduces period-wise inequality, but is unranked according to whole-stream inequality for any ranking above. An addition 5 choices involve a transfer that increases period-wise inequality while reducing stream-wise inequality under discounting dominance. The final 3 choices pay a positive amount in exactly one period, and this period is the same for each person and stream. This replicates the conditions of a static choice environment while keeping the decision interface the same as the other choices. The remaining 10 choices are what we call *perturbation choices*, which we construct by introducing a small inefficiency in the transfer (i.e. a transfer of $2 from player 1 to player 2 would become a transfer of $2 from player 1 and a transfer of $1.90 to player 2) in an existing choice. These perturbation choices allow us to determine if participants dislike inequality enough to trade (another player's) money to reduce inequality.

In addition to being social planners for other participants, we also asked participants to make several binary decisions about payment streams for themselves. These choices are designed to measure if the participants find that particular dominance notions adequately describe which streams are better. As such, the streams in these choices do not involve any transfers; instead, they simply present one dominant stream and one dominated stream according to either vector dominance, cumulative dominance, or discounting dominance.

The choices participants make for themselves and the single period choices both help us determine why participants' behavior might deviate from the predictions of model. Participants who choose dominant streams for themselves but make fail to select inequality-reducing options for others may not be inequality averse in general. Similarly, participants who choose the inequality-averse option in the single-period setting but fail to select dominant streams for themselves may not believe these rankings adequately rank streams.

1.4 Results

1.4.1 Choosing for oneself

We begin our analysis with the second part of the experiment, where subjects were choosing between different payment streams for themselves.

Descriptive Statistics

Table 1.1 describes the percentage of participants or percentage of choices where participants selected the dominant stream for themselves.

The results indicate that these dominance notions are indeed different in practice. Participants selected the vector dominant option essentially every time it was available, but selected the discounting dominant option only slightly[11] more frequently than half of the time.

We next build upon the results of table 1.1 using a linear probability model (LPM) and a logit model. We present the result in table 1.2. The regressions contain indicator variables for each dominance type, utilize both fixed effects and standard error clustering at the participant level, and can also disaggregate the effects of cumulative dominance by the sum of the streams.

Result 1 *Participants choose dominant options more often when they are vector dominant than cumulatively dominant, and more often when they are cumulatively dominant than discounting*

[11] The p-value of the z test against the null hypothesis of choosing this option exactly half of the time is 0.0702.

Table 1.1. Proportion of dominant choices participants selected for themselves.

	vector	cumulative	discounting
proportion of dominant choices	0.9870	0.8125	0.5410
proportion of participants making the dominant choice every time	0.9830	0.4661	0.2203
proportion of participants making the dominant choice at least half of the time	1	0.9915	0.7373

Table 1.2. Change in likelihood of choosing the dominant option when it is cumulatively dominant or discounting dominant (baseline is vector dominant). Standard errors included in parenthesis and are clustered at the participant level. * is significance at 90%, ** is significance at 95%, and *** is significance at 99%.

	LPM	LPM	LPM	LPM	Logit	Logit	Logit	Logit
cumulative, aggregate	-.1745 (.0158)***		-.1696 (.0157)***		-2.8643 (.6679)***		-2.8330 (.6687)***	
discounting, aggregate	-.4460 (.0275)***	-.4460 (.0275)***	-.4289 (.0282)***	-.4340 (.0282)***	-4.1664 (.7333)***	-4.1664 (.7333)***	-4.0575 (.7352)***	-4.0763 (.7382)***
cumulative, different total payment		-.0201 (.0167)**		-.0244 (.0111)**		-.9558 (.5815)*		-.9872 (.5809)*
cumulative, same total payment		-.2755 (.0224)***		-.2670 (.0226)***		-3.4279 (.6840)***		-3.3653 (.6884)***
constant (vector dominance)	.9870 (.0091)	.9870 (.0091)	.9782 (.0131)	.9808	4.3307 (.7075)	4.3307 (.7075)		-4.0763 (.7382)
Log Likelihood					-679.3026	-636.7355	-490.1338	-447.775
(pseudo) R^2	.1240	.1652	.1240	.1652	.1196	.1748	.1452	.1452
-451.84031								
N	1,330	1,330	1,330	1,330	1,330	1,330	1,246	1,246
# clusters	118	118	118	118	118	118	106	106
Participant FE	No	No	Yes	Yes	No	No	Yes	Yes

dominant. However, participants choose the cumulatively dominant option far more often when the total payment of the streams is different than when it is the same.

The first half of this result is not particularly surprising. The second part, however, indicates that cumulative dominance does not adequately capture how participants view streams. This result suggests that participants evaluated streams using a heuristic based on total payment received rather than a consideration of the distribution of payments across time. Additionally, this result suggests that participants can somewhat distinguish between the different types of streams. This indicates that a failure to choose streams with lower intertemporal inequality for others does not stem from a failure to identify which streams are better than others.

1.4.2 Choosing for others

We now move to the main section of the experiment, where the participants acted as social planners for others. This is a variant of a dictator game, where the dictator's monetary

	Your payment in two weeks	Your payment in four weeks	Your payment in six weeks
Option 1	$4	$8	$2
Option 2	$7	$2	$5

Figure 1.2. Discounting dominant choice 2

	Your payment in two weeks	Your payment in four weeks	Your payment in six weeks
Option 1	$3	$7	$3
Option 2	$5	$3	$5

Figure 1.3. Discounting dominant choice 2

payoff is not determined by the dictator's actions. Participants with no concerns for inequality are indifferent between the non-perturbation choices they face in this portion of the experiment, so they should attempt to complete this portion of the experiment as soon as possible. We find that behavior in this portion of the experiment is *not* affected by completion time. We also would like to examine behavior in unperturbed-perturbed pairs of choices to search for nondeliberate (random) choice. Figures 4 and 5 show an example of an unperturbed-perturbed pair of choices. Of our 118 participants, 22% make inequality averse choices more often in the perturbed choices than they do in the unperturbed variants. For example, these participants would pick option 2 in the choice depicted by Figure 4 and switch to option 1 in the choice from Figure 5. These participants make statistically similar choices to the other participants when choosing for themselves, but behave very differently when acting as social planners—where their choices are statistically indistinguishable from choosing uniformly at random. Our analysis below will include both the whole sample as well as the subsample of participants that excludes those identified as choosing uniformly at random. (henceforth referred to simply as 'subsample').

	Player 1's payment in two weeks	Player 1's payment in four weeks	Player 1's payment in six weeks		Player 2's payment in two weeks	Player 2's payment in four weeks	Player 2's payment in six weeks
Option 1	$8	$3	$1		$6	$4	$2
Option 2	$12	$3	$1		$2	$4	$2

Figure 1.4. Unperturbed choice

	Player 1's payment in two weeks	Player 1's payment in four weeks	Player 1's payment in six weeks		Player 2's payment in two weeks	Player 2's payment in four weeks	Player 2's payment in six weeks
Option 1	$8	$3	$1		$5.9	$4	$2
Option 2	$12	$3	$1		$2	$4	$2

Figure 1.5. Perturbed choice

We first focus on the set of decisions in which inequality notions agreed. This means that each choice had a period-wise inequality averse option, while some choices had a discounting dominant, cumulatively dominant, or vector dominant whole-stream inequality averse option (when a choice had an inequality averse option, this option was the same across each type of inequality aversion). Recall that our dominance notions are nested—when a stream vector dominates another, it also cumulatively dominates it. When a stream cumulatively dominates another, it also discounting dominates it. Since our participants seemed to understand this nesting for their own outcomes, the natural prediction is that we should see declining inequality aversion as we move from static to period-wise to discounting dominant to cumulatively dominant to vector dominant. As table 1.3 demonstrates, we do not observe this monotonicty in the data.

While the type of dominance ranking reasonably predicted how participants would choose for themselves, it does not adequately predict how they will choose for others. This is confirmed

Table 1.3. Aggregate proportion of inequality averse choices across decision environments. Stars represent significant differences in proportions between the general sample and subsample, * at the 90% level, ** at the 95% level, and *** at the 99% level.

dominance type	all sample	subsample
single-period	.7566 (.0198)	.8324 (.0199)***
vector dominance	.6985 (.0278)	.71 (.0321)
cumulative dominance	.6898 (.0190)	.7682 (.0201)***
discounting dominance	.7260 (.0237)	.8030 (.0245)***
only period-wise dominance	.6674 (.0217)	.7443 (.0233)***

Table 1.4. Multiple specifications comparing the proportion of inequality averse decisions made to proportion in single-period settings. FE are individual fixed effects, all SE are clustered at the individual level for logit specifications, * is significant at 90%, ** is significant at 95%, *** is significant at 99%

	LPM	LPM	LPM	LPM	Logit	Logit	Logit	Logit
Vector	-.0578 (.0363)	-.0442 (.0362)	-.1224 (.0347)***	-.1126 (.0352)***	-.2925 (.1837)	-.2745 (.2194)	-.7073 (.1948)***	-.7390 (.2239)***
Cumulative	-.0665 (.0259)**	-.0665 (.0259) **	-.0642 (.0264)**	-.0642 (.0264)**	-.3335 (.1347)**	-.4043 (.1628)**	-.4046 (.1748)**	-.4561 (.1947)
Discounting	-.0303 (.0308)	-.0304 (.0308)	-.0294 (.0305)	-.0294 (.0305)	-.1584 (.1610)	-.1909 (.1944)	-.1973(.2042)	-.2208 (.2295)
Period-wise	-.0890 (.0318)***	-.0890 (.0318)***	-.0881 (.0316)***	-.0881 (.0316)***	-.4364 (.1592)***	-.5309 (.1933)***	-.5341 (.1954)***	-.6051 (.2197)
Constant (static)	.7564 (.0283)	.7546 (.0187)	.8324 (.0252)	.8311 (.0175)	1.1328 (.1536)		1.6026 (.1802)	
Log Likelihood					1,301.9919	-875.8778	-848.9060	-587.8545
(pseudo) R^2	.0049	.0048	.0092	.0091	.0041	.0074	.0087	.0128
N	2,160	2,160	1,608	1,608	2,160	2,160	1,608	1,410
# clusters	118	118	88	88	118	118	88	77
Subsample only	No	No	Yes	Yes	No	No	Yes	Yes
Participant FE	No	Yes	No	Yes	No	Yes	No	Yes

by the results of table 1.4, which presents several specifications on both the full sample and subsample of participants who are choosing deliberately.

Result 2 *Participants select the inequality-averse option in the static settings and discounting dominant settings with the same frequency. They selected the inequality-averse option in vector dominant, cumulatively dominant, and period-wise settings with the same frequency as each other, but lower frequency than in static or discounting dominant settings.*

The previous analysis considered only cases where the lower inequality choice under discounting dominance was also the lower inequality choice under period-wise dominance. We also have choices for which one option was lower inequality under discounting dominance but higher under period-wise dominance, and choices where one option has lower inequality under discounting dominance but the options are not ranked according to period-wise dominance.

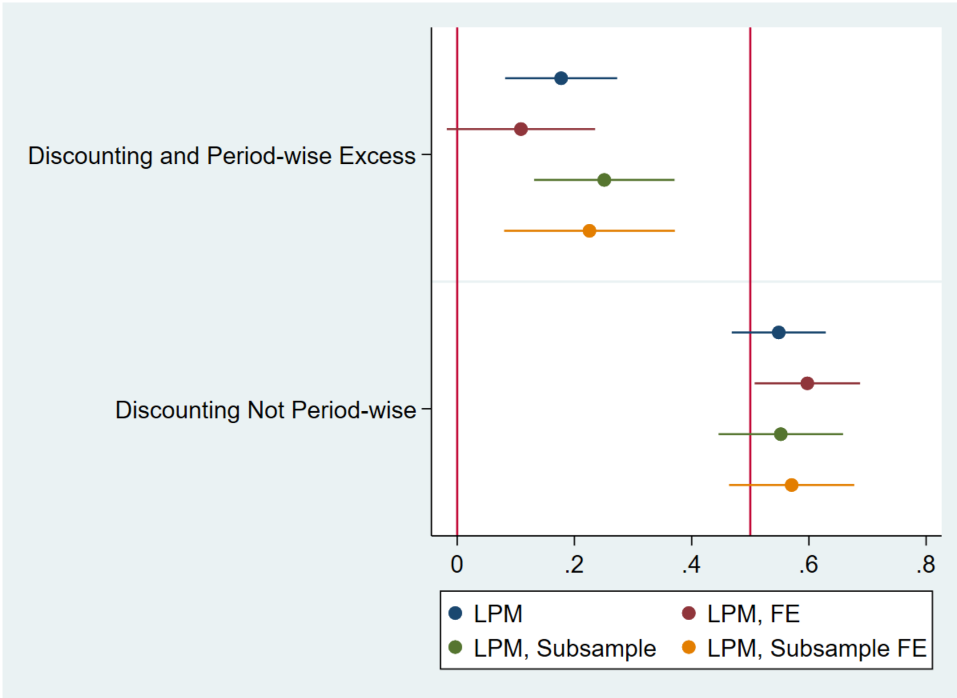


Figure 1.6. Regression results of the difference in likelihood of choosing the inequality averse option under discounting and period-wise dominance versus just discounting dominance.

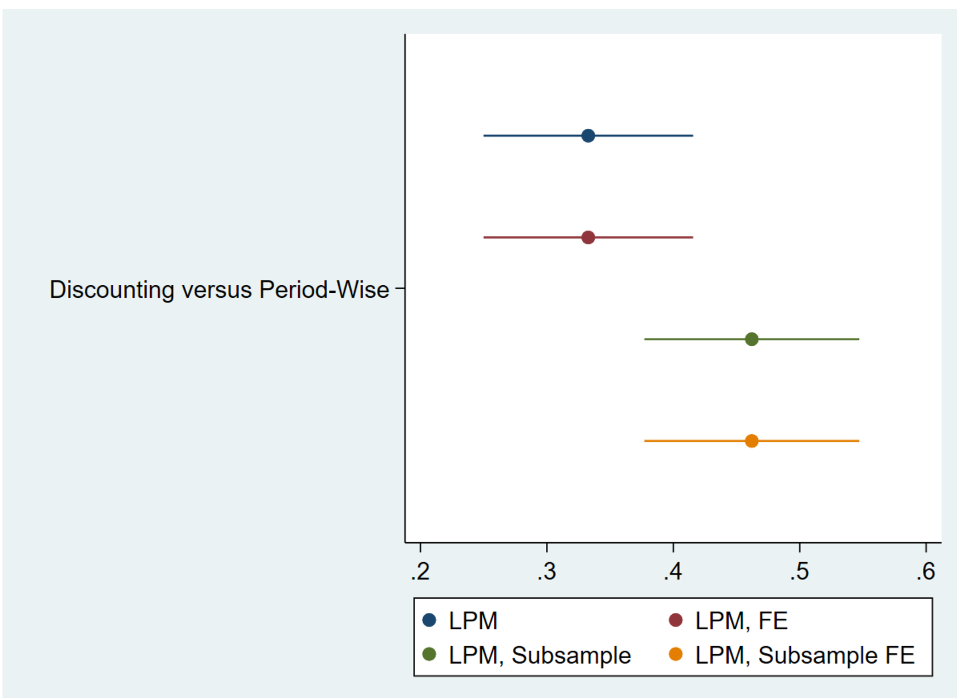


Figure 1.7. Regression results of the difference in likelihood of choosing the lower inequality option under discounting dominance when the other option has lower inequality under period-wise dominance.

Result 3 *Participants choose the discounting dominant option more when it is also period-wise dominant than when it is just discounting dominant. They choose the discounting dominant option less when the other option is period-wise dominant.*

This result is puzzling. Participants choose the inequality averse option under discounting dominance the most frequently out of any dynamic setting, but will not choose it when it is not also period-wise dominant. Further, when the other option has lower period-wise inequality, participants flock to that option.

1.5 Conclusion

The central theme of this paper revolves around the complex nature of measuring inequality, especially when viewed through the lens of time. Our primary contribution lies in extending the defining principle of inequality changes to an intertemporal environment. The main difficulty for the generalization lies in the way future payoffs could be treated: we propose three ways that differ by the access to different saving/borrowing technologies.

One of the insights from our study is the prevalence of period-wise inequality aversion over stream-wise inequality aversion. This highlights a tendency that the subjects think of inequality on a period-by-period basis, rather than by aggregation of all incoming payments. This emphasis on period-wise inequality aversion suggests a more nuanced understanding of intertemporal income inequality, which has substantial implications for policy design and public perception. Our findings reveal that people have opinions and biases about policies even when they're not directly affected. This observation is crucial for policymakers who need to understand and anticipate public reactions to their decisions. One of the potential explanations for our findings on discounted-dominating streams could be the lack of use of borrowing/lending by students, who constitute the typical subject pool in many economic experiments.

Furthermore, our findings present a challenge to the conventional application of social Exponential Discounted Utility (EDU) in intertemporal decision-making. The preference patterns observed in our experiments suggest that the standard EDU model might not accurately capture how individuals value future payments, particularly when it comes to inequality aversion. This discrepancy opens new avenues for resolving the 'impossibility result' noted by Zuber (2011) and Jackson and Yariv (2015). Moreover, our laboratory experiments have raised several intriguing questions, particularly concerning the distinct behavioral patterns among different groups. Understanding how these groups differ in their perception and valuation of intertemporal transfers remains an open question, requiring further investigation.

Chapter 1 is co-authored with Wolanski, Adrian and is currently being prepared for submission for publication. The dissertation author is one of primary authors of this chapter.

Chapter 2

Consistent Social Choice

2.1 Introduction

Many individuals make choices that have consequences for others. This paper examines whether social decision-makers make similar choices for different groups of others. Knowing when a social decision-maker's preferences are related across groups can be critical when the decision-maker is chosen through some selection process. Consider the case of a politician running for a higher office. Besides touting their record while serving in lower office, candidates often highlight how they interact with their family, friends, and other members of their community in an effort to convey how they will make choices for all of their constituents. But can we draw inferences about the candidate's behavior on a larger social domain from their choices on a smaller social domain? Such inferences would be useful, but require the presence of a link between preferences on these different social domains.

