UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Heidelberg University

Alfred Weber Institute of Economics

Dissertation:

The Context of Choice: Three Essays on the Malleability of Preferences, Cultural Norms, and Institutional Design

A Thesis Submitted to Obtain the Degree of:

Doctor of Philosophy (Ph.D.)

Presented by:

Ming Dai

Supervisor: Prof. Dr. Christiane Schwieren

January, 2026

Dissertation

Submitted to the
Alfred Weber Institute of Economics
Heidelberg University, Germany
for the degree of
Doctor of Philosophy (Ph.D.)

Presented by
Ming Dai

Supervisor
Prof. Dr. Christiane Schwieren

Subject
The Context of Choice: Three Essays on the Malleability of Preferences, Cultural Norms, and Institutional Design

---

Oral examination: January, 2026

Acknowledgements

Three years ago, I began my journey in Heidelberg and stepped into the vibrant, demanding life of a PhD candidate. Pursuing a doctorate has long been a cherished dream—driven by my passion for solving complex problems and a deep sense of family pride. Coming to Germany marked not just a new chapter, but a transformative experience of self-discovery.

To my beloved parents, thank you for your unwavering support. You encouraged me to explore the world beyond and have been the foundation of my progress. I am deeply grateful to my doctoral supervisor, Professor Christiane—a generous and passionate supervisor. She gave me the intellectual freedom to grow and, most importantly, the opportunity to begin my PhD at Heidelberg University. Her guidance has been a constant source of inspiration and clarity throughout my research.

To my colleagues, thank you for making collaboration both joyful and enlightening. I will always remember our summer evenings, sharing beers under golden sunsets—moments of true camaraderie. To my friends Yangbiao, Yao, and Yifan, fellow travelers from China and companions in Heidelberg: your presence brought warmth and comfort to this foreign land. Our conversations, journeys, and laughter have meant more than words can express.

This dissertation would not have been possible without the dedication of my collaborators. Professor Cui, whose scientific passion continues to inspire me, was the catalyst behind my Heidelberg research. The initial spark of our project came from an unexpected source—his own courtship story with his wife—which became the foundation of our work on signaling mechanisms. It's a reminder that deep scholarship can arise from life's most poetic moments.

Professor Jonathan, a researcher of rare wisdom, offered pivotal guidance through our many discussions. His insistence on highlighting the role of social information reshaped our approach and helped us address key limitations in existing models. From these scholars, I have learned not only methodological rigor, but the spirit of intellectual curiosity that drives meaningful research.

I also wish to thank Tamas, Leon, Sara, Vinicius, Mai, and Cosima—visionary collaborators whose sharp insights and generous feedback shaped this work. Beyond their academic brilliance, they brought energy and joy to my daily life. I especially cherish those amber summer afternoons with Tamas and Leon, where clinking steins carried both bold hypotheses and shared humanity.

My sincere thanks go to the administrative staff, especially Katharina, who helped me navigate many travel procedures with patience and care.

I would like to express my sincere gratitude to the China Scholarship Council for providing me with the scholarship, which has supported my doctoral studies and covered my living expenses during this stage.

The winds of Heidelberg whisper along the Neckar's banks and reach the hill of Würzburg, echoing the chimes of Milan's cathedral. My footsteps have traced Rome's winding streets, while Freiburg's summers linger like warm honey. The scent of Strasbourg's patisseries mingles with the solemnity of Gauss's grave in Göttingen, and Marx's home in Trier stands as a quiet witness to history. I wandered through Luxembourg's compact grandeur and watched Zaandam's windmills spin tales of centuries past. Amidst these journeys, I still cherish the memory of strolling with her through Heidelberg's bustling afternoons and sharing quiet dinners together, moments that gave rhythm and warmth to the passage of time.

"Long have I held lofty aspirations,
Now I return to Jinggang Mountain.
A thousand miles to seek this familiar land,
Its old face now wears a new look.
Orioles sing and swallows dance everywhere,
Streams murmur gently,
And high roads pierce the clouds.
Beyond Huangyangjie,
No need to fear the once-treacherous path.

Thunder roars,
Banners fly,
This is the world of man.
Thirty-eight years have passed,
A fleeting moment in time.
We may scale the heavens to seize the moon,
Or plunge into the seas to capture turtles,
And return in triumph with laughter.
Nothing in the world is insurmountable,
As long as we dare to climb."

Declaration and Statement on AI Tool Usage

The author affirms that the conceptualization, idea development, theoretical hypotheses, experimental design and implementation, data analysis, figure generation, and the substantive writing of this thesis were conducted independently without the assistance of any artificial intelligence (AI) tools.

AI was used solely for minor language refinement during the final editing stage, such as improving grammar, clarity, and fluency. Specifically, the author used Microsoft Copilot (MacOS version 26.0.1, October 2025), ChatGPT (GPT-4.5, via the web interface), and Gemini (Gemini 2.5 Pro, via the web interface). These tools were employed exclusively to polish sentence structure and enhance readability. No AI-generated content contributed to the research's intellectual substance, structure, or interpretation.

The author declares that the use of AI in this thesis does not constitute a conflict of interest with any individual, institution, or commercial entity. AI tool usage in this thesis adheres to ethical standards concerning transparency, data integrity, and responsible deployment. No personal, sensitive, or proprietary data were input into AI systems during the editing process. All interactions with AI tools were conducted in compliance with publicly available terms of use, and no content generated by these tools was accepted without author verification. The author retains full accountability for the accuracy, originality, and integrity of the final manuscript.

This declaration is made to ensure full transparency regarding the role of AI tools in the preparation of this thesis.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

Traditional economic models rest on a core assumption: individuals possess stable and exogenous preferences (Arrow, 2020; Samuelson, 1938b; Schildberg-Hörisch, 2018; Stigler and Becker, 1977). This assumption facilitates analytical tractability and predictive power within rational choice frameworks. However, it overlooks the complexity and context-dependence of real-world decision-making. Individuals do not act in isolation; their choices are embedded in specific social information, cultural norms, and institutional arrangements. These contextual factors shape cognitive processes, value formation, and the evaluation of alternatives.

This thesis asks: To what extent, and through which mechanisms, do social and institutional contexts systematically influence individual and group economic decisions? This question challenges the classical assumption of preference stability and connects to key themes in behavioral economics, social psychology, and institutional economics. Behavioral research shows that framing, defaults, and social norms can significantly alter choices (Giuliani et al., 2023; Kimbrough and Vostroknutov, 2016; Krupka and Weber, 2009; List, 2007; Thaler and Sunstein, 2021). Social psychology highlights the roles of peer pressure, identity, and social comparison (Festinger, 1954; Frey and Meier, 2004; Giletta et al., 2021; Haslam et al., 2009; Wu et al., 2023). Institutional economics emphasizes how legal rules, property rights, and governance structures shape incentives and constraints (Alt and Shepsle, 1998; Greif and Kingston, 2011; North, 1990; Udah and Ayara, 2014).

By empirically isolating the specific mechanisms through which social information, cultural norms, and institutional rules operate, this thesis contributes to a more nuanced economic framework. Rather than rejecting traditional models, it refines them by demonstrating how preferences and choices are conditional on their environment. This experimentally grounded approach situates behavior within concrete settings, offering precise insights into when and why preferences evolve. Ultimately, these findings help bridge the gap between abstract theory and observed behavior, providing micro-foundations for designing more effective, context-aware policies.

This thesis advances a central claim: context is not noise—it is a core determinant of economic decision-making. This view challenges the conventional economic paradigm that treats social information, cultural norms, and institutional mechanisms as extraneous factors, when in fact they shape preferences, influence behavior, and aggregate choices. Economic decisions do not occur in a vacuum; they are embedded in social structures and institutional environments. Ignoring these contextual variables means overlooking the generative forces behind behavior.

To systematically examine how context affects economic decision-making, the thesis develops a three-level analytical framework that progresses from the micro to the macro, from the individual to the institutional level:

1.1 The Malleability of Individual Preferences: The Measurement of Changes in Distributional Preferences

Chapter 2 begins at the micro level by challenging the classical assumption of stable preferences. Drawing on a lab experiment, it shows that individual preferences—regarding allocation, risk, and fairness—change systematically in response to social cues such as the maximum giving of all participants, group feedback, and normative prompts. These findings suggest that preferences are socially sensitive and dynamically formed through interaction. The results contribute to behavioral economics by highlighting the role of social influence and offer a foundation for modeling preference formation.

1.2 Cultural Variation in Social Norms: How Culture Shapes Normative Signaling and Cooperation in a Repeated Public Goods Game

Chapter 3 shifts focus from immediate social information to deeper cultural norms. Culture is not merely a backdrop but a generative structure that shapes behavioral expectations. Social norms across cultures influence cooperation, strategic choices, and conflict resolution within groups. Using cross-cultural experiments and comparative analysis, this chapter shows how culture functions as an implicit institution that regulates economic behavior. For example, norms emphasizing hierarchy and collectivism may foster coordination and compliance, while those stressing individual rights and competition may encourage strategic independence. This analysis underscores the cultural embeddedness of economic behavior and informs culturally sensitive institutional design.

1.3 Institutional Aggregation of Complex Decision-Making: Group Decision Rules and Intertemporal Risk Choices

Chapter 4 extends the analysis to formal institutional settings, examining how voting rules—specifically majority versus unanimity—aggregate diverse preferences in risky and intertemporal contexts. Through a laboratory experiment, the chapter establishes a causal link between institutional constraints and collective outcomes. Central findings reveal a context-dependent effect: while the unanimity rule induces a "cautious shift" in pure risk settings, it mitigates the discounting effect in intertemporal contexts, effectively preserving long-term value against delay. Exploratory analyses further suggest that these outcomes are driven by the asymmetric influence of risk-averse members and distinct coordination dynamics (median compromise vs. modal convergence). These results highlight the critical role of voting mechanisms in balancing risk-taking with the sensitivity to discounting.

Chapter 5 concludes with a discussion on avenues for future research. It gives an outlook on new directions and potential experiments on human behavior. Future research should prioritize cross-disciplinary methodologies to unravel these complexities, ultimately advancing predictive models of human behavior in evolving socio-technological landscapes.

Chapter 2 The Measurement of Changes in Distributional Preferences

Authors

Ming Dai, Chi Cui & Jonathan Alevy[1]

Abstract

In this chapter, we examine whether social information reshapes not only giving behavior but also the structure of individual distributional preferences. We employ a two-stage dictator game with varying budget lines to estimate CES utility parameters that capture trade-offs between selfishness and fairness (α) and between efficiency and equality (ρ). In the treatment group, participants observed the maximum donation made by others in the previous round; the control group received no such information. Our results show that social information significantly increases giving when the price of giving is low and induces systematic changes in CES parameters. Notably, some individuals switch preference type—from equality- to efficiency-oriented. These findings demonstrate the malleability of preferences and the endogenous, context-dependent nature of distributional norms.

Keywords

Distributional Preferences; GARP; Dictator Game; Maximum Information

[1] Ming Dai is a PhD candidate in economics at the Alfred Weber Institute of Economics, Heidelberg University, Germany. Chi Cui is an associate professor in economics at the School of Economics and Management, Northeast Normal University, China. Jonathan Alevy is an associate professor in economics at the College of Business and Public Policy, University of Alaska Anchorage, the United States.

2.1 Introduction

Distributional preferences—how individuals allocate resources between themselves and others—are central to behavioral and experimental economics (Duffy and Kornienko, 2010; Goldstein et al., 2008; Murphy and Banerjee, 2015; Schultz et al., 2002; Shang and Croson, 2009). While traditional models typically treat these preferences as stable and exogenous (Stigler and Becker, 1977), a growing literature shows that they are malleable—particularly sensitive to contextual factors such as social information (Dvorak and Fischbacher, 2024; Fehr and Charness, 2025; O'Garra et al., 2019; Potters and Xu, 2020). This raises a fundamental yet underexplored question: Does social information reshape not just observed behavior, but the underlying structure of distributional preferences?

From a theoretical perspective, if preferences are shaped by social cues and the surrounding environment rather than fixed, standard frameworks in welfare analysis, mechanism design, and behavioral prediction warrant re-examination. Addressing this gap calls for integrating concepts such as preference endogeneity and context-dependent utility (Bowles, 1998; Hausman, 2011; Kahneman and Tversky, 1984; Loewenstein et al., 2003; Pattanaik and Xu, 2015; Thomadsen et al., 2018; Tversky and Simonson, 1993). In practice, social norm interventions—such as ranked donation displays, neighbor-based energy reports, and normative messaging—are widely used in policy design and behavioral nudges (Allcott, 2011; Bicchieri, 2016; Martin and Randal, 2008; Reinstein and Riener, 2012; Xu et al., 2023). If such interventions alter behavior without changing underlying preferences, both their long-term effectiveness and their ethical justification come into question.

A substantial body of experiments documents that social information affects behavior in dictator games. Cues such as recipient identity (Eckel and Grossman, 1996), facial expressions (Weiß et al., 2021), donation rankings (Duffy and Kornienko, 2010), moral framing (Brañas-Garza, 2007), and peer behavior (Potters and Xu, 2020) often increase average giving. Yet responses vary widely: some individuals react positively (O'Garra and Sisco, 2020), while others show negative or negligible effects (Gächter et al., 2017; O'Garra et al., 2019). This heterogeneity underscores the need to understand the mechanisms—in particular, whether social information reshapes the underlying structure of distributional preferences.

Andreoni and Miller (2002) pioneered an approach to identifying distributional preferences by varying budget constraints in dictator games, revealing trade-offs between selfishness and fairness and between efficiency and equality. Their use of the Generalized Axiom of Revealed Preference (GARP) enabled internal-consistency checks, while utility-based parameterization captured individual heterogeneity. However, the design was static and excluded social information, limiting insights into dynamic preference change. In contrast, Fisman et al. (2015a) examined how the 2008 recession shifted Americans' preferences toward self-interest and efficiency, but their study lacked a control group and did not test for transitions across preference types.

In this chapter, we investigate how micro-level social information interventions dynamically influence distributional preferences. Specifically, we ask whether exposure to social information induces preference-type transitions—for example, from selfish to fair, or from efficiency- to equality-oriented. We implemented a two-stage dictator game: in Stage I, all participants made ten allocation decisions under varying budget constraints without informational cues; in Stage II, the treatment group viewed the previous round's maximum donation by all dictators, while the control group received no such information. Preferences were modeled using a constant elasticity of substitution (CES) utility function with two parameters, α and ρ, capturing trade-offs between selfishness and fairness, and between equality and efficiency, respectively.

Consistent with the findings of Andreoni and Miller (2002), dictator giving is responsive to its price. Further, the bulk of our subjects' behavior can be rationalized by the functional forms common to consumer theory. We elaborate on these results in Section 2.4.2. With respect to the introduction of social information, the experiment yielded three main findings. First, social information increased average giving when the cost of giving was low; at high costs, the effect sometimes reversed. Second, CES parameters shifted systematically following the intervention, indicating that the weight on distributional preference parameters was responsive to the social information treatment. Third—and most notably—we documented individual-level preference-type transitions: some participants moved from exhibiting selfish preferences to behavior consistent with perfect substitutes or Cobb–Douglas utility, suggesting a structural reshaping of preferences.

This chapter makes three contributions. First, it provides direct evidence that distributional preferences respond dynamically to social cues. We go beyond mean shifts in giving by identifying structural changes in CES utility parameters and tracking within-subject transitions across preference types. Second, we extend the static framework of Andreoni and Miller (2002) to a dynamic, within-subject design. Combining baseline elicitation with peer-based social information and varying the price of giving allows us to isolate behavioral and structural responses and jointly identify trade-offs between selfishness and fairness, and between efficiency and equality. Third, we uncover a novel price-contingent interaction: social cues increase generosity when giving is low-cost but dampen it when giving is high-cost. This pattern suggests that efficiency considerations can mediate responses to social stimuli, helping reconcile mixed findings in prior work (Fisman et al., 2015a; Potters and Xu, 2020). Rather than simply reducing selfishness, social influence can reorient preferences toward efficiency—sometimes at the expense of equality.

The remainder of the chapter is organized as follows: Section 2.2 describes the experimental design. Section 2.3 states the hypotheses. Section 2.4 presents the main results. Section 2.5 concludes with broader implications.

2.2 Experimental Design

We begin by outlining the overall structure of the experiment, followed by a detailed description of the treatment conditions. The design builds upon Andreoni and Miller (2002) and Fisman et al. (2007). A total of 218 participants were randomly recruited from the subject pools of Northeast Normal University and Jilin University in Changchun, China, and randomly assigned to the treatment and control conditions. The project received approval from the Ethics Review Committee at Northeast Normal University. The experiment was programmed and implemented using *z-Tree* (Fischbacher, 2007).

2.2.1 Main Experiment

At the beginning of the experiment, participants received written instructions and were given 15 minutes to review them and complete comprehension exercises.[2] To ensure clarity, experimenters also read the instructions aloud and answered questions individually. Participants were informed that one experimental token would be exchanged for 0.3 RMB (approximately €0.037).

During the experiment, participants allocated resources between themselves and an anonymous other. As shown in Figure 2.1, subject i chose allocations along the linear budget constraint between self and other. We denote *self* and *other* by s and o, with monetary payoffs π_s and π_o. The feasible payoff set is $\pi = (\pi_s, \pi_o)$. The interface presents a two-dimensional budget set characterized by the linear constraint $p_s\pi_s + p_o\pi_o = M$, with M the dictator's endowment and the relative price $p = \frac{p_o}{p_s}$. Thus, giving one additional token to the other costs the subject p tokens of own payoff. Participants selected their preferred allocation either by clicking on the budget line or by adjusting a slider to choose (π_s, π_o), with π_s retained by the self and π_o allocated to the other. After each choice, they clicked "Confirm Selection" to proceed to the next round

The game lasted 20 rounds, with budget lines varying in relative prices from 0.33 to 4 and intercepts adjusted to generate diverse choice sets (see Table 2.1). This table presents the budget parameters, including the X- and Y-intercepts, relative prices, and normalized budget constraints. To maintain engagement and avoid deterministic repetition, intercepts were randomly shifted by $\pm 5\%$ from the standard budget lines.

At the end of the experiment, participants completed a demographic questionnaire. They were then randomly paired, with one member of each pair designated as the decision-maker. One decision from the decision-maker's 20 rounds was randomly selected for payment to both members. All participants also received a fixed 10 RMB show-up fee (approximately €1.22).

[2] Further details regarding the experimental instructions and comprehension exercises are provided in the Appendix A.1.3.

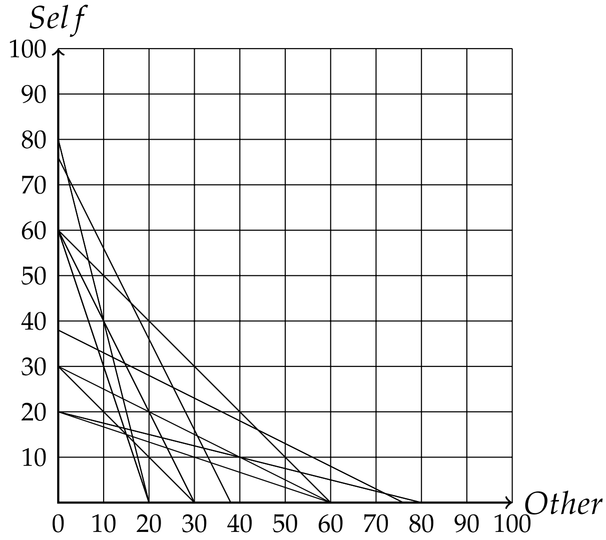


Figure 2.1: Standard Budget Lines

Table 2.1: Standard Normalized Budget Lines ($M = 1$)

Rounds	X-intercept	Y-intercept	Relative Price	Budget Functions
1	38	76	2	$1 = x/38 + y/76$
2	20	60	3	$1 = x/20 + y/60$
3	20	60	3	$1 = x/20 + y/60$
4	30	60	2	$1 = x/30 + y/60$
5	60	60	1	$1 = x/60 + y/60$
6	80	20	0.25	$1 = x/80 + y/20$
7	30	30	1	$1 = x/30 + y/30$
8	20	80	4	$1 = x/20 + y/80$
9	30	60	2	$1 = x/30 + y/60$
10	76	38	0.5	$1 = x/76 + y/38$
11	30	30	1	$1 = x/30 + y/30$
12	60	30	0.5	$1 = x/60 + y/30$
13	60	60	1	$1 = x/60 + y/60$
14	30	60	2	$1 = x/30 + y/60$
15	80	20	0.25	$1 = x/80 + y/20$
16	76	38	0.5	$1 = x/76 + y/38$
17	20	60	3	$1 = x/20 + y/60$
18	20	80	4	$1 = x/20 + y/80$
19	60	20	0.333	$1 = x/60 + y/20$
20	38	76	2	$1 = x/38 + y/76$

Note: The X-intercept and Y-intercept denote the maximum tokens allocable to the recipient and kept by the dictator, respectively. The budget-line slope reflects the relative price of giving.

2.2.2 Treatments

The experiment consists of two treatments: the Baseline Treatment (BT) and the Maximum Information Treatment (MIT) (see Figure 2.2). In BT, participants play a 20-round dictator game with varying allocation scenarios across rounds, receiving no additional information at any point. MIT uses the same sequence of budget lines as BT but incorporates a two-stage design. Stage I (Rounds 1-10) in MIT is identical to BT, with no information cues. In Stage II (Rounds 11-20), participants are shown the maximum donation made by any dictator in the preceding round.[3] For example, after completing $Decision_{10}$, MIT participants move to Stage II and, before making $Decision_{11}$, view the maximum giving observed in $Round_{10}$. This procedure repeats through $Round_{20}$. In Figure 2.2, light blue rectangles denote Stage I, while light orange (or green) rectangles denote Stage II.

Importantly, participants in MIT received identical instructions to those in BT. The information regarding maximum giving was introduced only after participants in MIT completed the first 10 rounds (Stage I). This design ensures that participants were unaware of the social information intervention during Stage I, thereby ruling out anticipatory reciprocity or strategic giving aimed at maximizing future payoffs.

In Stage I, participants make decisions ($Decision_i$, $i = 1, \ldots, 10$) based solely on the budget line and are shown only their own earnings. In Stage II ($Decision_j$, $j = 11, \ldots, 20$), participants first view the maximum amount allocated to others in the previous round ($Max.\ Giving_{j-1}$), then make their allocation decision based on the current budget line.

The two treatments could be summarized as follows:

i) ***Baseline Treatment (BT).***—Participants receive no information about the maximum giving by any dictator in either Stage I or Stage II.

ii) ***Maximum Information Treatment (MIT).***—Stage I mirrors BT, with no social information. In Stage II, participants are shown the maximum amount allocated to recipients by any dictator in the previous round before making each decision.

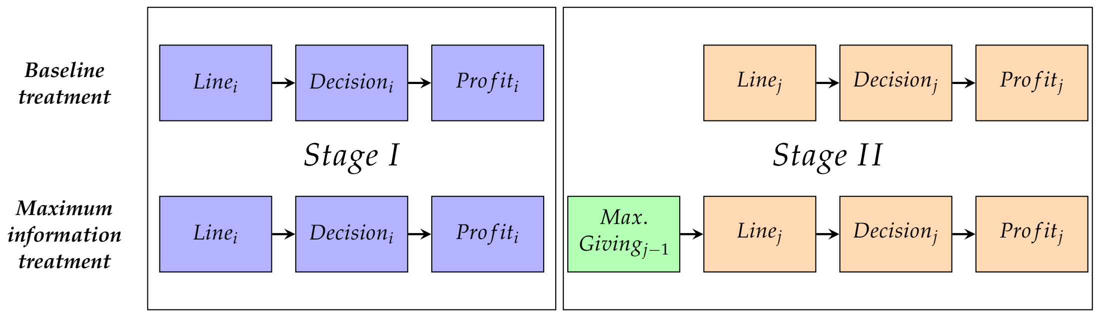


Figure 2.2: Treatments and Experimental Procedure

[3] Although BT participants receive no information in any round, we divide BT into the same two stages during analysis to isolate the causal effect of social information. The decision tasks in Stage I are repeated in Stage II.

By comparing participants' giving shares between Stage I and Stage II, we identify within-subject changes in allocation behavior attributable to the introduction of social information. A difference-in-differences measure between-subjects identifies the impact of social information controlling for learning and experience.

From the CES utility function estimates, we recover the parameters α and ρ, representing, respectively, the trade-off between selfishness and fairness (α) and the trade-off between efficiency and equality (ρ). Comparing these parameters in Stage II across BT and MIT allows us to assess how social information reshapes the underlying structure of distributional preferences.

2.3 Hypotheses

Building on the experimental design and theoretical framework, we develop a set of testable hypotheses addressing three dimensions: (i) behavioral responses to social information, (ii) structural shifts in distributional preferences, and (iii) mechanisms and boundary conditions. This framework enables us to assess not only whether social information affects giving, but whether the behavior is interpreted in the context of the underlying preference structure.

2.3.1 Behavioral Response Hypotheses

A large literature shows that social information about others' generosity increases prosocial behavior (Goldstein et al., 2008; Shang and Croson, 2009). Such information serves as a descriptive norm (Agerström et al., 2016; Cialdini et al., 1990; Goeschl et al., 2018), providing a behavioral reference point for what is commonly done in similar contexts.

Social comparison theory (Festinger, 1954; Frey and Meier, 2004; Samek et al., 2020) further posits that individuals are motivated to align their behavior with that of others, particularly when peers exhibit clearly prosocial actions. In a repeated game—such as our experimental design—this conformity motive becomes especially salient. In our study, the presence of information about the maximum giving made by dictators in the previous round heightens the perceived normative level of giving. However, the norm may not be prescriptive, since it is known that other participants gave less than the maximum amount. While the social information signals a higher behavioral standard, social comparison to the group may counteract that standard. In Stage I, both BT and MIT receive no social information, and thus we expect no significant difference in average giving between treatments. In Stage II, however, only participants in MIT receives social information (the maximum giving made by any dictator). We therefore expect higher average giving in MIT than in BT during Stage II, as well as an intra-individual increase in giving from Stage I to Stage II within MIT. In contrast, BT (no social information) should exhibit stable giving behavior across stages.

Taken together, these considerations lead to the following hypotheses:
Hypothesis 1a (Within-Subject Giving Response to Social Information):
Exposure to social information about maximum giving increases participants' giving in Stage II relative to Stage I within MIT.

Hypothesis 1b (Between-Subject Causal Effect of Social Information):
Participants in MIT will give more in Stage II than those in BT, indicating a causal effect of social information on giving.

Even if social information increases individuals' prosocial motivation, its translation into actual behavior is constrained by the cost structure of the environment. Andreoni and Miller (2002) show that variations in the price of giving significantly affect individuals' corner versus interior choices. When the cost of giving is low, prosocial behavior is more likely to occur; however, when the cost is high, individuals tend to shift toward self-interest or efficiency considerations, which can suppress the effect of social norms.

From a psychological perspective, violating a perceived social norm incurs internal costs, such as guilt or discomfort (Kimbrough and Vostroknutov, 2016). The presence of social information may reduce this psychological burden by legitimizing norm-conforming behavior and reinforcing one's willingness to act prosocially. However, whether this information leads to behavioral change depends not only on the strength of the norm, but also on the perceived feasibility of complying with it. If the monetary cost of giving exceeds the internal cost of norm violation, individuals may prefer to act selfishly, despite the resulting psychological discomfort.