In this paper, we provide an axiom that characterizes the internal consistency of a family of preference relations across domains of different agents followed by an experimental test of the relevance of this condition. This axiom, which we call *joint consistency*, requires preferences to remain the same when agents receiving choice-independent outcomes are added to or removed from the domain of agents the decision-maker chooses for. Combined with the usual axioms of completeness, transitivity, and continuity, joint consistency implies an additively separable utility representation with domain-invariant Bernoulli functions. This representation yields two additional domain changes for which preferences are invariant: 1) changing the level of a choice-independent outcome and 2) adding an agent with an identical choice-dependent outcome for each existing agent with a choice-dependent outcome.

To demonstrate these domain changes more concretely, consider the example in table 2.1. Dave is the mayor of a small town and is constructing a public park[1]. The facility can include either a dog park or a playground, in addition to the standard park features. In the first choice environment, Dave's decision only affects two other agents—Anne and Betty. In the second environment, a new constituent named Charles moves to town. Charles cares only about the general park and not about whether it has a dog park or playground, so his outcome is independent of which product Dave chooses to construct. In the third environment, the domain of agents is the same as in the second environment; the only difference is Charles experiences a different level of choice-independent outcome. Finally, the fourth environment replaces Charles with two new agents, Evelyn and Fiona. Evelyn experiences the same outcome as Anne while Fiona experiences the same outcome as Betty. In order for Dave's preferences to be jointly consistent, he must choose to build the same project in each of these four choice environments[2].

Joint consistency and its implications are similar to some existing properties, namely the sure-thing principle in Savage (1954) and coordinate independence in Wakker (1988). In the example from table 2.1, Dave obeys the sure-thing principle if he builds the same project in choice 1 and choice 2, and obeys coordinate independence if he builds the same project in choice 2 and choice 3. Joint consistency is stronger than these existing properties because it requires that neither the level *nor the presence* of an agent's choice-independent outcome affects the planner's decision. In the example from table 2.1, this results in Dave choosing the same project in each choice environment.

[1]Construction of public goods is a scenario where joint consistency is more likely to arise naturally. Granting new agents access to a public good has a small impact on the experience of previous users, so the choice of which public good to construct should not depend on the presence of people who are indifferent between the different projects.

[2]The planner making the same choice in environment 4 requires an additional assumption of anonymity; that agents who receive the same outcome are viewed the same. This property naturally arises in laboratory settings.

Table 2.1. Four different choices faced by Dave, our social planner. The first two choices differ only by the presence of Charles, whose outcome is independent of the choice Dave makes. The second and third choices differ only by the level of the choice-independent outcome Charles receives. The first and fourth choices differ by the presence of Anne′ and Betty′, who each receive the same outcome as Anne and Betty respectively. We study when Charles selects the same option in each of these choices.

Choice 1	Outcome for Anne	Outcome for Betty		
Dog park	x	w		
Playground	y	z		
Choice 2	Outcome for Anne	Outcome for Betty	Outcome for Charles	
Dog park	x	w	c	
Playground	y	z	c	
Choice 3	Outcome for Anne	Outcome for Betty	Outcome for Charles	
Dog park	x	w	c'	
Playground	y	z	c'	
Choice 4	Outcome for Anne	Outcome for Betty	Outcome for Evelyn	Outcome for Fiona
Dog park	x	w	x	w
Playground	y	z	y	z

We investigate the empirical relevance of our criterion using an online laboratory experiment with 200 participants. Each participant acts as a social planner making binary decisions for groups of participants, across the various domain types described above[3]. Overall rates of joint consistency are high, including among choices where the planner's outcome is unaffected by their decision. Interpreting joint consistency as a measure of attention, this result indicates that participants often still behave purposefully when their choices only affect other people. Additionally, we find limited differences in consistency across the types of domain changes. Despite this success, we find a few substantive departures from the theory.

First, there is more consistency between choices that differ by exactly one of the three domain changes described in table 2.1 than between choices differing by compounded domain changes. This suggests inconsistency stems from a cognitive difficulty in parsing the domain changes. Second, we find adding or removing the *planner* results in lower consistency than adding or removing another agent (i.e. there is less consistency among choice 1 and choice 2

[3]The example in table 2.1 involves the decision-maker only choosing for others. We consider both this type of environment, as well as environments where the decision-maker is one of the agents affected by their decision.

when Charles is replaced with Dave). This demonstrates that social planners make systematically different choices depending on their presence in the domain of agents. Additionally, we analyze the direction of inconsistencies and find that: 1) inconsistent planners are less inequality averse than consistent planners, especially over more complex domain changes, and 2) inconsistent planners exhibit diminishing inequality aversion as domain size increases. We propose two possible causes for this result: 1) inconsistent preferences over efficiency and inequality of outcomes and 2) cognitive complexity[4]. Our current design limits our ability to distinguish these causes, but we believe this is a productive area for future research.

Many works explore the consistency of decision rules from a theoretical perspective. Polman and Wu (2020) provides a comprehensive overview of the literature on decisions under uncertainty made for oneself and others, and Thomson (2011) offers a thorough overview of which allocation rules adhere to different notions of consistency. While there have many theoretical treatments of preference consistency, there has been very limited empirical work on the topic. Zame et al. (2020) is the only empirical work we are aware of, which focuses on the differences between social and personal preferences in the face of uncertainty. We eliminate uncertainty to focus on the social choice component of these decisions, in addition to investigating choices on different sizes of social domains rather than comparing choices in personal and social domains.

While there is limited empirical work on consistency, there is a vast set of empirical work on social preferences[5]. One difference between our study and much of this literature is that we consider both environments, one where the decision-maker's choices affect their own outcome and one in which they do not. Two important papers in this area are Traub et al. (2009) and Hong et al. (2015), which both find systematic differences in how planners choose between equity and

[4]The literature studying when choices may differ from underlying preferences due to mistakes contains recent work by Esponda and Vespa (2023), Nielsen and Rehbeck (2022), Oprea (2020), Oprea (2023), and Enke et al. (2023).

[5]Harsanyi (1961) pioneered this literature introducing the ultimatum game, which Kahneman et al. (1986), Forsythe et al. (1994), and Andreoni and Miller (2002) refined into the modified dictator game. See Engel (2011) for a comprehensive review of dictator game literature.

efficiency when their choices affect their own outcome and when they do not. We find this same result in our data, which manifests as much higher consistency within domains that keep the planner's presence constant rather than across those domains.

The rest of the paper is structured as follows. Section 2 provides the theoretical components of this paper. Section 3 provides details on the experimental design. Section 4 shares our main findings from the experiment, and section 5 concludes the paper.

2.2 Theory

Let $\mathscr{N} = \{1, ..., n\}$ be the set of all agents a social planner makes decisions for. Then $A \in \mathscr{P}(\mathscr{N})$ and $B \in \mathscr{P}(\mathscr{N})$ are two domains that contain some of the agents from $\mathscr{N}$, where $\mathscr{P}(\mathscr{N})$ refers to the power set of $\mathscr{N}$. For each domain $A \in \mathscr{P}(\mathscr{N})$, $x_A \in \mathbb{R}^{|A|}$ refers to a social allocation in A. The social planner has a preference relation $\succeq_A$ over such social allocations that satisfies the following standard axioms:

2.2.1 Axioms and Definitions

Axiom 1. *A preference relation $\succeq_A$ is complete if for all $x_A, y_A \in \mathbb{R}^{|A|}$, either $x_A \succeq_A y_A$ or $y_A \succeq_A x_A$.*

Axiom 2. *A preference relation $\succeq_A$ is transitive if for all $x_A, y_A, z_A \in \mathbb{R}^{|A|}$, if $x_A \succeq_A y_A$ and $y_A \succeq_A z_A$, then $x_A \succeq_A z_A$.*

Axiom 3. *A preference relation $\succeq_A$ is continuous if both $E = \{x_A | x_A \succeq_A y_A\}$ and $F = \{x_A | y_A \succeq_A x_A\}$ are closed for all $y_A \in \mathbb{R}^{|A|}$.*

Given that Euclidean space is complete and separable, these axioms imply the existence of a real-valued utility function that represents $\succeq_A$ (Debreu (1954)).

For $A \in \mathscr{P}(\mathscr{N})$ and $B \in \mathscr{P}(\mathscr{N})$, s.t. $A \subset B$, by $(x_A; c_{B \setminus A})$ we denote a collection of payments $x \in \mathbb{R}^{|A|}$ to ordered agents in A, and a collection of payments $c \in \mathbb{R}^{|B|-|A|}$ to ordered agents that are in B, but not in A.

When it is clear from the context which domain we are referring to, we will drop the subscripts and write x, instead of x_A. Additional notation: for $x \in \mathbb{R}^n$ and $c \in \mathbb{R}$, we would denote by $(x_{-i}c_i)$ an n-dimensional vector x, whose i-th component was replaced by c.

An additional aspect of social preferences is how the decision-maker treats different agents receiving the same payment profiles. We define the following pair of anonymity conditions, which require (most) agents receiving the same payoff to be treated equally. In the second anonymity condition, we acknowledge the possibility that the decision-maker treats themselves differently from others but requires that they treat the rest equally when they receive the same payoff[6]. These axioms are not necessary for the main theoretical result, but they naturally arise in our experimental setting and provide a direct method for extending the domain of preferences.

Axiom 4. *A preference relation $\succeq$ that has an additively separable utility representation $U(x) = \sum_i u_i(x_i)$ satisfies anonymity among everyone if $x_i = x_j \implies u_i(x_i) = u_j(x_j)$.*

Axiom 5. *A preference relation $\succeq$ that has an additively separable utility representation $U(x) = u_{dm}(x_{dm}) + \sum_{i \neq dm} u_i(x_i)$ satisfies anonymity among others if $x_i = x_j, i, j \neq dm \implies u_i(x_i) = u_j(x_j)$.*

The final piece of background that we require before introducing our consistency condition is *coordinate independence*, which states that an alternative is preferred to another if and only if changing the level of one choice-independent component by the same amount in both alternatives does not change the preference ranking. Wakker (1988) shows that a preference relation admits an additively separable utility function if and only if it satisfies coordinate independence, axioms 1-3, and some mild regularity conditions. We will use this property to construct the utility representation for our consistency property.

Definition 7. A preference relation $\succeq$ satisfies coordinate independence, if for all $x, y \in \mathbb{R}^n$, $1 \leq i \leq n$, $c, d \in \mathbb{R}$, we have $(x_{-i}c_i) \succeq (y_{-i}c_i) \iff (x_{-i}d_i) \succeq (y_{-i}d_i)$.

[6] By construction, both anonymity axioms are equivalent whenever the decision-maker is not receiving a direct payment.

We now present the main definition of consistency which we develop in this paper. Intuitively, this condition says that if an alternative is preferred on a smaller domain, then it should still be preferred when we expand the domain by adding additional agents whose outcome is independent of the decision-maker's choice.

Definition 8. For $A \in \mathscr{P}(\mathscr{N})$ and $B \in \mathscr{P}(\mathscr{N})$, s.t. $A \subset B$, a pair of preference relations $\succeq_A$ and $\succeq_B$ satisfies *pair-wise consistency* if $\forall c \in \mathbb{R}^{|B|-|A|}; \forall x, y \in \mathbb{R}^{|A|}$ we have $x \succeq_A y \iff (x_A; c_{B\setminus A}) \succeq_B (y_A; c_{B\setminus A})$.

Definition 9. A family of preference relations $\{\succeq_A\}_{A \in \mathscr{P}(\mathscr{N})}$, satisfies *joint consistency* if $\forall A, B \in \mathscr{P}(\mathscr{N}) : A \subset B;\ \ \succeq_A$ and $\succeq_B$ satisfy pairwise consistency.

While similar to coordinate independence, joint consistency refers to how different preference rankings relate to each other when adding an agent with a choice-independent outcome. Coordinate independence only deals with how a single preference ranking behaves when an agent's choice-independent outcome changes, not with the relationship between preferences over different sets of agents.

2.2.2 Examples of Jointly Consistent Preferences

One example of jointly consistent preferences is a variation of Rawlsian preferences, which we call *Rawlsian lexicographic preferences* and denote by $\succeq_{RL}$. Standard Rawlsian preferences aim to make the lowest agent's payoff as high as possible, and are mathematically given by $x \succeq_R y \iff \min_j x_j \geq \min_j y_j$. These preferences are not jointly consistent; if $\min_j x_j \neq min_j y_j$, then either $x_j \succ y_j$ or $y_j \succ x_j$, but if $c < \min\{\min_j x_j, \min_j y_j\}$ then $(x_j, c) \sim_R (y_j, c)$. In other words, a Rawlsian planner is indifferent between all options when we add an agent with a choice-independent payoff lower than the previous minimum payoff. A planner with Rawlsian Lexicographic preferences aims to make the *lowest choice-dependent payment* as high as possible. Mathematically, these preferences are described by: $x \succeq_{RL} y \iff x_{(m)} > y_{(m)}$ where $m = \min k$ such that $x_{(k)} \neq y_{(k)}$, and $x_{(k)}$ is the kth order statistic of the allocation x. Like standard lexicographic preferences, Rawlsian lexicographic preferences are not continuous and do not admit a utility representation. Our next example, however, admits a familiar utility representation.

A second example of jointly consistent preferences is *consistent utilitarian preferences*, which we denote by $\succeq_{CU}$. A planner with such preferences maximizes a weighted sum of agent-domain-specific individual utilities where the marginal rate of substitution between any two agents is the same on any domain containing both agents. Mathematically, $x \succeq_{CU} y \iff \sum_{i\in A} u_{A,i}(x_i) \geq \sum_{i\in A} u_{A,i}(y_i)$ and $\frac{u'_{B,j}}{u'_{B,i}} = \frac{u'_{A,j}}{u'_{A,i}}$ whenever $i, j \in A \cap B$. The restriction that the marginal rate of substitution between agents is the same across domains prevents arbitrary changes in the Bernoulli function between domains, which is critical for joint consistency. Equivalently, we could instead require that the Bernoulli functions be unique up to a *jointly cardinal transformation* (the multiplication of every utility index by the same positive constant). We provide this alternative characterization because we can then suppress the choice domain in the notion: $x \succeq_{CU} y \iff \sum_{i\in A} u_{A,i}(x_i) \geq \sum_{i\in A} u_{A,i}(y_i) \forall A \in \mathscr{P}(\mathscr{N})$ where the u_i are unique up to jointly cardinal transformation. It is reassuring that there are utilitarian preferences that are jointly consistent, given their importance in many economics models and empirical applications.

2.2.3 Theoretical Results

Under some mild regularity conditions, joint consistency has remarkable implications for the utility formulation of families of social preferences:

Proposition 7. *$\{\succeq_A\}_{A\in\mathscr{P}(\mathscr{N})}$ is jointly consistent and each $\succeq_A$ satisfies completeness, transitivity, and continuity, if and only if $\{\succeq_A\}_{A\in\mathscr{P}(\mathscr{N})}$ has a consistent additively separable representation: $U_A(x) = \sum_{i\in A} u_{A,i}(x_i)$ for all $A \in \mathscr{P}(\mathscr{N})$, where u_i are unique up to a jointly cardinal transformation.*

Proof. The 'if' direction is straightforward. To show the 'only if' direction, we first show that joint consistency implies coordinate independence for all $A \in \mathscr{P}(\mathscr{N})$.

First, choose any $A, B \in \mathscr{P}(\mathscr{N})$, in such a way that $A \subset B$, $|B| = |A| + 1$. Choose arbitrary vectors $(x_A; c_{B \setminus A}), (y_A; c_{B \setminus A}) \in \mathbb{R}^{|B|}$, and assume without loss of generality that $(x_A; c_{B \setminus A}) \succeq_B (y_A; c_{B \setminus A})$.

Consider the $|A|-$dimensional vectors x_A and y_A. Since $\succeq_A$ is complete, either $x_A \succeq_A y_A$ or $y_A \succeq_A x_A$. If $y_A \succeq_A x_A$, then by pairwise consistency it must be the case that $(y_A; c_{B \setminus A}) \succeq_B (x_A; c_{B \setminus A})$; $\forall c \in \mathbb{R}$. This is only possible when both options are indifferent. so we must have $(y_A; c_{B \setminus A}) \succeq_B (x_A; c_{B \setminus A})$. Alternatively, $x_A \succeq_A y_A$, which by pairwise consistency implies that $(x_A; c_{B \setminus A}) \succeq_B (y_A; c_{B \setminus A})$; $\forall c \in \mathbb{R}$.

We have demonstrated that $(x_A; c_{B \setminus A}) \succeq_B (y_A; c_{B \setminus A})$ implies $(x_A; d_{B \setminus A}) \succeq_B (y_A; d_{B \setminus A})$, which proves that joint consistency implies coordinate independence of each $\succeq_A$. Using a theorem from Wakker (1988), each preference relation $\succeq_A$ then has an additively separable utility representation that is unique up to a jointly cardinal transformation. Now, consider an additively separable representation for $\succeq_{\mathscr{N}}$, $U(x) = \sum_{i \in \mathscr{N}} u_{A,i}(x_i)$. Because $\succeq_A$ is a jointly consistent family, for any $A \in \mathscr{P}(\mathscr{N})$ and any $i, j \in A$ we must then have $\frac{u'_{\mathscr{N},j}}{u'_{\mathscr{N},i}} = \frac{u'_{A,j}}{u'_{A,i}}$. □

An alternative framing is that consistent utilitarian preferences are the unique preference that satisfying the decomposition in the following propositions and corollary, which demonstrate explicitly how to relate utility representations across domains.

Proposition 8. *If $\{\succeq_A\}_{A \in \mathscr{P}(\mathscr{N})}$ is jointly consistent and each $\succeq_A$ satisfies completeness, transitivity, and continuity; $\forall B \in \mathscr{P}(\mathscr{N})$ such that $B = i \cup k$, there exists U_i, U_j, U_B, that respectively represent $\succeq_i, \succeq_j$, and $\succeq_B$, such that $U_B = U_i + U_j$.*

Proof. By the previous theorem, there exist u_i and u_j that represent $\succeq_i$ and $\succeq_j$. Additionally, there exists some function U_B that represents $\succeq_B$ and is additively separable. Therefore, $U_B = f(u_i) + g(u_j)$ for some increasing functions f and g.

Define $U_i = f(u_i)$ and $U_j = g(u_j)$; therefore $U_B = U_i + U_j$ where U_i, U_j, and U_B respectively represent $\succeq_i, \succeq_j$, and $\succeq_B$ □

Proposition 9. *If $\{\succeq_A\}_{A\in\mathscr{P}(\mathscr{N})}$ is jointly consistent and each $\succeq_A$ satisfies completeness, transitivity, and continuity; $\forall A \subset B \in \mathscr{P}(\mathscr{N})$ such that $B = A \cup \{k\}$, there exists U_A, U_B, U_k, that respectively represent $\succeq_A, \succeq_B$, and $\succeq_k$, such that $U_B = U_A + U_k$.*

Proof. Fix $i \in A$ and $k \in B \setminus A$, and define $C = i \cup k$. By proposition 1, $\succeq_C$ has an additively separable representation $U_C = u_i + u_k$ (note that u_k also represents $\succeq_k$ by proposition 2).

By proposition 1, there exists an additively separable function $U_A = u_i + \sum_{j\neq i\in A} u_j$ that represents $\succeq_A$ where u_i is the same in both U_A and U_C. Define $U_B = u_i + \sum_{j\neq i\in A} u_j + u_k = u_A + u_k$.

The above U_B is additively separable and is jointly consistent because $\frac{u'_{B,i}}{u'_{B,j}} = \frac{u'_{A,i}}{u'_{A,j}}$ for all $i, j \in A$ and $\frac{u'_{B,i}}{u'_{B,k}} = \frac{u'_{C,i}}{u'_{C,k}}$ (by construction we also have $\frac{u'_{B,j}}{u'_{B,k}} = \frac{u'_{D,j}}{u'_{D,k}}$ such that $j, k \in D$). U_B therefore also represents $\succeq_B$.

□

By repeated application proposition 3, we obtain the following corollary.

Corollary 1. *If $\{\succeq_A\}_{A\in\mathscr{P}(\mathscr{N})}$ is jointly consistent and each $\succeq_A$ satisfies completeness, transitivity, and continuity; $\forall A \subset B \in \mathscr{P}(\mathscr{N})$, there exists U_A, U_B, $U_{B\setminus A}$, that respectively represent $\succeq_A, \succeq_B$, and $\succeq_{B\setminus A}$, such that $U_B = U_A + U_{B\setminus A}$.*

In addition, if preferences satisfy either anonymity condition we can extend $\{\succeq_A\}$ to include additional agents not in $\mathscr{N}$ as long as the new agent's payoff profile is identical to an agent already in the domain.

Proposition 10. *Suppose that $\{\succeq_A\}_{A\in\mathscr{P}(\mathscr{N})}$ are jointly consistent, anonymous, continuous, transitive, and complete. Suppose that agent $k \notin \mathscr{N}$ but there exists some agent $j \in \mathscr{N}$ such that $x_k = x_j$ for all options x. Then the family of preferences$\{\succeq_B\}$ for $B \in \mathscr{P}(\mathscr{N} \cup k)$ described by $u_{B,\mathscr{N}\cup\{k\}} = u_{B,\mathscr{N}}$ for $B \in \mathscr{P}(\mathscr{N}) \cap \{k\}^c$ and $u_{B,\mathscr{N}\cup\{k\}} = u_{B,\mathscr{N}} + u_j$ for B in $\mathscr{P}(\mathscr{N} \cup k)$ where j satisfies $x_k = x_j$ for all x is a jointly consistent family of preferences over $\mathscr{P}(\mathscr{N} \cup k)$*

2.2.4 Commentary

There are two important implications of these results. First, they inform the experimental design we describe in the following section. Second, they demonstrate how important consistency (or the lack of it) is for empirical applications. Joint consistency implies external validity of preference measurements—that the measurement of a planner's preferences do not depend on the particular domain of agents included in the measurement. In particular, the failure of joint consistency has strong implications for the class of planners with additively separable utility representations. Joint consistency guarantees that measured Bernoulli functions do not vary (with the possible exception of a common scale factor) with the domain of measurement. Any measured Bernoulli functions for an inconsistent planner are specific only to the domain of measurement, and carry no information about the planner's preferences on other domains. This limits the value of such preference measurement for conducting welfare analysis, policy, analysis, etc.

2.3 Experimental Design

We recruited 200 participants using Prolific.com for an online laboratory experiment to test for joint consistency of preferences. Each participant made decisions for 31 binary choices over allocations to themselves and to other people, resulting in 31,400 pairs of decisions[7]. Each participant received \$2.50 for their participation; 10% received an additional bonus based on a decision selected at random. The median completion time was 10 minutes, and the average payment with the bonus was \$4, yielding an average hourly wage of[8] \$24.

Questions were presented in two blocks: one where the participant makes decisions that affect their payment and the payments of others, and a second where they only make decisions

[7]Due to a coding error, we are only able to use 26 of the 31 decisions for the first 60 of our participants.

[8]We find no evidence that the time a participant takes to complete the study affects joint consistency. On average, participants taking an additional minute to complete the study had a lower joint consistency rate of 0.15%. We fail to reject the null hypothesis that this effect is 0, with a p-value for the t-test of 0.528.

that affect the payments of others[9]. The order of the blocks and the order of questions within the blocks were randomized[10].

Table 2.2. Basic choice our participants made. There are three variants; the decision-maker could be player 1 (receiving the lower payment profile), player 2 (receiving the higher payment profile), or not a player and receive no payment from this decision.

Choice 1	Payment to Player 1	Payment to Player 2
option 1	$5	$25
option 2	$10	$15

2.3.1 The Basic Choices

There were three 'basic' choices our participants made, which involved only the choice of allocating $5 to person 1 and $25 to person 2, or allocating $10 to person 1 and $15 to person[11] 2. The difference between these three choices is the payment to the decision-maker. In one variant of the choice, they are not receiving either of the payments they are allocating. In the other two variants, they are receiving one of the payments they are allocating (either the poorer payment profile or the richer payment profile). Table 3.1 provides a description of these decision environments. While none of the decisions made in these basic choices are comparable to each other to test for joint consistency, every other decision our participants made is comparable to the decision in one of these three choices.

2.3.2 Additional Terminology

We refer to option 1 in all of the choices as 'inefficiency averse', and option 2 as 'inequality averse'[12]. We refer to the profile receiving $5 or $10 as the 'poorer' profile, and the

[9]The role of the other players was assigned to other different participants in the study at random each round, so at no point did any participant know exactly who was receiving the payments they were allocating to other people.

[10]The order of the blocks was not randomized for the first 60 participants. Because of this and the aforementioned coding error, we repeat all of our analysis dropping these individuals in section 4.5 and we find largely the same results.

[11]These numbers were selected to generate a tradeoff between efficiency and equality; we use multiples of $5 to make the choices clear for participants to understand and minimize rounding.

[12]We randomly presented the options to participants; we have standardized the order here for clarity of exposition.

Table 2.3. Comparison of choices that differ by the addition of adding agents with a choice-independent outcome. The decision-maker faces multiple variants where they are either player 1, player 2, player 3, or are not one of the players and receive no payment, and variants where the value of c varies between 5, 18, and 25.

Choice 1	Payment to Player 1	Payment to Player 2		
option 1	\$5	\$25		
option 2	\$10	\$15		
Choice 2	Payment to Player 1	Payment to Player 2	Payment to Player 3	
option 1	\$5	\$25	\$c	
option 2	\$10	\$15	\$c	
Choice 3	Payment to Player 1	Payment to Player 2	Payment to Player 3	Payment to Player 4
option 1	\$5	\$25	\$c	\$$\hat{c}$
option 2	\$10	\$15	\$c	\$$\hat{c}$

profile receiving either \$25 or \$15 as the 'richer' profile. When the decision-maker is making choices involving their own monetary payments, we refer to the option that provides them with the higher monetary payment as either the 'selfish' option or the individually payment-maximizing option (i.e. option 2 is selfish when the decision-maker is poorer, but option 1 is selfish when the decision-maker is richer)[13].

2.3.3 Simple Change 1: Addition of Players with Choice-Independent Outcomes

For each of the three basic decisions, we change the domain of choices by adding agents who receive a choice-independent outcome to each option. Under the assumption of joint consistency, a decision-maker chooses the same option in both choices under such a domain transformation. Table 2.3 illustrates this comparison explicitly; choice 1 is a basic decision, choice 2 adds an additional agent receiving a choice-independent outcome, and choice 3 adds two additional agents receiving choice-independent outcomes. A decision-maker who is jointly consistent chooses option 1 in one choice if and only if they choose option 1 in all choices.

[13]These attributions are not theoretical predictions but are useful for describing the direction of inconsistencies.

Table 2.4. Comparison of choices that differ only by duplication of agents with variable payments. The decision-maker is not one of the players receiving payment here, as that would violate anonymity.

Choice 1	Payment to Player 1	Payment to Player 2				
option 1	$5	$25				
option 2	$10	$15				
Choice 2	Payment to Player 1	Payment to Player 2	Payment to Player 3	Payment to Player 4		
option 1	$5	$25	$5	$25		
option 2	$10	$15	$10	$15		
Choice 3	Payment to Player 1	Payment to Player 2	Payment to Player 3	Payment to Player 4	Payment to Player 5	Payment to Player 6
option 1	$5	$25	$5	$25	$5	$25
option 2	$10	$15	$10	$15	$10	$15

2.3.4 Simple Change 2: Changing the level of a choice-independent outcome

Besides changing the set of agents in the choice domain, we can change the outcomes the agents receive. Table 2.3 again illustrates these changes. A decision-maker who is jointly consistent chooses option 1 in choice 2 for one level of c if and only if they choose option 1 in choice 2 for any level of c.

2.3.5 Simple Change 3: Duplications

In addition to changing the domain by adding agents receiving a choice-independent outcome, we test for joint consistency by duplicating all choice-dependent outcome profiles and assigning them to new agents. Table 2.4 compares a basic choice to a choice with one duplication and a to a choice with two duplications. A decision-maker who is jointly consistent and satisfies anonymity chooses option 1 in one choice if and only if they choose option 1 in all choices[14].

2.3.6 Combining the Domain Changes and Testable Predictions

Joint consistency of preferences implies that decisions are invariant under each of the domain expansions described above and under the composition of these changes. Table 2.5 demonstrates several environments formed by the combination of these domain changes. Joint

[14]Our experimental design enforces anonymity among others, so we can test for joint consistency using duplications whenever the decision-maker is not one of the agents receiving a choice-dependent outcome profile. We, therefore, design our experiment to have duplications only when the decision-maker is not one of the agents receiving a choice-dependent outcome profile.

Table 2.5. Comparison of choices that can differ by both the addition of agents with choice-independent payments or the duplication of the choice-dependent payment profiles. The decision-maker faces variants where they are either player 6 or not one of the players and receive no payment, and where *c* varies between 5, 18, and 25.

Choice 1	Payment to Player 1	Payment to Player 2				
option 1	$5	$25				
option 2	$10	$15				
Choice 2	Payment to Player 1	Payment to Player 2	Payment to Player 3			
option 1	$5	$25	$c			
option 2	$10	$15	$c			
Choice 3	Payment to Player 1	Payment to Player 2	Payment to Player 3	Payment to Player 4		
option 1	$5	$25	$5	$25		
option 2	$10	$15	$10	$15		
Choice 4	Payment to Player 1	Payment to Player 2	Payment to Player 3	Payment to Player 4	Payment to Player 5	
option 1	$5	$25	$5	$25	$c	
option 2	$10	$15	$10	$15	$c	
Choice 5	Payment to Player 1	Payment to Player 2	Payment to Player 3	Payment to Player 4	Payment to Player 5	Payment to Player 6
option 1	$5	$25	$5	$25	$c	$ĉ
option 2	$10	$15	$10	$15	$c	$ĉ

consistency implies that the decision-maker chooses option 1 in one choice in table 2.5 if and only if they choose option 1 in all choices in that table.

In our experiment, this means there are three categories of choices based on the individual outcome experienced by the decision-maker: the poorer payment profile, the richer payment profile, or a choice-independent profile (this includes when the decision-maker receives no payment). Each participant made 7 choices where they received the richer payment profile (generating 21 pairs of comparable decisions), 7 choices when they received the poorer payment profile, and 17 choices where their outcome was independent of their choice (generating 136 pairs of comparable decisions[15]). This results in 178 pairs of decisions per participant.

2.4 Results

2.4.1 Descriptive Statistics

The average proportion of consistent choices is 0.8141, and the median proportion of consistent pairs among participants is 0.8764. We find no difference in consistency from gender,

[15]The coding error for 60 of our participants resulted in only 12 choices from this block appearing correctly, so this set of participants only has 66 comparable pairs.

Table 2.6. Proportion of consistent decision pairs by different choice environments, with the more consistent option displayed on the left along with stars to indicate statistically significant differences in consistency. * is significant at 90%, ** is significant at 95%, *** is significant at 99%, **** is significant at 99.99%.

condition	proportion consistent
Adding 1 choice-independent outcome	0.8672 (0.0040)
Changing level of 1 choice independent outcome	0.8662 (0.0051)
Duplication	0.8284 (.0094)
combined change	0.7789 (.0031)

age, employment status, length of time spent on the experiment, or survey experience[16]. We include the sample-balance results in table B.1 in the appendix. We next examine the proportion of consistent choice pairs across different types of domain changes and present the results in table 2.6. There are two primary results.

Result 1 Adding an agent with a choice-independent outcome and changing the level of a choice-independent outcome have the same proportion of consistent choices

Our first result is consistent with the theory—that the different types of domain changes should be associated with the same rate of consistency. While it appears that combined changes and duplications have lower consistency rates, there are additional experimental parameters that may be driving these results.

2.4.2 Regression Results

We next build upon the results of table 2.6 using a linear probability model (LPM) and a logit model. We present the result in table 2.7. The regressions contain indicator variables for each domain change analyzed in 2.6, utilize both fixed effects and standard error clustering at the participant level, and include controls for addition parameters of the experimental design.

[16]Survey experience has a statistically significant but economically insignificant effect on consistency. A participant would need to complete an additional 100,000 surveys to have a 1.59% increase in their consistency percentage. The maximum number of surveys completed in our sample was 7,063.

Table 2.7. Multiple specifications comparing effect of domain changes on consistency. The baseline is changing only the level of a choice-independent outcome when the decision-maker is not paid. All regressions computed with standard errors clustered at the participant level, along with participant-level fixed effects and additional controls where indicated. * is significant at 90%, ** is significant at 95%, *** is significant at 99%.

	LPM	LPM	Logit	Logit	LPM	LPM	Logit	Logit
adding choice-independent outcome	.0009 (.0061)	.0006 (.0063)	.0087 (.0545)	.0053 (.0613)	-.0035 (.0059)	-.0036 (.0060)	-.0330 (.0529)	-.0383 (.0624)
duplication	-.0379 (.0133)***	-.0358 (.0133)***	-.2937 (.0966)***	-.3316 (.1163)***	.0065 (.0115)	.0087 (.0115)	.0357 (.0825)	.0651 (.0996)
combined change	-.0873 (.0120)***	-.0826 (.0114)***	-.6086 (.0774)***	-.6990 (.0871)***	-.0181 (.0071)**	-.0157 (.0070)**	-.1316 (.0531)**	-.1376 (.0632)**
DM paid on one only					-.0454 (.0128)***	-.0494 (.0129)***	-.2613 (.0783)***	-.3691 (.0966)***
DM paid on both					.0045 (.0199)	-.0038 (.0189)	.0290 (.1301)	-.0373 (.1517)
DM richer					.0611 (.0193)***	.0623 (.0192)***	.4801 (.1549)***	.5817 (.1799)***
DM poorer					.1020 (.0171)***	.1033 (.0170)***	.9410 (.1719)***	1.0978 (.1906)***
Regression constant	.8663 (.0106)	.8635 (.0075)	1.8681 (.0913)		.8185 (.0171)	.8219 (.0112)	1.5171 (.1146)	
Log Likelihood			-14,981.969	-11,071.051			-14,692.688	-10,843.016
(pseudo) R^2	.0116	.0116	.0125	.0178	.0226	.0226	.0257	.0380
N	31,400	31,400	31,400	23,010	31,400	31,400	31,400	23,010
# Clusters	200	200	200	145	200	200	200	145
Participant FE	No	Yes	No	Yes	No	Yes	No	Yes

Result 2 There is less consistency among choices involving a combination of changes, but consistency among choices involving duplication is the same as the other simple changes when controlling for additional experimental parameters.

Choices involving duplications were more likely to have the decision-maker receiving payment on one choice but on the other, which has a significant negative effect on consistency. Additionally, choices where the decision-maker has a choice-dependent outcome only occur when duplications are not present. When controlling for these factors, duplications have the same effect on consistency as the other two basic types of domain changes. However, combining changes still has a significant negative effect on consistency. Additionally, the effects of the controls provide insight on the effect of an additional change.

Result 3 There are no differences in consistency when the decision-maker receives a choice-independent payment on both choices when that payment is positive or when it is 0. However, there is a significant decrease in consistency when comparing choices where the decision-maker receives a choice-independent positive outcome in one choice and 0 in another choice.

Our participants view adding themselves to a domain differently than adding another agent to the domain. This is inconsistent with the theory; joint consistency requires that adding an additional agent with a choice-independent payoff does not change preferences regardless of the identity of the added agent. While this indicates that participants make systematically different choices depending on their inclusion in the domain, they make consistent choices within the regime.

2.4.3 Directions of Inconsistencies

One conclusion we drew from the results in table 2.6 was that our participants made systematically different choices in different environments. To further support this conclusion, we compare the likelihood of selecting the inequality-averse choices on different environments between participants who are fully consistent and those who are not. We present the results of this analysis in table 2.8, which compares the proportions between these different groups for a variety of decision environments.