We therefore hypothesize that the effect of social information is asymmetric: it enhances giving behavior when the cost of giving is low but may be weakened or even nullified when the cost is high.

Hypothesis 1c (Price Sensitivity Effect):
The positive effect of social information on giving is stronger when the relative price of giving is low. When the relative price of giving is high, the motivational effect of social information is attenuated or reversed.

2.3.2 Distributional Preference Hypotheses

To capture trade-offs in resource allocation, we follow Andreoni and Miller (2002) and Fisman et al. (2015a, 2007) and model distributional preferences with a CES utility function:

$$u_s(\pi_s, \pi_o) = \left[\alpha \cdot (\pi_s)^\rho + (1-\alpha) \cdot (\pi_o)^\rho\right]^{\frac{1}{\rho}}$$

The parameters $\alpha \in [0,1]$ and $\rho \in (-\infty, 1]$ are structural indicators of distributional preferences. α is the weight on own versus other's payoffs: higher α denotes more self-regarding preferences; lower α reflects greater concern for the other's payoff. ρ governs the elasticity of substitution between own and other payoffs ($\sigma = \frac{1}{1-\rho}$), capturing the extent to which individuals trade off equality for efficiency. As $\rho \to 1$, payoffs become highly substitutable (efficiency-oriented); as ρ decreases, they become more complementary, indicating stronger equality concerns.

We classify preferences by parameter values:

i) ***Perfect substitutes:*** $\rho = 1$, so $u_s(\pi_s, \pi_o) = \alpha\pi_s + (1-\alpha)\pi_o$. With $\alpha = 0.5$, $u_s(\pi_s, \pi_o) = 0.5\pi_s + 0.5\pi_o$, implying sum-maximization.
ii) ***Cobb-Douglas:*** $\rho \to 0$, yielding $u_s(\pi_s, \pi_o) = \pi_s^{\alpha}\pi_o^{1-\alpha}$.
iii) ***Leontief:*** $\rho \to -\infty$, yielding $u_s(\pi_s, \pi_o) = \min\{\pi_s, \pi_o\}$; in this limit, the weight α becomes irrelevant.
iv) ***Selfishness:*** $\alpha = 1$, so $u_s(\pi_s, \pi_o) = \pi_s$. For α close to 1, behavior is effectively selfish regardless of ρ.

In traditional economic models, preference parameters are assumed to be stable and exogenous. A substantial behavioral literature, however, shows that preferences are context-dependent and can evolve with changes in the environment or information (Bowles, 1998; Hausman, 2011; Kahneman et al., 1986; Loewenstein et al., 2003). Social information—behavioral cues about others—as well as contextual cues can shift perceptions of what constitutes a normatively appropriate allocation, prompting adjustments in decision principles (Agerström et al., 2016; Bicchieri, 2005, 2016; Goeschl et al., 2018; Krupka and Weber, 2009; List, 2007). Recent evidence aligns with this view: Fisman et al. (2015a) document that exposure to macroeconomic shocks increases emphasis on efficiency relative to equality, while Inukai et al. (2022) show that large-scale events such as the COVID-19 pandemic affect altruism and tolerance for inequality. Together, these findings underscore that distributional preferences are malleable and responsive to context.

In our experiment, subjects in MIT observe the previous round's maximum donation by all dictators. This signal functions both as a behavioral reference point and as an activation of prevailing social norms. Accordingly, participants may place less weight on own payoffs—lower α—and place greater emphasis on the recipient's payoff or total welfare—higher ρ. These adjustments should produce systematic shifts in allocation patterns between the no-information and information conditions.

In the Baseline Treatment (BT), no social information is provided in either Stage I or Stage II, implying no informational shock and, theoretically, no change in underlying preferences: $\alpha_1 \approx \alpha_2$ and $\rho_1 \approx \rho_2$.[4] By contrast, for MIT we expect parameter shifts consistent with our hypotheses. According to Hypothesis 1a, $\alpha_2 < \alpha_1$ (greater generosity) from Stage I to Stage II. As suggested by Hypothesis 1c, the cost of giving moderates responses to social information, implying $\rho_2 > \rho_1$, i.e., a stronger orientation toward efficiency in Stage II.

[4] The parameters (α_1, ρ_1) and (α_2, ρ_2) are estimated using the data from Stages I and II, respectively.

Hypothesis 2a (Stability of CES Parameters in BT:)
In the absence of social information (BT), participants' CES utility parameters (α and ρ) will remain stable between Stage I and Stage II.

Hypothesis 2b (Social Information and Parameter Shifts):
Participants in MIT will exhibit significant changes in CES parameters between stages. Specifically, α will decrease from Stage I to Stage II, indicating reduced weight on self-payoffs, while ρ will increase, reflecting a stronger preference for efficiency.

In a between-subjects comparison, neither BT nor MIT receives social information in Stage I, so we expect no significant differences in α or ρ across treatments. In Stage II, MIT receives social information, which should raise donation levels. Consequently, MIT's α_2 in Stage II should be significantly lower than BT's in Stage II, and, under low-cost conditions, MIT's ρ_2 in Stage II should be significantly greater than BT's in Stage II.

Hypothesis 2c (Between-Subject Stability of CES Parameters:)
In Stage I, where no information is provided, α and ρ do not differ significantly between BT and MIT. In Stage II, exposure to social information (MIT) leads to significantly lower α and higher ρ relative to the no-information condition (BT).

It is important to note that preference types are defined by specific ranges of α and ρ. Changes in either parameter do not necessarily indicate a discrete shift in type; they may simply reflect a directional movement toward greater fairness or efficiency. However, if individuals are highly sensitive to social information, the resulting shifts in α or ρ may be large enough to cross threshold values, producing a switch in preference type. As a result, the likelihood of preference-type switching is expected to be greater in MIT Stage II than in BT Stage II.

Hypothesis 2d (Type Transition Hypothesis):
A nontrivial proportion of participants will transition from one preference type (e.g., selfish) to another (e.g., fairness-oriented or efficiency-oriented) between Stage I and Stage II in MIT. The incidence of preference-type switching from Stage I to Stage II will be significantly higher in the MIT group than in the BT group.

2.3.3 Mechanism & Moderation Hypothesis

Beyond estimating average treatment effects, it is essential to examine why social information shapes behavior and under what conditions—or for whom—these effects are most pronounced. To address this, we advance one auxiliary hypothesis concerning the underlying mechanisms and potential moderators. This aim to clarify how social information operates through individual decision-making processes and the heterogeneous pathways by which it exerts influence.

One potential driver of behavioral change is endogenous learning through repeated exposure. In repeated decision tasks, individuals may gradually adjust their behavior as they become more familiar with the game's structure (Camerer et al., 2004; Charness and Rabin, 2002; Engle-Warnick and Slonim, 2006; Erev and Roth, 1998; Ioannou and Romero, 2014). In the dictator game, even without external stimuli, participants may display more consistent or fairer allocations over time simply due to greater familiarity with budget constraints and the decision environment (Andreoni and Miller, 2002; Fisman et al., 2007).

Alternatively, behavioral changes may result from exogenous informational inputs—specifically, social information that changes individuals' preferences or perceived normative obligations (Goeschl et al., 2018; Goldstein et al., 2008; Shang and Croson, 2009). In our two-stage dictator game, only participants in MIT receive such social information in Stage II: the maximum amount given by any dictator in the previous round. This provides a salient normative reference point (Bicchieri, 2016; Goeschl et al., 2018; Krupka and Weber, 2009), potentially updating beliefs about what is socially expected or desirable.

Accordingly, if we observe no significant difference in giving behavior between MIT and BT in Stage I (before social information is introduced) but find that MIT participants give significantly more than BT participants in Stage II, we may infer that social information—rather than mere learning or task repetition—is the primary driver of the behavioral shift.

Hypothesis 3 (Learning versus Information Hypothesis):
Any observed increase in giving between stages in BT is attributable to experience or learning, whereas any increase in MIT is driven by social information and learning.

2.4 Results

2.4.1 Statistics Overview

The experiment was conducted at the Laboratory of Economics and Management and the Key Laboratory for Applied Statistics, both at Northeast Normal University, over five semesters: Winter 2017, Summer and Winter 2018, Summer 2023, and Summer 2024.[5] A total of 218 participants were recruited as dictators from the subject pools of Northeast Normal University and Jilin University, with 98 assigned to BT and 120 to MIT.

Table 2.2 reports the demographic balance across treatment groups. Gender composition does not differ significantly: 41% of BT participants and 34% of MIT participants are male (Pearson's χ^2 test: $\chi^2(1) = 1.02$, $p > 0.1$). The majority identify as Han Chinese—87% in BT and 88% in MIT—with no significant difference between treatment groups (Pearson's χ^2 test: $\chi^2(1) = 0.03$, $p > 0.1$).[6] Educational attainment is similarly balanced: in BT, 86%

[5] Data collection was interrupted for three years due to the COVID-19 pandemic. In Section 2.4.2, we assess the impact of the pandemic on distributional preferences. We control for potential pandemic-related shifts to isolate the causal effect of social information.

[6] According to the Seventh National Population Census, Han Chinese account for 91% of China's

are undergraduates and 14% are graduate students (master's or PhD), compared with 90% and 10% in MIT (Pearson's χ^2 test: $\chi^2(1) = 0.94$, $p > 0.1$). Family background characteristics also align closely: in BT, 38% are only children, 42% have one sibling, and 20% have two or more siblings; in MIT, the corresponding shares are 33%, 45%, and 16%. These differences are statistically insignificant (Wilcoxon Mann-Whitney test: $z = 0.62$, $p > 0.1$).

Table 2.2: Treatment Balance

Treatments	BT		MIT		*p*-value
	Mean	S.D.	Mean	S.D.	
Female	0.59	0.49	0.66	0.48	0.31
Han Chinese	0.87	0.34	0.88	0.33	0.87
Undergraduate	0.86	0.35	0.9	0.30	0.33
Siblings	0.95	1.02	0.82	0.83	0.54
Observations	98		120		–

Note: *Female*, *Han Chinese*, and *Undergraduate* are coded as dummy variables. *Female* equals 1 for female participants, 0 otherwise. *Han Chinese* equals 1 for Han ethnicity, 0 otherwise. *Undergraduate* equals 1 for undergraduate students, 0 otherwise. *Sibling* denotes the number of siblings in a participant's family.
$Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

2.4.2 Treatment Effect of Social Information

Parameter test

This section tests whether social information about the maximum donation affects dictators' giving, within and between subjects. We use two-sided t-tests to conduct two analyses: (i) compare giving shares, ($\frac{\pi_o}{\pi_o+\pi_s}$), within BT across Stage I and Stage II to assess learning; and (ii) compare the stage-to-stage change across treatments to isolate the causal effect of social information from potential learning. For robustness, all analyses are conducted on both the full sample (all observations) and the rational subsample (CCEI$\geq$0.8).[7]

Figure 2.3 reports the average values of giving shares for the full sample and the rational subsample across treatments and stages. Within BT, average giving shares rise from 0.30 in Stage I to 0.32 in Stage II (Two-sided *t*-test: $t = -2.50$, $p < 0.01$), consistent with a learning effect; reciprocity may also contribute. In the rational subsample, BT shows a similar pattern (Stage I vs. Stage II: 0.28 vs. 0.38; Two-sided *t*-test: $t = -3.24$, $p < 0.01$). Taken together, these within-subject results support Hypothesis 3: increases in BT reflect experience or learning rather than social information.

population, making our sample broadly representative of the national demographic structure. See: https://www.gov.cn.

[7] We calculated the Critical Cost Efficiency Index (CCEI) Afriat (1967, 1972) to measure participants' consistency with the Generalized Axiom of Revealed Preference (GARP) (Samuelson, 1938a; Varian, 1982). We classify participants as behaviorally consistent with GARP—and rationality—if their CCEI exceeds 0.8. See Appendix A.1.1 for more information.

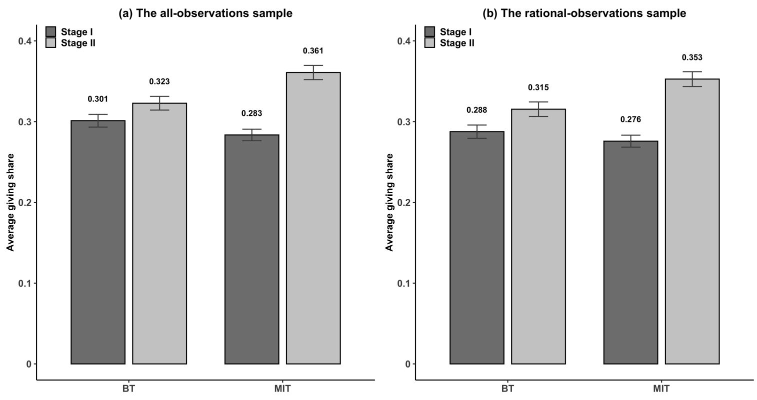


Figure 2.3: Average Giving Share: Full Sample and Rational Subsample

For MIT, in the full sample, the average giving share in Stage II is significantly greater than in Stage I (Two-sided t-test: $t = 6.82$, $p < 0.01$), reflecting the combined effects of learning and social information. To isolate the social-information effect, we implement a difference-in-differences comparison: the stage-to-stage change in giving is larger in MIT than in BT (Two-sided t-test: $t = 4.45$, $p < 0.01$). The rational subsample (CCEI$\geq$0.8) yields the same conclusion (Two-sided t-test: $t = 3.40$, $p < 0.01$). These results support Hypothesis 1a: exposure to social information raises giving shares in MIT Stage II relative to BT Stage II.

Figure 2.4 displays the distribution of giving shares across Stage I, Stage II, BT, and MIT for both the full sample and the rational subsample. Giving shares below 0.55 account for 75% of the observations. "Selfish" allocations, $\frac{\pi_o}{\pi_o+\pi_s} \leq 5\%$, comprise roughly 20%, while exactly equal splits account for 15%. These concentrations are consistent with focal distributional norms at zero giving and equal division.

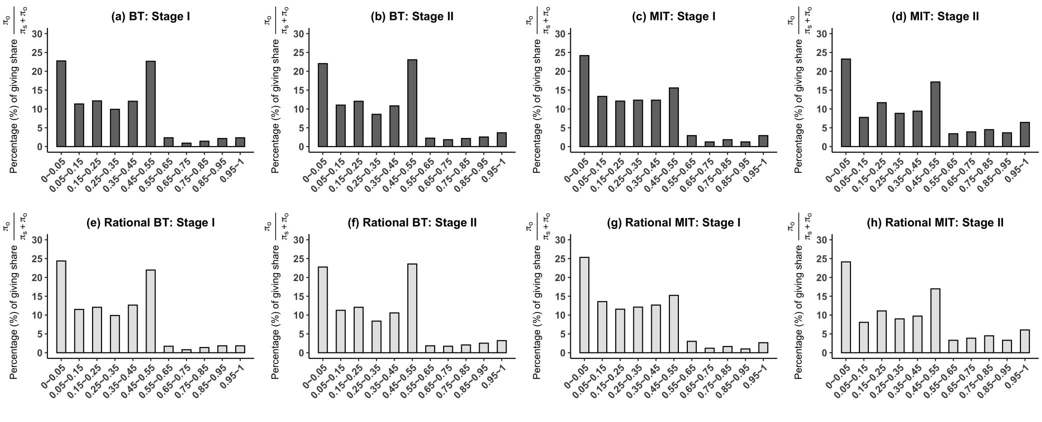


Figure 2.4: Distributions of Giving Shares: Full Sample and Rational Subsample

Panel Estimates on the Giving Share

Table 2.3 reports panel regressions of the giving share, $\frac{\pi_o}{\pi_o+\pi_s}$, defined as the fraction of tokens allocated to recipients. Each participant made 20 decisions; standard errors are clustered at the individual level. Column (1) shows that giving shares in Stage II are significantly greater than in Stage I across BT and MIT ($\beta_{Stage\ II} = 0.052$, $p < 0.01$), consistent with Section 2.4.3. Column (2), restricting to BT, finds a significant increase in Stage II ($\beta_{Stage\ II} = 0.022$, $p < 0.05$), which is indicative of learning or experience effects. The $MIT \times$ *Stage II* interaction is positive and significant ($\beta_{MIT \times Stage\ II} = 0.056$, $p < 0.01$), implying a greater increase in Stage II for MIT than for BT and, therefore, a causal effect of social information on generosity.

Table 2.3: Panel Estimates of the Giving Share ($\frac{\pi_o}{\pi_o+\pi_s}$)

Specifications	Dependent variable: Giving Share					
	Full Sample			Rational Subsample		
	(1)	(2)	(3)	(4)	(5)	(6)
MIT	0.010	-0.018	-0.017	0.013	-0.012	-0.010
	(0.021)	(0.020)	(0.020)	(0.023)	(0.022)	(0.021)
Stage II	0.052***	0.022**	0.022**	0.055***	0.028***	0.027***
	(0.007)	(0.009)	(0.009)	(0.007)	(0.009)	(0.009)
MIT×Stage II		0.056***	0.056***		0.049***	0.049***
		(0.012)	(0.012)		(0.013)	(0.013)
Constant	0.286***	0.301***	0.259***	0.274***	0.288***	0.251***
	(0.016)	(0.016)	(0.040)	(0.017)	(0.017)	(0.043)
Controls	No	No	Yes	No	No	Yes
Observations	4360	4360	4360	3920	3920	3920
Participants	218	218	218	196	196	196

Note: *MIT* and *Stage II* are dummy variables. Standard errors clustered at the subject level. Controls include gender, education, nationality, and the number of siblings.
$Significance levels indicate at * p<0.1, ** p<0.05, and *** p<0.01.

Results from the rational subsample (CCEI$\geq$0.8) closely mirror those from the full sample. Column (4) reports a significant coefficient of 0.055 ($p < 0.01$) for *Stage II*, comparable to the result in Column (1). In Column (5), the Stage II coefficient remains significant ($\beta_{Stage\ II} = 0.028$, $p < 0.01$), confirming a persistent learning effect. The $MIT \times$ *Stage II* interaction also remains significant ($\beta_{MIT \times Stage\ II} = 0.049$, $p < 0.01$), reinforcing the causal impact of social information. Together, social information robustly increases giving shares in both samples, even after accounting for deviations from rationality.

Given that our data span both pre- and post-pandemic periods, we examine whether the pandemic influences giving shares. As noted earlier, Stage II reveals a causal effect on giving shares, but the underlying source remains ambiguous. The observed increase may reflect social information, pandemic-related shifts, or both. If social information is the primary driver, the pandemic should show little correlation with giving. Alternatively, the two factors may interact, or the pandemic may dominate.

To assess the pandemic's impact, we regress the giving share on *MIT*, *Stage II*, *COVID-19*, and their interactions (see Table 2.4). Consistent with our empirical strategy, we report results for both the full sample (Columns (1)-(3)) and the rational subsample defined by CCEI$\geq$0.8 (Columns (4)–(6)).

Columns (1) and (2) estimate the social information effect separately for the pre- and post-pandemic periods. The $MIT \times Stage\ II$ interaction is positive and significant in the pre-pandemic sample ($\beta_{MIT \times Stage\ II} = 0.040$, $p < 0.05$), confirming that social information increases giving shares after accounting for learning.[8] In Column (3), the triple interaction $MIT \times Stage\ II \times COVID\text{-}19$ is not significant ($\beta_{MIT \times Stage\ II \times COVID\text{-}19} = 0.029$, $p > 0.1$), indicating that the magnitude of the social information effect does not differ across pandemic periods. In other words, the pandemic is not significantly correlated with giving shares, and the effect of social information remains dominant. Results from the rational subsample (CCEI$\geq$0.8) are consistent.

FINDING 1: ***The information on the maximum giving indeed results in a higher average giving share in MIT than in BT.***

The Price Effect on the Giving Share

This section examines how social information affects allocation decisions across budget lines with different relative prices (the cost of giving). First, we report average giving shares by price in Stage I and Stage II for BT and MIT and use two-sided *t*-tests to assess the price-giving relationship. Second, as a robustness check, we estimate panel regressions of the *Giving Share* on *Price*, *MIT*, *Stage II*, and their interactions.

Figure 2.5 presents average giving shares across prices in Stage I, Stage II, BT, and MIT. In the between-subject comparison, giving shares in Stage I are similar between BT and MIT across all price levels (see Figures 2.5(a)-(e)). In the within-subject comparison, MIT participants exhibit higher giving shares than those in BT at low prices in Stage II, while giving converges across treatments at higher prices (see Figures 2.5(b)-(f)). Table 2.5 reports two-sided *t*-test results comparing Stage II giving shares between treatments. At low prices (*Price*=0.25, 0.33, 0.5, and 1), MIT participants give significantly more than those in BT. At higher prices (*Price*=2 and 3), the difference is not significant, and at the highest price (*Price*=4), MIT participants are even more selfish than those in BT.

We next estimate panel regressions of the *Giving Share* on *Price*, *MIT*, and their interaction to test the robustness of the price effect. Tables 2.6 and 2.7 report results for the between-subject and within-subject comparisons, respectively.

Table 2.6 confirms a strong negative correlation between price and generosity: higher relative prices reduce the giving share. In Stage I, Columns (1)-(3) show this relationship is highly significant ($p < 0.01$). Column (3) suggests a steeper decline for MIT via the $MIT \times Price$ interaction, although this is not robust in the rational subsample (Column (9): $\beta_{MIT \times Price} = -0.019$, $p > 0.1$). In Stage II, price remains significantly negative (see Columns (4)-(6)).[9] In Column (6), the MIT coefficient is significant and positive ($\beta_{MIT} =$

[8] As discussed, a learning effect confounds the Stages I-II comparison within BT. The $MIT \times Stage\ II$ interaction isolates the social information effect by netting out round-based learning.

[9] Results are consistent in the rational subsample (Columns (10)-(12)) and remain robust to controls.

Table 2.4: Panel Estimates of the Giving Share ($\frac{\pi_o}{\pi_o+\pi_s}$) in BT, MIT, and pandemics

Specifications	Dependent variable: Giving share					
	Full Sample			Rational Subsample)		
	Pre-Cov19	Post-Cov19	Pooled	Pre-Cov19	Post-Cov19	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Without controls						
MIT	-0.033	0.0003	-0.033	-0.029	0.007	-0.029
	(0.031)	(0.278)	(0.020)	(0.031)	(0.030)	(0.031)
Stage II	0.029**	0.017	0.029**	0.029**	0.027**	0.029**
	(0.012)	(0.012)	(0.012)	(0.013)	(0.012)	(0.013)
COVID-19			0.00003			-0.011
			(0.032)			(0.034)
MIT×Stage II	0.040**	0.069***	0.040**	0.042**	0.057***	0.042**
	(0.018)	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)
Stage II×COVID-19			-0.012			-0.002
			(0.017)			(0.017)
MIT×COVID-19			0.034			0.037
			(0.041)			(0.043)
MIT×Stage II×COVID-19			0.029			0.015
			(0.024)			(0.025)
Constants	0.301***	0.301***	0.301***	0.294***	0.283***	0.294***
	(0.025)	(0.021)	(0.025)	(0.025)	(0.022)	(0.025)
Panel B: With controls						
MIT	-0.027	-0.003	-0.034	-0.026	0.002	-0.033
	(0.028)	(0.028)	(0.029)	(0.030)	(0.030)	(0.030)
Stage II	0.029**	0.017	0.029**	0.029**	0.027**	0.029**
	(0.012)	(0.012)	(0.012)	(0.013)	(0.012)	(0.013)
COVID-19			-0.004			-0.019
			(0.033)			(0.035)
MIT×Stage II	0.040**	0.069***	0.040**	0.042**	0.057***	0.042**
	(0.018)	(0.017)	(0.175)	(0.018)	(0.018)	(0.018)
Stage II×COVID-19			-0.012			-0.002
			(0.016)			(0.017)
MIT×COVID-19			0.033			0.040
			(0.041)			(0.043)
MIT×Stage II×COVID-19			0.029			0.015
			(0.024)			(0.025)
Constants	0.327***	0.319	0.345***	0.318***	-0.005	0.306***
	(0.079)	(0.223)	(0.053)	(0.088)	(0.043)	(0.056)
Observations	2000	2360	4360	1900	2020	3920
Participants	100	118	218	95	101	196

Note: *MIT*, *Stage II*, and *COVID-19* are indicator variables. Standard errors are clustered at the individual level. Columns (1) and (4) report results from the pre-pandemic sample, while Columns (2) and (5) use post-pandemic data. Columns (3) and (6) report results using the pooled data. Controls include gender, education, nationality, and the number of siblings.

$^{\$}$Significance levels indicate at * $p<0.1$, ** $p<0.05$, and *** $p<0.01$.

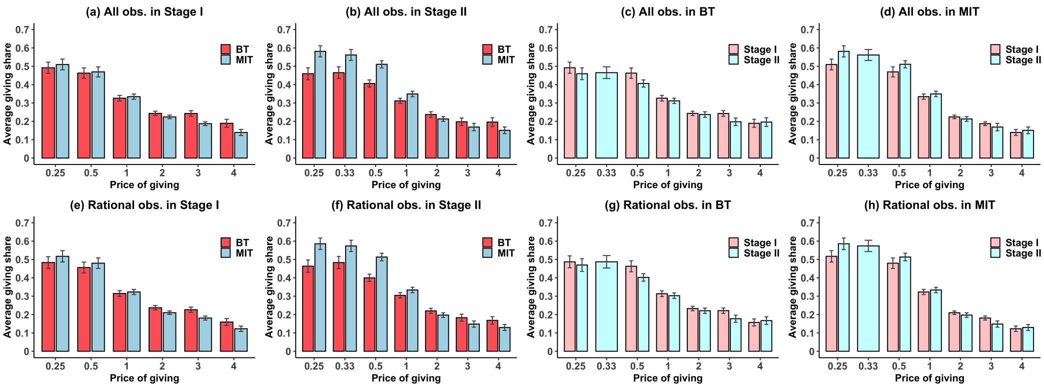


Figure 2.5: Average Giving Share by Price: Full Sample and Rational Subsample

Table 2.5: Two-Sided *t*-Tests of Giving Shares in Stage II: BT vs. MIT

Price	Full Sample			Rational Subsample		
	BT	MIT	Difference ($BT-MIT$)	BT	MIT	Difference ($BT-MIT$)
0.25	0.46	0.58	-0.12**	0.47	0.59	-0.12***
0.33	0.47	0.56	-0.09**	0.49	0.57	-0.08**
0.5	0.41	0.51	-0.10***	0.40	0.51	-0.11***
1	0.31	0.35	-0.04**	0.30	0.33	-0.03*
2	0.24	0.21	0.03	0.22	0.20	0.02
3	0.20	0.17	0.03	0.18	0.15	0.03
4	0.20	0.15	0.05*	0.17	0.13	0.04*

Note: Significance levels indicate at * $p<0.1$, ** $p<0.05$, and *** $p<0.01$.

0.106, $p < 0.01$), indicating that social information raises giving shares at the lowest price. The *MIT*×*Price* interaction is also significant and negative ($\beta_{MIT \times Price} = -0.047$, $p < 0.01$), implying that MIT participants' giving shares decline more sharply with price than those of BT participants. These findings are consistent with the earlier two-sided *t*-test results and support Hypothesis 1c.