Result 4 Individuals who are not fully consistent display significantly lower levels of inequality aversion than individuals who are fully consistent.

Inconsistent individuals overwhelmingly exhibit lower inequality aversion than consistent individuals. Additionally, inconsistent individuals appear to choose the inequality-averse option less frequently as domain size increases.

2.4.4 Direction of Inconsistencies as Size Changes

To further investigate changes in inequality aversion with domain size, we report the proportion of choices where the participants selected the inequality-averse choice on the smaller domain but not on the larger domain. We perform this analysis for each choice category investigated above, and report the results in table 2.9.

Result 5 When the domain changes through duplication or combined changes, inconsistencies are associated with decreasing inequality aversion. When the domain changes through adding choice-independent outcomes, there is no trend in the direction of inconsistencies.

Inconsistencies are systematic, which provides some insight into the mechanisms behind

Table 2.8. The proportion of decision pairs where participants selected the inequality-averse choice, broken down by different choice environments. The left column is the proportion among individuals who are fully consistent in the sample, the middle and right columns are the proportion among individuals who are not fully consistent. The left and middle columns are the proportion where individuals selected the inequality-averse option on the choice with a smaller domain (regardless of their choice on the larger domain), and the left and right columns are the proportion where individuals selected the inequality-averse option on the choice with a larger domain (regardless of their choice on the smaller domain). Stars indicate significant differences against the left column. * is significant at 90%, ** is significant at 95%, *** is significant at 99%, **** is significant at 99.99%.

Environment	Consistent	Inconsistent, Smaller	Inconsistent, Larger
Aggregate	.7383 (.0056)	.6009 (.0037)****	.5692 (.0038)****
Adding choice-independent outcome	.6648 (.0106)	.5735 (.0068)****	.5706 (.0069)****
Duplication	.7655 (.0203)	.6186 (.0141)****	.5781 (.0143)****
combined change	.7733 (.0067)	.6124(.0047)****	.5675 (.0048)****
DM choice-dependent outcome, aggregate	.6182 (.0120)	.5662 (.0075)***	.5501 (.0075)****
DM choice-dependent outcome, adding choice-independent outcome	.6182 (.0134)	.5629 (.0084)***	.5506 (.0084)****
DM choice-dependent outcome, combined change	.6182 (.0267)	.5793 (.0167)***	.5483 (.0169)**
DM choice-dependent outcome, richer, aggregate	.3272 (.0163)	.2497 (.0093)****	.2331 (.0091)****
DM choice-dependent outcome, richer, adding choice-independent outcome	.3273 (.0183)	.2483 (.0104)****	.2339 (.0101)****
DM choice-dependent outcome, richer, combined change	.3273 (.0365)	.2552 (.0209)*	.2299 (.0202)**
DM choice-dependent outcome, poorer, aggregate	.9091 (.0100)	.8828 (.0067)**	.8671 (.0073)***
DM choice-dependent outcome, poorer, adding choice-independent outcome	.9091 (.0112)	.8776 (.0079)**	.8672 (.0081)***
DM choice-dependent outcome, poorer, combined change	.9090 (.0223)	.9034 (.0142)	.8667 (.0163)***
DM choice-independent outcome, aggregate	.7816 (.0061)	.6127 (.0043)****	.5757 (.0044)****
DM choice-independent outcome, adding choice-independent outcome	.7588 (.0167)	.5943 (.0117)****	.6101 (.0116)****
DM choice-independent outcome, duplication	.7655 (.0203)	.6186 (.0141)****	.5781 (.0143)****
DM choice-independent outcome, combined change	.7880 (.0069)	.6153 (.0049)****	.5692 (.0050)****
DM paid on one domain, aggregate	.7838 (.0084)	.6127 (.0060)****	.5759 (.0061)****
DM paid on one domain, adding choice-independent outcome	.7702 (.0274)	.5985 (.0193)****	.6047 (.0193)****
DM paid on one domain, combined change	.7853 (.0089)	.6143 (.0063)****	.5727 (.0064)****
DM paid on both domains, aggregate	.6706 (.0089)	.5837 (.0057)****	.5640 (.0058)
DM paid on both domains, adding choice-independent outcome	.6363 (.0123)	.5658 (.0078)****	.5554 (.0078)****
DM paid on both domains, duplication	.7455 (.0240)	.6230 (.0164)****	.5724 (.0168)****
DM paid on both domains, combined change	.7005 (.0150)	.5991 (.0099)****	.5752 (.0010)****
DM paid on neither domain, aggregate	.8158 (.0118)	.6167 (.0086)***	.5675 (.0088)****
DM paid on neither domain, adding choice-independent outcome	.76 (.0302)	.6019 (.0211)***	.6444 (.0206)***
DM paid on neither domain, duplication	.8286 (.0368)	.6063 (.0275)****	.5937 (.0277)****
DM paid on neither domain, combined change	.8286 (.0133)	.6216 (.0101)****	.5459 (.0104)****

Table 2.9. Proportion Inequality ↓ is the proportion of choice pairs where we see decreasing inequality aversion as size increases; that is, where the decision-maker selected the inequality-averse option on the smaller domain but not on the larger domain. ** is significant at 95%, *** is significant at 99% (significance is in terms of difference from 50%, which would indicate no trending direction).

Environment	# Observations	Proportion Inequality ↓
Aggregate	4,450	.5611 (.0074)****
Adding choice-independent outcome	959	.5078 (.0161)
Duplication	278	.5863 (.0295)***
combined change	3213	.5749 (.0087)****
DM choice-dependent outcome, aggregate	604	.5779 (.0202)***
DM choice-dependent outcome, adding choice-independent outcome	473	.5454 (.0229)**
DM choice-dependent outcome, combined change	131	.6031 (.0427)**
DM choice-dependent outcome, richer, aggregate	362	.5497 (.0261)*
DM choice-dependent outcome, richer, adding choice-independent outcome	285	.5439 (.0295)
DM choice-dependent outcome, richer, combined change	77	.5715 (.0563)
DM choice-dependent outcome, poorer, aggregate	242	.5702 (.0318)**
DM choice-dependent outcome, poorer, adding choice-independent outcome	188	.5479 (.0363)
DM choice-dependent outcome, poorer, combined change	54	.6481 (.0650)**
DM choice-independent outcome, aggregate	3846	.5616 (.0080)****
DM choice-independent outcome, adding choice-independent outcome	486	.4712 (.0226)
DM choice-independent outcome, duplication	278	.5863 (.0295)***
DM choice-independent outcome, combined change	3082	.5736 (.0089)****
DM paid on one domain, aggregate	2235	.5545 (.0105)****
DM paid on one domain, adding choice-independent outcome	220	.4909 (.0337)
DM paid on one domain, combined change	2015	.5613 (.0111)****
DM paid on both domains, aggregate	769	.5488 (.0179)****
DM paid on both domains, adding choice-independent outcome	125	.4960 (.0447)
DM paid on both domains, duplication	224	.5982 (.0328)***
DM paid on both domains, combined change	420	.5381 (.0233)
DM paid on neither domain, aggregate	842	.5926 (.0169)****
DM paid on neither domain, adding choice-independent outcome	141	.4184 (.0415)*
DM paid on neither domain, duplication	54	.5370 (.0678)
DM paid on neither domain, combined change	647	.6352 (.0189)****
Duplication without choice-independent outcomes	54	.5370 (.0678)

them. One possible explanation is mechanical; participants may have inconsistent preferences that are decreasing in both inefficiency and inequality. We attempt to test this using choices from the category "duplication without choice-independent outcomes". Many standard inequality measures are *identical* under these types of choice comparisons[17]. The absolute efficiency loss (the difference between total payments received by all players in a choice from choosing the inequality-averse option), however, is higher in the inequality-averse option on the duplicated economy. A participant with decreasing preferences over efficiency loss and inequality measured by some index would therefore exhibit decreasing inequality aversion as the domain size increased in this choice category. While our observed probability is consistent with this explanation, our sample size is far too small to draw a firm conclusion about this hypothesis[18].

A second explanation centers on the complexity of the decision environments. Inconsistencies under duplication or combined changes appear to be systematic, while inconsistencies when adding choice-independent outcomes appear random. Duplication and combined changes are fundamentally more complex domain transformations than adding an agent with a choice-independent outcome, since the former entails adding and/or changing more than 2 numbers in each choice while the latter only adds 2 numbers to each choice. Additionally, it is far easier to determine which option pays a larger sum to all participants than to calculate an inequality index for each option. A decision-maker who experiences differential difficulty across these factors would be more likely to choose more efficient/higher inequality options as domain size increases.

Unfortunately, our current data and design are not well-equipped to completely disentangle these two explanations, nor the complexity costs of changing domain size versus composition. We believe, however, that this would be a productive area for future research.

[17] Examples of measures that are identical under duplication without adding choice-independent outcomes include but are not limited to: Gini index, Thiel index, Hoover index, Atkinson index, Palma ratio, variance, and average absolute difference. These measures depend only on the *distribution* of each payment profile within a choice; duplicating all payment profiles changes only the count of each payment profile without changing the distribution of profiles.

[18] Given the number of violations and proportion of each direction in this data set, we estimate we would need to recruit around 3,500 participants to effectively test this hypothesis using our current design.

Table 2.10. Prop Inequality Av is the proportion of choice pairs where the inequality averse choice is selected only after the indicated replacement is made. * is significant at 90%, ** is significant at 95%, *** is significant at 99% (significance is in terms of difference from 50%, which would indicate no trending direction).

	#Obs	Prop Inequality Av	Condition
total inconsistencies, same size	1,388		
Changing choice-independent level only	591		
Changing choice-independent only, replacing $25 with $18	173	.5607 (.0377)	more $18
Changing choice-independent only, replacing $25 with $5	175	.5371 (.0377)	more $5
Changing choice-independent only, replacing $18 with $5	243	.4733 (.0322)	more $5
combined change	797		
combined change, swapping DM without changing profiles	191	.4764 (.0361)	DM paid
combined change, swapping DM and $18 for $25	154	.6234 (.0390)***	more $18
combined change, swapping DM and$5 for $25	121	.6363 (.0437)***	more $5
combined change, swapping DM and $5 for $18	141	.5745 (.0416)*	more $5
combined change, swapping $18 for duplication	62	.5645 (.0629)	more dup
combined change, swapping $5 for duplication	45	.5556 (.0741)	more dup
combined change, swapping $18 and $25 for duplication	41	.5853 (.0769)	more dup
combined change, swapping $18 and $25 for $5	42	.5476 (.0769)	more $5
combined change, swapping DM aggregate	607	.4530 (.0202)**	DM paid
combined change, swapping DM without changing profiles	191	.4764 (.0361)	DM paid
combined change, swapping DM and $18 for $25	154	.3636 (.0387)***	DM paid
combined change, swapping DM and$5 for $25	121	.4711 (.0454)	DM paid
combined change, swapping DM and $5 for $18	141	.5035 (.0421)	DM paid

2.4.5 Direction of Inconsistencies When Size Does Not Change

There are two types of domain changes that do not change size—a simple change in the level of a choice-independent payoff, or a combined change that adds and removes different agents. There are four components of the domain that can change—the number of agents receiving a choice-independent payment of $5, the number of agents receiving a choice-independent payment of $18, the number of agents receiving a choice-independent payment of $25, and the number of duplications of the choice-dependent payment profiles. Table 2.10 depicts the direction of inconsistencies for all possible change categories.

Result 6 There is no trend in the direction of inconsistencies when changing only the level of a choice-independent outcome or when changing choice-independent outcomes to duplications. When the decision-maker receives a positive choice-independent outcome on one domain but a choice-independent 0 on the other, inequality aversion decreases as the level of choice-independent payments increases.

The first part of this finding agrees with our previous finding—that inconsistencies for simpler changes are less systematic than inconsistencies for complex changes. The second part of the result also agrees with our earlier finding that participants made systematically different choices depending on their inclusion in the domain. Ultimately, this result says that systematic inconsistencies remain when domain size does not change. If issues with cognitive complexity drive observed inconsistencies, this suggests complexity does not come solely from differences in the number of components the decision-maker has to evaluate.

2.4.6 Robustness Checks

Robustness of Including All Participants

We mentioned earlier that 60 of our participants had a slightly different experimental experience; they did not see the blocks of questions in a random order, and they did not see five of the choices displayed correctly. Since there may be concern about spillovers between the choices or other systematic differences between these participants and the others in our study, we repeat all of the analysis from section 4 dropping these participants from the sample. We present the results in the appendix. The results are qualitatively unchanged; while many of the numerical estimates of coefficients and proportions change slightly, none of the conclusions we drew above were due to this particular subset of participants.

Robustness Against Random Choice and Indifference

Joint consistency is a strong condition—choice data either satisfies or does not satisfy it, leaving no room to incorporate random choice or errors in decision-making. Since we are dealing with social choice data, it is also possible that some of our participants are choosing randomly because of indifference between options. The presence of any of these phenomena would bias our measures of inconsistency upwards, since any of them could cause inconsistent choices without inconsistent preferences. This section addresses these concerns.

Since joint consistency and anonymity imply an additively separable utility representation, one approach to dealing with the above ideas would be to make functional form assumptions on the Bernoulli functions and estimate a random utility model. Finding parameter values that reasonably fit the data would indicate participants follow a random variant of joint consistency rather than the deterministic version presented in section 2. However, we believe such a model is insufficiently parsimonious to yield a meaningful rejection of a random utility variant of joint consistency.

Instead, we propose a simple alternative random choice model. Suppose the decision-maker only chooses their preferred choice with probability $p \in [0.5, 1]$; with probability $1 - p$ the decision-maker instead selects their less preferred choice. When $p = 1$, the decision-maker's behavior is fully deterministic, and we can verify whether they are consistent or not. When $p = 0.5$, the decision-maker's choices are made uniformly at random and provide no information about their underlying preferences. Equivalently, a decision-maker with $p = 0.5$ can be thought of as someone who is indifferent between all options[19]. Because this interval is continuous, we can interpret a p estimate closer to 0.5 as providing less meaningful information about underlying preferences than a p estimate closer to 1. We use this model to analyze the 17 choices where participants' decisions have no impact on their own payment[20].

We can estimate what value of p is most likely to generate the distribution of consistency rates (percentage of consistent choices) observed in the data. In the appendix, we derive the following equation that relates p, the probability of random choice being in line with deterministic choice, to the expected consistency rate: $E[\text{consistency rate}] = 1 - 2p + 2p^2$. On the subset of choices where the participants are receiving a choice-independent payment (including no

[19]Our experimental interface randomizes the order of the questions and the order of the options presented. Therefore, a decision-maker who is indifferent between the options but is not actively choosing uniformly at random will appear to be choosing uniformly at random.

[20]We focus on these choices since we believe they are most likely to generate upward bias in inconsistency measures.

payment), our average consistency rate is 0.7897 with a standard deviation[21] of 0.0144. This consistency rate data and the above equation then provide an estimated choice probability value of $p = 0.8805$ with a standard deviation of 0.0189; the associated 95% confidence interval for p is then [0.8407.0.9183]. This estimate is substantially closer to 1 than to 0.5, indicating that there is indeed much we can learn about the consistency of underlying preferences despite potential randomness/errors of choice. Combined with trends in the direction of inconsistencies, this indicates that simple random choice is not responsible for most of our inconsistencies.

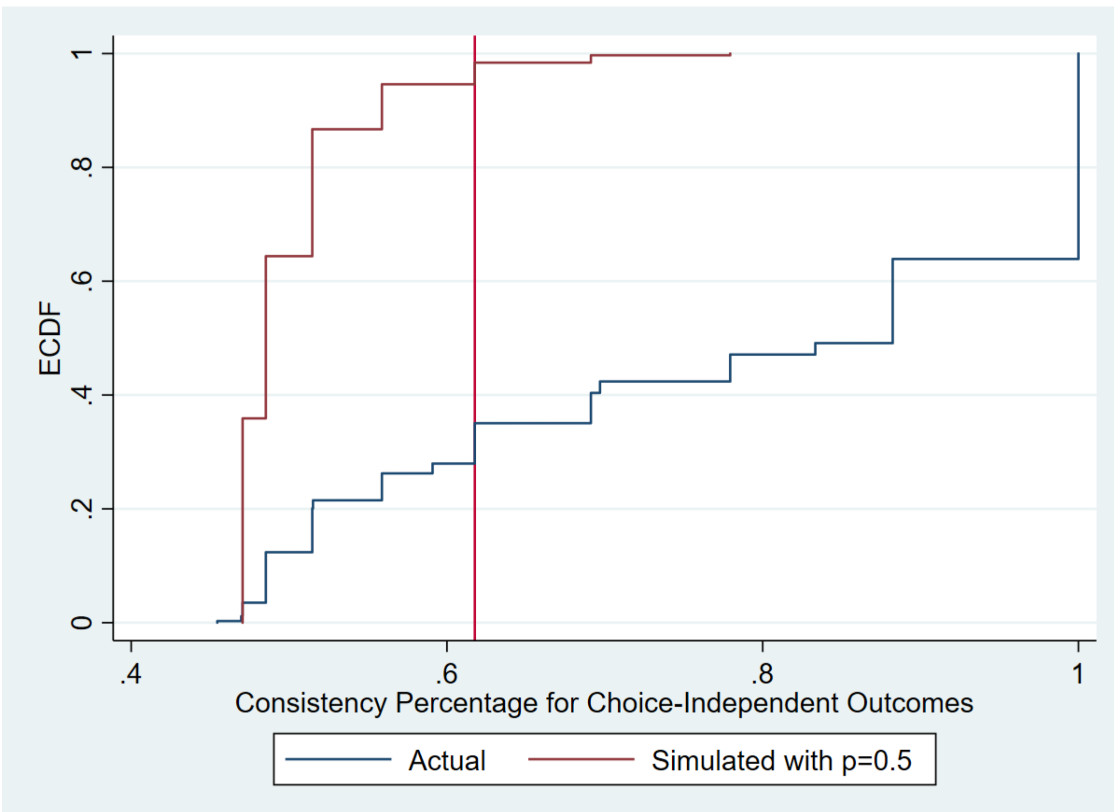


Figure 2.1. Differences in the distribution of consistency rate between our data and simulated data with $p = 0.5$. Only about 30% of our participants have a consistency rate that falls in the lower 98% of the simulated distribution.