Table 2.7 reports within-subject comparisons across stages. In BT, Columns (1) and (2) show a significantly negative price effect ($\beta_{Price} = -0.074$, $p < 0.01$), while the *Price*×*Stage II* interaction in Column (3) is insignificant, indicating no change in price sensitivity across stages. In contrast, MIT shows a significantly greater giving share in Stage II at the lowest price (Column (5): $\beta_{StageII} = 0.032$, $p < 0.01$). The *Price*×*Stage II* interaction in Column (6) is significant and negative ($\beta_{Price \times StageII} = -0.026$, $p < 0.01$), indicating greater sensitivity in Stage II. Results from the rational subsample match those from the full sample (see Columns (7)-(12)).

FINDING 2: ***The giving share is significantly negatively correlated with the relative price of giving in both BT and MIT. MIT exhibits greater sensitivity to price changes than BT.***

2.4.3 Parameters Estimates for the CES Utility Functions

In this section, we estimate individual parameters in a CES utility framework (Andreoni and Miller, 2002; Fisman et al., 2015a,b, 2007) using non-linear Tobit models. The CES utility function is defined as:

$$u_s\left(\pi_s, \pi_o\right) = \left[\alpha \cdot (\pi_s)^{\rho} + (1-\alpha) \cdot (\pi_o)^{\rho}\right]^{\frac{1}{\rho}}$$

where $\alpha \in [0, 1]$ measures the trade-off between selfishness and fairness—higher α indicates greater selfishness, while lower values reflect greater concern for fairness—and $\rho \in (-\infty, 1]$ measures the trade-off between efficiency and equality, with $\rho \to 1$ denoting efficiency orientation and lower values indicating stronger equality preferences.

The corresponding CES expenditure function is given by:

$$p_s \pi_s = \frac{g}{\left(\frac{p_s}{p_o}\right)^{\gamma} + g}$$

where $\gamma = \frac{\rho}{\rho - 1}$ and $g = \left(\frac{\alpha}{1-\alpha}\right)^{\frac{1}{\rho}}$. This generates the following individual-level specifications for each subject:

$$p^i_{s,n} \pi^i_{s,n} = \frac{g_n}{\left(\frac{p^i_{s,n}}{p^i_{o,n}}\right)^{r_n} + g_n} + \varepsilon^i_n$$

for $i = 1, ..., 20$, where $\varepsilon^i_n \sim N(0, \sigma^2_n)$. Because choices are censored at both ends of the

Table 2.6: Panel Estimates of the Giving Share ($\frac{\pi_o}{\pi_o+\pi_s}$) in Stages I and II

Specifications	Dependent variable: Giving Share											
	Full Sample						Rational Subsample					
	Stage I			Stage II			Stage I			Stage II		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Without controls												
Price	-0.086***	-0.086***	-0.073***	-0.101***	-0.101***	-0.075***	-0.091***	-0.091***	-0.080***	-0.110***	-0.110***	-0.085***
	(0.006)	(0.006)	(0.008)	(0.007)	(0.007)	(0.009)	(0.006)	(0.006)	(0.009)	(0.007)	(0.007)	(0.010)
MIT		-0.017	0.024		0.038	0.106***		-0.012	0.025		0.037	0.102***
		(0.020)	(0.032)		(0.023)	(0.035)		(0.022)	(0.035)		(0.025)	(0.039)
MIT×Price			-0.022*			-0.047***			-0.019			-0.044***
			(0.012)			(0.014)			(0.012)			(0.014)
Constants	0.452***	0.462***	0.439***	0.491***	0.470***	0.432***	0.452***	0.459***	0.438***	0.497***	0.476***	0.439***
	(0.016)	(0.019)	(0.022)	(0.018)	(0.022)	(0.025)	(0.017)	(0.021)	(0.025)	(0.019)	(0.024)	(0.028)
Panel B: With controls												
Price	-0.086***	-0.086***	-0.073***	-0.110***	-0.110***	-0.075***	-0.091***	-0.091***	-0.080***	-0.110***	-0.110***	-0.085***
	(0.006)	(0.006)	(0.008)	(0.007)	(0.007)	(0.010)	(0.006)	(0.006)	(0.009)	(0.007)	(0.007)	(0.010)
MIT		0.023	0.024		0.039*	0.107***		-0.013	0.024		0.037	0.103***
		(0.032)	(0.035)		(0.023)	(0.035)		(0.022)	(0.035)		(0.025)	(0.038)
MIT×Price			-0.022*			-0.047***			-0.019			-0.045***
			(0.012)			(0.014)			(0.012)			(0.014)
Constants	0.480***	0.490***	0.467***	0.540***	0.519***	0.482***	0.452***	0.458***	0.438***	0.508***	0.488***	0.451***
	(0.049)	(0.051)	(0.052)	(0.052)	(0.056)	(0.057)	(0.051)	(0.053)	(0.055)	(0.057)	(0.059)	(0.062)
Observations	2180	2180	2180	2180	2180	2180	1960	1960	1960	1960	1960	1960
Participants	218	218	218	218	218	218	196	196	196	196	196	196

Note: Standard errors are clustered at the subject level. Controls include gender, education, nationality, and the number of siblings.
[$]Significance levels indicate at * $p<0.1$, ** $p<0.05$, and *** $p<0.01$.

Table 2.7: Panel Estimates of the Giving Share ($\frac{\pi_o}{\pi_o+\pi_s}$) in BT and MIT

Specifications	Dependent variable: Giving Share											
	Full Sample						Rational Subsample					
	BT			MIT II			BT			MIT		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Without controls												
Price	-0.074***	-0.074***	-0.073***	-0.112***	-0.109***	-0.096***	-0.082***	-0.082***	-0.080***	-0.118***	-0.116***	-0.100***
	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Stage II		-0.009	-0.006		0.032***	0.076***		-0.007	0.001		0.029***	0.078***
		(0.008)	(0.013)		(0.008)	(0.014)		(0.007)	(0.014)		(0.008)	(0.014)
Price×Stage II			-0.001			-0.026***			-0.005			-0.029***
			(0.006)			(0.023)			(0.006)			(0.005)
Constants	0.435***	0.441***	0.439***	0.508***	0.489***	0.463***	0.439***	0.443***	0.438***	0.511***	0.493***	0.464***
	(0.023)	(0.023)	(0.022)	(0.023)	(0.022)	(0.023)	(0.026)	(0.026)	(0.025)	(0.025)	(0.024)	(0.024)
Panel B: With controls												
Price	-0.074***	-0.074***	-0.073***	-0.112***	-0.109***	-0.095***	-0.082***	-0.083***	-0.080***	-0.118***	-0.116***	-0.100***
	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Stage II		-0.009	-0.006		0.032***	0.076***		-0.007	0.001		0.029***	0.078***
		(0.008)	(0.013)		(0.008)	(0.014)		(0.007)	(0.014)		(0.008)	(0.014)
Price×Stage II			-0.002			-0.026***			-0.005			-0.029***
			(0.006)			(0.006)			(0.006)			(0.005)
Constants	0.411***	0.417***	0.415***	0.559***	0.539***	0.513***	0.364***	0.368***	0.363***	0.544***	0.527***	0.497***
	(0.084)	(0.083)	(0.084)	(0.051)	(0.052)	(0.053)	(0.089)	(0.089)	(0.089)	(0.055)	(0.055)	(0.056)
Observations	2180	2180	2180	2180	2180	2180	1960	1960	1960	1960	1960	1960
Participants	218	218	218	218	218	218	196	196	196	196	196	196

Note: Standard errors are clustered at the subject level. Controls include gender, education, nationality, and the number of siblings.
$Significance levels indicate at * $p<0.1$, ** $p<0.05$, and *** $p<0.01$.

budget constraint, we estimate γ and g for each subject by non-linear Tobit maximum likelihood, imposing $0 \leq p_s \pi_s \leq 1$ (Andreoni and Miller, 2002).

In the analysis that follows, we first compare the estimated CES parameters $\hat{\alpha}_n$ and $\hat{\rho}_n$ within subjects using the Wilcoxon Signed-rank test. We then compare parameters between subjects using the Wilcoxon Mann-Whitney test and finally test robustness with OLS and probit specifications.

Non-Parametric Test of CES Utility Parameters

Based on the non-linear Tobit estimates, Table 2.8 reports the average values of $\hat{\alpha}_n$ and $\hat{\rho}_n$ for Stage I and Stage II in BT and MIT. For BT (all observations), $\hat{\alpha}_n$ averages 0.69 in Stage I and 0.71 in Stage II, a difference that is not significant (Wilcoxon Signed-Rank test: $z = -0.53$, $p > 0.1$). MIT shows a similar pattern: no significant change in $\hat{\alpha}_n$ between stages (Wilcoxon Signed-Rank test: $z = -0.95$, $p > 0.1$).

Table 2.8: Mean Estimates of $\hat{\alpha}_n$ and $\hat{\rho}_n$

Treatments	Full Sample				Rational Subsample			
	Mean $\hat{\alpha}_n$		Mean $\hat{\rho}_n$		Mean $\hat{\alpha}_n$		Mean $\hat{\rho}_n$	
	Stage I	Stage II	Stage I	Stage II	Stage I	Stage II	Stage I	Stage II
BT	0.69	0.71	-3.64	-5.53	0.70	0.70	-3.75	-5.70
	(0.265)	(0.277)	(11.013)	(17.154)	(0.253)	(0.283)	(11.439)	(17.779)
MIT	0.67	0.68	-3.20	-1.64	0.67	0.69	-3.24	-1.58
	(0.244)	(0.199)	(11.975)	(8.589)	(0.237)	(0.192)	(12.364)	(8.744)

Note: Standard deviations, clustered at the subject level, are reported in parentheses.

Turning to $\hat{\rho}_n$, BT shows no significant stage-to-stage change (Wilcoxon Signed-Rank test: $z = -0.11$, $p > 0.1$). By contrast, in MIT, $\hat{\rho}_n$ is significantly higher in Stage II than in Stage I (Wilcoxon Signed-Rank: $z = -3.07$, $p < 0.01$), indicating that dictators exposed to social information become more efficiency-oriented (less equality-oriented). Results for the rational subsample mirror those for the full sample. Thus, MIT supports Hypothesis 2b; the absence of change in α does not.

Between-subject comparisons use Wilcoxon Mann-Whitney tests. In Stage I, BT and MIT do not differ in $\hat{\alpha}_n$ ($p > 0.1$). In Stage II, $\hat{\alpha}_n$ is marginally higher in MIT (Wilcoxon Mann-Whitney test: $z = 1.71$, $p < 0.1$), but this difference is insignificant in the rational subsample ($z = 1.19$, $p > 0.1$), indicating limited robustness. For $\hat{\rho}_n$, no significant difference between treatments is observed in Stage I ($z = -1.41$, $p > 0.1$), but in Stage II, MIT's value exceeds that of BT ($z = -2.12$, $p < 0.05$), consistent with a stronger efficiency orientation under social information. This result is robust in the rational subsample ($z = -1.90$, $p < 0.1$).

Overall, the difference in $\hat{\rho}_n$ in Stage II between treatments supports Hypothesis 2c: there is no treatment effect in Stage I, but $\hat{\rho}_n$ is significantly higher in Stage II for MIT than that for BT. Results for $\hat{\alpha}_n$ do not support Hypothesis 2c.

Regression Analysis of CES Utility Parameters

We regress $\hat{\alpha}_n$ and $\hat{\rho}_n$ to assess treatment effects on estimated preference parameters (see Tables 2.9-2.11). Following Fisman et al. (2015b), we adopt quantile regression to reduce sensitivity to extreme values, given the high skewness of $\hat{\rho}_n$ (see Table 2.10). We also estimate a probit model with an indicator variable equal to 1 if $\hat{\rho}_n \geq 0$, reflecting whether a subject is efficiency-oriented (see Table 2.11).

Table 2.9 reports OLS regressions of $\hat{\alpha}_n$ across Stage I, Stage II, BT, and MIT. Columns (1) and (2) show no significant correlation between $\hat{\alpha}_n$ and MIT in either stage. Columns (3) and (4) similarly find no significant difference between BT and MIT in Stage I or Stage II. Results for the rational subsample are consistent (see Columns (5)-(8)). Restricting to CCEI$\geq$0.8 and $\hat{\rho}_n \leq 1$, Column (9) shows no difference in Stage I, while Column (10) reports a significant decline in $\hat{\alpha}_n$ for MIT in Stage II ($\beta_{MIT} = -0.082$, $p < 0.01$), consistent with the earlier results of non-parametric tests. The contrast between Columns (9)-(12) and Columns (1)-(8) arises from excluding observations that violate the CES assumption ($\rho \leq 1$).[10] Under these stricter conditions, the purely social information effect emerges.

Table 2.10 presents quantile regressions of $\hat{\rho}_n$ at the 25th, 50th, and 75th percentiles, excluding $\hat{\rho}_n > 1$ to satisfy the GARP condition. Columns (1)-(2) and Columns (5)-(6) report results for Stage I. For Stage II, $\hat{\rho}_n$ is significantly and positively correlated with MIT at the 50th and 75th percentiles, indicating that dictators in MIT are more efficiency-oriented (less equality-oriented) than those in BT.

Beyond the quantile regressions, Table 2.11 reports probit estimates of the efficiency-orientation indicator ($\hat{\rho}_n > 0$). Columns (1) and (2) show no significant within-subject effects for BT or MIT. In Column (4), MIT participants in Stage II are 38.4% more likely to be efficiency-oriented than BT participants ($p < 0.05$), consistent with the quantile regression results. Findings for the rational subsample (CCEI$\geq$0.8) mirror those of the full sample (Columns (5)-(8)).

FINDING 3: ***Exposure to information on maximum giving leads dictators to place greater weight on efficiency and less on equality in their distributional preferences.***

2.4.4 Changes in Distributional Preferences

Table 2.12 reports estimated parameters from a CES utility function used to classify preference types in Stage II for BT and MIT. Several participants switch between types, indicating meaningful changes in distributional preferences. In BT, some transitions likely reflect instability rather than genuine shifts in preferences, which we interpret as learn-

[10] Fisman et al. (2007) used 50 budget lines to generate rich choice data; here, we used 10 per stage to limit participant fatigue, which may increase violations of model assumptions.

Table 2.9: OLS Estimates of $\hat{\alpha}_n$

Specifications	Dependent variable: $\hat{\alpha}_n$											
	Full Sample				Rational Subsample				Rational Subsample and $\hat{\rho}_n \leq 1$			
	Satge I	Satge II	BT	MIT	Satge I	Satge II	BT	MIT	Satge I	Satge II	BT	MIT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Without controls												
MIT	-0.019	-0.023			-0.033	-0.003			-0.024	-0.082***		
	(0.035)	(0.033)			(0.035)	(0.035)			(0.029)	(0.029)		
Stage II			0.016	0.014			-0.008	0.022			0.022	-0.013
			(0.039)	(0.029)			(0.041)	(0.029)			(0.031)	(0.026)
Constants	0.689***	0.706***	0.689***	0.670***	0.704***	0.697***	0.704***	0.671***	0.736***	0.761***	0.744***	0.703***
	(0.027)	(0.028)	(0.027)	(0.022)	(0.027)	(0.030)	(0.027)	(0.023)	(0.023)	(0.023)	(0.021)	(0.018)
R-squared	0.0015	0.0023	0.0009	0.0009	0.0044	0.0000	0.0002	0.0027	0.0040	0.0426	0.0032	0.0012
Panel B: With controls												
MIT	-0.023	-0.025			-0.036	-0.006			-0.032	-0.086***		
	(0.035)	(0.033)			(0.035)	(0.035)			(0.029)	(0.029)		
Stage II			0.016	0.014			-0.008	0.022			0.021	-0.014
			(0.039)	(0.029)			(0.040)	(0.029)			(0.029)	(0.026)
Constants	0.719***	0.754***	0.783***	0.700***	0.756***	0.795***	0.812***	0.744***	0.846***	0.868***	0.951***	0.775***
	(0.078)	(0.075)	(0.092)	(0.063)	(0.082)	(0.077)	(0.101)	(0.064)	(0.067)	(0.069)	(0.079)	(0.057)
R-squared	0.0107	0.0103	0.0287	0.0052	0.0205	0.0152	0.0339	0.0095	0.0531	0.0657	0.1098	0.0163
Observations	218	218	196	240	196	196	174	218	175	175	158	211

Note: Standard errors are clustered at the subject level. Controls include gender, education, nationality, and the number of siblings.
$Significance levels indicate at * $p<0.1$, ** $p<0.05$, and *** $p<0.01$.

Table 2.10: Quantile Estimates of $\hat{\rho}_n$

Specifications	Dependent variable: $\hat{\rho}_n$ ($\hat{\rho}_n \leq 1$)							
	Full Sample				Rational Subsample			
	Stage I (1)	Stage I (2)	Stage II (3)	Stage II (4)	Stage I (5)	Stage I (6)	Stage II (7)	Stage II (8)
25th percentile								
MIT	1.681	2.173*	0.839	0.615	0.639	1.424	0.474	0.672
	(1.672)	(1.266)	(0.709)	(1.368)	(1.669)	(1.325)	(0.862)	(1.289)
Constants	-2.640	-4.810**	-1.364**	-2.591*	-1.536	0.164	-0.978	-2.961
	(1.602)	(2.348)	(0.682)	(1.447)	(1.652)	(0.235)	(0.834)	(5.978)
50th percentile								
MIT	0.241	0.049	0.254	0.339**	0.267	0.164	0.168	0.291*
	(0.173)	(0.211)	(0.185)	(0.141)	(0.182)	(0.235)	(0.154)	(0.165)
Constants	-0.219	-0.378	-0.098	-0.618***	-0.188	-1.009*	-0.011	-0.569
	(0.205)	(0.371)	(0.093)	(0.206)	(0.150)	(0.570)	(0.099)	(0.377)
75th percentile								
MIT	0.001	0.154	0.221**	0.249***	0.006	0.149**	0.210*	0.252**
	(0.119)	(0.136)	(0.100)	(0.088)	(0.097)	(0.073)	(0.106)	(0101)
Constants	0.382***	-0.171	0.309***	-0.224	0.386***	-0.182	0.329***	-0.299
	(0.095)	(0.345)	(0.045)	(0.173)	(0.033)	(0.337)	(0.070)	(0.183)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	186	186	186	186	173	173	173	173

Note: All subjects here do not contain those observations with $\hat{\rho}_n > 1$. Controls include gender, education, nationality, and the number of siblings.
[$]Significance levels indicate at * p<0.1, ** p<0.05, and *** p<0.01.

Table 2.11: Probit Estimates of $\hat{\rho}_n$ ($\hat{\rho}_n \leq 1$)

Specifications	Dependent variable: Indicator for $\hat{\rho}_n \geq 0$							
	Full Sample				Rational Subsample			
	BT	MIT	Stage I	Stage II	BT	MIT	Stage I	Stage II
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Without controls								
Stage II	7.62e-17	0.141			4.76e-17	0.101		
	(0.202)	(0.172)			(0.209)	(0.179)		
MIT			0.243	0.384**			0.221	0.322*
			(0.187)	(0.188)			(0.194)	(0.194)
Constants	-0.208	0.035	-0.208	-0.208	-0.120	0.100	-0.120	-0.120
	(0.143)	(0.121)	(0.143)	(0.143)	(0.148)	(0.126)	(0.148)	(0.148)
R-squared	0.0000	0.0023	0.0066	0.0164	-0.0000	0.0012	0.0055	0.0116
Panel B: With controls								
Stage II	0.002	0.147			0.003	0.105		
	(0.207)	(0.175)			(0.213)	(0.183)		
MIT			0.297	0.462**			0.289	0.416**
			(0.190)	(0.192)			(0.198)	(0.201)
Constants	-1.223**	-0.700*	-0.884**	-1.235***	-1.052*	-0.954**	-0.864*	-1.514***
	(0.533)	(0.369)	(0.424)	(0.419)	(0.544)	(0.398)	(0.442)	(0.449)
R-squared	0.0609	0.0420	0.0436	0.0691	0.0525	0.0614	0.0481	0.0822
Observations	158	214	186	186	146	200	173	173

Note: Stage II and MIT are indicator variables. The sample excludes observations with $\hat{\rho}_n > 1$. Controls include gender, education, nationality, and the number of siblings.
$^{\$}$Significance levels indicate at * p<0.1, ** p<0.05, and *** p<0.01.

ing. The estimated elasticity of substitution for Weak Leontief preferences is $\sigma = -0.74$, indicating strong complementarity between π_s and π_o. For Weak Selfishness and Weak Perfect Substitutes, the elasticities are $\sigma = -2.63$ and $\sigma = -3.02$, respectively, suggesting both have relatively flat indifference curves, with Perfect Substitutes being slightly flatter (Andreoni and Miller, 2002).

Table 2.12: Estimates of Parameters (Standard Errors) for CES Utility Function for the Three Weak Types (Andreoni and Miller, 2002)

	Weak Selfish	Weak Leontief	Weak Perf. Subst.
$g = \left(\frac{\alpha}{1-\alpha}\right)^{\frac{1}{1-\rho}}$	20.18	1.60	2.54
	(5.586)	(0.081)	(0.311)
$\gamma = -\frac{\rho}{1-\rho}$	-1.64	0.26	-2.02
	(0.265)	(0.067)	(0.188)
α	0.758	0.654	0.576
ρ	0.621	-0.350	0.669
σ	-2.64	-0.74	-3.02
s.e.-self	0.222	0.179	0.244
	(0.011)	(0.009)	(0.014)
ln *likelihood*	-107.62	52.12	-69.58
Number of cases	380	230	242

In MIT, 29% of participants (30 out of 102) changed their distributional preference type between Stage I and Stage II, compared with 11% (8 out of 73) in BT.[11] These shifts in BT may reflect noise from repeated decisions rather than genuine preference changes, although learning is not ruled out. A one-sided Fisher's exact test confirms a significant treatment difference in preference shifts ($p < 0.01$), indicating that social information in MIT substantially increases the likelihood of type changes. This finding supports Hypothesis 2d.

Table 2.13 reports probit estimates of preference-type switching, controlling for individual characteristics. Column (1) indicates that MIT participants are 69% more likely to switch preference types than BT participants ($p < 0.01$). This effect remains statistically significant with the inclusion of additional controls, confirming the robustness of the results (see Column (2)). These findings align with earlier tests.

FINDING 4: ***In the individual-level analysis, MIT participants exhibit a significantly higher rate of preference switching from Stage I to Stage II than BT participants.***

Next, we examine the shifts in parameters $\hat{\alpha}$ and $\hat{\rho}$, which determine the direction of transitions between distributional preference types (see Figures 2.6 and 2.7). Figure 2.6 plots $\hat{\alpha}$ for participants with changed preferences, along with reference lines $y = x$, $y = 0.5$, and $x = 0.5$ to aid interpretation. Points above the $y = x$ line indicate an increase in $\hat{\alpha}$

[11] Appendix A.1.2 documents a shift in preferences from one type to another. Further details are provided therein.

Table 2.13: Probit Estimates of Preference Switching

Specifications	Dependent variable: Indicator for Preference Switching	
	(1)	(2)
MIT	0.687***	0.704***
	(0.236)	(0.246)
Constants	-1.229***	-0.277
	(0.196)	(0.461)
Controls	No	Yes
R-squared	0.0497	0.0844
Observations	175	175

Note: The dependent variable is a binary indicator equal to 1 if the subject switches preference types, and 0 otherwise. Controls include gender, education, nationality, and the number of siblings.
$Significance levels indicate at * p<0.1, ** p<0.05, and *** p<0.01.

from Stage I to Stage II; points below the line indicate a decrease. Among these participants, 63% (19 out of 30) show a decrease in $\hat{\alpha}$, suggesting a shift toward fairness over self-interest. The remaining 37% exhibit an increase, placing more weight on self-interest. The heterogeneity—where $\hat{\alpha}$ increases for some subjects but decreases for others—explains the insignificant treatment effect in Stage II. The opposing shifts offset one another, resulting in a negligible change in aggregate $\hat{\alpha}$.

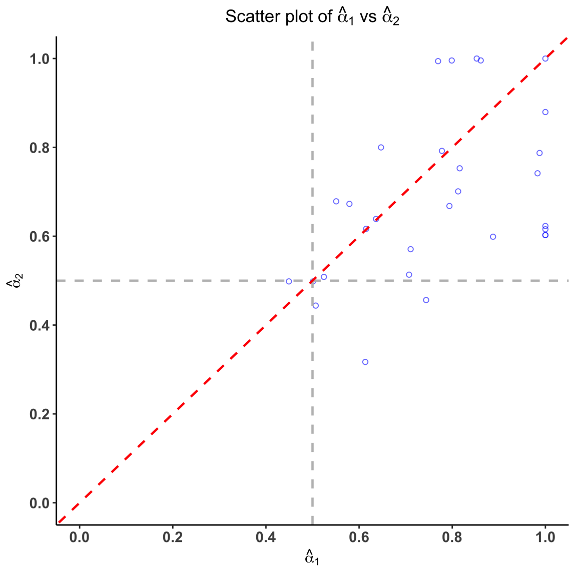


Figure 2.6: Scatterplot of Parameter $\hat{\alpha}$: Participants with Preference Changes in the MIT

Figure 2.7 shows the corresponding changes in $\hat{\rho}$, with reference lines $y = x$, $y = 0$, and $x = 0$. Specifically, 70% (21 out of 30) of participants show an increase in $\hat{\rho}$, suggesting a stronger focus on efficiency in Stage II. The remaining 30% show a decrease, indicating a shift toward equality. These patterns suggest that social information leads most participants to prioritize fairness in allocation while simultaneously increasing sensitivity to efficiency trade-offs.

To examine the relationship between changes in $\hat{\alpha}$ and $\hat{\rho}$, we estimate a probit model with

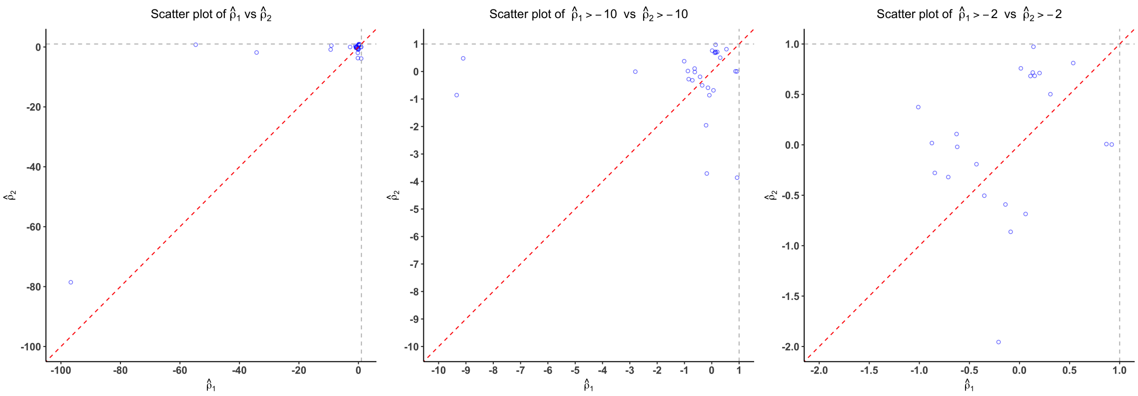


Figure 2.7: Scatterplot of Parameter $\hat{\rho}$: Participants with Preference Changes in the MIT

$\Delta\rho$ as the dependent variable (equal to 1 if $\hat{\rho}_2 - \hat{\rho}_1 > 0$, and 0 otherwise). The key independent variable, $\Delta\alpha$, equals 1 if $\hat{\alpha}_2 - \hat{\alpha}_1 > 0$, and 0 otherwise. Table 2.14, Column (1) reports a significant and negative correlation between $\Delta\alpha$ and $\Delta\rho$ ($\beta_{\triangle\alpha} = -1.601$, $p < 0.01$); the result remains robust with controls (Column (2): $\beta_{\triangle\alpha} = -1.571$, $p < 0.01$). This negative correlation suggests a tradeoff: participants who shift toward fairness tend to increase emphasis on efficiency, and vice versa. As illustrated in Figures 2.6 and 2.7, changes in utility parameters reflect a dual tendency—greater concern for fairness and heightened attention to efficiency—highlighting the nuanced impact of social information on distributional preferences.