In addition to estimating a sample-wide choice probability, we estimate individual choice probabilities to identify which participants might not be choosing purposefully. To do this, we simulate the distribution of consistency rates for $p = 0.5$ and identify which participants

[21] The standard deviation is estimated via bootstrap.

have consistency rates that fall close to the center of this distribution. A comparison of these distributions is presented in figure 2.1. Only about 30% of our population have consistency rates that lie in the lower 98% of the distribution generated if our population was choosing uniformly at random. The remaining participants have consistency rates unlikely to come from such choice patterns. We then repeat the analysis of section 4 including an additional control for agents we identify as not choosing purposefully, and present the results in section 7.2.3. Our primary results remain: compound changes are associated with lower consistency, inconsistent decision-makers are less inequality averse, and decision-makers become less inequality averse/more inefficiency averse as domain size increases. Our main conclusions are not driven by participants with low consistency rates, but rather by the participants who behave purposefully. Additionally, the fact that trends in the direction of inconsistencies remain after removing many noisy observations further indicates our conclusions are not driven by simple random choice. Instead, there are systematic reasons for observed inconsistencies which require further research to understand fully.

2.5 Conclusion

In this paper, we proposed a criterion that relates families of social preferences across domains of different sizes and compositions. Combined with standard regularity assumptions, this criterion produces an additively separable utilitarian representation with domain-invariant Bernoulli functions. A planner following the criterion has invariant preferences under three types of domain changes, in addition to any combination or composition of these changes.

We then examine the empirical relevance of this criterion using an online laboratory experiment. While rates of joint consistency were generally high, there were systematic differences across types of domain changes. Importantly, participants were more consistent between domains differing by a single simple domain change than a combination or composition of changes, suggesting that the difficulty of decisions may play a role in inconsistent behavior. We also observe higher rates of consistency when the decision-maker's presence in the domain is constant between choices, indicating participants make systematically different choices depending on their inclusion. In addition, we document that inconsistent individuals are less inequality-averse than consistent individuals and that inequality aversion decreases as domain size increases. These trends indicate that inconsistencies are systematic, and again may be due to choice complexity rather than an accurate reflection of underlying preferences.

One promising direction for future research is a more thorough investigation of the mechanisms driving our observations. We provide suggestive evidence that cognitive limitations play a substantial role in our observed inconsistencies, but this requires more experiments with different designs to confirm this hypothesis. Additionally, understanding precisely where cognitive limitations may enter the decision process may be critical for understanding choice more broadly. Recent work by Oprea (2023) and Enke et al. (2023) suggest that classical behavioral anomalies in many settings may come from similar cognitive limitations rather than a disparate set of non-standard preferences. If our results are also driven by cognitive limitations, disentangling how much is unique to our particular setting and how much could be generalized to other choice environments would further support this emerging literature.

Chapter 2 is co-authored with Wolanski, Adrian and is currently being prepared for submission for publication. The dissertation author is one of primary authors of this chapter.

Chapter 3

Extending Social Choice Consistency Experiments with Large Language Models

3.1 Introduction

Human decisions are inherently complex and context-dependent. Traditional laboratory economic experiments, which aim to study decision-making under controlled conditions, often face restrictions due to confidentiality concerns. These limitations prevent the inclusion of private information in the experiments, hindering the ability to capture some nuances of real-world decision-making scenarios.

Experimental Economics had three experimental settings up to this day - laboratory, field, and online. With the AI advancements, a new way of doing experimental research emerges – the use of Large Language Models (LLMs) either as a participant, or for generating different participants in experiments Shapira et al. (2024); Phelps and Russell (2023).

There are certain benefits of using Large Language Models instead of human subjects. Firstly, LLM-based experiments significantly reduce the costs associated with recruiting and compensating human subjects, even compared to online experiments. Such experiments are also easily scalable, enabling researchers to perform extensive experiments without the time constraints of human-based studies. By using LLMs, researchers can simulate interactions between agents with varied backgrounds and characteristics, enhancing the external validity of their findings. And finally, such experimental designs could allow us to study questions that could not be attempted with previous means for various reasons, such as IRB constraints.

However, the use of LLMs is not without its challenges. The goal of LLMs is for a computer to generate natural language responses to prompts provided in a natural language. These responses rely on giving the programs access to large training data sets. LLMs are not designed to reproduce human behavior directly. Further, their programming and their training sets are changing rapidly with every new model iteration. For these reasons, it is not clear if LLMs currently provide useful insights into the behavior of human subjects, nor is it clear that a LLM will provide the same response next month that it provides today.

This paper tests choice consistency of LLM generated data. It compares consistency of data obtained from human subjects to that generated by LLMs. If one wishes to claim that LLM data can be a proxy for data generated by human subjects, choice consistency is a necessary condition. The test is weak because choice consistency is only one descriptive statistic available from the data. This paper studies the degree to which LLM generated responses in one situation are similar to LLM generated responses in a similar situation and compares this ratio to the analogous ratio derived from human-subject data from Wolanski and Baranov (2023). It does not study other ways in which LLM data may diverge from human-subject data. Whether LLM generated data can fully replace human-subject data requires a more ambitious analysis that is outside the scope of this paper. The reports we describe here are a first step.

Even if LLMs do not fully succeed as proxies for human behavior, this paper offers another potential upside. It illustrates how the use of LLMs provides a very cost-effective way to obtain large datasets, which can be valuable for purposes beyond research, such as teaching experimental design or data analysis.

The main findings of the paper indicate that LLMs exhibit behavior regarding social choice consistency similar to that of real human subjects. Additionally, social choices made my LLMs are highly influenced by environmental changes, including close friendships and detailed personal descriptions of participants. Specifically, LLM-generated agents tend to be more inconsistent in their social choices when considering friends or individuals with shared features such as education or occupation. Moreover, they often reverse their choices when cooperation is possible, such as receiving a lower pay-off to increase a friend's pay-off. Conversely, they tend to be more egoistic when friends receive more money and the decision-maker cannot affect the outcome.

The paper is structured as follows. Firstly, we discuss what is currently known about social choice, Experimental Economics, and the involvement of LLMs in these areas. Then, we adjust the concept of social choice consistency of Wolanski and Baranov (2023) and explain how we measure it using surveys involving LLM-generated economic agents. Finally, we present the results of simulations involving LLMs, analyze how these results compare to experiments with human subjects, and explore how additional factors affect consistency.

3.2 Literature review

3.2.1 Social choice

In the economic literature, the study of consistency of decisions across different environments is primarily theoretical. Thomson (2011) analyzes how various allocation principles conform to different notions of consistency, Polman and Wu (2020) provides an overview of decision-making on behalf of oneself and others in risky environments. Zame et al. (2020) studies consistency experimentally but focuses on how social and personal preferences differ under uncertainty. Lastly, Wolanski and Baranov (2023) strips away from risk and uncertainty and explores how varying sizes of social domains might affect decisions.

We also contribute to the literature on social preferences. Traub et al. (2009) and Hong et al. (2015) both identify consistent differences in planners' decisions between equity and efficiency depending on whether their choices impact their outcomes or not, explaining one potential cause of inconsistency between choices. Our work is also related to Andreoni and Petrie (2004), which attempts to remove a part of confidentiality from actual lab experiments by showing subjects pictures of players involved in their decisions.

Finally, there is a growing literature that explains inconsistencies among decisions as mistakes, including, but not limited to, works by Esponda and Vespa (2023), Nielsen and Rehbeck (2022), Oprea (2020), Oprea (2023), and Enke et al. (2023). Asking for further elaborations from LLM-based agents about why a particular decision was made could allow us to see if this explanation remains valid.

3.2.2 LLMs for simulation

The current use of Large Language Models (LLMs) in economics involves using one as a Research Assistant Charness et al. (2023), which can help with experimental design, better wording of the questions, easier data analysis, and documentation production. However, LLMs show emergent capabilities in text understanding, zero- and few-shot classification, question answering, and natural language reasoning Brown et al. (2020); Wang et al. (2019). The prominent way to utilize LLMs is to use instruction prompting Ouyang et al. (2022) where the user describes a task for a language model, e.g., to find mistakes in the provided text or to solve some problem. One of the most influential applications of such instructions is a role-playing prompt Park et al. (2023) that adds a set of behavioral patterns to the initial task given to the LLM. Role-playing prompts open a new door in the computational social sciences, allowing us to simulate human behavior realistically Park et al. (2023).

LLMs are good at representing the average human because it is trained on a vast amount of textual data from the internet Dillion et al. (2023), including user-generated ones. It is applied to various tasks. Binz and Schulz (2023) show that GPT-3 Brown et al. (2020) answer psychological questions similar to human subjects. To evaluate the reasoning abilities of LLMs, Aher et al. (2023) propose Turing experiments. Turing experiments are a couple of economic, social, or psychological tasks: ultimatum game, garden path sentences, Milgram shock experiment, and wisdom of the crowds. Akata et al. (2023) additionally apply LLMs to the scenario of the repeated game. Ghaffarzadegan et al. (2023) extend classic agent-based modeling with LLMs to empower the simulation of real-human subject behavior. Motoki et al. (2023) propose strategies to generate survey data, and Kim and Lee (2023) employ LLM fine-tuning with survey data to align LLMs with human behavior and further advance simulations. Shapira et al. (2024); Phelps and Russell (2023) show emergent economical behavior.

One of the important problems is that LLM answers are poorly aligned with human expectations, possibly generating dangerous or offensive content. So, fine-tuning methods Ouyang et al. (2022); Lee et al. (2023); Rafailov et al. (2023) were proposed. These methods lead to the limited representation of humans by modern aligned models Santurkar et al. (2023); Harding et al. (2023); Tjuatja et al. (2023); Heyman and Heyman (2023); Crockett and Messeri (2023). Also, some studies still show limited human representation in some economic scenarios Fan et al. (2023). Thus, our goal is to validate whether LLMs' behavior is aligned with human subject behavior in social choice and further extend the experiment to obtain data about more societal compositions with possibly sensitive info.

3.3 Methodology

Our main goal is to study human and LLM social choice behavior in the different domains. As the first step, we build on the experiment of Wolanski and Baranov (2023), which studies differences in decisions made in different domains across which it is natural to expect similar behavior. In particular, the experiment measures whether the subject always picks a more equal allocation or goes for a more efficient one when independent players get introduced, when they get their skin in the game, and when their decision affects a larger group of people.

Further, we conduct a series of experiments with a simulated participant to study choice reversals:

1. We start with a default experiment when we do not provide any information about players involved in the question. This experiment aims to check whether the behavior of LLMs is aligned with human behavior observed in Wolanski and Baranov (2023), and thus, whether we can use LLMs as participants in our study.

2. We extend the default prompt by describing the players for whom the decision-maker selects an option. The goal is to study whether people are prone to changing their decisions to favor someone with similar features like a job or education.

3. We extend the default prompt with the friendship between the decision maker and the first player. Thus, we further deepen the severe intent of the decision maker to change the behavior to favor another person.

3.3.1 Social choice consistency

Following Wolanski and Baranov (2023), we say that a pair of decisions is consistent if the subject chooses the same type of allocation – either the one with lower inequality or higher efficiency. For each pair of the decisions made by the agent, we can say whether they are consistent with each other or not. If the agent's choices are consistent for each pair, we say that it satisfies pair-wise consistency. The way the decisions are constructed, pair-wise consistency is a generalization of a common property – independence of irrelevant alternatives – since we compare environments that differ by introducing new agents with payoffs independent of the choice.

3.3.2 Behavioral experiment

The basic decision is described in Table 3.1. Our experimental decisions are designed to capture trade-offs between inequality and efficiency. Option 1 offers a higher sum of payments to Player 1 and Player 2, but the payoffs are more spread out. Option 2 has $5 less in total to be paid out, with the final payouts much closer to each other. The more the agent cares about inequality, the more likely they should be to choose a more equal Option 2.

There are three variations for each of the proposed decisions: the decision-maker could be an outside observer, a player receiving a lower payment, or a player receiving a higher payment.

These decisions get changed in two ways: by adding more players with payoffs independent of the choice and by duplicating player 1 and player 2 (effectively having 4 agents, where 2

Table 3.1. Basic choice our participants made. There are three variants; the decision-maker could be player 1 (receiving the lower payment profile), player 2 (receiving the higher payment profile), or not a player and receive no payment from this decision.

Choice 1	Payment to Player 1	Payment to Player 2
Option 1	$25	$5
Option 2	$15	$10

have the same payoffs as player 1 in the basic decision, and the other 2 have the same payoffs as player 2).

Our experiment has three treatments. The first one follows the standard Wolanski and Baranov (2023) setup and is designed to align our model with existing data on real subjects. It is used as a test that the model produces results which could be used to some degree to model average human behavior. The second treatment introduced descriptions of players for whom the model makes decisions. While economic experiments must be anonymous, most social interactions are not. The use of AI here allows us to study the impact of these social links on decisions. The third treatment introduces a layer of social relationships between different agents in the sample. In particular, we tell the LLM agent that it is friends with player 1 in the experiment.

3.3.3 LLM as a participant

To conduct behavioral experiments, the OpenAI GPT-4-1106-preview[1] Achiam et al. (2023) model is used with prompts described in Figure 3.1.

Words inside figure brackets (e.g., {income}) refer to the different variables. Education stands for the level of education of the participants or players, e.g., high school, bachelor's, master's, etc. Income can be of two types: below average and above average. The job sector can be private, government, self-employed, or unknown. These variables are filled using a sample from the US Census dataset Kohavi (1996). Players' descriptions are empty for default and friendship experiments and filled with income, job, and education information from the US

[1]https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo

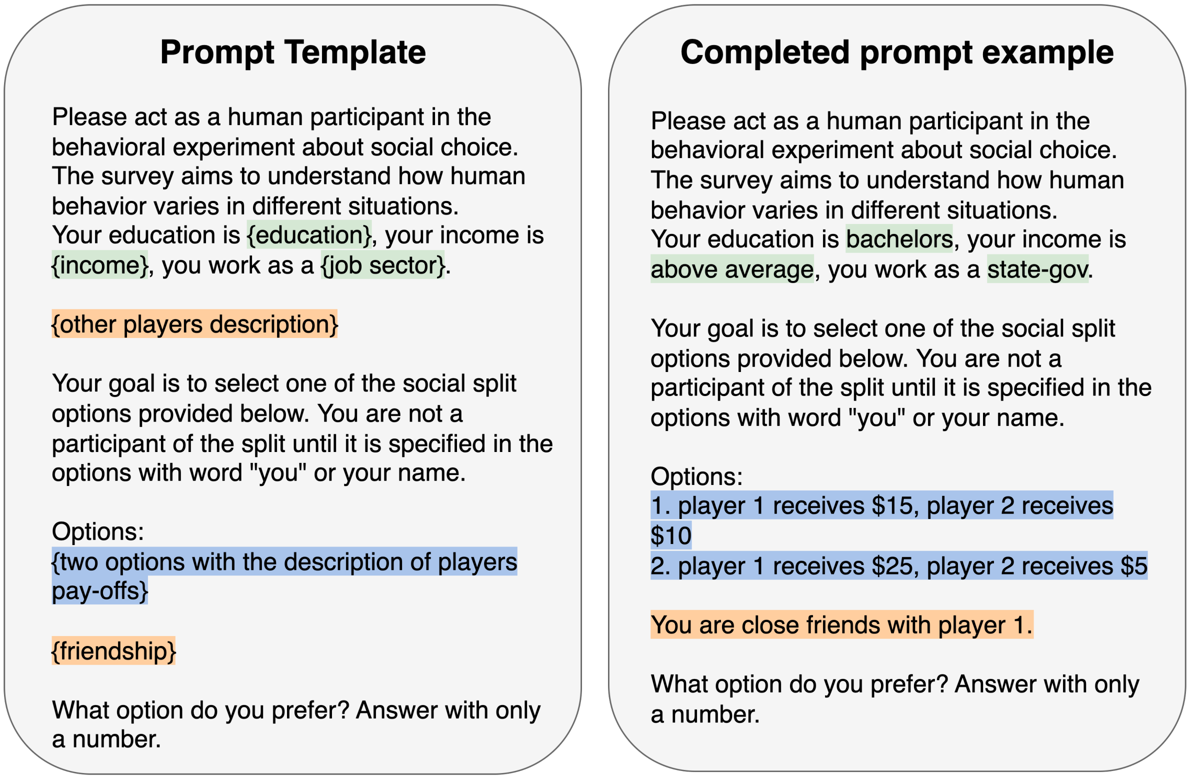


Figure 3.1. Prompt template for the survey. Words inside figure brackets (e.g., {income}) refer to the different variables. Variables highlighted in green are the parameters of a participant. Orange fields are optional variables for extended experiments (with players' descriptions or with friendship). Blue one stands for the options of the question.

Census for the experiment with players' descriptions.

Options are presented in the subsection 3.3.2 and describe the payoffs for participants of the split. To receive more robust results, we generate responses with LLM three times in a row for each question and each person. The answer to analyze is selected by the simple majority. All questions were asked in three different environments: default, with other players' descriptions, and with friendship links.

All of our experiments cost about $300 combined, much less than hiring human participants using prolific.com (which asks to pay each participant at least minimum wage), a commonly used method for conducting economic research online.

3.3.4 Metrics

To understand whether the behavior of humans is consistent across different cases, we use the metric called consistency rate. Further, we study the changes in agents' answers after providing additional information to them (players' descriptions and friendship connections to any of the players).

Consistency rate

Previously, we have defined *pairwise consistency* as a preference for a similar option across different domains, i.e., if a person prefers inequality in one domain, they should prefer it in others and vice-versa. Each question in our survey has precisely two options: one option is inequality-averse, and another is inefficiency-averse. So, the choice consistency between two questions means that a participant selects the same option.