FINDING 5: ***When preferences shift, the selfishness-fairness tradeoff substantially substitutes for the efficiency-equality tradeoff. After receiving information on others' maximum giving, participants place greater weight on both fairness and efficiency when navigating these trade-offs.***

Table 2.14: Probit Estimates of $\Delta\rho$

Specifications	Dependent variable: Indicator for $\Delta\rho > 0$	
	(1)	(2)
$\Delta\alpha$	-1.601***	-1.571***
	(0.566)	(0.593)
Constants	1.252***	1.216***
	(0.393)	(0.604)
Controls	No	Yes
R-squared	0.2577	0.2940
Observations	30	30

Note: $N = 30$; The dependent variable $\Delta\rho$ is equal to 1 when $\hat{\rho}_2 - \hat{\rho}_1 > 0$; otherwise it is equal to 0. The independent variable $\Delta\alpha$ is equal to 1 when $\hat{\alpha}_2 - \hat{\alpha}_1 > 0$; otherwise, its value is 0. Controls include gender, education, nationality, and number of siblings.
[$]Significance levels indicate at * p<0.1, ** p<0.05, and *** p<0.01.

2.5 Conclusion and Discussion

This chapter investigated whether social information affects not only giving in dictator games but also the deeper structure of distributional preferences. Building on the framework of Andreoni and Miller (2002), we implemented a two-stage design with varying budget lines and estimated CES utility parameters at the individual level. Our analysis produced three main findings. First, social information raises average giving shares, particularly when the relative cost of giving is low, but the effect weakens or even reverses under high-cost conditions. Second, the treatment group shows systematic parameter shifts across stages, while the control group remains stable. Third, we observed preference-type switching at the individual level, with some participants moving from selfishness toward fairness- or efficiency-oriented preferences.

Taken together, these results demonstrate that distributional preferences are more malleable than traditionally assumed. While prior studies established that social cues can shape behavior, our evidence extends this literature by showing that preference structures themselves can shift in response to information. The ability to detect such changes at the level of CES parameters underscores the value of structural estimation in experimental economics.

Although our design was not explicitly aimed at identifying causal mechanisms, several theoretical perspectives are consistent with the observed patterns. Social comparison may encourage individuals to align with perceived standards of generosity in the short run (Festinger, 1954; Krupka and Weber, 2009). Norm internalization suggests that repeated exposure to generous choices can embed prosocial expectations (Bicchieri, 2016). State-dependent preference theory further implies that contextual shocks such as social information may update the evaluative standards individuals apply to fairness and efficiency, resulting in persistent changes in preference structure. The moderating role of cost conditions indicates that these mechanisms are bounded by trade-offs between equity and efficiency.

These insights carry important implications for both theory and practice. They challenge the conventional economic assumption that preferences are stable primitives, highlighting the need for models that treat preferences as endogenous and context-dependent. For policymakers, these findings imply that social information interventions can be highly effective if they successfully reshape underlying preferences; however, such strategies may prove counterproductive in high-cost environments. Future research should explore alternative forms of social information (e.g., average versus maximum contributions), directly measure normative beliefs, and examine heterogeneity in responsiveness. By showing that social information can reshape both actions and preferences, this chapter offers a foundation for advancing behavioral economic theory and designing more effective interventions.

Chapter 3 Strategic Misrepresentation vs. Intrinsic Compliance: How Culture Shapes Normative Signaling and Cooperation in a Repeated Public Goods Game

Authors

Chi Cui, Ming Dai & Christiane Schwieren[1]

Abstract

In this chapter, we use a lab experiment to investigate how sending a signal of following a social norm impacts cooperative behavior in a repeated public goods game across cultures. Using a cross-cultural design with Chinese (collectivist) and German (individualist) samples, we implemented three treatments: a baseline treatment, an internalization-of-norm treatment, and a signaling treatment. We found that culture mediated the role of the normative signaling mechanism, and the German sample but not the Chinese sample exhibited strategic behavior (misrepresentation). The signaling mechanism had a heterogeneous effect on contributions across groups and cultures. Despite the misrepresentation, the signaling mechanism did not reduce aggregate cooperation but rather improved efficiency. Cultural differences in cheating behavior and norm-conformity behavior were evident. The Germans exhibited higher rate of cheating and responsiveness to signaling incentives, while the Chinese showed stronger intrinsic norm compliance. Our results validate low-cost normative signaling as an effective cooperation-enhancing mechanism while demonstrating culture's moderating role in norm conformity and cooperative behaviors.

Keywords

Cooperation; Signaling; Social Norms; Misrepresentation; Cross-Cultural Experiment

[1] Chi Cui is an associate professor in economics at the School of Economics and Management, Northeast Normal University, China; Ming Dai is a PhD candidate in economics at the Alfred Weber Institute of Economics, Heidelberg University, Germany; Christiane Schwieren is a full professor in economics at the Alfred Weber Institute of Economics, Heidelberg University, Germany.

3.1 Introduction

How to sustain cooperation has been considered one of the most important issues in collective action, together with solving the problem of free-riding behavior (Andreoni, 1995; Dawes et al., 1977; Fehr and Gächter, 2000; Isaac and Walker, 1988; Masclet et al., 2003; Yamagishi, 1986). To sustain cooperation among strangers, classical theory posits costly signaling as a credible commitment device for initiating cooperation (Bird and Power, 2015; Hawkes, 1991; Smith and Bird, 2000; Soler, 2012; Spence, 1974a). However, costly signaling might be unsustainable due to the fact that costs – especially in repeated interactions with strangers—might become prohibitively high (Ostrom, 1990). Social norms, in contrast, work as a kind of low-cost signal that everyone can afford (Posner, 2009), but allow for strategic misrepresentation, leading to a negative effect on cooperation. Thus, it is unclear whether a normative signal can help to maintain cooperation in a repeated game, despite the vulnerability to strategic misrepresentation and thus, the possibility for exploitative behavior.

Compliance with social norms depends on a group's moral identity (Aquino and Reed II, 2002) and the degree of norm internalization (Bicchieri, 2005; Gintis, 2003; Kimbrough and Vostroknutov, 2016). Groups with high moral standards perceive norm adherence as intrinsically virtuous and thus motivated conduct (Kohlberg, 1981). Even when recognizing potential gains from strategic behavior, they prioritize reinforcing the utility derived from norm conformity over profit maximization. The degree of intrinsically motivated norm conformity is correlated with culture, and thus cultural differences may mediate normative signaling (Almås et al., 2025; Cappelen et al., 2025; Croson and Buchan, 1999; Falk et al., 2018; Gächter et al., 2010; Gelfand, 2012; Gelfand et al., 2006, 2011; Lewis et al., 2009; Oosterbeek et al., 2004).

This study describes a cross-cultural experiment on cooperative signaling and strategic misrepresentation comparing a German and a Chinese sample. Based on models of cultural differences between China and Germany, we expect the two samples to differ both in intrinsic norm-following behavior and in (deceptive) signaling of norm-following behavior. Both Confucian tradition and socialist collectivist values prioritize communal obligations, leading to the expectation that the Chinese sample is more intrinsically norm-following and less strategic (Gunnthorsdottir et al., 2023; Heine and Ruby, 2010; Triandis, 2018; Xiao, 2021; Zhou et al., 2025). Recent literature, however, showing that Chinese samples show higher strategic intelligence (Mizrahi et al., 2020) than Western European samples, so predictions are not entirely clear. The German sample represents individualism, rooted in the Enlightenment, prioritizing personal autonomy. Thus, we would expect the individualistic sample to exhibit norm-breaking more often, weighing monetary costs and benefits of norm compliance. Where norm-compliance serves individual outcome maximization, it will be observed, while otherwise norm-breaking will be seen frequently.

We use a rule-following game in which participants can send a normative signal before playing a repeated public goods game, in a similar way as Kimbrough and Vostroknutov (2016). The matching principle in the repeated public goods game is based on the behavior in the rule-following game, thus allowing participants—at least in one treatment—to signal norm-following behavior. We varied whether participants were informed about the matching principle. In the internalization-of-norms treatment, participants were not informed about the matching principle for the public goods game, and norm-following behavior in the rule-following game could therefore not be strategic but should be based entirely on intrinsic motivation. In the so-called signaling treatment, in contrast, participants were aware of the matching principle and could therefore use the rule-following game to signal norm-following behavior and thus potentially cooperative behavior.

The signal could, however, also be used as strategic misrepresentation. Participants could send a norm-following signal to end up in a more cooperative group, where they could make higher profits by free-riding. We compare the two treatments to a baseline treatment where participants were randomly assigned to groups for the public goods game. By comparing the two treatments (signaling and internalization), we can distinguish between a signaling effect that can be affected by strategic misrepresentation and an effect of norm internalization on cooperation. The two different samples allow us to understand intercultural differences with respect to strategic behavior and intrinsic norm-following, which is important for intercultural economic interactions.

We obtained four main results from the experiment. First, we did find significant cultural differences with respect to the use of signaling. The Chinese sample did not strategically misrepresent norm-following behavior in the signaling treatment, while participants in the German sample often did misrepresent their norm-following tendencies. Second, we also found a cultural difference in cheating behavior and general norm conformity. The German sample showed more cheating behavior, while the Chinese sample showed more general rule-following behavior. Third, due to the possibility of strategically misrepresenting norm-following behavior, groups in the signaling treatment consisted of a mix of (intrinsic) rule-followers and rule-breakers, while in the setting where disguise was not possible, a split by norm-following tendency led to significant differences in cooperative behavior. Fourth, the German sample was more conditionally cooperative than the Chinese sample, which explains the difference in cooperative behavior between cultures.

Our study contributes to the literature on cooperation, norms, and cross-cultural behaviors. We experimentally validated Posner (2002)'s theory of low-cost social norm signaling and demonstrated its feasibility and effectiveness. Our findings suggest that even if there is a possibility for strategic misrepresentation, this did not offset the aggregate welfare gains from sending a normative signal. While Kimbrough and Vostroknutov (2016) demonstrated that rule-followers sustained cooperation, they did not find a feasible implementation mechanism. Our experiment institutionalizes social norm signals into a measurable, low-cost classification system and demonstrated that it is robust cross-cultural applicability.

The remainder of the study is organized as follows: Section 3.2 reviews the relevant literature. Section 3.3 presents the experimental design and procedure. Section 3.4 outlines hypotheses. Section 3.5 provides the experimental results, and Section 3.6 concludes and discusses.

3.2 Literature Review

A significant number of studies exist on the signaling model, mainly concerning the theoretical framework and its practical application in the labor market (Bedard, 2001; de Haan et al., 2011; Heller and Kessler, 2024; Heywood and Wei, 2004; Spence, 1974a, 1976, 2002; Waldman, 2016). The model was first proposed by Spence (1974a) to explain educational signaling in job matching in labor markets. The core function of the model lies in distinguishing between high-quality and low-quality signalers, thereby enabling the screening and matching of heterogeneous workers (Coles et al., 2013; Kübler et al., 2008; Spence, 1974a,b).

The application of Signaling Theory has expanded to include research on biological evolution and human social cooperation (Bird and Power, 2015; Gintis et al., 2001; McAndrew, 2021; Roberts, 1998; Soler, 2012; Wright, 1999; Zahavi, 1977). This body of work, also called Costly Signaling Theory, posits that signals imposing high costs can indicate the honesty of the signaler. In such scenarios, the cost of misrepresentation is high, thereby deterring low-quality signalers from misrepresentation attempts (Hawkes, 1991; Hill and Hurtado, 2017; Smith, 2004; Smith and Bird, 2000; Soler, 2012). Among the behaviors that can be costly signals are public philanthropy (Bereczkei et al., 2010; Haley and Fessler, 2005; Smith and Bird, 2000), risk-taking and heroism (Hawkes, 1991; Smith, 2004; Wilson and Daly, 1985), conspicuous consumption (Griskevicius et al., 2007; Kruger and Kruger, 2018; Saad, 2007), and religious commitment (Chvaja et al., 2019; Potz, 2022; Sosis, 2003).

Sustaining such costly signals in repeated games becomes very expensive, and if the cumulative cost of signaling exceeds the benefits of cooperation, it may ultimately lead to the unsustainability of cooperation (Stewart and Plotkin, 2014; Wang, 2021). On the other hand, the ability to pay for a signal is not evenly distributed among the community members. The exclusion of individuals with cooperative potential who cannot afford a signal reduces overall welfare and the efficiency of the cooperative system (BliegeBird and Smith, 2005; Gintis et al., 2001; Ostrom, 1990).

A low-cost alternative could be social norms. Following social norms communicates trustworthiness and a low discount rate to others, indicating that the signaler is a reliable cooperative partner (Eriksson, 2019; Posner, 2009; Sliwka, 2007). As Posner (2009) argues, following social norms serves as a positive signal, enabling receivers to more readily identify suitable partners and initiate cooperation. Nevertheless, low-cost normative signals may induce a misrepresentative strategy, wherein free riders engage in exaggerated conformity to social norms in order to mask themselves and be selected into cooperative settings, from which they can then profit by free riding. This behavior may decrease overall trust and cooperation and poses a challenge to the theoretical framework of low-cost normative signaling (Liu et al., 2024; Szolnoki and Perc, 2014; Wang et al., 2020). This study employs a novel experimental design to empirically validate Posner's theoretical framework and quantify the impact of a misrepresentative strategy on cooperation dynamics.

Another strand of literature that this study contributes to is the literature on cross-cultural differences in cooperation (Burton-Chellew and West, 2021; Gächter and Herrmann, 2009; Gächter et al., 2010; Spadaro et al., 2022; Wong and Hong, 2005) and norm-related behavior (Chen-Xia et al., 2023; Gelfand, 2019, 2012; Gelfand et al., 2011; Kimbrough et al., 2024; Mu et al., 2015; Oh, 2013). Social norms embody moral judgments, which are developed and internalized during childhood (Johnson, 1962; Nunner-Winkler, 2007). Individuals with high moral internalization of social norms perceive normative compliance as intrinsically justifiable, exhibiting resistance to modifying their conformity levels in response to potential opportunistic gains. Generally, research finds a positive correlation between the degree of norm conformity and cooperation (Fehr and Fischbacher, 2004; Kimbrough and Vostroknutov, 2016). Culture significantly moderates norm conformity (Morris et al., 2015). A strand of literature documents the difference in norm conformity between individualism and collectivism, with substantial evidence indicating a greater level of conformity to behavioral, ritualistic, and face-maintenance norms in collectivism (Abadeer, 2015; Jetten et al., 2002; McAuliffe et al., 2003; Saracevic et al., 2022; Ting-Toomey, 1988; Triandis, 2018). Similarly, "tightness-looseness" models posit that tight societies enforce stricter normative conformity and impose more severe sanctions for deviation than loose societies (Chua et al., 2019; Gelfand, 2019; Gelfand et al., 2006, 2011; Uz, 2015). There is, however, no clear agreement about these effects. Other studies, for example, documented that individualistic cultures exhibited greater norm conformity than collectivism with regards to emotional norms (Vishkin et al., 2023), foundations of morality (Graham et al., 2011), and legal injunctions (Tyler, 2006). This empirical discord underscores the persistent controversy regarding culture's moderating effect on normative compliance. To address this, we employed a cross-cultural experimental design with a Chinese sample (representing collectivism) and a German sample (representing individualism). The comparative design systematically examines cultural differences in social norm conformity using a rule-following game framework.

Based on the above discussion, we expect that culture may impact cooperative behavior by moderating the effectiveness of social norms as signals. Existing evidence indicates that tight cultures are more sensitive to normative signals and that cooperative behavior in such cultures relies more heavily on norm consistency; in contrast, loose cultures are more tolerant of norm violations, leading to weaker signaling effects (Gelfand et al., 2011). However, other studies show that descriptive norms are more effective at promoting cooperation in individualism, whereas injunctive norms are more influential in collectivism (Cialdini et al., 1990). On the other hand, culture can also directly affect cooperation rates. Since collectivism emphasizes conformity more than individualism, the level of cooperation is generally higher in collectivistic countries (Gunnthorsdottir et al., 2023). Conversely, Western societies (e.g., the U.S. and the U.K.) tend to show higher levels of cooperation among strangers than non-Western societies, even though they are generally thought of as individualistic (e.g., Morocco and Turkey) (Weber et al., 2023). Consequently, some studies suggest that individualism exhibits higher cooperation tendencies than collectivism (Cadsby et al., 2007; Koch and Koch, 2007). This study aims to explore, through a cross-cultural experimental design (China vs. Germany), whether social norms function effectively as low-cost signals across cultural contexts, and under what conditions this signaling effect may be disrupted by strategic misrepresentation. Finally, this study seeks to examine how culture moderates the relationship between the normative signaling mechanism and cooperation, thereby filling a gap at the intersection of signaling theory, cultural psychology, and cooperation research.

3.3 Experimental Design and Procedure

We begin with an overview of the main experiment's structure, followed by a detailed description of the treatments. The design allows us to disentangle the effects of norm internalization from those driven by strategic considerations.

3.3.1 Setting and Basic Structure of the Experiment

The experiment was conducted at the Economic Management Laboratory and the Key Laboratory for Applied Statistics of MOE, Northeast Normal University, during the winter semester of 2018, the summer semester of 2019, and the winter semester of 2021. The extended timeline reflects delays caused by the COVID-19 outbreak in December 2019. To examine cross-cultural differences, sessions were also run at the AWI Lab, Heidelberg University, during the summer and winter semesters of 2024 and the summer semester of 2025.

A total of 480 participants from China and Germany were randomly assigned to treatments within each country. Each session included 16 participants, lasted approximately 45 minutes, and was conducted entirely on computers using *z-Tree* (Fischbacher, 2007). The experimental exchange rate of one token was equal to 0.1 RMB (€0.012) for Chinese subjects and €0.07 for German subjects. Average earnings were 35.5 RMB (€4.2) in China, including a 15 RMB (€1.8) show-up fee, and €9.0 in Germany, including a €3.0 show-up fee.[2]

[2] From 2018 to 2024, the statutory minimum hourly wage in Changchun, Jilin Province, China averaged 19 RMB (€2.29). The average hourly payoff for participants in our Chinese sample exceeds this threshold. In 2024, the minimum hourly wage in Baden-Württemberg, Germany was €12.41.

The experiment consisted of three stages: (i) a dice game, (ii) a rule-following (RF) game, and (iii) a repeated public goods game. After completing the three games, participants filled out a brief demographic survey. Final payoffs equaled the sum of earnings from the three games.

Stage 1: The Dice Game

The dice game was designed to measure group-level cheating. Prior research finds a positive correlation between a group's honesty and its norm-conformity behavior (Amato et al., 2020; Gächter and Schulz, 2016; Gross and De Dreu, 2021). Honest individuals tend to act more pro-normatively, though norm-conformity does not necessarily imply honesty in this task.

Our design followed Fischbacher and Föllmi-Heusi (2013) and Gächter and Schulz (2016). Participants rolled a six-sided die twice in an opaque cup (see Figure 3.1) and reported the outcome of the first roll and its corresponding payoff via a computer interface. Rolls were private and unobservable to others. Payoffs were 5, 10, 15, 20, or 25 tokens for reported rolls of 1 through 5, and 0 tokens for a roll of 6 (see Table 3.1). Because people often wish to preserve an honest self-image, outright lying about a roll may be psychologically costly, whereas "rule-bending" may not. For example, a participant might report the higher of the two rolls rather than the first roll as instructed.[3] To balance earnings across the three games, we increased the dice-game payoffs to four times those in the design of Gächter and Schulz (2016). Participants were instructed to enter the first roll and its corresponding payoff, with final earnings from this stage equal to the reported payoff multiplied by 0.1 RMB or €0.07.

Table 3.1: Payoff Structure of the Dice Task

Number	1	2	3	4	5	6
Tokens	5	10	15	20	25	0

Stage 2: The Rule-Following Game

The rule-following (RF) game was designed to measure participants' willingness to comply with a simple environmental norm. Following Kimbrough and Vostroknutov (2016, 2018), we directly elicited norm-adherence in this task. As shown in Figure 3.2, each participant received 100 "rubbish balls" and could allocate them either to a trash can (*Box A*) or the ground (*Box B*). Participants were explicitly told the norm: *"People should put all rubbish balls into the trash can (Box A)."* This norm is straightforward, widely understood, and salient in both China and Germany, where environmental protection is a prominent

The effective hourly earnings in our experiment are close to this level, indicating that participant compensation aligns with minimum wage standards in both regions.

[3] In the design of Gächter and Schulz (2016), reporting the better of two rolls yields the "justified dishonesty" benchmark: claims of 6 should occur in $1/36 \approx 2.8\%$ of cases; claims of 1 in $3/36 \approx 8.3\%$; and claims of 2, 3, 4, and 5 in 13.9%, 19.4%, 25%, and 30% of cases, respectively.

Figure 3.1: Experimental Apparatus: Die and Cup

public policy.

Payoffs differed by choice: each ball in *Box A* earned 0.25 tokens, while each ball in *Box B* earned 0.4 tokens. Allocating all balls to *Box A* indicated complete rule-following; allocating all balls to *Box B* indicated complete rule-breaking. Total earnings equaled the sum of payoffs from both boxes. For example, subject i allocates X balls to *Box A* and $100 - X$ balls to *Box B*, with a sum of payoffs equal to $40 - 0.15X$ tokens. The task is artificial in that violations carry no penalties. Nonetheless, Kimbrough and Vostroknutov (2016, 2018) documented substantial rule-following behavior in a similarly structured environment. This allows us to measure participants' willingness to follow the rule in the absence of external enforcement (i.e., punishment or reward).

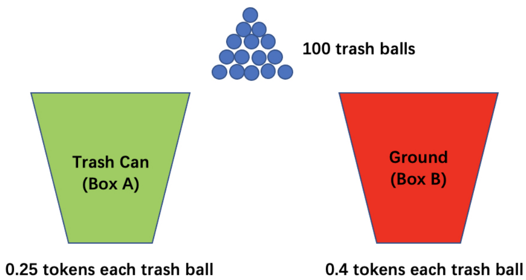


Figure 3.2: The Rule-Following Game

Stage 3: The Repeated Public Goods Game

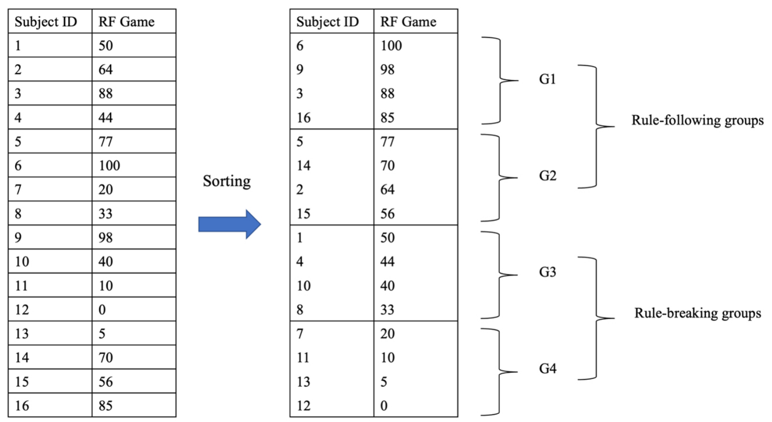
The repeated public goods game (PGG) measured cooperative behavior (Ledyard et al., 1994). In each of ten rounds, participants received an endowment of 20 tokens and were assigned to fixed groups of four. Each participant had a private account and a public account and decided how to allocate their tokens between the two. If a participant contributed X tokens to the public account and their three group members contributed Y, A, and B tokens, respectively, that a participant's payoff for the round was $20 - X + 0.4 \cdot (X + Y + A + B)$. After each round, participants were informed of their own earnings and the contributions of others in their group. Final earnings from this stage were equal to one payoff randomly selected from ten rounds, multiplied by the country-specific exchange rate.

3.3.2 Experimental Procedure

The experiment featured three treatments: the baseline treatment (BT), the internalization of norms treatment (INT), and the signaling treatment (ST) (see Figure 3.4). Participants were randomly assigned to treatments. Across 25 sessions, each with 16 participants, subjects were further assigned to four groups of four (G1, G2, G3, and G4) for the repeated PGG.

Baseline Treatment (BT): Before the dice game, experimenters explained that participants would roll a die in an opaque cup and demonstrated the procedure. To ensure exactly two rolls, the experimenter guided the process step-by-step: "When I say 'Start', you roll the die in the cup; when I say 'Stop', you stop rolling. Check the number and remember it. We will do it twice." After rolling twice, participants were not allowed to do it again. They then received instructions for the RF game, in which they allocated 100 rubbish balls between *Box A* (trash can) and *Box B* (ground). Finally, participants were randomly assigned to groups of four in the ten-round PGG, played with fixed group members. Thus, G1-G4 should not differ systematically with respect to cheating, rule-following, and cooperation.

Internalization of Norms Treatment (INT): The dice and RF games were identical to those in BT, but group assignments in the PGG were based on RF game behavior, following Kimbrough and Vostroknutov (2016, 2018). Participants were ranked by the number of balls placed in *Box A* and assigned to G1-G4 in descending order (see Figure 3.3). This grouping rule was not disclosed to participants. Groups remained fixed throughout the whole experiment.



Subject ID	RF Game
1	50
2	64
3	88
4	44
5	77
6	100
7	20
8	33
9	98
10	40
11	10
12	0
13	5
14	70
15	56
16	85

Subject ID	RF Game
6	100
9	98
3	88
16	85
5	77
14	70
2	64
15	56
1	50
4	44
10	40
8	33
7	20
11	10
13	5
12	0

Figure 3.3: An Example of the Group Assignment Procedure in the INT and ST

From Figure 3.3, G1 and G2 allocated more balls to *Box A* than G3 and G4 did, indicating a higher propensity to follow the norm. For simplicity, we refer to G1 and G2 as rule-following (RF) groups and G3 and G4 as rule-breaking (RB) groups. Participants were only told, *"All participants will be assigned to groups of four."* They were unaware of the relationship between the RF game and the PGG, eliminating incentives to manipulate RF game behavior. Thus, observed differences in cooperation reflect intrinsic norm internalization: the stronger the internalization, the more consistently participants followed the rule.