All questions are of three types:

1. a participant gets no or equal pay-off.
2. a participant receives a richer pay-off than other players.
3. a participant receives a poorer pay-off than other players.

These groups suppose different answers by the user, so we create answer pairs only inside these groups. Overall, we have 1747 unique pairs for a person.

Given the pairs of questions and participant's answers, we can calculate whether each choice is pairwise consistent. Finally, we calculate the consistency rate for a participant as a rate of the pairwise consistent answers and take the average over all users.

To better understand different properties of behavior, we define several properties for each pair of questions. These properties are binary and describe the questions. Below, we present the list of the properties and their name in the results table in brackets;

- The decision-maker participates in the social choice and receives a monetary incentive at least in one of the questions (DM paid).
- The decision-maker does not participate in the social choice and does not receive a monetary incentive at least in one of the questions (DM not paid)
- The decision-maker participates in the social choice and receives equal monetary incentive between options at least in one of the questions (DM paid a constant)
- The decision-maker participates in the social choice and receives equal monetary incentive between options in both questions (DM paid a constant both)
- The decision-maker participates in the social choice and receives equal monetary incentive between options in only one question from the pair (DM paid a constant only one)
- The questions have the same number of players (same size)
- The questions have different numbers of players (different sizes)
- At least one question has no duplication (no duplication)
- At least one question has duplication (duplication included)
- The decision-maker participates in the social choice and receives different monetary incentives between options (DM paid non-constant).
- DM not paid or DM paid a constant (DM constant or no payment)
- The decision-maker participates in the social choice and receives poorer monetary incentive on average in both questions (DM paid poorer)
- The decision-maker participates in the social choice and receives a richer monetary incentive on average in both questions (DM paid richer).

Choice reversals

To analyze the effect of additional descriptions, we do the same analysis as previously but within additional properties of questions and descriptions. The list of additional properties for the descriptions is the following:

- **Same education at least once**: the decision-maker has a similar education level as one of the players
- **Same income at least once**: the decision-maker has a similar income level as one of the players
- **Same job sector at least once**: the decision-maker has a similar job sector as one of the players
- **Same education exactly once**: the decision-maker has a similar education level with precisely one player
- **Same income exactly once**: the decision-maker has a similar income level with precisely one player
- **Same job sector exactly once**: the decision-maker has a similar job sector with precisely one player
- **Same education, poorer**: the decision-maker has a similar education level with precisely one player who receives poorer monetary incentive on average
- **Same education, richer**: the decision-maker has a similar education level with precisely one player who receives richer monetary incentive on average
- **Same income, poorer**: the decision-maker has a similar income level with precisely one player who receives poorer monetary incentive on average

- **Same income, richer**: the decision-maker has a similar income level with precisely one player who receives richer monetary incentive on average
- **Same job sector, poorer**: the decision-maker has a similar job sector level with precisely one player who receives poorer monetary incentive on average
- **Same job sector, richer**: the decision-maker has a similar job sector level with precisely one player who receives richer monetary incentive on average
- The rest of the properties are logical and between the properties for the default experiment and the extended one

Here is the list of additional properties for the friendship:

- **Friend paid poorer**: the friend (player 1) receives poorer monetary incentive on average
- **Friend paid richer**: the friend (player 1) receives richer monetary incentive on average
- **Friend paid constant**: the friend (player 1) receives equal monetary incentives in both questions
- The rest of the properties are logical and between the properties for the default experiment and the extended one

3.4 Results

Table 3.2 shows the choice consistency results for the default question properties. We observed that the choices in role-playing scenarios appear to be a little bit more stable compared to human answers of Wolanski and Baranov (2023), who report an average consistency rate of 0.8141. The main reason is the changes in the questionnaire. This consistency is notably influenced by the presence of players' descriptions of individuals or friendships.

Table 3.2. Choice consistency rate across default properties (standard deviation is lower than 0.01 for all metrics)

	Default	Descriptions	Friendship
DM paid	0.82	0.70	0.74
DM not paid	0.86	0.76	0.89
DM paid a constant	0.77	0.66	0.79
DM paid a constant both	0.69	0.59	0.71
DM paid a constant only one	0.80	0.69	0.83
Same size	0.87	0.73	0.89
Different sizes	0.83	0.73	0.70
No duplication	0.83	0.65	0.68
Duplication included	0.82	0.74	0.69
DM paid non-constant	0.89	0.74	0.68
DM constant or no payment	0.82	0.72	0.85
DM paid poorer	0.97	0.69	0.98
DM paid richer	0.85	0.77	0.53

Table 3.3. Choice consistency rate across properties specific for the questions with players' descriptions (standard deviation is lower than 0.01 for all metrics)

	Descriptions	DM low income	DM high income	DM is a government worker	DM is a private worker
Same education at least once	0.75	0.76	0.69	0.74	0.74
Same income at least once	0.74	0.74	0.71	0.74	0.73
Same job sector at least once	0.75	0.75	0.71	0.80	0.73
Same education exactly once	0.73	0.74	0.68	0.74	0.72
Same income exactly once	0.67	0.67	0.69	0.72	0.66
Same job sector exactly once	0.71	0.72	0.67	0.80	0.67
Same education, poorer	0.85	0.85	0.87	0.86	0.87
Same education, richer	0.71	0.74	0.61	0.70	0.70
Same income, poorer	0.90	0.97	0.70	0.91	0.91
Same income, richer	0.76	0.81	0.58	0.84	0.73
Same job sector, poorer	0.87	0.90	0.76	0.79	0.90
Same job sector, richer	0.67	0.69	0.60	0.81	0.62
DM paid poorer AND Same education, richer	0.79	0.90	0.49	0.93	0.76
DM paid poorer AND Same income, richer	0.68	0.71	0.55	0.66	0.66
DM paid poorer AND Same job sector, richer	0.64	0.67	0.56	1.00	0.59
DM paid richer AND Same education, poorer	0.88	0.97	0.64	0.89	0.88
DM paid richer AND Same education, poorer	0.83	0.83	0.83	0.85	0.84
DM paid richer AND Same income, poorer	0.78	0.77	0.81	0.83	0.72
DM paid richer AND Same job sector, poorer	0.85	0.87	0.76	0.78	0.89
DM paid constant AND Same education, richer	0.92	1.00	0.86	1.00	0.91
DM paid constant AND Same education, richer	0.67	0.72	0.55	0.80	0.64
DM paid constant AND Same income, richer	0.85	0.84	0.88	0.82	0.85
DM paid constant AND Same income, richer	0.70	0.71	0.64	0.68	0.68
DM paid constant AND Same job sector, poorer	0.85	0.85	0.89	0.80	0.85
DM paid constant AND Same job sector, poorer	0.68	0.69	0.62	0.77	0.63

Table 3.3 shows the additional question properties for the experiment with players' descriptions. We observe distinct patterns regarding income levels and occupational backgrounds, shedding light on the dynamics of decision-making consistency within these contexts.

When the DM income falls below the average, we observe a propensity towards increased consistency in behavior, mainly when parameters are identical to those of other players. Specifically, when the DM's average payoff is lower or equal to that of others, their choices tend towards uniform distributions; otherwise, they lean towards unequal distributions.

Conversely, if the DM's income surpasses the average, there is a behavior divergence across most cases. The prevailing tendency is towards unequal distribution.

Government workers exhibit slightly higher consistency, particularly towards uniform distributions, except when the opportunity for higher earnings is present. This implies that individuals in public service may prioritize fairness or equality in their decision-making, albeit with some sensitivity to potential financial gains.

Table 3.4. Choice consistency rate across properties specific for the questions with friendship (standard deviation is lower than 0.01 for all metrics)

	Friendship	DM low income	DM high income	DM is a government worker	DM is a private worker
Friend paid poorer, both	0.84	0.82	0.94	0.91	0.84
Friend paid richer, both	0.88	0.88	0.87	0.90	0.88
Friend paid constant, both	0.98	0.98	0.98	0.98	0.98
Friend paid poorer	0.27	0.27	0.25	0.25	0.27
Friend paid richer	0.78	0.78	0.77	0.79	0.78
Friend paid constant	0.90	0.90	0.91	0.92	0.90
DM paid poorer AND Friend paid richer, both	0.98	0.99	0.94	0.99	0.97
DM paid richer AND Friend paid poorer, both	0.84	0.82	0.94	0.91	0.84
DM paid a constant AND Friend paid richer, both	0.71	0.71	0.73	0.75	0.71
DM not paid AND Friend paid richer, both	0.71	0.71	0.73	0.75	0.71
DM not paid AND Friend paid constant, both	0.98	0.98	0.98	0.98	0.98

Table 3.4 presents the additional results for the properties specific to the experiment about friendship.

When the peer's average payoff is lower, and the DM's income exceeds the average, or when the player is a government worker or self-employed, there is a heightened tendency towards consistently favoring equality in decision-making. This suggests that individuals in relatively advantageous positions, whether in income or occupational status, are more inclined to prioritize equitable outcomes when their peers face lower average payoffs.

On the other hand, when the DM's income surpasses the average, and the potential payoff is low, while the peer's payoff is high, there is a slight decrease in the consistency of choices towards equality. This indicates that when individuals with above-average incomes are confronted with a scenario where their potential gains are limited, but their peer stands to gain significantly more, their commitment to equality wavers slightly.

Table 3.5. Choice reversal rate across different question types after adding friendship with player 1.

DM receive payment	Question type	Friend type	DM paid >Friend	Reversal probability	Part from all reversals	Inequality (default)	Inequality (friend)
False	equal	equal	False	0.06	0.05	0.05	0.01
		richer	False	0.00	0.00	0.00	0.00
True	richer	poorer	False	0.73	0.67	0.82	0.09
		richer	False	0.01	0.01	0.99	1.00
	poorer	richer	False	0.03	0.05	0.02	0.01
	equal	richer	False	0.10	0.14	0.03	0.08
			True	0.34	0.09	1.00	0.66

After introducing connections with the first player, the analysis of choice reversal using Table 3.5 reveals two main reasons driving these shifts.

- In most cases, DM tends to reduce inequality. This also usually increases the resulting pay-off for a friend but reduces it for the DM
- DM does not change their choice if DM has no pay-off, but a friend can get a higher pay-off, sacrificing equality.

One of the benefits of using LLM as a participant is the ability to provide detailed answers. Due to privacy reasons and possibly a lot of required time for a participant, we can not always ask for it from a human participant. We present the sample explanations in the Appendix C.1. To generate this explanation, we replace the ”Answer with only a number” in the prompt with the ”Describe your choice in a few sentences. Please be honest and do not try to look better than you are”. The explanations are consistent with our previous analysis, showing that participants are more egoistic when they can maximize their pay-off and a pay-off of the friend together but prone to equality when monetary incentives differ.

3.5 Conclusion

Large Language Models are prominent tools for Behavioral and Experimental Economics. They allow easy extension of experiments with human subjects with role-playing. Our analysis reveals that LLMs exhibit behavior agreeing with real humans regarding social choice consistency.

This result suggests that LLMs can serve as valuable tools for simulating human behavior in extended experiments, providing researchers with a cost-effective and efficient means of conducting large-scale studies. Moreover, we show that LLMs are potent tools for interpretation. LLM-based participants can generate high-quality reasoning about why they conduct specific behaviors. Such reasoning may also be applied to explain the behavior of real humans.

Our investigation uncovers the significant influence of environmental factors on social choice consistency. Specifically, we found that social choices are highly sensitive to changes in the environment, such as the presence of close friendships and shared characteristics among participants. Additionally, we found intriguing patterns in decision-making behavior, particularly regarding cooperation and altruism. We observe that individuals often exhibit altruistic tendencies when cooperation is possible, even at the expense of their payoff. However, we also found evidence of egoistic behavior, particularly when individuals can not affect their constant outcome.

Both findings highlight the necessity of extensive research on choice consistency in more complex environments like different organizational structures or even social networks to improve our understanding of human incentives and logic about social welfare decisions. Additionally, further refinement of the LLMs as experimental tools provides the ability to simulate possible scenarios and help people make fair and Pareto-efficient decisions.

Chapter 3 is co-authored with Kisilev Dmitrii, Marchenkov Valerii, and Makarov Ilya, and is currently being prepared for submission for publication. The dissertation author is one of primary authors of this chapter.

Appendix A

Supplementary material for Chapter 1

A.1 Omitted proofs

Proposition 0: If $I^A R^C I^B$, then there exists ε such that $\sum_{t \leq \tau} \varepsilon_t \leq 0$ for all t and $I^D R^V I^B$, where $I^D = I^A + \varepsilon$.

Proof of Proposition 0: Let $t_1 = \min t$ such that $I_t^A < I_t^B$. Since $I^A R^C I^B$, $I_{t_1}^B - I_{t_1}^A \leq \sum_{t=1}^{t_1} I_t^A - I_t^B$. We can then define $I_t^D = I_t^B$ for $t < t_1$ and $I_{t_1}^D = I_{t_1}^A + \sum_{t=1}^{t_1} I_t^A - I_t^B$.

Let $t_k = \min_{t > t_{k-1}} t$ such that $I_{t_k}^A < I_{t_k}^B$. We can then define $I_t^D = I_t^B$ for $t_{k-1} < t < t_k$ and $I_{t_k}^D = I_{t_k}^A + \sum_{t=1}^{t_1} I_t^A - I_t^B$.

Once all the cases where $I_t^A < I_t^B$ are covered, the resulting I^D by construction will be vector dominated by I^A and it will vector dominate I^B.

Proof of Proposition 1: Let $\varepsilon = \frac{1}{4}(I^A - I^B)$. We will show that

$$I^A R (I^A - \varepsilon) \implies (\frac{3I^A + I^B}{4}) R (\frac{2I^A + 2I^B}{4}) R (I^B + \varepsilon) \implies (\frac{I^A + 3I^B}{4}) R I^B$$

for $R = R^V, R = R^C$, and $R = R^D$.

For $I^A R^V I^B$ we have $I_t^A \geq I_t^B$, which implies that $I_t^A \geq (\frac{3I_t^A + I_t^B}{4}) \geq (\frac{2I_t^A + 2I_t^B}{4}) \geq (\frac{I_t^A + 3I_t^B}{4}) \geq I_t^B$.

For $I^A R^C I^B$ we have $\sum_{\tau=1}^t I_t^A \geq \sum_{\tau=1}^t I_t^B$ for all t, which implies that

$$\sum_{\tau=1}^{t} I_t^A \geq \frac{3\sum_{\tau=1}^t I_t^A + \sum_{\tau=1}^t I_t^B}{4} \geq \frac{2\sum_{\tau=1}^t I_t^A + 2\sum_{\tau=1}^t I_t^B}{4} \geq \frac{\sum_{\tau=1}^t I_t^A + 3\sum_{\tau=1}^t I_t^B}{4} \geq \sum_{\tau=1}^{t} I_t^B$$

for all t.

For $I^A R^D I^B$ we have $\sum_{t=1}^T \delta^T I_t^A \geq \sum_{t=1}^T \delta^t I_t^B$ for all $\delta \in (0,1)$, which implies that $\sum_{t=1}^T \delta^T I_t^A \geq \frac{3\sum_{t=1}^T \delta^T I_t^A + \sum_{t=1}^T \delta^T I_t^B}{4} \geq$
$\geq \frac{2\sum_{t=1}^T \delta^T I_t^A + 2\sum_{t=1}^T \delta^T I_t^B}{4} \geq \frac{1\sum_{t=1}^T \delta^T I_t^A + 3\sum_{t=1}^T \delta^T I_t^B}{4} \geq \sum_{t=1}^T I_t^B$ for all $\delta \in (0,1)$.

Proof of Proposition 2: Let τ be the minimum of $t \geq 0$ such that I^A is richer than I^B at period τ. If no such τ exists, switch the roles of A and B. Let τ' be the minimum of $t > \tau$ such that B is weakly richer than A in period τ'. If $I^A R^C I^B$, instead define $\tau' = \tau + 1$. Let $\varepsilon_\tau = \frac{1}{4} \cdot \min_{\tau \leq t < \tau'} w_t^{I^A} - w_t^{I^B}$, $\varepsilon_{\tau'} = -\varepsilon_\tau$, and $\varepsilon_t = 0$ otherwise. By construction, $I^A - \varepsilon$ is still richer than $I^B + \varepsilon$ at $\tau \leq t < \tau'$ and $I^B + \varepsilon$ is still richer than $I^A - \varepsilon$ at τ', but the difference in wealth is smaller for $\tau \leq t < \tau'$ and unchanged for $t \geq \tau'$ and $t < \tau$, so a planner who follows generalized Pigou-Dalton principle for period-wise inequality prefers $(I^A - \varepsilon, I^B + \varepsilon)$ to (I^A, I^B).

Proof of Proposition 3: For whole-stream inequality measures, if $cR(c-\varepsilon)R(c+\varepsilon)Rc$, then $\varepsilon = 0$. Since all of the stream dominance relations we discussed are partial orders, they are transitive, and therefore $cR(c-\varepsilon)R(c+\varepsilon)Rc$ implies $(c+\varepsilon)R(c-\varepsilon)$ and therefore that $c - \varepsilon = c + \varepsilon$, so $\varepsilon = 0$

For period-wise inequality, if $(I^A, I^B) = (c, c)$, then $w_t^{I^A} = w_t^{I^B}$ for all t. Therefore if ε satisfies $w_\tau^{I^A} \geq w_\tau^{I^A - \varepsilon} \geq w_\tau^{I^B + \varepsilon} \geq w_\tau^{I^B}$, or $w_\tau^{I^B} \geq w_\tau^{I^B + \varepsilon} \geq w_\tau^{I^A - \varepsilon} \geq w_\tau^{I^A}$ for all τ then $\varepsilon = 0$

Proof of Proposition 4: The proof is by construction. For all three rankings, it is enough to take $\varepsilon = \frac{I^A - I^B}{3}$.

For R^V, the result is trivial.