Signaling Treatment (ST): The dice game in ST was identical to that in BT and INT. After completing the dice game, participants were instructed to play an RF game, followed by a ten-round PGG. They were explicitly told that their group assignment in the PGG would be based on their behavior in the RF game, which was the same as the rule in INT. As stated in the RF game instructions, the relationship between the RF game and the repeated PGG was made explicit to participants as follows:

> *"This game relates to the next game. Your decision in this game will determine your assignment in the next game. In other words, your performance in distributing trash balls between Box A and Box B will determine with whom you will be matched in this game. For example, you will be assigned to groups with three other participants who have a number of trash balls put in Box A similar to yours."*

Thus, all participants understood the matching principle in the repeated PGG: they could be grouped with others whose RF game decisions were similar to theirs. Participants were motivated to send a pro-normative signal in order to be matched with "good types", such as potentially cooperative partners. The grouping rule was identical to that in INT, meaning that rule-followers (rule-breakers) would be matched with other rule-followers (rule-breakers). Consequently, participants had a clear motivation to increase norm-conformity in the RF game to improve their chances of being paired with higher-quality partners.

The three treatments could be summarized as follows:

- ***Baseline treatment (BT).***—Grouping in the PGG was random and unrelated to RF game behavior.
- ***Internalization of norms treatment (INT).***—Groups in the PGG were formed by ranking participants in descending order of RF game performance (sorted match), but this rule was not disclosed to participants
- ***Signaling treatment (ST).***—Groups in the PGG were formed using the same sorted-match rule as in INT, but participants were informed that they would be matched with others whose RF game behavior was similar to their own, allowing them to send a pro-normative signal.

By comparing PGG behavior in BT and INT, we are able to identify the causal effect of the grouping principle (random match vs. sorted match). By comparing INT and ST, we are able to isolate the pure signaling effect (sorted match without vs. with explicit signaling).

These treatments also enable cross-cultural comparisons between China and Germany. By comparing the internalization effect across countries, we test for differences in norm conformity and cooperation. Similarly, comparing the signaling effect across countries reveals whether cooperation under a signaling mechanism varies by culture.

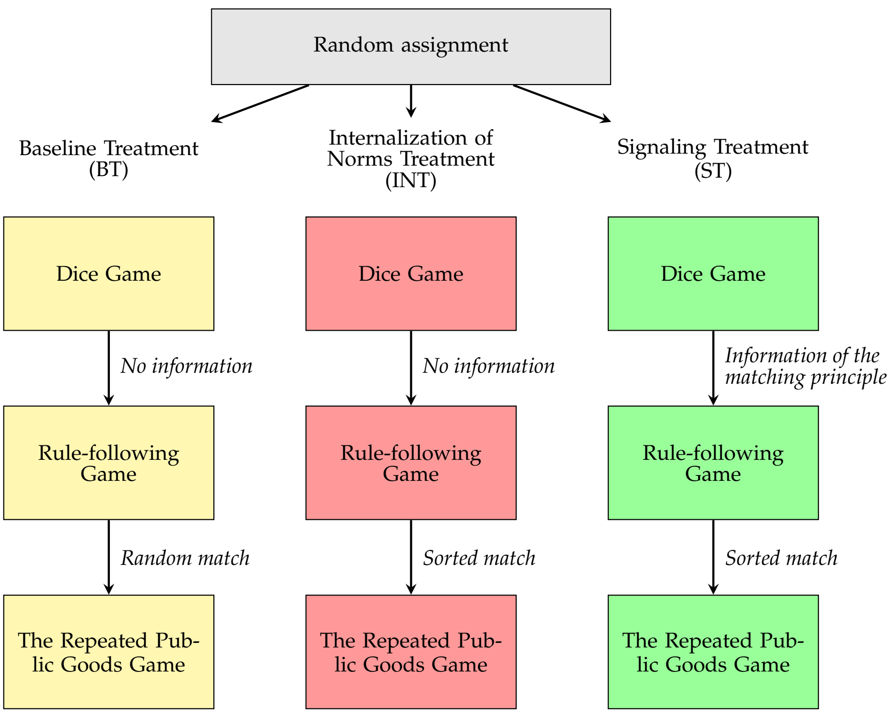


Figure 3.4: Overview of Experimental Treatments and Procedures

3.4 Hypotheses

Previous research suggests that norm-conforming behavior can predict the degree of dishonesty (Gächter and Schulz, 2016; Gross and De Dreu, 2021). Gross and De Dreu (2021) employed both a die-rolling task and a rule-following task to examine the relationship between individuals' cheating behavior and their tendency to follow rules. They found a significant negative correlation: participants who adhered more closely to the rule-following task reported lower average payoffs in the die-rolling task, indicating less dishonest reporting.

Similarly, Gächter and Schulz (2016) conducted cross-societal experiments in 23 countries, demonstrating a robust correlation between the prevalence of rule violations and intrinsic honesty. They constructed a *Prevalence of Rule Violations* (PRV) index using 2008 country-level data on corruption, tax evasion, and fraudulent politics. Their findings showed that intrinsic honesty was stronger among participants from low-PRV countries than among those from high-PRV countries.

Taken together, these studies indicate that rule-following is negatively correlated with dishonest behavior. Rule-followers may be more resistant to social pressures that encourage deceit, and through their displays of honesty, may signal that cheating is socially unacceptable. We formulate Hypothesis 1 as follows:

Hypothesis 1 (Rule-Following and Dishonesty):
Within the internalization-of-norms treatment, cheating behavior is negatively correlated with

rule-following behavior.

Posner (2009) argued that adherence to social norms can function as a signal that fosters cooperation. However, as Spence (1974a) noted, the signaling mechanism faces the problem that low-quality players may have an incentive to misrepresent their type when the cost of doing so is low. In the rule-following (RF) task, the cost of such misrepresentation is the foregone profit from choosing to follow the rule rather than break it. According to Kimbrough and Vostroknutov (2016), the utility structure of the RF game implies that violating a rule—especially when it yields a significant profit—generates psychological disutility. Intrinsic rule-followers, who have higher norm sensitivity than rule-breakers, therefore experience greater psychological costs from such violations.

When the matching principle is not disclosed, participants will reveal their intrinsic propensity to follow rules: rule-followers and rule-breakers act according to their preferences or profit-maximizing motives, without engaging in strategic misrepresentation. Under the signaling mechanism, by contrast, participants are informed about the matching principle in the repeated public goods game (PGG) and its relationship to the RF task. Rational participants may then adopt a misrepresentation strategy—pretending to follow the rule—to secure placement with more cooperative partners in the repeated PGG. For intrinsic rule-breakers, this can be a profitable tactic in subsequent interactions. Rule-followers, however, are less likely to alter their behavior, as their high baseline propensity to follow rules leaves little scope for further improvement.

Hypothesis 2 (Strategic Misrepresentation Effect):
Rule-breakers in the signaling treatment are expected to allocate more rubbish balls to the trash can than their counterparts in the internalization-of-norms treatment or the baseline treatment.

According to Hypothesis 1, in the internalization-of-norms treatment and the baseline treatment, rule-followers are expected to cheat less than rule-breakers. This implies that their self-reported payoffs in the die-rolling task should be lower than those of rule-breakers. In behavioral terms, rule-followers are close to full honesty, whereas rule-breakers are farther away from it and closer to justified dishonesty. The cheating patterns of the two groups (rule-following groups and rule-breaking groups) should therefore be clearly distinguishable.

From Hypothesis 2, we infer that when the signaling mechanism is introduced (signaling treatment), some rule-breakers will adopt a *misrepresentation strategy*—pretending to follow the rule—to gain entry into the rule-following group. Relative to the internalization-of-norms treatment, strategic misrepresentation alters the sorting mechanism of compliance types. Intrinsic rule-breakers secure a higher rank, thereby entering rule-following groups. This effectively crowds intrinsic rule-followers out into rule-breaking groups. This mixing of high-quality and low-quality players within the same group is expected to raise the cheating rate among those groups composed of self-identified rule-followers, shifting their behavior away from full honesty and toward justified dishonesty. For rule-breaking groups, by contrast, cheating behavior may move closer to complete honesty than otherwise expected, though it still remains closer to justified dishonesty. Overall, under the signaling treatment, the gap in cheating behavior between rule-following and rule-breaking groups should narrow relative to that observed in the internalization-of-norms treatment.

Hypothesis 3 (Convergence of Cheating Behavior Under Signaling):
In the signaling treatment, the cheating behavior of rule-following groups is expected to converge toward that of rule-breaking groups. Moreover, rule-following groups in the signaling treatment are more likely to engage in cheating than their counterparts in the internalization-of-norms treatment.

The internalization-of-norms treatment followed the exact design of Kimbrough and Vostroknutov (2016). In this setting, both rule-following and rule-breaking groups began the first period of the public goods game (PGG) with similar contribution levels, reflecting initial adherence to the PGG norm (Kimbrough and Vostroknutov, 2016). We expect that the difference in first-period contributions between the two groups is small and statistically insignificant because they do not have beliefs about others' RF behavior before playing the PGG.

Under the signaling treatment, by contrast, participants were informed that their group assignment in the repeated PGG would be based on the degree of rule following in the RF game. Participant *i* thus formed a belief about the types of others in his/her group, expecting them to have similar RF behavior and, by extension, similar first-period PGG contributions. Moreover, knowledge of the matching principle could trigger an *identity effect*, whereby perceived similarity strengthens cooperative intentions.

Consequently, in the signaling treatment, rule-following groups are expected to begin with relatively high first-period contributions, whereas rule-breaking groups might contribute little or even free-ride. The resulting gap in first-period contributions between rule-following and rule-breaking groups will be significantly greater than in the internalization-of-norms treatment.

Hypothesis 4 (Divergence of Initial Contributions Under Signaling):
In Period 1, the gap in public goods contributions between groups in the signaling treatment will be larger than the corresponding gap in the internalization-of-norms treatment.

In the internalization-of-norms treatment, Kimbrough and Vostroknutov (2016) found that rule-following groups sustained higher public goods (PG) contributions over time than rule-breaking groups, as the former adhered to the *norm of conditional cooperation*. We expect our internalization-of-norms treatment to produce results broadly consistent with theirs, but with one key difference in the contribution trajectory. In K&V's design, participants were not told the exact number of PGG rounds, which led to stable or non-decreasing contributions. In our design, participants knew they would play exactly 10 rounds, so we expect that contributions even in the rule-following groups will gradually decline over time.

When strategic misrepresentation of types occurs in the signaling treatment, the rule-following group will contain both intrinsic rule-followers and misrepresentative rule-breakers, while the rule-breaking group would similarly be mixed. In such cases, intrinsic rule-followers might continue contributing η in line with conditional cooperation, whereas intrinsic rule-breakers would tend toward selfish strategies. This mixing would likely reduce contributions in the rule-following group and increase them in the rule-breaking group, relative to the internalization-of-norms treatment.

If no strategic misrepresentation occurs, intrinsic rule-followers and rule-breakers will remain in their respective groups. In this scenario, the signaling treatment will resemble the internalization-of-norms treatment in group composition, but with an additional *information effect*. Rule-followers will still contribute η, but the identity/similarity cue could further boost their contributions. Conversely, rule-breakers might contribute less than in the internalization-of-norms treatment, anticipating that they are matched with lower-quality partners.

From this reasoning, we derive the following hypotheses:
Hypothesis 5a (null):
When strategic misrepresentation occurs, the rule-following group in the signaling treatment will contribute less than in the internalization-of-norms treatment, while the rule-breaking group will contribute more.

Hypothesis 5b (alternative):
If strategic misrepresentation does not occur, the rule-following group in the signaling treatment will contribute more than in the internalization-of-norms treatment, while the rule-breaking group will contribute less.

To explore the mechanisms through which culture moderates cooperation, we elicited beliefs regarding social norms and measured conditional cooperation in Chinese and German samples. This approach allows us to disentangle the sources of cross-cultural heterogeneity. Existing literature suggests that "tight" cultures are more sensitive to normative signals, making cooperative behavior heavily dependent on norm compliance. Conversely, "loose" cultures exhibit greater tolerance for norm violations, thereby attenuating the signaling effect (Gelfand et al., 2011; Herrmann et al., 2008). Specifically, injunctive norms exert a stronger influence in collectivist contexts (Cialdini et al., 1990).

Collectivism is characterized by interdependence and the prioritization of group goals (Markus and Kitayama, 1991; Oyserman et al., 2002; Triandis, 1995, 2001). In this cultural context, cooperation is not merely a personal choice but a moral obligation to maintain group harmony and "face". By contrast, individualistic cultures emphasize autonomy and individual rights (Gelfand et al., 2011). While valuing fairness, individualists view cooperation as a contractual act based on volition rather than a mandatory group directive. To avoid social exclusion or shame, collectivists tend to hold higher baseline expectations for cooperation (Gunnthorsdottir et al., 2023). For instance, Parks and Vu (1994) find that individuals from collectivist cultures exhibit higher initial cooperation rates in social dilemmas, driven by strong group-oriented norms.

Hypothesis 6a (Culture and Normative Beliefs):
Beliefs regarding cooperative norms are stronger in collectivist cultures than in individualistic cultures.

However, stronger “normative obligations” in collectivism do not imply stronger “conditional cooperation”. On the contrary, individualists typically exhibit stricter conditional cooperative behavior.

Interactions in individualistic societies often occur among strangers, relying on universal principles of fairness and high “general trust” (Allik and Realo, 2004; Fukuyama, 1996; Gheorghiu et al., 2009; Yamagishi and Yamagishi, 1994). Consequently, interactions are viewed as equity-based exchanges. If a counterpart defects, individualists swiftly withdraw cooperation to restore equity. This sensitivity to input-output equivalence leads to strong conditional cooperation (i.e., a high sensitivity to changes in others’ contributions).

Conversely, cooperation in collectivist societies is built on relationship-specific “assurance” (Yamagishi and Yamagishi, 1994). First, to maintain harmony, collectivists may tolerate minor defections rather than engage in immediate “tit-for-tat” retaliation, resulting in a flatter conditional cooperation slope (Koch and Koch, 2007; Leung, 1997; Ohbuchi et al., 1999). Second, in typical anonymous laboratory experiments lacking in-group cues, collectivist cooperation may decline rapidly due to a lack of assurance, manifesting as a lower baseline of trust (Buchan et al., 2002; Huff and Kelley, 2003; Yamagishi et al., 1998; Yamagishi and Yamagishi, 1994). For example, Koch and Koch (2007) compare subjects from the U.S. (individualist), Austria, and Japan (collectivist/mixed) and find that while conditional cooperation exists universally, it is most robust and prevalent among American subjects.

Hypothesis 6b (Cultural Moderation of Conditional Cooperation Intensity):
Individualists exhibit a stronger tendency toward conditional cooperation than collectivists.

3.5 Results

In this section, we analyze data from a total of 480 participants, comprising a Chinese sample and a German sample. In China, 368 participants were recruited from Northeast Normal University and randomly assigned to one of three treatments: 96 to the baseline treatment (BT), 144 to the internalization-of-norms treatment (INT), and 128 to the signaling treatment (ST). In Germany, 112 participants were recruited from Heidelberg University and randomly assigned to one of three treatments: 48 to BT, 32 to INT, and 32 to ST.

We analyze the data with respect to the hypotheses by addressing two key analytical goals. First, we identify the causal effect of the sorted match by comparing the Internalization-of-Norms (INT) and Baseline (BT) Treatments, and the causal effect of the signaling mechanism by comparing the Signaling (ST) and Internalization-of-Norms (INT) Treatments. Second, we examine cross-cultural differences in behavioral outcomes between the Chinese and German samples across these treatments.

3.5.1 Cheating Behaviors in the Dice Game

Difference in Cheating Behaviors between Treatments and Countries

The dice game was conducted first for all participants and thus was unaffected by either the RF task or the repeated public goods game (PGG). Consequently, no systematic differences in cheating behavior across treatments are expected within each sample. Figures 3.5(a)-(e) display the distributions of self-reported payoffs in the dice game for the baseline (BT), internalization-of-norms (INT), and signaling (ST) treatments in both the Chinese and German samples. Possible self-reported payoffs were 0, 5, 10, 15, 20, and 25. We used Pearson's χ^2 tests to confirm that these differences are not statistically significant ($p > 0.1$). Cheating behavior was comparable across treatments, confirming that the random assignment was successful.

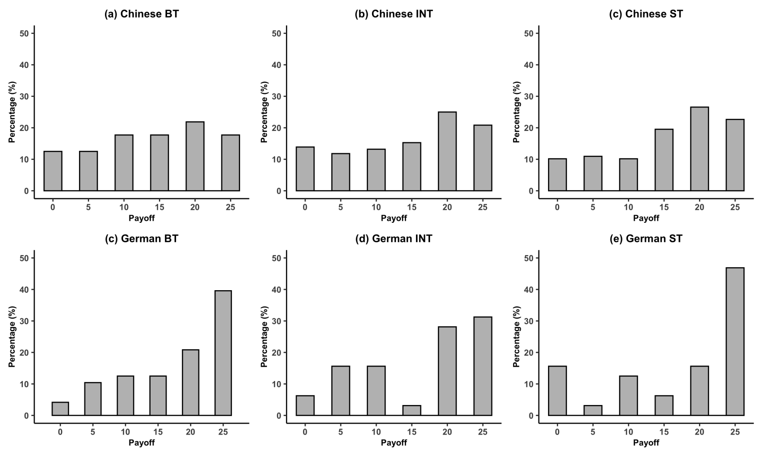


Figure 3.5: Distribution of Self-Reported Payoffs in the Dice Game by Treatment

The dataset also allows a cross-cultural comparison of cheating behavior. Visual inspection of the figures indicates that the German sample's payoff distribution is more right-skewed than that of the Chinese sample. Pearson's χ^2 tests support this observation: the German sample reports significantly higher payoffs than the Chinese sample ($p < 0.1$), implying a greater propensity to cheat among German participants.

Testing the Relationship between Cheating Behavior and Rule-Following Behavior

In this section, we test Hypothesis 1, which predicts a negative correlation between cheating behavior and rule-following behavior (Amato et al., 2020; Gächter and Schulz, 2016; Gross and De Dreu, 2021). Specifically, self-reported payoffs in the dice game are expected to be negatively correlated with the number of rubbish balls allocated to the trash can (*Box A*).

Figure 3.6 displays the distribution of self-reported payoffs in the dice game for the Chinese and German samples across treatments. We used Spearman's rank correlation to test the relationship between cheating behavior (self-reported payoff) and rule-following behavior (number of rubbish balls in the trash can) in INT. For the Chinese sample, the correlation is negative but statistically insignificant (Spearman's $\gamma = -0.10$, $p > 0.1$).[4] For the German sample, the correlation is also insignificant in INT (Spearman's $r = -0.20$, $p > 0.1$) and BT (Spearman's $r = -0.19$, $p > 0.1$). Overall, in both cultural contexts, cheating behavior does not significantly predict rule-following behavior—contrary to Hypothesis 1, which anticipated a negative correlation (Gächter and Schulz, 2016; Gross and De Dreu, 2021).

FINDING 1: ***There was no significant correlation between cheating behavior and rule-following behavior in either the Chinese or German samples.***

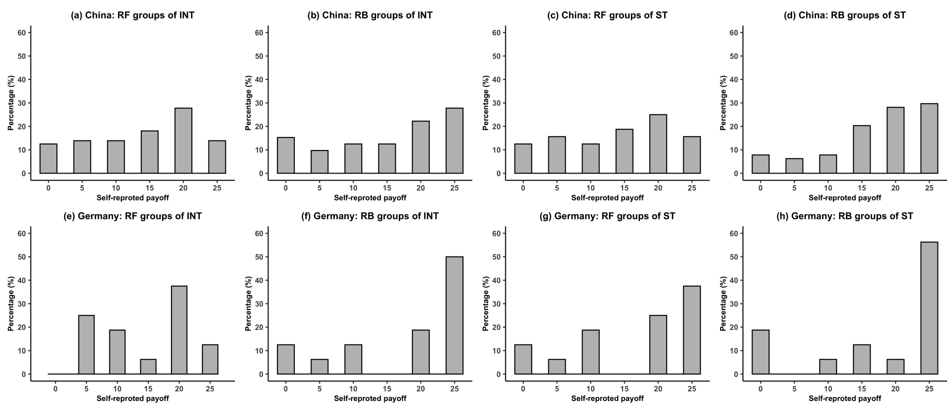


Figure 3.6: Distributions of Self-Reported Payoffs: Rule-Following vs. Rule-Breaking Groups in the INT and ST

3.5.2 Behaviors in the Rule-Following Task

The average number of rubbish balls allocated to *Box A* serves as a measure of participants' propensity to follow norms related to environmental protection (i.e., norm-conformity behavior). Figures 3.7(a)-(b) display the averages by treatment and by rule-

[4] For the baseline treatment (BT), the correlation is also insignificant: Spearman's $\gamma = -0.07$, $p > 0.1$, indicating that cheating behavior does not predict rule-following behavior in the Chinese sample.

following (RF) and rule-breaking (RB) groups.

In this section, we first use two-tailed t-tests to examine the causal effect of the signaling mechanism and to compare norm-conforming behavior across cultures. We then estimate Ordinary Least Squares (OLS) regressions on the RF-task performance to verify the robustness of these results.

Causal Effect and Testing Misrepresentation

Figures 3.7(a)-(b) show the rule-following behavior of the Chinese and German samples across treatments, respectively. For the Chinese sample, the average number of rubbish balls allocated to *Box A* was 53 in BT, 51 in INT, and 52 in ST (see Figure 3.7(a)). Two-tailed t-tests indicate no statistically significant differences in rule-following behavior between treatments ($p > 0.1$), and in particular, ST does not differ significantly from INT. For the German sample, the corresponding averages were 25 in BT, 35 in INT, and 47 in ST. Although the ST mean exceeds that of INT, the difference is not statistically significant ($p > 0.1$).

We further compared rule-following behavior between ST and INT separately for rule-following (RF) and rule-breaking (RB) groups (see Figures 3.7(c)-(d)). In the Chinese sample, the RF groups averaged 80 rubbish balls in INT and 81.2 in ST (Two-tailed t-tests: $t = -0.38$, $p > 0.1$), while the RB groups averaged 22.6 in INT and 23.0 in ST (Two-tailed t-tests: $t = -0.10$, $p > 0.1$). These results provide no evidence in support of Hypothesis 2 for Chinese participants, suggesting that the signaling mechanism had no measurable effect on their rule-following behavior.

In the German sample, the RF groups in ST allocated 10 more rubbish balls to *Box A* than did the RF groups in INT, but this difference was not statistically significant (Two-tailed t-tests: $t = -1.05$, $p > 0.1$). By contrast, the RB groups in ST allocated 13.4 more rubbish balls than did the RB groups in INT, a difference that was statistically significant (Two-tailed t-tests: $t = -2.04$, $p < 0.05$). This finding is consistent with Hypothesis 2, indicating that the German rule-breakers engaged in strategic misrepresentation by increasing their rule-following behavior in the RF task.

We further test Hypothesis 3, which offers an indirect approach to detect misrepresentation under the signaling mechanism.[5] This test serves as a complement to Hypothesis 2, which provides direct evidence and has already been validated in the German sample. Consequently, a null result for Hypothesis 3 does not invalidate our observation of misrepresentation via direct evidence.[6]

[5] Hypothesis 3 is theoretically predicated on the joint validity of Hypotheses 1 and 2. While our experimental result did not support Hypothesis 1, we note that the necessary mechanism (Hypothesis 2) is already validated by direct evidence in key subsamples (e.g., the German sample). Consequently, we assert that further empirical verification of Hypothesis 3 remains warranted and analytically relevant.

[6] Hypothesis 2 is independent of the other hypotheses; consequently, its test results do not compro-

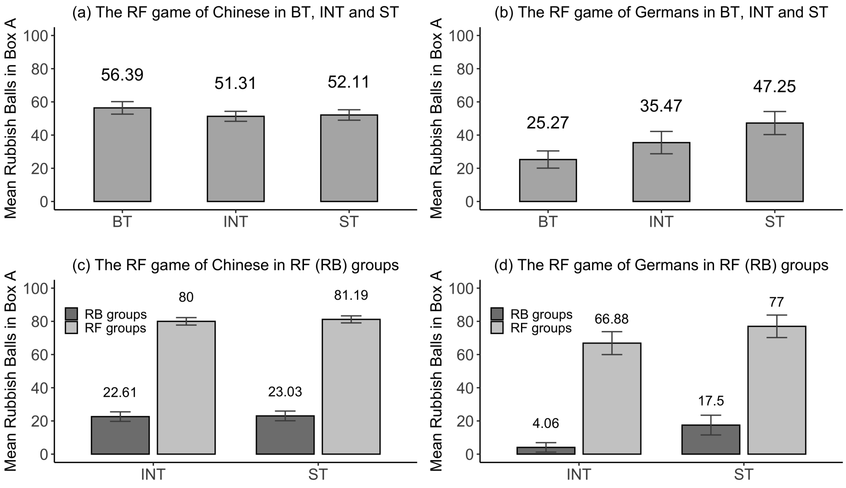


Figure 3.7: Average Number of Balls Allocated to Box A

For the test of Hypothesis 3, we again use the data from the dice game to compare the cheating rates of rule-following and rule-breaking groups across treatments. If misrepresentation occurs in the signaling treatment (ST), some participants—more prone to cheating than intrinsic rule-followers—may enter the rule-following group. This would increase the prevalence of cheating (i.e., reduce honesty) in rule-following groups under ST relative to the internalization-of-norms treatment (INT).

For the Chinese sample, self-reported payoffs in the rule-following group were similar between ST and INT. However, the rule-breaking group in ST reported significantly higher payoffs than in INT (Two-tailed *t*-test: $t = -1.51$, $p < 0.1$), suggesting that ST contained a greater proportion of dishonest participants. In other words, dishonest participants in ST behaved in a more rule-breaking manner than their counterparts in INT. One possible explanation is that common knowledge of the matching principle led these participants to prioritize maximizing profits in the RF task over sending a positive signal.

For the German sample, the mean self-reported payoff in the rule-following group was 16.6 (*s.d.*=9.3) in ST and 14.7 (*s.d.*=7.4) in INT, a difference that was not statistically significant (Two-tailed *t*-test: $t = -0.63$, $p > 0.1$). Similarly, the rule-breaking group's payoffs were comparable between ST and INT, indicating no treatment-level difference in cheating behavior for the German sample.

Cultural Difference in the Rule-following Behavior

We now turn to the analysis of cultural differences between the Chinese and German samples. In BT and INT, participants had no incentive to misrepresent their rule-following behavior because they had no knowledge of the sorted match in the PGG based on the performance in the RF game. Thus, rule-following behavior in these treatments reflects participants' intrinsic propensity to follow rules. On average, Chinese participants allo-

mise its validity.

cated 34 more rubbish balls to *Box A* than German participants in BT, and 16 more in INT. Two-tailed t-tests confirm that the Chinese sample was significantly more rule-following than the German sample in both BT and INT ($p < 0.01$).

We then examined how the signaling mechanism interacted with cultural background using OLS regressions of the number of rubbish balls in *Box A* on a *Germany* dummy, an *ST* dummy, their interaction term (*Germany* × *ST*), and a constant (see Table 3.2). Columns (1) and (2) restrict the sample to RF and RB groups, respectively, while Column (3) includes all participants of INT and ST.

In Column (1) (RF groups), the Germany dummy is negative and significant, indicating that German RF groups in INT were less rule-following than their Chinese counterparts. The *Germany* × *ST* interaction is insignificant, suggesting that the difference in rule-following between INT and ST does not differ significantly between Chinese and German RF groups. In Column (2) (RB groups), the *Germany* dummy is again negative and significant, showing that German RB groups were less rule-following than Chinese RB groups. Here, the *Germany* × *ST* interaction is positive and significant ($p < 0.1$), indicating that the difference in rule-following between INT and ST was greater for German RB groups than for Chinese RB groups. This suggests that the signaling mechanism influences German RB groups in ST, but not their Chinese counterparts. As for Column (3), at the treatment-level comparison (all participants), the difference between ST and INT in the Chinese sample does not differ significantly from that in the German sample. This implies that, overall, the signaling mechanism had no statistically detectable effect in either country—possibly due to limited statistical power in the German sample.

FINDING 2: ***In the German sample, the signaling mechanism led the rule-breaking group to allocate significantly more rubbish balls to the trash can in ST than in INT, whereas no such effect was observed in the Chinese sample.***

3.5.3 Repeated Public Goods Game

In this section, we first analyze public goods (PG) contributions to identify the causal effects of treatment differences, and then examine how cultural background interacts with cooperative behavior under the signaling mechanism.

Causal Effect on Cooperation

We identify the causal effect of the sorted-matching principle by comparing INT with BT, and the causal effect of the signaling mechanism by comparing ST with INT. Figures 3.8(a)-(b) report average public goods (PG) contributions by treatment and by rule-following (RF) and rule-breaking (RB) groups.

The causal effect of the sorted-matching principle is significant. For the Chinese sample, average contributions were 31% in BT, 34% in INT, and 33% in ST (see Figure 3.8(a)). A

Table 3.2: Panel Estimates of Rule-Following Behavior

Specifications	RF groups (1)	RB groups (2)	All (3)
Germany	-13.125*	-18.549***	-15.837**
	(7.135)	(4.022)	(7.318)
ST	1.188	0.420	0.804
	(3.105)	(4.113)	(4.362)
Germany×ST	8.938	13.017*	10.977
	(9.982)	(7.666)	(10.526)
Constant	80***	22.611***	51.306***
	(2.262)	(2.885)	(3.018)
R-squared	0.0403	0.0564	0.0174
Observations	168	168	336

Note: The dependent variable is the number of rubbish balls placed in *Box A*. *Germany*, *ST*, and *Germany×ST* are dummy variables. *Germany* equals 1 for observations from the German sample and 0 for the Chinese sample. *ST* equals 1 for the signaling treatment and 0 for the internalization-of-norms treatment. *RF groups* and *RB groups* refer to the rule-following and rule-breaking groups, respectively. The regression sample includes both RF and RB groups from INT and ST.
$^{\$}$Significance levels indicate at * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

two-tailed t-test indicates that contributions in INT were significantly higher than those in BT ($t = -2.30$, $p < 0.05$), which indicates a positive causal effect of sorted matching. For the German sample, average contributions were 26% in BT, 33% in INT, and 35% in ST (see Figure 3.8(b)). The difference between BT and INT ($BT - INT$) is significantly less than zero (Two-tailed t-test: $t = -3.06$, $p < 0.01$). In both countries, this effect arises because RF groups—comprising a higher proportion of rule-followers—sustain higher PG contributions than RB groups, and the elevated contributions of RF groups offset declines in RB groups (Kimbrough and Vostroknutov, 2016).

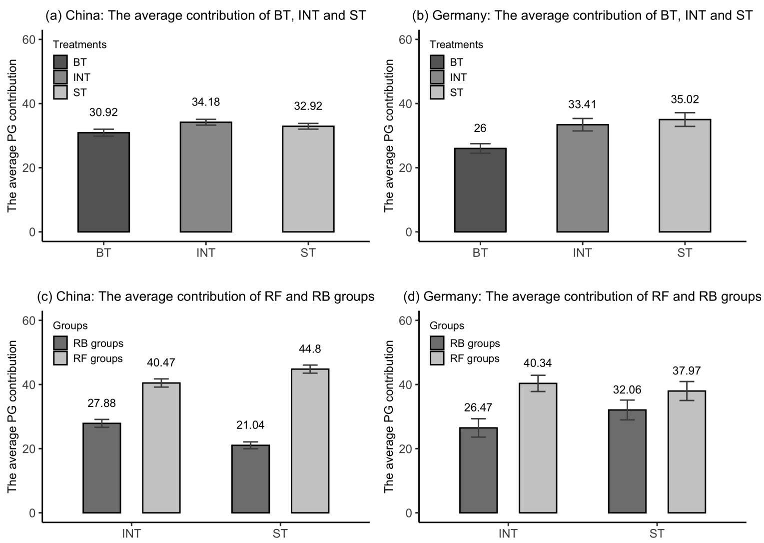


Figure 3.8: Average Contributions in the Public Goods Game by Treatment and Group Type

To estimate the causal effect of the signaling mechanism, we treat INT as the control group and ST as the treated group. At the treatment level, comparing INT to ST, two-tailed t-tests reveal no significant effect for either country (Chinese: $t = 0.99$, $p > 0.1$; German: $t = -0.56$, $p > 0.1$). However, the signaling mechanism exhibits heterogeneous effects across RF and RB groups. For the Chinese sample, RF groups in ST contributed 4.3% more than their INT counterparts (Two-tailed t-test: $t = -2.39$, $p < 0.05$), whereas RB groups in ST contributed 6.8% less than those in INT (Two-tailed t-test: $t = 4.16$, $p < 0.01$). This pattern is consistent with Hypothesis 5b, as no evidence of misrepresentations was also found in the analyses described in Section 3.5.2. Given knowledge of the matching principle, RF-group participants are more willing to cooperate, while RB-group participants adopted more selfish strategies. For the German sample, contributions in RF groups were similar between ST and INT (Two-tailed t-test: $t = 0.61$, $p > 0.1$). In contrast, RB groups in ST contributed 5.6% more than those in INT (Two-tailed t-test: $t = -1.33$, $p < 0.1$). This finding aligns with Hypothesis 5a: due to strategic misrepresentations, RB groups in ST—comprising a mix of intrinsic rule-breakers and misrepresentative "rule-followers"—achieved higher contributions than RB groups in INT, which consisted solely of rule-breakers.

Figure 3.9 presents the public goods (PG) contributions for BT, INT, and ST, as well as for rule-following (RF) and rule-breaking (RB) groups, in the Chinese and German samples. For the Chinese sample, the first-period PG contribution in ST (51.7%) is significantly higher than in INT (42.8%) (Two-tailed t-test: $t = 2.22$, $p < 0.05$). For the German sample, the first-period contribution in ST also exceeds that in INT, but the difference is not statistically significant (Two-tailed t-test: $t = 0.99$, $p > 0.1$). These results indicate that the Chinese sample's first-period behavior is consistent with Hypothesis 4, whereas the German sample's is not.

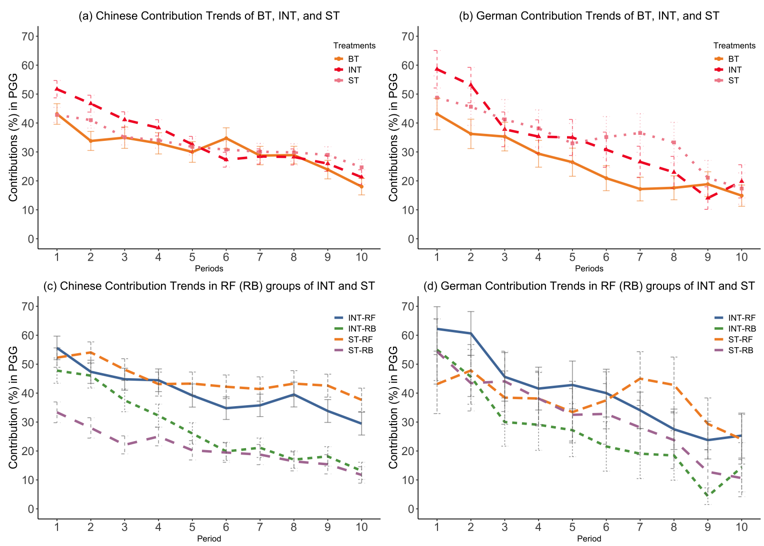


Figure 3.9: Trends in Public Goods Contributions by Treatment and Group Type

To provide statistical support, we analyzed PG contributions across periods using panel regressions with standard errors clustered at the group level. Table 3.3 reports results separately for RF and RB groups in the Chinese and German samples. These regressions report panel estimates of PG contributions on *Period*, an *ST* indicator, and a *Period* × *ST* interaction. In this analysis framework:

•The coefficient on *Period* captures the PG contribution trend in INT for RF (or RB) groups.
•The coefficient on *ST* measures the difference in first-period PG contributions between ST and INT.
•The coefficient on *Period* × *ST* reflects the difference in PG contribution trends between ST and INT.

Table 3.3: Panel Estimates of Contribution Trends: Rule-Following and Rule-Breaking Groups by Nationality

Specifications	Chinese		Germans	
	RF groups (1)	RB groups (2)	RF groups (3)	RB groups (4)
Period	-2.013***	-3.372***	-3.879***	-3.174***
	(0.409)	(0.571)	(1.132)	(0.655)
ST	4.969	-15.391**	-5.177	18.188
	(6.143)	(6.157)	(15.790)	(19.736)
Period×ST	0.544	1.367*	2.339*	-1.307
	(0.669)	(0.742)	(1.326)	(1.060)
Constant	47.906***	47.458***	51.615***	38.521***
	(4.279)	(5.126)	(9.085)	(7.445)
Groups	46	46	12	12
Observations	1840	1840	480	480

Note: *N*=1840 (46 groups×10 periods×4 players); *N*=480 (12 groups×10 periods×4 players). Standard errors clustered at the group level. RF (RB) groups refer to rule-following (rule-breaking) groups. ST represents the signaling treatment when it equals 1; internalization of norms treatment otherwise.
$Significance levels indicate at * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

As for the Chinese sample, Columns (1) and (2) present results for RF and RB groups, respectively. In Column (1), the *Period* × *ST* interaction is insignificant, indicating no difference in contribution trends between ST and INT for RF groups. In Column (2), the *Period* × *ST* interaction is positive and significant, showing that RB groups in ST see a slower decline in contributions than those in INT. The coefficient on ST is negative and significant, indicating that RB groups in ST begin with lower first-period contributions than their INT counterparts. This suggests that, given beliefs about partner quality, RB-group participants in ST adopted more selfish strategies at the outset of the repeated PGG. By the final period, contributions in RB groups converged, averaging 13.1% in ST and 11.7% in INT.

In the German sample, Columns (3) and (4) report results for RF and RB groups, respectively. In Column (3), the *Period* × *ST* interaction is positive and significant, indicating that RF groups in ST exhibit a significantly slower decline in PG contributions than those in INT. In Column (4), there is no significant difference in contribution trends between ST and INT for RB groups. These results suggest that, among German participants, information about the matching principle (ST) influenced RF groups but not RB groups.

FINDING 3: ***For both the Chinese and German samples, average contributions in INT exceeded those in BT, while no significant difference was observed between ST and INT at the treatment level. However, the signaling mechanism reduced contributions in Chinese RB groups, whereas it increased contributions in German RB groups.***

Cultural Difference in Cooperation

In this section, we first analyze cultural differences in public goods (PG) contributions between Chinese and German participants. Because the repeated PGG in BT involved random assignment, we can identify cultural differences in PG contributions without the influence of sorted matching. Comparing Chinese INT with German INT allows us to assess cultural differences under sorted matching, while comparing ST across the two samples reveals how the signaling mechanism interacts with cultural background.

In BT, the average contribution of the German sample was 5.2% lower than that of the Chinese sample, and this difference was statistically significant (Two-tailed t-test: $t = 2.884$, $p < 0.01$), indicating greater cooperativeness among the Chinese participants. In INT, the average contribution difference between the two samples was only 0.77%, which was not statistically significant (Two-tailed t-test: $t = 0.36$, $p > 0.1$). This is because RF and RB groups in INT exhibited similar contribution levels across cultures. In ST, the average contribution difference between samples was 2.097%, also not statistically significant (Two-tailed t-test: $t = -1.01$, $p > 0.1$). This reflects the heterogeneous impact of the signaling mechanism on RF and RB groups across cultures. Specifically, RF groups contributed 44.8% in the Chinese sample versus 37.9% in the German sample (Two-tailed t-test: $t = 2.33$, $p < 0.05$), while RB groups contributed 21.0% in the Chinese sample versus 32.1% in the German sample (Two-tailed t-test: $t = -4.19$, $p < 0.01$).

We further employed panel regressions, clustering standard errors at the group level, to examine cultural differences in RF and RB groups within INT and ST (see Table 3.4). In INT, Column (1) shows that the *Period* $\times$ *Germany* interaction is negative and significant for RF groups, indicating that German RF groups experienced a steeper decline in contributions over time than Chinese RF groups. Column (2) shows no significant difference in contribution trends for RB groups between the two cultures. In the signaling treatment (ST), Column (3) finds no significant difference in contribution trends for RF groups between Chinese and German samples. Column (4) reports a significantly negative *Period* $\times$ *Germany* interaction for RB groups, indicating that German RB groups in ST experienced a faster decline in contributions than their Chinese counterparts.

FINDING 4: ***In the internalization-of-norms treatment, German RF groups' contribu-***

tions declined more rapidly over time than those of Chinese RF groups. In the signaling treatment, German RB groups' contributions declined more rapidly than those of Chinese RB groups.

Table 3.4: Panel Estimates of Public Goods Contributions in the INT and ST

Specifications	INT		ST	
	RF groups	RB groups	RF groups	RB groups
	(1)	(2)	(3)	(4)
Period	-2.347***	-3.935***	-1.469***	-2.005***
	(0.531)	(0.790)	(0.539)	(0.482)
Germany	10.495	1.873	-6.438	24.640
	(8.207)	(12.115)	(13.438)	(18.263)
Period×Germany	-1.932*	-0.597	-0.071	-2.476***
	(1.019)	(1.064)	(0.865)	(0.949)
Constant	53.379***	49.523***	52.875***	32.068
	(5.764)	(6.643)	(4.478)	(3.466)
Groups	22	22	20	20
Observations	880	880	800	800

Note: N=880 (22 groups × 10 periods × 4 players); N=800 (20 groups×10 periods×4 players). Standard errors clustered at the group level. *RF* and *RB* refer to rule-following (rule-breaking) groups. *Germany* represented the German sample when it was equal to 1, and 0 for the Chinese sample.
$Significance levels indicate at * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Elicited Norm and Conditional Cooperation

To further examine cultural differences in cooperation, we applied the method of Krupka and Weber (2009) to elicit beliefs about the cooperative norm in the public goods (PG) game. Figures 3.10(a)-(i) display the elicited norms for each treatment, for rule-following (RF) and rule-breaking (RB) groups.

The figures reveal clear differences between the Chinese and German samples. In BT, the elicited appropriateness ratings for the German sample exceeded those of the Chinese sample when the amount allocated to the private account was less than 10 (out of 20). However, as shown in Section 3.5.3, the German sample's average PG contribution was lower than that of the Chinese sample, indicating a divergence between the elicited norm and actual contributions. This separation between stated norms and behavior was also evident for the German sample in INT. Under the signaling mechanism (ST), however, the elicited norms of the German and Chinese samples were similar.

When comparing RF and RB groups under INT, German participants (both RF and RB) expressed stronger beliefs in the cooperative norm than their Chinese counterparts when the private account allocation was less than 10. Under ST, no significant difference emerged in the elicited norms of RF groups between the two cultures. In contrast, German RB groups expressed weaker beliefs in cooperative norms than Chinese RB groups when the private account allocation was equal to or greater than 8 (out of 20) (see Figure 3.10(i)). This pattern differs from the INT comparison because, in ST, RB groups comprised a mix of high-quality and low-quality players. Consequently, these results are inconsistent with Hypothesis 6a, which predicted that cooperative norms would be stronger in collectivist cultures than in individualistic ones.

FINDING 5: ***The German sample expressed stronger beliefs in the cooperative norm than the Chinese sample. However, for the Germans, there was a notable separation between the elicited cooperative norm and actual PG contributions.***

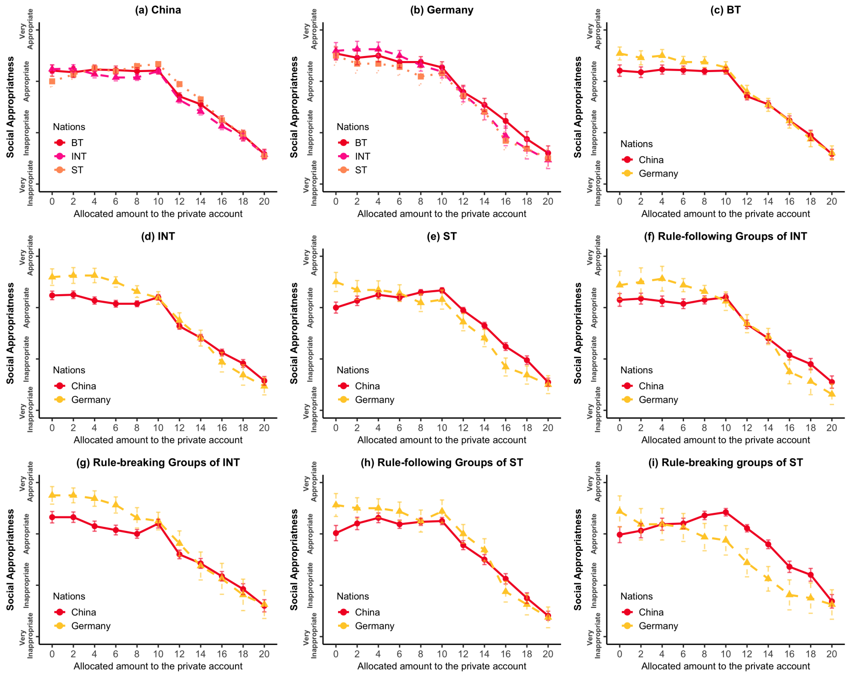


Figure 3.10: Elicited Cooperative Norms: Chinese vs. German Samples

We analyzed conditional cooperation to further explore cultural differences in cooperative behavior. Figures 3.11(a)-(i) show contributions conditional on the average contributions of others.

By comparing BT between the Chinese and German samples, the Wilcoxon Mann-Whitney test shows no significant difference in conditional contributions ($p > 0.1$). When the sorted-matching design (INT) is introduced, however, the German sample exhibits significantly higher conditional cooperation than the Chinese sample ($p < 0.05$).

A closer look at INT reveals that both rule-following (RF) and rule-breaking (RB) groups in the German sample are more conditionally cooperative than their Chinese counterparts. This greater responsiveness to others' contributions helps explain why PG contributions in the German RF groups declined more rapidly over time than in Chinese RF groups.

Under the signaling mechanism (ST), conditional cooperation patterns shift. For RF groups, there is no significant cross-cultural difference (Wilcoxon Mann-Whitney test: $z = -0.25$, $p > 0.1$). For RB groups, however, the Germans participants display significantly higher conditional cooperation than Chinese participants ($p < 0.1$). These results are consistent with Hypothesis 6b.

FINDING 6: ***In the internalization-of-norms treatment, the German sample was more conditionally cooperative than the Chinese sample. In the signaling treatment, German RB groups were more conditionally cooperative than Chinese RB groups.***

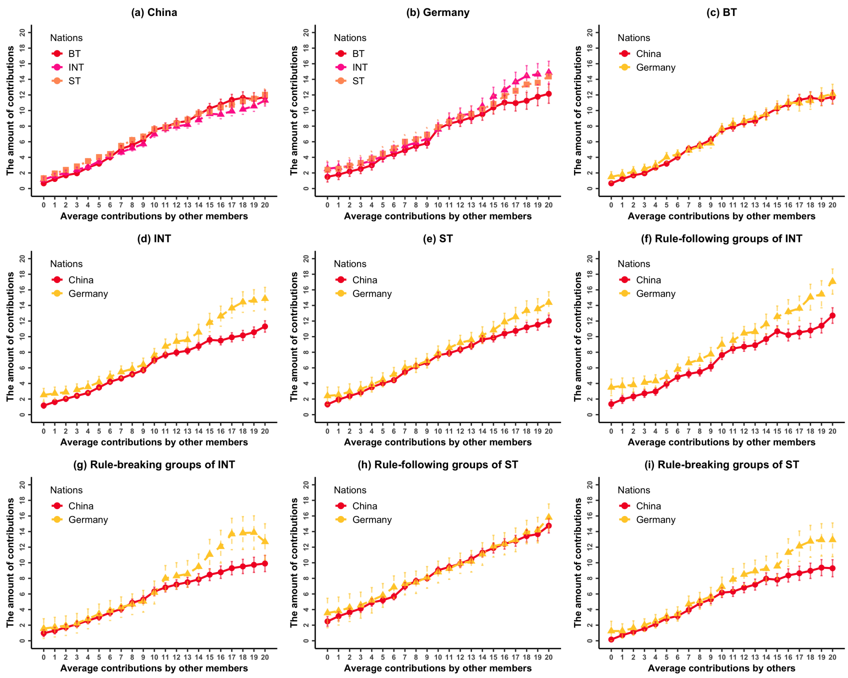


Figure 3.11: Conditional Cooperation: Chinese vs. German Samples

3.6 Conclusion and Discussion

This study examined whether, and through which channels, a low-cost norm-following signal shapes cooperation in a repeated public goods game, using Chinese and German university participants. Consistent with prior work (Kimbrough and Vostroknutov, 2016), individuals classified as rule-followers in a rule-following task contributed more than rule-breakers across cultures. Introducing an explicit signaling mechanism altered the gap between these groups in heterogeneous ways.

Under the signaling treatment, we observe evidence consistent with misrepresentative signaling in the German sample but not in the Chinese sample. Aggregate cooperation did not fall when signaling was available. Compared to the internalization-of-norms condition, signaling was associated with higher early round contributions among rule-followers and lower contributions among rule-breakers; the latter decline was partly offset by the attenuation of the typical downward trend in rule-breaking groups observed without signaling. A plausible interpretation is that participants who increased rule-following in the RF task under the signaling treatment also held stronger beliefs about partner cooperativeness (Bicchieri, 2005; Gächter and Schulz, 2016), which sustained contributions despite strategic motives.

These results indicate that making simple, inexpensive norm-following actions observable can operate as a coordination device that raises cooperation relative to the baseline without signaling. The mechanism appears sufficiently robust to avoid efficiency losses even when some compliance is strategic, consistent with the view (Posner, 2002) that norm adherence conveys credible social information in collective action settings.

For policy and institutional design, the findings suggest a practical implication. Environments that render norm-following visible and verifiable provide a low-cost lever to foster cooperation alongside established instruments such as communication, sanctions, or partner selection (Balliet et al., 2014; Fehr and Gächter, 2000; Ostrom, 1990). The impact is unlikely to be uniform: subgroups that differ in baseline norm-following respond differently, which argues for tailoring implementation to the composition of participants. Designing settings in which cooperative intent can be credibly signaled through inexpensive acts may improve group outcomes without relying on strong enforcement.

Several limitations qualify these conclusions. The evidence derives from laboratory settings; the German and Chinese samples did not playing under completely identical conditions; and social identities, reputational spillovers, and network structure—factors likely to mediate signaling in the field—were not explicitly manipulated (Henrich et al., 2010). Future work could incorporate belief elicitation, repeated measures of norm-following, and richer social context treatments to distinguish strategic signaling from preference change and to assess external validity in naturally occurring groups.

Taken together, norm-following signals are not a universal remedy for cooperation failures, but the evidence supports their role as a practical, low-cost complement in the toolkit for sustaining contributions in repeated social dilemmas across diverse cultural settings.

Chapter 4 Group Decision Rules and Intertemporal Risk Choices

Authors

Ming Dai

Abstract

In this chapter, I conduct a lab experiment to study how two voting rules, majority and unanimity, systematically influence group decisions in intertemporal risk contexts. Building on the canonical risk task of Holt and Laury (2002), I implement two conditions: a risky choice (RC) task and a risky-intertemporal choice (RIC) task, in which payoffs are delayed by four weeks to jointly capture the interaction of risk and the discounting effect. Results show that in the pure risk condition, the unanimity rule induces a shift toward greater group-level risk aversion relative to individual decisions, whereas group decisions under the majority rule closely track the average preferences of individuals. When a delay is introduced, groups under the unanimity rule appear less sensitive to the discounting effect of delay, retaining a higher willingness to accept delayed risk compared to those under the majority rule. Furthermore, the effect of decision rules on selecting moderate-risk proposals (compromise behavior) varies by task context: it is more pronounced in the RC task but diminishes in the RIC task due to opposing shifts in group behavior. These findings provide micro-level evidence for the institutional foundations of collective choice, suggesting that voting rules are not merely mechanisms for preference aggregation; they also condition the influence of individual heterogeneity and shape the trajectory of group coordination.

Keywords

Group Decisions; Risky and Intertemporal Choice; Majority; Unanimity; Voting

4.1 Introduction

Group decision-making plays a central role in many social and economic environments. International bodies such as the United Nations Security Council (Strand and Rapkin, 2011; Wallensteen and Johansson, 2015) and the European Council (Häge, 2008; Heisenberg, 2005; Warntjen, 2010), corporate boards (Bainbridge, 2002), and even households (Corfman and Lehmann, 1987) routinely make collective choices rather than relying on individuals (Black, 1948; Maciejovsky et al., 2013; Patterson and Caldeira, 1988; Poole and Daniels, 1985; Tropman, 2013; Wright et al., 1990). A substantial body of literature suggests systematic behavioral differences between individuals and groups (Baker et al., 2008; Del Carpio et al., 2022; Glätzle-Rützler et al., 2021; Masclet et al., 2009; Shupp and Williams, 2008). In particular, findings often indicate that groups can be more risk-averse and more patient than individuals when facing uncertainty or intertemporal trade-offs (Baillon et al., 2016; Charness et al., 2007; Kugler et al., 2012).

Many real-world decisions, however, involve the interaction of risk and time discounting. Risky outcomes often materialize only after a delay, creating "intertemporal-risk" environments (Abdellaoui et al., 2011; Anderhub et al., 2001; Andreoni and Sprenger, 2012; Baucells and Heukamp, 2012; Epper and Fehr-Duda, 2024; Noussair and Wu, 2006; Somasundaram and Eli, 2022). A corporate board evaluating a long-term investment must weigh both probabilistic payoffs and the fact that these payoffs occur months or years later. In such a setting, the safe option is typically both sooner and certain, whereas the risky option is later and uncertain. Because delayed payoffs are discounted, the expected value of the risky option is lower in the delayed task. As a result, decisions in these environments reflect a complex trade-off between time discounting and risk-taking rather than simple patience. How groups jointly evaluate this interaction of risk and delay—and how institutional rules shape this process—remains an open question.

The outcome of such complex decisions is influenced not only by individual preferences but also by the institutional rules governing the process (Baillon et al., 2016; Castore and Murnighan, 1978; Gillet et al., 2009; Guttman, 1998; Hastie and Kameda, 2005; Miller, 1985; Romme, 2004; Tsebelis, 2013; Williams and Taormina, 1993). Majority and unanimity rules, in particular, present distinct trade-offs regarding efficiency, minority protection, and information aggregation (Asheim et al., 2006; Dougherty et al., 2014; Dougherty and Edward, 2005; Glynia et al., 2025; Macé and Treibich, 2024; Miller, 1985; Tsebelis, 2013). While prior work has examined the behavioral implications of these rules in static risk or preference-aggregation contexts (Baharad et al., 2020; Dougherty et al., 2014; Glynia et al., 2025; Nehring et al., 2016; Velden et al., 2007), the interaction between these rules and decisions involving both risk and delay remains less understood. This chapter contributes to this literature by studying how majority and unanimity rules shape group choices in intertemporal-risk environments, and how preference heterogeneity mediates collective outcomes under each rule.