For R^C, $I^A R^C I^B \iff I^A - I^B$ cumulatively dominates a vector of zeros. A proposed epsilon preserves cumulative dominance and makes the difference smaller.

For R^D, we refer the reader to Theorem 1 of Chambers and Echenique (2018).

A.2 Experimental instructions

This study takes place in two parts. The first part of the study examines peoples' decisions about monetary payments received by other people. For this, we have designed a series of choices of payment schedules that will be received by other people.

You will be presented with a pair of payment schedules and asked which of the schedules you would rather be given to two other participants, known to you as 'player 1' and 'player 2'. These two other participants have just been chosen at random from the participants in the session, before any decisions have been made. While these groupings are random, no one will be making decisions for a participant who is also making decisions for them. You will make decisions for the same pair of participants in each choice, but either participant could be player 1 or player 2 for a given choice.

After you have made all of your decisions for part 1, one of your choices will be selected at random and become your decision-that-counts for part 1–there is a 50 percent chance that your decision-that-counts for part 1 will be paid to other participants and a 50 percent chance that your decision-that-counts will for part 1 will not be paid to anyone (this is because each of you is making decisions for two other participants, which means that you are having decisions made for you by two other people. Rather than paying both of these decisions, only one of these decisions is paid. We will explain this in more detail below). After you have made all of your decisions for part 1, and when prompted to by the computer, you will then flip over the card in front of you. This card contains three pieces of information. First, each card has a dedicated space for you to write the decision-that-counts from part 1. Second, each card lists which other participants (identified by ID number) you made decisions for and who was labelled as 'player 1' and who was 'player 2' in the decision-that-counts. This means you will learn which participants you made decisions for only after you have made all of your decisions, and at no point will any participant learn who made decisions for them (we include the ID numbers only so that we can re-distribute the cards appropriately). Third, each of these cards has either HEADS or TAILS written on the second line. These designations have been randomly assigned to the cards, but have been done in such a way that there is exactly one HEADS card with each participant ID number and exactly one TAILS card with each participant ID number.

You will record your decision-that-counts in the dedicated space on the card. The experimenters will then collect all of the cards, sort them into the HEADS group and the TAILS group, and then flip a coin. The experimenters will then distribute the cards from the group chosen by the coin flip to the participants listed on them, and the decision-that-counts listed on them will be that participant's payment for part 1. The decisions listed on cards from the other group will not be paid to anyone.

The payment schedule from the decision-that-counts listed on a participant's card will be their payment from part 1. This means that there is a 50 percent chance that your decision-that-counts from part 1 will be paid to two other participants in this session and a 50 percent chance that your decision-that-counts from part 1 will not be paid to anyone. As such, it is in your best interest to truthfully state which option you would rather be distributed to the other participants in each choice, since that decision could be paid to other participants. Each participant will receive exactly one card, and therefore each of you will receive exactly one payment schedule from the decision-that-counts in part 1. The purpose of these cards is to ensure that each participant leaves the experiment knowing their payment schedules.

Appendix B

Supplementary material for Chapter 2

B.1 Sample-balance Table

Table B.1. Linear regression of the effect of different demographic variables on consistency percentage of the decision-maker, with standard errors clustered at the participant level. * is significant at 90%, ** is significant at 95%, *** is significant at 99%. We fail to reject the null hypothesis that the male and female coefficients are identical, with the F-test producing a p-value of 0.6427.

Dependent Variable	Effect
male	-.0377 (.0552)
female	-.0499 (.0577)
age	-.0007 (.0010)
employed	-.0156 (.0246)
# of surveys completed	.0000164 (.00000769)**
time spent on experiment	-.00003 (.00004)
regression constant	.8720 (.0618)
R^2	.0310

B.2 Robustness Checks

B.2.1 Results from Section 4.6.1

This section repeats the analysis in section four but without the first 60 participants who saw 5 incorrect questions which were dropped from the analysis, and did not see the blocks in a

Table B.2. A version of table 2.7 which excludes the first 60 participants. The table presents multiple specifications comparing the effect of different choice environments on consistency. All regressions computed with with standard errors clustered at the participant level, and with participant-level fixed effects where indicated. * is significant at 90%, ** is significant at 95%, *** is significant at 99%. All conclusions are the same as in table 2.7.

	LPM	Logit	LPM	Logit
adding choice-independent outcome	.0004 (.0061)	.0044 (.0529)	-.0036 (.0059)	-.0332 (.0529)
duplication	-.0358 (.0134)***	-.2807 (.0983)***	.0087 (.0115)	.0522 (.0824)
combined change	-.0827 (.0114)***	-.5780 (.0751)****	-.0156 (.0070)**	-.1135 (.0527)**
DM paid on one only			-.0494 (.0129)***	-.2901 (.0777)***
DM paid on both			.0038 (.0189)	-.0303 (.1231)
DM richer			.0623 (.0192)***	.4904 (.1547)***
DM poorer			.1032 (.0170)****	.9519 (.1715)****
control for 60 participants	.0412 (.0239)*	.3031 (.1850)	.0356 (.0238)	.2672 (.1860)
regression constant	.8551 (.0129)	1.7928 (.1010)	.8146 (.0180)	1.4889 (.1203)
Log Likelihood		-14,861.112		-14,669.928
(pseudo) R^2	.0134	.0145	.0239	.0272
N	31,400	31,400	31,400	31,400
# Clusters	200	200	200	200

random order. There are no qualitative changes to any of our results.

B.2.2 Results from Section 4.6.2

Derivation of Relationship Between Consistency Percentage and Probability of Choice

Suppose that the decision-maker follows the random choice model described in section 4.6.2, and chooses the actual preferred option with probability p and the less preferred option with probability $1-p$. Each choice is a Bernoulli trial with probability p, so a collection of n realizations can then be described as a vector of length-n composed of 0's and 1's. For joint consistency, we are interested in the probability that any two components of this vector are equal. The probability that the vector contains k ones and $n-k$ zeroes is:

$$Pr(k \text{ ones and } n-k \text{ zeros}) = p^k \cdot (1-p)^{n-k} \cdot \binom{n}{k}$$

Assuming both n and k are greater than 1, we can calculate the resulting percentage of consistent choices conditional on n and k as:

Table B.3. A version of table 2.8 which excludes the first 60 participants. This table details the proportion of decision pairs where participants selected the inequality-averse choice, broken down by different choice environments. The left column is the proportion among individuals who are fully consistent in the sample, the middle and right columns are the proportion among individuals who are not fully consistent. The left and middle columns are the proportion where individuals selected the inequality-averse option on the choice with a smaller domain (regardless of their choice on the larger domain), and the left and right columns are the proportion where individuals selected the inequality-averse option on the choice with a larger domain (regardless of their choice on the smaller domain). Stars indicate significant differences against the left column. * is significant at 90%, ** is significant at 95%, *** is significant at 99%, **** is significant at 99.99%.

Environment	Consistent	Inconsistent, Smaller	Inconsistent, Larger
Aggregate	.7966 (.0059)	.5981 (.0041)****	.5624 (.0042)****
Adding choice-independent outcome	.7359 (.0123)	.5717 (.0079)****	.5624 (.0080)***
Duplication	.8286 (.0212)	.6190 (.0158)****	.5757 (.0161)****
combined change	.8188 (.0069)	.6070 (.0051)****	.5610 (.0052)****
DM choice-dependent outcome, aggregate	.6857 (.0143)	.5590 (.0088)****	.5419 (.0089)****
DM choice-dependent outcome, adding choice-independent outcome	.6857 (.0160)	.5560 (.0099)****	.5425 (.0099)****
DM choice-dependent outcome, combined change	.6857 (.0320)	.5714 (.0197)	.5397 (.0199)***
DM choice-dependent outcome, richer, aggregate	.4000 (.0213)	.2304 (.0106)****	.2229 (.0105)****
DM choice-dependent outcome, richer, adding choice-independent outcome	.4000 (.0239)	.2286 (.0118)****	.2230 (.0117)****
DM choice-dependent outcome, richer, combined change	.4000 (.0478)	.2381 (.0240)***	.2222 (.0234)***
DM choice-dependent outcome, poorer, aggregate	.9714 (.0073)	.8876 (.0080)****	.8610 (.0087)****
DM choice-dependent outcome, poorer, adding choice-independent outcome	.9714 (.0081)	.8833 (.0089)****	.8619 (.0097)****
DM choice-dependent outcome, poorer, combined change	.9714 (.0162)	.9047 (.0165)**	.8571 (.0197)****
DM choice-independent outcome, aggregate	.8286 (.0062)	.6093 (.0047)****	.5683 (.0047)****
DM choice-independent outcome, adding choice-independent outcome	.8286 (.0177)	.6007 (.0132)****	.5993 (.0133)****
DM choice-independent outcome, duplication	.8286 (.0212)	.6190 (.0158)****	.5757 (.0161)****
DM choice-independent outcome, combined change	.8286 (.0070)	.6096 (.0052)****	.5626 (.0053)****
DM paid on one domain, aggregate	.8286 (.0087)	.6047 (.0065)****	.5702 (.0066)****
DM paid on one domain, adding choice-independent outcome	.8286 (.0285)	.6038 (.0213)****	.6019 (.0214)****
DM paid on one domain, combined change	.8286 (.0091)	.6049 (.0068)****	.5670 (.0069)****
DM paid on both domains, aggregate	.7445 (.0103)	.5795 (.0067)****	.5544 (.0068)****
DM paid on both domains, adding choice-independent outcome	.7061 (.0146)	.5605 (.0092)****	.5463 (.0092)****
DM paid on both domains, duplication	.8286 (.0260)	.6254 (.0193)****	.5667 (.0197)****
DM paid on both domains, combined change	.7782 (.0170)	.5944 (.0116)****	.5636 (.0117)****
DM paid on neither domain, aggregate	.8286 (.0118)	.6183 (.0088)****	.5619 (.0090)****
DM paid on neither domain, adding choice-independent outcome	.8286 (.0319)	.6095 (.0238)****	.6262 (.0236)****
DM paid on neither domain, duplication	.8286 (.0368)	.6063 (.0275)****	.5937 (.0277)****
DM paid on neither domain, combined change	.8286 (.0136)	.6216 (.0101)****	.5459 (.0104)****

Table B.4. A version of table 2.9 which excludes the first 60 participants. Proportion Inequality ↓ is the proportion of choice pairs where we see decreasing inequality aversion as size increases; that is, where the decision-maker selected the inequality-averse option on the smaller domain but not on the larger domain. ** is significant at 95%, *** is significant at 99% (significance is in terms of difference from 50%, which would indicate no trending direction).

Environment	# Observations	Proportion Inequality ↓
Aggregate	3782	.5664 (.0081)****
Adding choice-independent outcome	722	.5249 (.0186)
Duplication	221	.5928 (.0330)***
combined change	2839	.5749 (.0093)****
DM choice-dependent outcome, aggregate	428	.5631 (.0240)***
DM choice-dependent outcome, adding choice-independent outcome	338	.5503 (.0271)*
DM choice-dependent outcome, combined change	90	.6111 (.0514)**
DM choice-dependent outcome, richer, aggregate	242	.5248 (.0321)
DM choice-dependent outcome, richer, adding choice-independent outcome	193	.5181 (.0360)
DM choice-dependent outcome, richer, combined change	49	.5510 (.0711)
DM choice-dependent outcome, poorer, aggregate	186	.6129 (.0357)***
DM choice-dependent outcome, poorer, adding choice-independent outcome	145	.5931 (.0408)**
DM choice-dependent outcome, poorer, combined change	41	.6829 (.0191)**
DM choice-independent outcome, aggregate	3354	.5668 (.0086)****
DM choice-independent outcome, adding choice-independent outcome	384	.5026 (.0255)
DM choice-independent outcome, duplication	221	.5928 (.0330)***
DM choice-independent outcome, combined change	2749	.5737 (.0094)****
DM paid on one domain, aggregate	1980	.5495 (.0112)
DM paid on one domain, adding choice-independent outcome	185	.5027 (.0368)
DM paid on one domain, combined change	1795	.5543 (.0117)****
DM paid on both domains, aggregate	566	.5767 (.0208)***
DM paid on both domains, adding choice-independent outcome	92	.5435 (.0519)
DM paid on both domains, duplication	167	.6108 (.0377)***
DM paid on both domains, combined change	307	.5570 (.0284)**
DM paid on neither domain, aggregate	808	.6064 (.0172)****
DM paid on neither domain, adding choice-independent outcome	107	.4673 (.0482)
DM paid on neither domain, duplication	54	.5370 (.0678)
DM paid on neither domain, combined change	647	.6253 (.0189)****
Duplication without choice-independent outcomes	54	.5370 (.0679)

$$1 - \text{consistency rate} = Pr(1 \text{ one and } 1 \text{ zero}) = 1 - \frac{n-k}{n} \cdot \frac{k}{n-1} \cdot 2$$

Combining the two terms together, we can calculate the total expected consistency rate given probability p to be:

$$\text{Expected consistency rate} = 1 - \sum_{k=0}^{n} p^k (1-p)^{n-k} \cdot \frac{2 \cdot (n-2)!}{(k-1)! \cdot (n-k-1)!} = 1 - 2p + 2p^2$$

Using this equation and the delta method, we can obtain an estimator for p based on the observed sample mean and sample variance (computed via bootstrap) of the consistency percentage.

Data and Analysis from Section 4.6.2

This section presents the detailed analysis from the above section, which repeats our initial analysis accounting for individuals with sufficiently low consistency rates that they appear to choose uniformly at random. There are no qualitative changes to the conclusions.

Table B.5. A version of table 2.7 which accounts for individuals whose consistency is so low they are classified as choosing uniformly at random. Multiple specifications comparing effect of different choice environments. All regressions computed with with standard errors clustered at the participant level, and with participant level fixed effects where indicated. * is significant at 90%, ** is significant at 95%, *** is significant at 99%. All conclusions are the same as in table 2.7.

	LPM	Logit	LPM	Logit
adding choice-independent outcome	.0008 (.0061)	.0076 (.0589)	-.0035 (.0059)	-.0369 (.0595)
duplication	-.0373 (.0135)***	-.3277 (.1108)***	.0072 (.0115)	.0476 (.0945)
combined change	-.0861 (.0115)***	-.6854 (.0832)***	-.0173 (.0070)**	-.1444 (.0597)**
DM paid on one only			-.0467 (.0128)***	-.3163 (.0900)***
DM paid on both			.0019 (.0181)	.0133 (.1361)
DM richer			.0615 (.0192)***	.5458 (.1717)***
DM poorer			.1025 (.0169)****	1.0487 (.1849)****
control for low consistency	-.3075 (.0124)****	-1.8504 (.1071)****	-.3073 (.0124)****	-1.8794 (.1069)****
regression constant	.9511 (.0101)	2.6468 (.1206)	.9051 (.0146)	2.2770 (.1369)****
Log Likelihood		-13,078.505		-12855.812
(pseudo) R^2	.1371	.1327	.1480	.1475
N	31,400	31,400	31,400	31,400
# Clusters	200	200	200	200

B.3 Experimental Instructions

B.3.1 General instructions:

This study examines peoples' decisions about monetary payments received by yourself and other people. You will be presented with a pair of options to be paid out to a randomly generated group of participants of this study, and asked which of the proposed options you would like to see paid out. Once we have collected data from all participants, we will select one decision at random to become the decision-that-counts. We will then randomly select 10% of participants to receive a bonus payment based on the decision-that-counts; this means that your decision-that-counts can affect the payment that you receive or that other people receive. As such, you should answer each question honestly, since your answers can affect your payment and the payments of other people. These bonus payments can range from \$5 to \$25. At no point will you see the choices other participants made for you, nor will any other participants see the choices you made for them.

Your total payment for the study consists of two components: your participation payment

Table B.6. A version of table 2.9 which excludes participants with consistency rates so low they are classified as choosing uniformly at random. This table details the proportion of decision pairs where participants selected the inequality-averse choice, broken down by different choice environments. The left column is the proportion among individuals who are fully consistent in the sample, the middle and right columns are the proportion among individuals who are not fully consistent. The left and middle columns are the proportion where individuals selected the inequality-averse option on the choice with a smaller domain (regardless of their choice on the larger domain), and the left and right columns are the proportion where individuals selected the inequality-averse option on the choice with a larger domain (regardless of their choice on the smaller domain). Stars indicate significant differences against the left column. * is significant at 90%, ** is significant at 95%, *** is significant at 99%, **** is significant at 99.99%.

Environment	Consistent	Inconsistent, Smaller	Inconsistent, Larger
Aggregate	.7384 (.0056)	.6197 (.0047)****	.5987 (.0048)****
Adding choice-independent outcome	.6648 (.0106)	.5951 (.0086)****	.5855 (.0086)
Duplication	.7655 (.0203)	.6340 (.0178)****	.6245 (.0179)****
combined change	.7733 (.0068)	.6301 (.0059)****	.6023 (.0060)****
DM choice-dependent outcome, aggregate	.6182 (.0120)	.5867 (.0095)**	.5630 (.0095)***
DM choice-dependent outcome, adding choice-independent outcome	.6182 (.0134)	.5833 (.0106)**	.5639 (.0107)***
DM choice-dependent outcome, combined change	.6182 (.0267)	.6000 (.0211)	.5593 (.0214)*
DM choice-dependent outcome, richer, aggregate	.3273 (.0163)	.26 (.0119)***	.2319 (.0115)****
DM choice-dependent outcome, richer, adding choice-independent outcome	.3273 (.0183)	.2583 (.0133)***	.2333 (.0129)****
DM choice-dependent outcome, richer, combined change	.3273 (.0365)	.2667 (.0269)	.2259 (.0255)**
DM choice-dependent outcome, poorer, aggregate	.9091 (.0100)	.9133 (.0077)	.8941 (.0084)
DM choice-dependent outcome, poorer, adding choice-independent outcome	.9091 (.0112)	.9083 (.0088)	.8944 (.0093)
DM choice-dependent outcome, poorer, combined change	.9091 (.0224)	.9333 (.0152)	.8926 (.0188)
DM choice-independent outcome, aggregate	.7817 (.0061)	.6309 (.0054)****	.6108 (.0055)****
DM choice-independent outcome, adding choice-independent outcome	.7588 (.0167)	.6182 (.0147)****	.6283 (.0146)****
DM choice-independent outcome, duplication	.7655 (.0178)	.6340 (.0178)****	.6245 (.0179)****
DM choice-independent outcome, combined change	.7880 (.0069)	.6328 (.0062)****	.6061 (.0063)****
DM paid on one domain, aggregate	.7838 (.0085)	.6311 (.0075)****	.6078 (.0076)****
DM paid on one domain, adding choice-independent outcome	.7702 (.0274)	.6200 (.0243)****	.6250 (.0242)***
DM paid on one domain, combined change	.7853 (.0089)	.6323 (.0079)****	.6060 (.0080)****
DM paid on both domains, aggregate	.6706 (.0089)	.6050 (.0072)****	.5824 (.0073)****
DM paid on both domains, adding choice-independent outcome	.6364 (.0126)	.5877 (.0098)***	.5698 (.0099)****
DM paid on both domains, duplication	.7455 (.0240)	.6370 (.0206)***	.6167 (.0209)****
DM paid on both domains, combined change	.7005 (.0150)	.6222 (.0124)****	.5908 (.0126)****
DM paid on neither domain, aggregate	.8158 (.0118)	.6301 (.0109)****	.6179 (.0110)****
DM paid on neither domain, adding choice-independent outcome	.76 (.0302)	.6209 (.0265)***	.6567 (.0259)****
DM paid on neither domain, duplication	.8286 (.0368)	.6256 (.0347)***	.6462 (.0342)***
DM paid on neither domain, combined change	.8286 (.0136)	.6329 (.0127)****	.6049 (.0129)****

Table B.7. A version of table 2.9 which excludes the participants with consistency rates so low they are classified as choosing uniformly at random. Proportion Inequality ↓ is the proportion of choice pairs where we see decreasing inequality aversion as size increases; that is, where the decision-maker selected the inequality-averse option on the smaller domain but not on the larger domain. ** is significant at 95%, *** is significant at 99% (significance is in terms of difference from 50%, which would indicate no trending direction).

Environment	# Observations	Proportion Inequality ↓
Aggregate	1697	.5657 (.0120)****
Adding choice-independent outcome	423	.5366 (.0242)
Duplication	117	.5299 (.0461)
combined change	1157	.5799 (.0145)****
DM choice-dependent outcome, aggregate	298	.6074 (.0283)***
DM choice-dependent outcome, adding choice-independent outcome	236	.5890 (.0320)***
DM choice-dependent outcome, combined change	62	.6774 (.0594)***
DM choice-dependent outcome, richer, aggregate	202	.5941 (.0346)***
DM choice-dependent outcome, richer, adding choice-independent outcome	159	.5849 (.0391)**
DM choice-dependent outcome, richer, combined change	43	.6279 (.0737)*
DM choice-dependent outcome, poorer, aggregate	96	.6354 (.0491)***
DM choice-dependent outcome, poorer, adding choice-independent outcome	77	.5974 (.0559)*
DM choice-dependent outcome, poorer, combined change	19	.7895 (.0935)**
DM choice-independent outcome, aggregate	1399	.5568 (.0132)***
DM choice-independent outcome, adding choice-independent outcome	187	.4706 (.0365)
DM choice-independent outcome, duplication	117	.5299 (.0461)
DM choice-independent outcome, combined change	1095	.5744 (.0149)****
DM paid on one domain, aggregate	777	.5611 (.0178)***
DM paid on one domain, adding choice-independent outcome	74	.4865 (.0581)
DM paid on one domain, combined change	703	.5690 (.0187)***
DM paid on both domains, aggregate	322	.5621 (.0276)**
DM paid on both domains, adding choice-independent outcome	53	.5283 (.0686)
DM paid on both domains, duplication	91	.5604 (.0520)
DM paid on both domains, combined change	178	.5730 (.0370)*
DM paid on neither domain, aggregate	300	.5400 (.0288)
DM paid on neither domain, adding choice-independent outcome	60	.4000 (.0632)
DM paid on neither domain, duplication	26	.4231 (.0969)
DM paid on neither domain, combined change	214	.5935 (.0336)
Duplication without choice-independent outcomes	26	.4231 (.0969)

of $2.50, and a 10% chance for a bonus payment ranging from $5 to $25 which depends on your choices, the choices of others, and random chance.

Once you have read and understood these instructions, please answer the following brief comprehension check.

Based on the instructions, what kinds of decisions will you be asked to make in this survey?

- ○ Choosing payments for yourself and other people
- ○ Choosing your desired salary
- ○ Rating restaurants between one and five stars
- ○ None of the other choices

Figure B.1. Comprehension check

B.3.2 Treatment

This study contains 34 questions, and is split into two parts. For one part of this study, you will be making decisions for some number of other participants in the study. For each decision, a different set of other participants in the study has been selected at random and paired with you for that decision. A participant you are paired with in a decision is referred to as **”player N”**, where N is a number (i.e. player 1, player 2, player 3, etc). At no point will you learn the identity of any other participants assigned you made decisions for, nor the identity of any participants who made decisions for you.

An example of the decision interface is provided below. Please read the interface; we will then ask two comprehension checks to make sure you understand the interface. **These questions are only examples designed to ensure you understand the interface; the outcome of the following choices will not be paid to anyone.**

The middle row describes the payments if you choose option 1, and the bottom row describes the payments earned if you choose option 2. The middle column describes the payments that would be earned by you, and the right column describes the payments that would be earned by player 1.

	Payment to player 1	Payment to player 2
Option 1	**$26**	**$14.81**
Option 2	**$17**	**$19.40**

Figure B.2. Example 1

If you chose option 1 in this decision, and this decision was selected to become the decision-that- counts, what would be player 1's bonus payment from this study?

If you chose option 1 in this decision, and this decision was selected to become the decision-that-counts, what would be the bonus payment to player 2?

You have successfully completed the comprehension check, and may now begin the study. We would like to remind you of three pieces of information.

All decisions you make for the remainder of the study have a chance of becoming the bonus payments paid to you or to other people in the study, so you should answer each question honestly.

There is no right or wrong answer for any of these decisions. We are interested in studying your preferences.

Some decisions may be repeated.

For this part of the study, you will be making decisions for some number of other participants in this study. For each decision, a different set of other participants in the study has been selected at random and paired with you for that decision. A participant you are paired with in a decision is referred to as ”player N”, where N is a number (i.e. player 1, player 2, player 3, etc). **Remember: all decisions you make in the study have a chance of becoming the bonus payments paid to you or to other people in the study, so you should answer each question honestly.**

Appendix C

Supplementary material for Chapter 3

C.1 Sample explanations of participants' behavior generated by LLM

Table C.1. Sample explanations on the question when DM receives richer pay-off at average and reverse behavior towards lower inequality after introducing the friend increasing the friend's pay-off.

Question	Typical answer without a friend	Typical answer with a friend
1. you receive $25, player 1 receives $5 2. you receive $15, player 1 receives $10	I would prefer option 1, where I receive $25 and player 1 receives $5. My choice is based on my personal financial situation, which is described as below average. As a state-gov employee, I might have a stable job, but the income does not necessarily meet all my needs or wants. Therefore, when presented with an opportunity to increase my income, even in a hypothetical scenario, my instinct is to maximize my gain. This choice also reflects a common economic behavior where individuals prioritize their own financial well-being, especially when the information about the other player is limited. I do not have information about player 1's financial situation, so I am making a decision based on my own needs. It is a practical and self-interested choice, not necessarily one that is aimed at achieving fairness or equality.	Given that player 1 is a close friend, I would prefer option 2, where I receive $15 and player 1 receives $10. My reasoning is based on the value I place on our friendship and the desire to maintain a sense of fairness and balance in our relationship. Even though my income is below average, the difference of $10 between the two options is not significant enough to justify causing potential harm to the friendship by taking a larger share for myself. By choosing a more equitable split, I'm reinforcing the trust and mutual respect that underpins our friendship, which I value more than the extra $10 I could receive in option 1.

Table C.2. Sample explanations on the question when DM does not receive any pay-off and reverse behavior towards higher inequality after introducing the friend increasing the friend's pay-off.

Question	Typical answer without a friend	Typical answer with a friend
1. player 1 receives $15, player 2 receives $10 2. player 1 receives $25, player 2 receives $5	Given the information provided and considering my background as someone with a Bachelor's degree working in State-gov with a below-average income, my preference would lean towards option 1, where player 1 receives $15 and player 2 receives $10. My logic is based on a sense of fairness and a preference for a more equitable distribution of resources. Although option 2 provides a higher total sum of money, the disparity between what player 1 and player 2 receive is significantly greater. Given my own experiences with a below-average income, I am more sensitive to income disparities and would prefer a split that is less unequal. Therefore, I would choose the first option as it seems fairer and lessens the gap between the two players' outcomes. This choice reflects a value for social equity over maximizing personal gain.	Given that I am close friends with player 1, my preference would lean towards the option that benefits them the most, which is option 2, where player 1 receives $25 and player 2 receives $5. My decision is influenced by my personal relationship with player 1, as I naturally have a bias towards wanting to see my friend benefit more from the situation. Additionally, since my income is below average, I might empathize with the desire to maximize financial gain when the opportunity arises, which could further justify my choice. It's important to acknowledge that this choice is less equitable for player 2, but my personal connection to player 1 is the primary factor driving my decision.

Bibliography

Achiam, J., S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, et al. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Aher, G. V., R. I. Arriaga, and A. T. Kalai (2023, 23–29 Jul). Using large language models to simulate multiple humans and replicate human subject studies. In A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett (Eds.), *Proceedings of the 40th International Conference on Machine Learning*, Volume 202 of *Proceedings of Machine Learning Research*, pp. 337–371. PMLR.

Akata, E., L. Schulz, J. Coda-Forno, S. J. Oh, M. Bethge, and E. Schulz (2023). Playing repeated games with large language models. *arXiv preprint arXiv:2305.16867*.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The economic journal 100*, 464–477.

Andreoni, J. and J. Miller (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica 70*, 737–753.

Andreoni, J. and R. Petrie (2004). Public goods experiments without confidentiality: a glimpse into fund-raising. *Journal of public Economics 88*(7-8), 1605–1623.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of economic theory 2*, 244–263.

Basili, M., P. Casaca, A. Chateauneuf, and M. Franzini (2017). Multidimensional pigou–dalton transfers and social evaluation functions. *Theory and Decision 83*, 573–590.

Billot, A. and X. Qu (2022). Perfect altruism breeds time consistency. *Working paper*.

Binz, M. and E. Schulz (2023). Using cognitive psychology to understand gpt-3. *Proceedings of the National Academy of Sciences 120*(6), e2218523120.

Bosmans, K., L. Lauwers, and E. Ooghe (2009). A consistent multidimensional pigou–dalton transfer principle. *Journal of Economic Theory 144*, 1358–1371.

Brown, T., B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems 33*, 1877–1901.

Chambers, C. P. and F. Echenique (2018). On multiple discount rates. *Econometrica 86*, 1325–1346.

Charness, G., B. Jabarian, and J. A. List (2023). Generation next: Experimentation with ai. Technical report, National Bureau of Economic Research.

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

Clark, A. E. and C. d'Ambrosio (2015). Attitudes to income inequality: Experimental and survey evidence. *Handbook of income distribution 2*, 1147–1208.

Crockett, M. and L. Messeri (2023). Should large language models replace human participants?

Debreu, G. (1954). Representation of a preference ordering by a numerical function. *Decision processes 3*, 159–165.

Dillion, D., N. Tandon, Y. Gu, and K. Gray (2023). Can ai language models replace human participants? *Trends in Cognitive Sciences 27*(7), 597–600.

Engel, C. (2011). Dictator games: A meta study. *Experimental economics 14*, 583–610.

Enke, B., T. Graeber, and R. Oprea (2023). Complexity and time. *Working Paper*.

Esponda, I. and E. Vespa (2023). Contingent thinking and the sure-thing principle: Revisiting classic anomalies in the laboratory. *Review of Economic Studies (Forthcoming)*.

Fan, C., J. Chen, Y. Jin, and H. He (2023). Can large language models serve as rational players in game theory? a systematic analysis. *arXiv preprint arXiv:2312.05488*.

Feng, T. and S. Ke (2018). Social discounting and intergenerational pareto. *Econometrica 86.5*, 1537–1567.

Forsythe, R., J. L. Horowitz, N. E. Savin, and M. Sefton (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior 6*, 347–369.

Ghaffarzadegan, N., A. Majumdar, R. Williams, and N. Hosseinichimeh (2023). Generative agent-based modeling: Unveiling social system dynamics through coupling mechanistic models with generative artificial intelligence. *arXiv preprint arXiv:2309.11456*.

Gini, C. (1921). Measurement of inequality of incomes. *The economic journal 31*, 124–126.

Harding, J., W. D'Alessandro, N. Laskowski, and R. Long (2023). Ai language models cannot replace human research participants. *Ai & Society*, 1–3.

Harsanyi, J. C. (1961). On the rationality postulates underlying the theory of cooperative games. *Journal of Conflict Resolution 5*, 179–196.

Hayashi, T. (2016). Consistent updating of social welfare functions. *Social Choice and Welfare 46.3*, 569–608.

Heyman, T. and G. Heyman (2023). The impact of chatgpt on human data collection: A case study involving typicality norming data. *Behavior Research Methods*, 1–8.

Hong, H., J. Ding, and Y. Yao (2015). Individual social welfare preferences: An experimental study. *Journal of Behavioral and Experimental Economics 57*, 89–97.

Jackson, M. O. and L. Yariv (2015). Collective dynamic choice: the necessity of time inconsistency. *American Economic Journal: Microeconomics 7.4*, 150–178.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1986). Fairness and the assumptions of economics. *Journal of Business 59*, S285–S300.

Kim, J. and B. Lee (2023). Ai-augmented surveys: Leveraging large language models for opinion prediction in nationally representative surveys. *arXiv preprint arXiv:2305.09620*.

Kohavi, R. (1996). Census Income. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5GP7S.

Kolm, S.-C. (1977). Multidimensional egalitarianisms. *The Quarterly Journal of Economics 91*, 1–13.

Lasso de la Vega, C., A. Urrutia, and A. de Sarachu (2010). Characterizing multidimensional inequality measures which fulfil the pigou–dalton bundle principle. *Social Choice and Welfare 35.2*, 319–329.

Lee, H., S. Phatale, H. Mansoor, K. Lu, T. Mesnard, C. Bishop, V. Carbune, and A. Rastogi (2023). Rlaif: Scaling reinforcement learning from human feedback with ai feedback. *arXiv preprint arXiv:2309.00267*.

Maasoumi, E. (1986). The measurement and decomposition of multi-dimensional inequality. *Econometrica 54*, 991–997.

Millner, A. and G. Heal (2018). Time consistency and time invariance in collective intertemporal choice. *Journal of Economic Theory 176*, 158–169.

Miyagishima, K. (2022). Time consistent fair social choice. *Theoretical Economics Forthcoming*.

Motoki, F. Y. S., J. Monteiro, R. Malagueño, and V. Rodrigues (2023). Charting new territory: Using chatgpt to enhance survey instruments for management accounting research.

Nielsen, K. and J. Rehbeck (2022). When choice are mistakes. *American Economic Review 102.7*, 2237–2268.

Oprea, R. (2020). What makes a rule complex? *American Economic Review 110.2*, 3913–3951.

Oprea, R. (2023). Simplicity equivalents. *Working Paper*.

Ouyang, L., J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, et al. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems 35*, 27730–27744.

Park, J. S., J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein (2023). Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–22.

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

Polman, E. and K. Wu (2020). Decision making for others involving risk: A review and meta-analysis. *Journal of Economic Psychology 77*, 200–218.

Rafailov, R., A. Sharma, E. Mitchell, S. Ermon, C. D. Manning, and C. Finn (2023). Direct preference optimization: Your language model is secretly a reward model. *arXiv preprint arXiv:2305.18290*.

Santurkar, S., E. Durmus, F. Ladhak, C. Lee, P. Liang, and T. Hashimoto (2023). Whose opinions do language models reflect? *arXiv preprint arXiv:2303.17548*.

Savage, L. J. (1954). Foundations of statistics. *John Wiley & Sons*.

Shapira, E., O. Madmon, R. Reichart, and M. Tennenholtz (2024). Can large language models replace economic choice prediction labs? *arXiv preprint arXiv:2401.17435*.

Theil, H. (1967). Economics and information theory.

Thomson, W. (2011). Consistency and its converse: an introduction. *Review of Economic*

Design 15, 257–291.

Tjuatja, L., V. Chen, S. T. Wu, A. Talwalkar, and G. Neubig (2023). Do llms exhibit human-like response biases? a case study in survey design. *arXiv preprint arXiv:2311.04076*.

Traub, S., C. Seidl, and U. Schmidt (2009). An experimental study on individual choice, social welfare, and social preferences. *European Economic Review 53*, 385–400.

Tsui, K.-y. (1999). Multidimensional inequality and multidimensional generalized entropy measures: An axiomatic derivation. *Social Choice and Welfare 16*, 145–157.

Wakker, P. (1988). The algebraic versus the topological approach to additive representations. *Journal of Mathematical Psychology 32*, 421–435.

Wang, A., Y. Pruksachatkun, N. Nangia, A. Singh, J. Michael, F. Hill, O. Levy, and S. Bowman (2019). Superglue: A stickier benchmark for general-purpose language understanding systems. *Advances in neural information processing systems 32*.

Wolanski, A. and E. Baranov (2023). Consistent social choice. *Working paper*.

Zame, W. R., B. Tungodden, E. Sørensen, S. Kariv, and A. W. Cappelen (2020). Linking social and personal preferences: Theory and experiment. *Working Paper*.

Zuber, S. (2011). The aggregation of preferences: can we ignore the past? *Theory and decision 70.3*, 367–384.