To explore these dynamics, I implement two treatments: i) the Majority Treatment, and ii) the Unanimity Treatment. Each treatment includes two core tasks: a risky choice (RC) task and a risky-intertemporal choice (RIC) task. The RC task follows the multiple price list (MPL) method of Holt and Laury (2002), in which subjects choose between safe payoffs and risky payoffs. The RIC task is a modification of the RC task, requiring subjects to choose between safe-and-soon payoffs and risky-and-delayed payoffs. This design is intended to capture how decision-makers trade off delay against risk when outcomes are shifted into the future. Using a within-subject design, I compare individual and group decisions to assess changes in risk attitudes and intertemporal risk evaluations. With a between-subject design, I compare group decisions in risky and risky-intertemporal conditions across decision rules.

The main results are as follows. First, in the pure risk condition, groups under the unanimity rule tend to exhibit greater risk aversion than individuals, while group decisions under the majority rule closely mirror individual preferences. Second, when a delay is introduced, groups under the unanimity rule appear less sensitive to the discounting effect of delay compared to individual decisions. Unlike the majority rule, where group choices track the individual tendency to discount delayed risky payoffs, the unanimity rule preserves a higher willingness to accept delayed risk. Third, the effect of decision rules on the selection of moderate-risk proposals is context-dependent. In the RC task, the unanimity rule leads to a higher likelihood of moderate-risk proposals relative to the majority rule. This institutional difference diminishes in the RIC task, where the combined cost of risk and delay shifts group behavior: the selection of moderate-risk proposals increases under the majority rule and slightly declines under the unanimity rule, yielding similar outcomes across decision rules.

This chapter makes three contributions to the study of group decision-making under intertemporal risk. First, it provides experimental evidence on the interaction between institutional voting rules and the nature of the decision context. While existing work typically studies risk and time separately or focuses on individual choices (Anderhub et al., 2001; Andreoni and Sprenger, 2012; Baucells and Heukamp, 2012; He et al., 2012; Noussair and Wu, 2006; Rockenbach et al., 2007; Viscusi et al., 2011; Zhang and Casari, 2012), this study sheds light on the limits of institutional influence. I find that the efficacy of voting rules is context-dependent: while the unanimity rule induces a robust "cautious shift" in pure risk tasks, this institutional effect is attenuated when delay is introduced. This reveals that the interplay between risk and delay creates a distinct decision environment where the standard predictions of voting power theories may not fully apply.

Second, it delineates the boundary conditions of preference aggregation by distinguishing between risk preferences and sensitivity to delay. I show that voting rules (particularly unanimity) are highly effective in empowering risk-averse members to veto risky options in static contexts. However, this study finds that when payoffs are delayed, the cognitive bias of time discounting tends to overshadow these strategic incentives. This suggests that the "cost of delay" acts as a dominant constraint, compressing the strategic space for voting rules to operate. This finding contributes to the behavioral political economy literature by identifying when institutions matter most: they effectively mediate conflicts regarding risk but are less effective in counteracting biases stemming from time discounting (Andreoni and Sprenger, 2012; Jackson and Yariv, 2015).

Third, it uncovers the strategic mechanisms of consensus formation, highlighting the role of "pre-emptive compromise". Contrary to the view that unanimity leads to negotiation failure or fatigue, my analysis documents high coordination efficiency. I provide evidence that the unanimity rule acts as a coordination device that clarifies the focal point. Aided by communication, group members anticipate the necessity of full agreement and strategically converge on moderate, safe options immediately, bypassing prolonged bargaining (Karpowitz et al., 2012; Wabara, 2021). This contrasts with the majority rule, which favors modal convergence often at the expense of minority preferences (Fleck and Hanssen, 2013; Horwitz, 1966; Maletz, 2002; Sartori, 2016). These insights refine my understanding of how procedural design shapes not just the outcome, but the strategic path to consensus.

The remainder of the study is organized as follows. Section 4.2 reviews the relevant literature. Section 4.3 presents the experimental design. Section 4.4 outlines the theoretical hypotheses. Section 4.5 presents the results and data analysis. Section 4.6 concludes with a discussion.

4.2 Literature Review

Research on individual and group decision-making has yielded substantial insights in both economics and psychology (Abdellaoui et al., 2011; Bateman and Munro, 2005; Bornstein et al., 2004; Bougheas et al., 2013; Glätzle-Rützler et al., 2021; Gonzalez-Jimenez and Müller, 2025; Kocher and Sutter, 2005; Masclet et al., 2009; Shupp and Williams, 2008; Zhang and Casari, 2012). Early experiments focused primarily on measuring individual risk and time preferences (Anderson and Stafford, 2009; Bagaïni et al., 2025; Harrison et al., 2005; Holt and Laury, 2002; Schildberg-Hörisch, 2018), laying the groundwork for understanding patience and risk tolerance. As scholars increasingly recognized that real-world decisions often occur in group contexts, attention shifted toward group behavior. A growing body of evidence suggests that groups often exhibit greater risk aversion and patience in intertemporal choices than individuals (Del Carpio et al., 2022; Glätzle-Rützler et al., 2021; Masclet et al., 2009; Shupp and Williams, 2008). This "cautious shift" implies that group preferences are not merely the aggregation of individual ones but are likely shaped by interaction, coordination, and institutional constraints.

Risk and time are not independent dimensions; in practice, decisions often involve their interaction. Investment, policy, and household decisions typically entail both uncertainty and delay, forming multidimensional "delayed risk" contexts (Abdellaoui et al., 2011;

Anderhub et al., 2001; Andreoni and Sprenger, 2012; Baucells and Heukamp, 2012; Noussair and Wu, 2006), many of which arise in group decision settings. Existing research at the individual level has explored a complex relationship between risk aversion and time discounting. Since delayed payoffs are typically discounted, the subjective value of a risky option diminishes as its realization is postponed. For instance, Anderhub et al. (2001) report a negative correlation between risk and time preferences, noting that subjects exhibiting high risk aversion also tend to discount future payoffs more heavily. This suggests that avoiding a delayed risky option may not reflect simple impatience, but rather a rational response to the reduced expected utility caused by the combined cost of risk and delay. Similarly, Noussair and Wu (2006) observe systematic differences in risk choices between immediate and delayed payoffs, finding that subjects often display higher risk tolerance when lotteries are delayed. These findings indicate that decision-makers actively trade off the disutility of delay against the disutility of risk. However, this literature focuses primarily on individual choices; how group interplay and decision rules jointly shape this interaction between discounting and risk-taking in collective contexts remains less understood.

At the group level, institutional decision rules are widely recognized for their influence in shaping collective outcomes (Baker et al., 2008; Brunette et al., 2015; Morton, 1987; Shupp and Williams, 2008; Zhang and Casari, 2012). Majority voting and unanimity are the two most commonly studied rules, attracting sustained attention across political science, social psychology, and economics (Dougherty et al., 2014; Miller, 1985; Tsebelis, 2013). These rules have been applied to a wide range of group decision contexts, including jury deliberation (Duggan and Martinelli, 2001; Feddersen and Pesendorfer, 1998), minority exclusion (Miller, 1985; Ohtsubo et al., 2004; Velden et al., 2007), decision efficiency (Dougherty and Edward, 2005; Nemeth, 1977; Stasson et al., 1991), and process optimization (Compte and Jehiel, 2017; Guttman, 1998; Maggi and Morelli, 2006).

The literature suggests that the majority rule facilitates decision efficiency but may risk marginalizing minority preferences, potentially resulting in a "tyranny of the majority" (Fleck and Hanssen, 2013; Horwitz, 1966; Maletz, 2002; Miller, 1985; Ohtsubo et al., 2004; Sartori, 2016). By contrast, the unanimity rule effectively grants equal veto power to all members, offering stronger protection for minority interests but often at the expense of decisiveness (Baillon et al., 2016; Karpowitz et al., 2012; Velden et al., 2007; Wabara, 2021). This safeguard can be strategically misused, especially when minority members act out of self-interest, impeding group decisions (Velden et al., 2007). Because decision rules differ in how preferences are expressed and aggregated, they generate systematic variation in intertemporal risk choices. Majority-based decisions tend to reflect dominant preferences, while the unanimity rule fosters compromise that often converges on the median preference (Baillon et al., 2016; Brunette et al., 2015). In intertemporal-risk contexts, this creates a specific tension: delaying a risky payoff reduces its present value due to discounting. Groups must therefore decide whether the potential high return is sufficient to offset the cost of waiting. While groups generally exhibit caution, the mechanism by which they resolve the trade-off between immediate certainty and delayed uncertainty under different voting rules remains an open question.

Group heterogeneity further complicates the collective decision process. Heterogeneity in risk preferences influences group decisions; the proportion of risk-averse and risk-seeking members affects not only the final choice but also the direction and pace of negotiation. Evidence suggests that homogeneous groups are more prone to polarization (Isenberg, 1986; Moscovici and Zavalloni, 1969; Sia et al., 2002; Sunstein, 2002) because individuals in such groups are exposed primarily to information that reinforces their initial positions. When a majority faction exists, the group's final decision tends to align with the majority preference (Asch, 1951; Banerjee, 1992; Deutsch and Gerard, 1955; Levitan and Verhulst, 2016; Sherif, 1936; Wei et al., 2019). This holds not only under the formal majority rule. Even under the unanimity rule, such a structure (e.g., 2-to-1) generates significant social conformity pressure, compelling the minority to either concede or face greater persuasive pressure to reach consensus. Highly heterogeneous groups tend to converge toward a moderate position due to negotiation delays (Black, 1948; Karpowitz et al., 2012; Nemeth, 1977). In particular, under the unanimity rule, each member's veto power compels extended negotiation. To prevent impasse, members must concede extreme proposals (e.g., the most risk-averse and risk-seeking ones). The group thus tends to compromise on the median proposal (moderate-risk), often the only mutually acceptable outcome.

This dynamic parallels findings in political science on ideological homogeneity and policy extremity, highlighting an interaction between group composition and institutional rules. Moreover, personality traits have been shown to shape patterns of conformity and voting behavior within groups (Altemeyer, 1981; Mondak, 2010; Rammstedt and John, 2007; Roccas et al., 2002), indicating that group decisions reflect not only preference aggregation but also psychological factors. The experiment incorporates measures of personality traits as control variables to account for individual heterogeneity.

4.3 Experimental Design

4.3.1 Choice Tasks

The experiment was conducted at the Alfred Weber Institute of Economics, Heidelberg University, during the Winter 2023/24, Summer 2024, and Winter 2024/25 semesters. The experiment was programmed in *oTree* (Chen et al., 2016).

Each participant completed two types of tasks: a Risky Choice (RC) task and a Risky-Intertemporal Choice (RIC) task. The risky choice task adapts the multiple price list design of Holt and Laury (2002) to elicit risk preferences. The RIC task modifies this design by imposing a four-week delay on the payment of all potential outcomes. This within-subject design allows me to compare decisions under risk with those involving both risk and an intertemporal trade-off, for both individuals and groups.

Risky Choice (RC) Task: This task uses a multiple price list format, displayed in Table 4.1, to elicit subjects' risk preferences. In each of the ten rows, participants choose between a safe lottery (*Option A*) and a risky lottery (*Option B*).

- ***Option A*** offers a certain payoff of 8.0 points.
- ***Option B*** offers a high payoff of 12.0 points with probability p and a low payoff of 4.0 points with probability $1 - p$.

Table 4.1: Payoff for the Risky Choices (RC) Task

No.	Option A (*Tomorrow*)	Or...	Option B (*Tomorrow*)
1	8.0 points with 100%	A ○ ○ B	12.0 points with 10%, otherwise 4.0 points
2	8.0 points with 100%	A ○ ○ B	12.0 points with 20%, otherwise 4.0 points
3	8.0 points with 100%	A ○ ○ B	12.0 points with 30%, otherwise 4.0 points
4	8.0 points with 100%	A ○ ○ B	12.0 points with 40%, otherwise 4.0 points
5	8.0 points with 100%	A ○ ○ B	12.0 points with 50%, otherwise 4.0 points
6	8.0 points with 100%	A ○ ○ B	12.0 points with 60%, otherwise 4.0 points
7	8.0 points with 100%	A ○ ○ B	12.0 points with 70%, otherwise 4.0 points
8	8.0 points with 100%	A ○ ○ B	12.0 points with 80%, otherwise 4.0 points
9	8.0 points with 100%	A ○ ○ B	12.0 points with 90%, otherwise 4.0 points
10	8.0 points with 100%	A ○ ○ B	12.0 points with 100%, otherwise 4.0 points

The probability p of the high payoff in *Option B* increases from 10% to 100% in 10% increments down the list, while the payoffs remain constant. A more risk-averse individual will switch from *Option A* to *Option B* later (i.e., at a high probability of the 12.0-point payoff). To mitigate any potential "present bias" or immediate demand effects, all payments for this task were paid on the day after the experiment ("*tomorrow*").

Risky-Intertemporal Choice (RIC) Task: The task maintains the same structure as the RC task but introduces a four-week payment delay to the risky lottery, as shown in Table 4.2. This modification presents participants with a choice between a sooner-certain outcome and a delayed-risky outcome.

- ***Option A*** offers a certain payoff of 8.0 points, paid out one day after the experiment.
- ***Option B*** offers a risky payoff (12.0 points with probability p or 4.0 points otherwise), paid out four weeks after the experiment.

To ensure the credibility of the delayed payments, participants were asked to collect their earnings in person from the institute's office. Payment dates were explicitly scheduled on business days, avoiding weekends and public holidays in Baden-Württemberg, Germany.

Table 4.2: Payoff for the Risky-Intertemporal Choice (RIC) Task

No.	Option A (*Tomorrow*)	Or...	Option B (*Four weeks later*)
1	8.0 points with 100%	A ○ ○ B	12.0 points with 10%, otherwise 4.0 points
2	8.0 points with 100%	A ○ ○ B	12.0 points with 20%, otherwise 4.0 points
		. . . etc.	
9	8.0 points with 100%	A ○ ○ B	12.0 points with 90%, otherwise 4.0 points
10	8.0 points with 100%	A ○ ○ B	12.0 points with 100%, otherwise 4.0 points

4.3.2 Experimental Procedure

Procedure

The experiment consisted of four parts,[1] completed in a fixed sequence (Figure 4.1). Participants first made decisions in the Risky Choice (RC) task, first individually (Part A) and then in three-person groups (Part B), including Participants A, B, and C. They then repeated this sequence for the Risky-Intertemporal Choice (RIC) task, again making decisions individually (Part C) and then in groups (Part D).

Coordination

Group decisions in Parts B and D followed a structured protocol lasting a maximum of five rounds. Each round consisted of three stages:

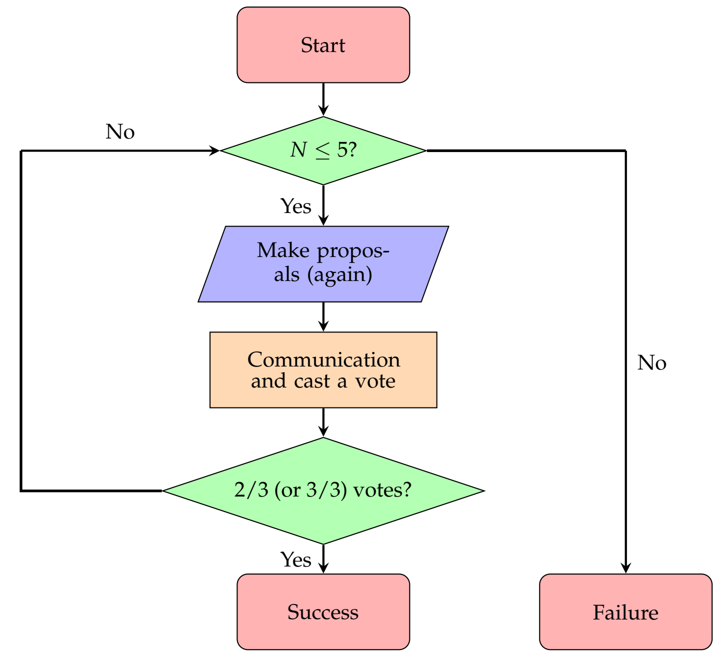
i) **Proposal:** Each group member simultaneously filled out his/her choice list and submitted his/her proposal to the group in Round 1. In Rounds 2-5, each member only needed to revise his/her proposal from the previous round and then submit the revised version to the group.

ii) **Communication:** After all group members had submitted their proposals, three proposals were displayed to the whole group. Members could then use an anonymous group chat to coordinate. Private messaging and offensive language were prohibited.

iii) **Voting:** After communicating, members cast votes on a single proposal, and the results were displayed to the entire group.

A proposal was accepted if it met the voting threshold required by the treatment. If a decision was reached, the task came to an end. Otherwise, the group proceeded to the next round. If no agreement was reached after five rounds, all group members earned zero for that task.

[1] Further information on the instructions and experimental procedure can be found in Appendix A.3.2.



Note: The two treatments differed in their decision-making threshold. The majority rule required the support of at least two out of three group members (a two-thirds majority), while the unanimity rule required the support of all three members (a consensus).

Figure 4.1: Coordination Procedure

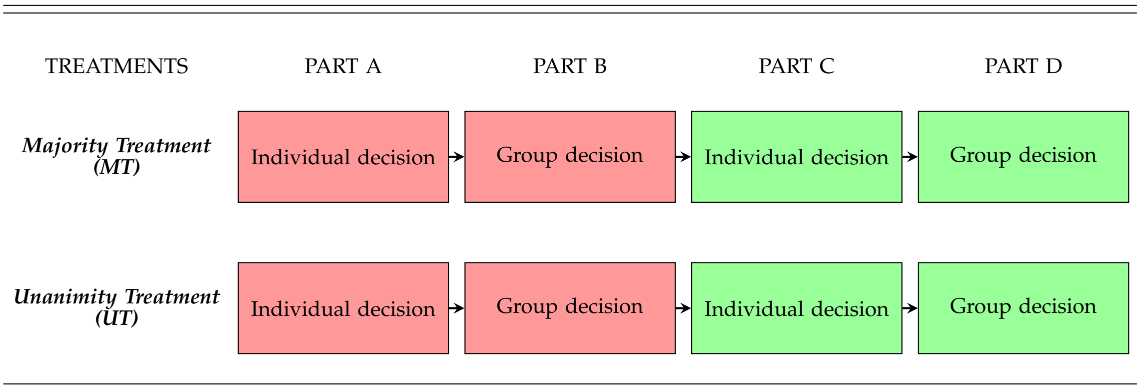
Treatments

I used a between-subject design with two treatments that differed only in the voting rule required for group decisions: Majority and Unanimity. Figure 4.2 presents the experimental procedure of the majority and unanimity treatments. In the figure, red and green rectangles represent RC and RIC tasks, respectively. The four-part experimental procedure was identical in both treatments. In Part B, participants were randomly assigned to groups of three. These groups remained fixed for the second group task in Part D.

i) ***Majority Treatment (MT):*** A group decision was reached if a proposal received at least two out of the three votes.
ii) ***Unanimity Treatment (UT):*** A group decision required unanimous support, meaning a proposal had to receive all three votes.

After completing all four tasks, participants filled out the Big Five Inventory survey (Rammstedt and John, 2007) and a brief demographic questionnaire. The entire session lasted approximately 25 minutes.

Participants received a € 3.00 show-up fee for their participation. Task earnings were calculated in experimental tokens, which were converted to cash at an exchange rate of 1 token = € 0.50. The average earning, including the show-up fee, was € 7.10.



Note: I employed a between-subject design, randomly assigning participants to either the Majority or the Unanimity treatment. Each participant completed all four parts of the experiment in a fixed sequence (Parts A, B, C, and D).

Figure 4.2: Experimental Design

I used a random incentive system to determine the final payment. At the end of the session, one of the four parts (A, B, C, or D) was randomly selected for payment. Then, one of the ten decision rows within that part was randomly selected. The participant's earnings were based on the outcome of this single decision. If a group task (Part B or D) was selected for payment and the group had failed to reach an agreement on the randomly chosen row, all members of that group earned zero tokens from the task and received only their show-up fee.

This experimental design enables both within- and between-subject comparisons. First, comparing the number of risky-soon or risky-delayed choices between individual and final group decisions captures the overall effect of group decision-making under majority or unanimity rules. Second, examining changes in these choices across decision rules reveals the treatment effect of institutional design on intertemporal risk preferences.

4.4 Hypotheses

In this section, I define $D^{j}_{i,k}$ as the number of risky choices (*Option B*) for task $i \in \{RC, RIC\}$ under the decision rule $j \in \{MT, UT\}$. The subscript k indicates the decision stage:

- $k = 1$ for the individual decision;
- $k = 2$ for a member's first proposal within their group;
- $k = 3$ for the final group decision after coordination.

For example, the terms map to the experimental parts under the majority rule (MT) as follows:

- $D^{MT}_{RC,1}$ and $D^{MT}_{RIC,1}$ are the individual decisions in Part A and Part C, respectively.
- $D^{MT}_{RC,2}$ and $D^{MT}_{RIC,2}$ are the first proposals by each member in Part B and Part D, respectively.

- $D^{MT}_{RC,3}$ and $D^{MT}_{RIC,3}$ are the final group decisions in Part B and Part D, respectively.

Similarly, $D^{UT}_{RC,1}$, $D^{UT}_{RIC,1}$, $D^{UT}_{RC,2}$, $D^{UT}_{RIC,2}$, $D^{UT}_{RC,3}$, and $D^{UT}_{RIC,3}$ are the corresponding variables under the unanimity rule (*UT*). The following hypotheses use this notation to specify predictions about individual and group behavior across treatments.

4.4.1 Individual versus Group Decisions

A large body of experimental evidence suggests that group decisions are less risky (more patient) relative to individual decisions (Baker et al., 2008; Del Carpio et al., 2022; Masclet et al., 2009; Shupp and Williams, 2008). This phenomenon, commonly referred to as the "cautious shift", may arise from mechanisms such as amplified concern over potential losses during group coordination and a preference for safer, more defensible choices. In the context of intertemporal risk, this caution extends to the evaluation of delayed outcomes. Since delayed risky payoffs are subject to time discounting, they carry a lower expected value compared to immediate equivalents. Consequently, groups may exhibit a stronger preference for the "safe and soon" option not merely due to risk aversion, but to avoid the combined disutility of uncertainty and delay.

Under the unanimity rule, each member effectively holds veto power. This implies that the presence of a highly risk-averse individual within the group can block any proposal perceived as excessively risky (Baillon et al., 2016; Bouton et al., 2018; Buchanan and Tullock, 1965; Karpowitz et al., 2012; Velden et al., 2007; Wabara, 2021). To reach consensus and avoid decision failure—which, in the context of this experiment, results in zero payoff—the group's final decision is likely to align with the preferences of the most cautious member.

Although the majority rule grants greater authority to the majority faction—allowing collective decisions to proceed without minority approval—the process of group interaction may still shift outcomes towards more conservative choices. This depends on whether the majority itself is risk-averse or inclined toward safer alternatives. Prior research suggests that even under the majority rule, group decisions may be less risky than individual choices (Baillon et al., 2016; Charness et al., 2007; Kugler et al., 2012; Masclet et al., 2009). Accordingly, this study posits that the cautious shift effect is present under both decision-making rules.

Hypothesis 1 (Individuals versus Groups):
In both the RC and RIC tasks, groups will choose fewer risky options than individuals under both decision rules, namely $D_{RC,1} > D_{RC,3}$ *and* $D_{RIC,1} > D_{RIC,3}$.

4.4.2 Decision Rules Moderate Group Decisions

The Risky Choice (RC) task

Differences in Group Decision Thresholds: The key distinction between majority and unanimity rules lies in the voting threshold required to reach a collective decision. Under the majority rule, a proposal requires agreement from two out of three members, whereas the unanimity rule requires consensus from all members. Consequently, the influence of minority and majority factions on group outcomes varies across the decision rules.

Evidence shows that group decisions under the unanimity rule are more risk-averse than those under the majority rule (Balliet et al., 2014; Kugler et al., 2012; Miller, 1985; Morone et al., 2017). The unanimity rule grants veto power to each member, meaning that if even one member dissents, a collective decision cannot be reached. The minority must therefore be persuaded or accommodated for the group to achieve consensus. For example, a single risk-averse individual can block riskier proposals, compelling the group to engage in extended negotiation and compromise (Morone et al., 2017). The resulting decision is often one that accommodates all members' preferences—typically deviating from the majority's proposal (Baillon et al., 2016; Bouton et al., 2018; Charness et al., 2007).

In contrast, the majority rule empowers risk-seeking members to form a "majority coalition", enabling them to override the preferences of highly risk-averse individuals. This dynamic may give rise to the "tyranny of the majority", whereby minority preferences are marginalized (Fleck and Hanssen, 2013; Horwitz, 1966; Maletz, 2002; Miller, 1985; Ohtsubo et al., 2004; Sartori, 2016). Under the majority rule, a risk-seeking minority only needs to persuade at least one member of the risk-averse majority to shift the collective decision toward the riskier option. Under the unanimity rule, however, the minority must persuade every member. This requirement is substantially more difficult and imposes significantly higher persuasion costs compared to the majority rule.

Consequently, in the absence of delay cost, the majority rule provides a more accessible pathway to riskier group choices, suggesting that the level of risk-taking will be higher than under the unanimity rule ($D^{MT}_{RC,3} > D^{UT}_{RC,3}$). However, the literature offers different evidence on this comparison; for instance, Kogan and Wallach (1967) suggest that under the unanimity rule, the diffusion of responsibility spreads the potential cost of failure across all agents, fostering psychological safety that promotes higher group risk-taking. In contrast, the majority rule allows dissenters to signal non-consent. This partitioning of accountability may dampen the majority's propensity to endorse extreme risks. Brunette et al. (2015) find group decisions under the majority rule to be less risky than those under the unanimity rule, underscoring the necessity of empirically testing this specific hypothesis.

I assume individual preferences are invariant across treatments due to random assignment, such that $D^{MT}_{RC,1} \approx D^{UT}_{RC,1}$. I define the pure group effect as the difference between the final group decision and the initial individual decision, $(D^{i}_{RC,3} - D^{i}_{RC,1})$, which isolates the effect of group interaction. Based on the premise that $D^{MT}_{RC,3} > D^{UT}_{RC,3}$, this leads to the testable implication that the group effect under the majority rule is greater than that under the unanimity rule.

Therefore, I formalize the hypothesis as follows:

Hypothesis 2a (Majority versus Unanimity in the RC Task):
In the RC task, the shift from individual to group decisions under the majority rule is greater than that under the unanimity rule, namely $D^{MT}_{RC,3} - D^{MT}_{RC,1} > D^{UT}_{RC,3} - D^{UT}_{RC,1}$.

The Risky-Intertemporal Choice (RIC) task

The introduction of temporal delay may reverse the relative effects of decision rules. This shift arises from fundamental differences in how each rule aggregates preferences when outcomes are subject to time discounting.

The unanimity rule mitigates the discounting effect. Prior research suggests that groups operating under unanimity are more likely to tolerate delay for higher returns (Ambrus et al., 2015; Denant-Boemont et al., 2017; Glätzle-Rützler et al., 2021; Jackson and Yariv, 2015; Kugler et al., 2012; Shapiro, 2010; Sutter, 2009). In the RIC task, the delayed risky option has a lower present value due to discounting. However, the unanimity rule, requiring universal assent for any decision, grants significant leverage to members who perceive the delayed return as valuable enough to offset the waiting cost. This institutional setting prevents the immediate rejection of the delayed option by members with high discount rates. Through this sustained deliberation, the group is forced to evaluate the delayed prospect more thoroughly, often converging on a compromise where the riskier, delayed option is accepted. Consequently, groups under the unanimity rule retain a higher willingness to accept delayed risk.

The majority rule reinforces the evaluation of immediate utility. Under the majority rule, the need for negotiation is substantially reduced. Since delayed payoffs are typically discounted, a majority of members may prefer the "safe and soon" payoffs to maximize immediate utility. The majority rule allows this coalition to quickly enforce their preference, disregarding the views of members who are willing to wait for uncertain future gains (Fleck and Hanssen, 2013; Horwitz, 1966; Maletz, 2002; Miller, 1985; Sartori, 2016). Without the institutional constraint to accommodate minority views, the tendency to discount future payoffs dominates the collective choice. Accordingly, I expect groups under the majority rule to exhibit a sharper reduction in risky choices when delay is introduced, compared to those under the unanimity rule.

The difference in decisions between the RIC and RC tasks indicates the sensitivity to discounting (the "delayed effect"). It is defined as the difference in the number of risky choices between the RIC and RC tasks.

$$\triangle_i^j = D_{RIC,i}^j - D_{RC,i}^j$$

where $i \in \{1,2,3\}$ and $j \in \{MT, UT\}$. A more negative $\triangle$ implies that the introduction of delay causes a larger drop in risk-taking, reflecting a stronger reaction to discounting. From the above discussion, group decisions under the majority rule ($\triangle_3^{MT}$) are expected to be more sensitive to delay than those under the unanimity rule ($\triangle_3^{UT}$), namely $\triangle_3^{MT} < \triangle_3^{UT} \Rightarrow D_{RIC,3}^{MT} - D_{RC,3}^{MT} < D_{RIC,3}^{UT} - D_{RC,3}^{UT}$. Given that $D_{RC,3}^{MT} > D_{RC,3}^{UT}$ (from the pure risk task), it follows that the gap will narrow or reverse in the RIC task, leading to $D_{RIC,3}^{MT} \leq D_{RIC,3}^{UT}$.

Hypothesis 2b (Majority versus Unanimity in the RIC Task):
In the RIC task, the average number of risky-intertemporal options under the majority rule is less than or equal to that under the unanimity rule, namely $D_{RIC,3}^{MT} \leq D_{RIC,3}^{UT}$.

Further Predictions on Delayed Effect

Based on the definition of the delayed effect, the value of $\triangle_i$ indicates the sensitivity to the discounting effect of delay for individual or group decisions.

- $\triangle_i \approx 0$ indicates minimal sensitivity to intertemporal framing ($D_{RIC,i} \approx D_{RC,i}$).
- $\triangle_i > 0$ indicates a preference for delay or negative discounting ($D_{RIC,i} > D_{RC,i}$).
- $\triangle_i < 0$ indicates that time discounting reduces the subjective value of the risky prospect ($D_{RIC,i} < D_{RC,i}$).

A four-week delay in payoffs reduces the present value of the potential reward, entailing a form of disutility that reflects individuals' time discounting rates (Green and Myerson, 2004; Kim et al., 2020). When this temporal disutility is compounded with the probabilistic disutility of risk, the expected utility of risky choices (*Option B*) is substantially reduced compared to the immediate condition. Consequently, rationality suggests that $D_{RIC,i}$ should be lower than $D_{RC,i}$.

On the other hand, commitments to future payments may carry credibility concerns (Andreoni and Sprenger, 2012; Doyle, 2013; Frederick et al., 2002; Noussair and Wu, 2006). Although the experimental design explicitly clarifies the payment procedure to enhance trust, participants may nonetheless perceive delayed payoffs as an additional, non-probabilistic source of uncertainty (transaction costs or default risk). This further diminishes the attractiveness of the delayed risky option. Accordingly, this study expects the number of risky choices in the RIC task to be generally lower than in the RC task due to the combined effect of discounting and transaction risk, namely $D_{RIC,i} < D_{RC,i} \Rightarrow \triangle_i < 0$.

Hypothesis 2c (Delayed Payments Reducing Risky Choices):
The number of risky choices in the RIC task is significantly smaller than that in the RC task, namely $D_{RIC,i} < D_{RC,i}$.

This study interprets $\triangle_1 = D_{RIC,1} - D_{RC,1}$ as the delayed effect of individual decisions and $\triangle_3 = D_{RIC,3} - D_{RC,3}$ as that of group decisions. The difference $\triangle_3 - \triangle_1$ captures the net impact of group decision-making on sensitivity to delay, isolating institutional effects from individual preferences.

Building on Hypothesis 2b, this study posits that the unanimity rule—requiring full agreement—induces deeper negotiation. This institutional constraint enables minority members (who may perceive higher future value) to exert influence, forcing the group to evaluate the delayed prospect more thoroughly. This mechanism may mitigate the discounting of delayed rewards, yielding a relatively modest decline in risky-intertemporal choices ($\triangle_3^{UT}$). In contrast, the majority rule allows members favoring immediate utility to dominate with minimal deliberation, amplifying the discounting effect and producing a sharper reduction in risky-intertemporal choices ($\triangle_3^{MT}$). From these discussions and Hypothesis 2c, it follows that $\triangle_3^{MT} < \triangle_3^{UT} < 0$. Given that individual-level delayed effects are invariant across treatments ($\triangle_1^{MT} \approx \triangle_1^{UT}$)[2], it follows that $\triangle_3^{UT} - \triangle_1^{UT} > \triangle_3^{MT} - \triangle_1^{MT}$.

Taken together, I predict that groups under the unanimity rule will exhibit significantly higher (or less negative) delayed effect values than those under the majority rule, indicating a stronger ability to resist the erosion of value caused by delay.

Hypothesis 2d (Delayed Effect Across Decision Rules):
The unanimity rule mitigates the negative impact of delay on risk-taking more effectively than the majority rule, namely $\triangle_3^{UT} - \triangle_1^{UT} > \triangle_3^{MT} - \triangle_1^{MT}$.

4.4.3 Heterogeneity of Risk Attitudes on Group Decisions

This study establishes risk-seeking behavior as the baseline (reference category) to examine how decision rules moderate the influence of risk-averse and risk-neutral members on collective decisions. Drawing on social choice theory—specifically the *Median Voter Theorem* (Black, 1948; Black et al., 1958) and *Veto Player Theory* (Tsebelis, 1995, 2002)—I derive the following predictions:

The unanimity rule grants every group member veto power (Buchanan and Tullock, 1965; Feddersen and Pesendorfer, 1998; Tsebelis, 2002). Under this institutional arrangement, collective decisions are constrained by a "weakest link" mechanism (Krick, 2017), where the threshold for consensus is determined by the most conservative member (Ambrus

[2] Due to the random assignment to each treatment, the individual decisions in the RC and RIC tasks are similar between treatments, namely $D_{RC,1}^{MT} \approx D_{RC,1}^{UT}$ and $D_{RIC,1}^{MT} \approx D_{RIC,1}^{UT}$. Thus, $D_{RIC,1}^{MT} - D_{RC,1}^{MT} \approx D_{RIC,1}^{UT} - D_{RC,1}^{UT} \Rightarrow \triangle_1^{MT} \approx \triangle_1^{UT}$.

et al., 2015; Buchanan and Tullock, 1965; König and Junge, 2008; Tsebelis, 1995). Relative to the risk-seeking baseline, risk-averse members possess a strong intrinsic motivation to exercise this veto to block high-risk proposals (Kahneman and Tversky, 1979; Masclet et al., 2009; Samuelson and Zeckhauser, 1988). This veto incentive is particularly potent in the RIC task: since the delay entails a discounting of future rewards, the subjective value of the risky option is further reduced. Consequently, risk-averse members are even more likely to perceive these delayed prospects as falling below their safety thresholds and block them.

In contrast, under the majority rule, the preferences of risk-averse members are susceptible to being overridden by coalitions formed by other members. According to the *Median Voter Theorem*, collective decisions under the majority rule converge to the preferences of the median member, effectively marginalizing individuals at the tails of the distribution (Black, 1948; Congleton, 2004). Consequently, provided risk-averse members do not constitute a majority, their resistance—whether driven by pure risk aversion or the devaluation of delayed outcomes—is rendered ineffectual by a coalition of risk-neutral and risk-seeking members capable of enforcing higher risk levels (Black, 1948; Buchanan and Tullock, 1965; Guinier and Carter, 1994).

Hypothesis 3a:
Under the unanimity rule, the negative impact of risk aversion on group risk-taking is the most pronounced compared to the risk-seeking baseline. This is because the unanimity rule grants risk-averse members veto power, allowing them to unilaterally block risky outcomes that deviate from their safety thresholds.

Risk-neutral members display distinct strategic alignments across decision rules, causing a shift in their positioning relative to the baseline. Under the majority rule, the decision-making process facilitates coalition formation. Here, risk-neutral members occupy the pivotal "median voter" position (Black, 1948; Black et al., 1958). Their alignment is driven by a rational evaluation of value. During deliberation, *Persuasive Arguments Theory* suggests that risk-neutral members are susceptible to arguments from risk-averse members (Burnstein and Vinokur, 1977; Rozin and Royzman, 2001; Vinokur and Burstein, 1974). This is particularly relevant in the RIC task: the introduction of delay discounts the expected value of the risky option. Risk-neutral members, who are typically sensitive to expected utility, perceive this "discounted risk" as objectively less attractive. Consequently, arguments emphasizing "safety" and "loss aversion" gain not only social validity but also economic weight (Kahneman and Tversky, 1979; Kameda and Davis, 1990; Tindale and Kameda, 2000; Tversky and Kahneman, 1991). Risk-neutral members are co-opted into a "prudent coalition", joining risk-averse members to counteract the aggressive strategies of the risk-seeking baseline. Statistically, this is reflected in a significantly negative coefficient for risk neutrality, which aligns closely in magnitude with risk aversion while diverging significantly from the risk-seeking baseline (Baker et al., 2008; Shupp and Williams, 2008).

Under the unanimity rule, the decision-making mechanism shifts from "coalition building" to a strict requirement for "consensus". This institutional constraint imposes high social costs on exercising veto power, including the risks of prolonged deadlock and group conflict (Banks and Duggan, 2013; Buchanan and Tullock, 1965). Although the delayed risky option carries a lower present value due to discounting, blocking it requires active obstruction. Unlike risk-averse members, who possess a strong intrinsic imperative to avoid potential losses, risk-neutral members lack the requisite intensity of preference to incur these obstruction costs. They face a trade-off between accepting a discounted risky payoff and paying the high transaction cost of a veto. Faced with the strong preferences of the risk-seeking baseline and the institutional pressure to reach an agreement, risk-neutral members are prone to strategic acquiescence (Feddersen and Pesendorfer, 1998). Lacking the internal motivation to actively block risk, they adopt a yielding strategy to facilitate consensus, effectively converging toward the risk-seeking baseline (De Dreu et al., 2000; Pruitt, 2013).

Hypothesis 3b:
Decision rules alter the strategic alignment of risk-neutral members relative to the risk-seeking baseline. Under the majority rule, risk-neutral members behave more like their risk-averse counterparts, forming a conservative coalition to moderate risk. Conversely, under the unanimity rule, risk-neutral members behave more like risk-seeking members, as consensus pressure drives them to acquiesce to higher risk levels.

4.4.4 Heterogeneity of Voting Pattern

First proposals (D_2) submitted by group members are categorized into three configurations: (i) two out of three proposals are the same (TTS), (ii) all proposals are the same (AS), and (iii) no proposals are the same (NS). These configurations may reflect distinct strategic orientations shaped by the decision environment. Each type of proposal can arise under both institutional treatments and task conditions. Decision rules are expected to exert heterogeneous effects on final group decisions across these configurations. Because AS indicates a unanimous first proposal and requires no further coordination, it is excluded from subsequent analysis.

TTS condition: Under the majority rule, final decisions require agreement from only a subset of members, incentivizing individuals to identify and align with the most probable "winning coalition"(Asch, 1951; Banerjee, 1992; Levitan and Verhulst, 2016; Wei et al., 2019). When first proposals follow a two-to-one configuration (TTS), the most direct and efficient strategy is to adopt the proposal already supported by two members. This behavior reflects path dependence and a coordination game dynamic, wherein the most commonly endorsed proposal becomes the natural focal point for resolution. The unanimity rule similarly promotes convergence toward the majority-supported proposal, as this configuration facilitates consensus with minimal negotiation costs. Accordingly, I expect comparable rates of selecting the majority-supported proposal across decision rules in the TTS condition.

Hypothesis 4a:
When facing the TTS condition, members are equally likely to select the majority-supported proposal under both decision rules.

Groups choose the majority-supported proposal as the final group decision in the TTS condition. One explanation is normative: individuals may adhere to the majority principle, whereby the member holding the minority-supported proposal adjusts their decision to align with the majority (Asch, 1951; Bicchieri, 2005; Davis, 1973; Deutsch and Gerard, 1955; Laughlin and Ellis, 1986). An alternative explanation is instrumental—majority-supported proposals may be intrinsically superior in terms of utility. In the RC task, safer proposals are attractive due to loss aversion. In the RIC task, this instrumental attraction is reinforced by time discounting: since delay reduces the expected value of risky options, the safer (and usually sooner) proposal becomes even more economically rational (Charness and Sutter, 2012; Vinokur and Burstein, 1974).

Let $D_{2,MA}^{TTS}$ and $D_{2,MI}^{TTS}$ indicate the number of risky options associated with the majority-supported and minority-supported proposals, respectively, in the TTS condition. Suppose the group selects the majority-supported proposal. The normative mechanism—adherence to the majority principle—can be distinguished from the instrumental account only if $D_{2,MA}^{TTS} \geq D_{2,MI}^{TTS}$. In this scenario, the majority proposal is not less risky (and in the RIC context, not less discounted) than the alternative. If the group still selects it, they are acting against the instrumental incentives of safety and immediate value, thereby isolating the normative motivation. By contrast, when $D_{2,MA}^{TTS} < D_{2,MI}^{TTS}$, the observed choice may indicate normative influence, instrumental reasoning (seeking safety or avoiding discounting), or a combination of both, but the underlying motivation cannot be clearly disentangled.

Hypothesis 4b (null):
If the majority-supported proposal is at least as risky as the minority-supported proposal ($D_{2,MA}^{TTS} \geq D_{2,MI}^{TTS}$) and is selected by the group, this outcome serves to isolate adherence to the majority principle.

Hypothesis 4c (alternative):
If the majority-supported proposal is less risky than the minority-supported proposal ($D_{2,MA}^{TTS} < D_{2,MI}^{TTS}$) and is selected, the underlying driver may be normative conformity, the intrinsic attraction of safer (and less discounted) proposals (instrumental explanation), or a combination thereof.

NS condition: In the NS condition, where all three members submit distinct proposals, no option holds an inherent advantage. Under the unanimity rule, avoiding impasse necessitates mutual concessions, driving the group toward the median proposal (moderate-risk). This natural compromise requires symmetric concessions from the most risk-seeking (or least delay-sensitive) and the most risk-averse (or most delay-sensitive) members, thereby maximizing the likelihood of collective acceptance. In the RIC context, this median proposal represents a balance point where the group trades off the cost of delay against the potential for higher returns.

Conversely, the majority rule requires agreement from only a subset of members, allowing individuals to prioritize proposals aligned with their own preferences. This dynamic increases the probability of deviation from the moderate-risk proposal, as a coalition can form around either the safe option (driven by discounting) or the risky option (driven by risk tolerance), bypassing the need for median compromise. Accordingly, I expect a higher incidence of selecting the moderate-risk proposal under the unanimity rule than under the majority rule in the NS condition.

Hypothesis 4d (Moderate-Risk Proposal):
When facing an NS condition, members under the unanimity rule are more likely to select the moderate-risk proposal than those under the majority rule.

4.5 Results

The results are organized into three parts. Section 4.5.1 confirms the balance of randomization. Section 4.5.2 reports the central analysis of treatment effects (Hypotheses 1 and 2). Finally, Sections 4.5.3 and 4.5.4 provide exploratory analyses of the mechanisms, focusing on individual heterogeneity (Hypothesis 3) and coordination dynamics (Hypothesis 4).

4.5.1 Overview Statistics

A total of 183 participants were recruited and randomly assigned to either the Majority Treatment ($N = 90$) or the Unanimity Treatment ($N = 93$). The experiment comprised a total of 22 sessions, with 10 sessions for MT and 12 sessions for UT. Table 4.3 summarizes participants' demographic characteristics. As shown in the table, the random assignment produced balanced treatment groups. The proportion of male participants was similar across the MT (45.6%) and UT (46.2%) conditions (Pearson's χ^2 test, $p > 0.1$). The mean age was also nearly identical (MT vs. UT: 26.9 vs. 26.7; Two-tailed t-test, $p > 0.1$). Furthermore, I find no significant differences in educational attainment (Wilcoxon Mann-Whitney test, $p > 0.1$) or ethnicity between the treatments (Pearson's χ^2 test, $p > 0.1$).

4.5.2 Individual and Group Decisions in the Tasks

This section presents the central findings regarding the causal impact of decision rules on collective outcomes. It begins by comparing individual and group decisions within each treatment, assessing whether the transition from individual to collective choice differs across the decision rules. It then analyzes the impact of delay across treatments, examining how the decision rule moderates the group's sensitivity to delay (the delayed effect).

Table 4.3: Summary Statistics

Specifications	Majority (MT)	Unanimity (UT)	All	MT v.s. UT (p value)
Male(=1)	0.456 (0.501)	0.462 (0.501)	0.459 (0.499)	0.926
Age	26.9 (8.398)	26.7 (8.660)	26.8 (8.509)	0.874
Education				0.737
Diploma	1.1%	3.2%	2.2%	
Bachelor	45.6%	34.4%	39.9%	
Master	33.3%	48.4%	40.9%	
Doctoral	16.7%	12.9%	14.8%	
Others	3.3%	1.1%	2.2%	
Ethnicity				0.489
White	82.2%	77.4%	79.8%	
Asian	15.6%	16.1%	15.9%	
Hispanic	1.1%	4.3%	2.7%	
Black	1.1%	2.2%	1.6%	
Observations	90	93	183	–

Note: Standard deviations in parentheses were calculated at the group level.
$Significance levels indicate at * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Changes between Individual and Group Decisions

Table 4.4 reports the average number of risky (or risky-delayed) options selected under each treatment. Specifically, D_1, D_2, and D_3 indicate the average number of risky options in individual decisions, first proposals, and final group decisions, respectively. The differences $D_2 - D_1$, $D_3 - D_2$, and $D_3 - D_1$ capture changes across decision stages.

Risky Choice (RC) Task. Under the majority rule (MT), the average number of risky options was 4.1 for D_1, 4.3 for D_2, and 4.3 for D_3, respectively. A matched-pair t-test revealed a significant increase from individual D_1 to the proposal stage D_2 ($p < 0.05$), suggesting that participants' first proposals were slightly riskier than their individual choices. However, final group decisions (D_3) did not differ significantly from individual preferences (D_1, $p > 0.1$).

Under the unanimity rule (UT), the average number of risky options was 4.4 for D_1, 4.3 for D_2, and 4.1 for D_3, respectively. The declines from D_1 to both D_2 and D_3 were marginally significant ($p < 0.1$), indicating that coordination under the unanimity rule tends to shift outcomes toward less risky decisions. Moreover, the reduction $D_3 - D_1$ was larger under UT than MT ($p < 0.1$), implying that the unanimity rule exerts a more pronounced moderating effect on risk-taking. This pattern likely reflects the requirement for full agreement, which may constrain group choices toward safe outcomes.

These findings provide partial support for Hypothesis 1, which states that group decisions (D_3) are less risky than individual ones (D_1). The unanimity rule appears to steer group decisions toward greater risk aversion, aligning with the expectation that consensus constraints favor safe options. In contrast, under the majority rule, lower coordination thresholds seem to preserve individual risk attitudes in collective resolutions.

Table 4.4: Average Number of Risky Choices by Treatment

Treatments	Tasks	Mean risky options			Mean differences		
		D_1	D_2	D_3	$D_2 - D_1$	$D_3 - D_2$	$D_3 - D_1$
MT	RC	4.12	4.3	4.27	0.18**	-0.03	0.14
		(0.912)	(0.794)	(0.980)	(0.079)	(0.131)	(0.156)
	RIC	3.16	3.3	3.2	0.14*	-0.1	0.04
		(1.099)	(1.037)	(1.472)	(0.099)	(0.196)	(0.229)
UT	RC	4.36	4.34	4.10	-0.01	-0.25*	-0.26*
		(0.890)	(0.841)	(1.193)	(0.068)	(0.153)	(0.171)
	RIC	3.31	3.31	3.39	0.00	0.08	0.08
		(1.128)	(0.977)	(1.334)	(0.085)	(0.159)	(0.187)

Note: Standard deviations for the average number of risky options were calculated at the group level, and standard errors for group differences were clustered accordingly. D_1, D_2, and D_3 represent the average number of risky options of individuals, the first group's decision and the final group's decision.
$Significance levels indicate at * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Risky-Intertemporal Choice (RIC) Task. Under MT, the average number of risky-delayed options was 3.2 for D_1, 3.3 for D_2, and 3.2 for D_3, respectively. As in the RC task, a marginally significant increase from D_1 to D_2 was observed ($p < 0.1$), while final decisions remained statistically indistinguishable from individual decisions. These results suggest that, under the majority rule, group decisions tend to align with individual preferences across both tasks. Under UT, the average number of risky-delayed options was 3.3 for both D_1 and D_2, and 3.4 for D_3. Matched-pair t-tests revealed no significant differences across stages, suggesting that the introduction of delay does not induce a further cautious shift under the unanimity rule.

These results of the RIC task do not provide evidence supporting Hypothesis 1. Under both the majority and unanimity rules, final group decisions do not exhibit statistically lower risk-taking compared to individual choices. This divergence from the RC task may be attributed to the discounting effect inherent in the RIC context. Since individuals already discount the value of the delayed risky option (reducing selections from 4.4 in the RC task to 3.3 here), the baseline for "caution" is already high. Notably, under the unanimity rule, the average number of risky-delayed options remains stable or slightly increases. This implies that while the unanimity rule tends to foster caution in pure risk settings, in intertemporal settings it appears to mitigate further devaluation of the delayed prospect, potentially helping to preserve the willingness to accept delayed risk against the erosion caused by discounting.

To examine the impact of decision rules on the change in risky (or risky-delayed) decisions from individuals to groups, I regressed the change variables ($D_{RC,3} - D_{RC,1}$ and $D_{RIC,3} - D_{RIC,1}$) on the majority treatment (MT) indicator and a set of controls. I employ OLS regressions as the primary estimation strategy throughout this chapter. Although the dependent variable is bounded between 0 and 10, I prioritize OLS because its coefficients provide a direct interpretation of marginal effects, facilitating straightforward comparison (Angrist and Pischke, 2009). To ensure robustness, I also estimated Ordered Probit models corresponding to each OLS specification presented in the chapter. These robustness checks, reported in Appendix A.3.1, yield qualitatively identical results to the OLS results; I thus prioritize the latter to simplify interpretation. Table 4.5 reports the results, with standard errors clustered at the group level.

From the table, Columns (1) and (2) present results from the RC task. The marginally significant and positive coefficient on *MT* reflects a directional difference in decision change between treatments. Specifically, the average shift is positive under MT ($D_3 - D_1 = 0.14$, $p > 0.1$) and negative under UT ($D_3 - D_1 = -0.26$, $p < 0.1$). The treatment effect is captured by $\beta_{MT} = 0.403$ ($p < 0.1$), suggesting a tendency for the final group decision under the majority rule to be riskier than that under the unanimity rule. This pattern aligns with the direction predicted by Hypothesis 2a. However, when additional controls are included in Column (2), the coefficient remains positive but becomes statistically insignificant ($\beta_{MT} = 0.377$, $p > 0.1$), indicating that this treatment effect is sensitive to individual heterogeneity and should be interpreted as a suggestive directional pattern rather than a robust causal estimate.

Table 4.5: OLS Estimates of Changes in Risky Choices ($D_3 - D_1$)

Specifications	Dependent variable: $D_3 - D_1$			
	RC task		RIC task	
	(1)	(2)	(3)	(4)
MT(=1)	0.403*	0.377	-0.031	-0.007
	(0.231)	(0.233)	(0.294)	(0.289)
Constant	-0.258	0.244	0.075	-0.936
	(0.170)	(1.064)	(0.186)	(1.127)
Observations	183	183	183	183
Controls	No	Yes	No	Yes
R-squared	0.0163	0.0559	0.0001	0.0777

Note: Standard errors are clustered at the group level. The term $D_3 - D_1$ denotes the difference in the number of risky-soon (or risky-delayed) choices between the final group decision and the corresponding individual decision. The variable "*MT*" is a binary indicator equal to 1 for the majority treatment and 0 for the unanimity treatment. Controls contain age, gender, education, and the Big Five personality traits.
$^{\$}$Significance levels indicate at * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Columns (3) and (4) report regressions for the RIC task, which incorporates both risk and delayed payoffs (see Table 4.5). The coefficient on *MT* is negative and insignificant, indicating the absence of a robust difference in $D_3 - D_1$ between treatments. Consequently, the data do not support Hypothesis 2b. This result reflects the observation that the transition from individual to group decisions ($D_3 - D_1$) is negligible under both decision rules (see Table 4.4). In other words, group decisions closely mirror individual preferences when risky payoffs are delayed. This suggests that in intertemporal risk contexts, the influence of voting rules may be attenuated by the salient discounting effect. As individuals appear to already incorporate the cost of delay into their valuations (resulting in a lower baseline of risky choices), there may be limited scope for voting rules to induce further shifts in risk attitudes.

Delayed Effect

This section examines how group decisions and decision rules impact the sensitivity to delay, offering an explanation for the significant treatment differences in the RC task ($D_{RC,3} - D_{RC,1}$) but the absence of such a difference in the RIC task ($D_{RIC,3} - D_{RIC,1}$) mentioned in the previous section. The dependent variable is the *Delayed Effect* ($\triangle_i^j = D^j_{RIC,i} - D^j_{RC,i}, i \in \{1, 3\}, j \in \{MT, UT\}$), which measures the erosion of risk-taking due to the discounting effect when a four-week payment delay is introduced.

To identify the impact of group interaction on discounting, I compared the delayed effect measured at the individual level ($\triangle_1^j = D^j_{RIC,1} - D^j_{RC,1}, j \in \{MT, UT\}$) with that of the final group decision ($\triangle_3^j = D^j_{RIC,3} - D^j_{RC,3}, j \in \{MT, UT\}$). The core of this analysis is the difference between these two measures ($\triangle_3^j - \triangle_1^j, j \in \{MT, UT\}$), which captures the net impact of group decision-making on the evaluation of delayed prospects after excluding individual preferences. This allows me to answer my central question in this section: Does the transition from individual to group decision-making alter the sensitivity to delay, and does this effect depend on the governing decision rule (majority versus unanimity)?

Figure 4.3 presents the main results on the delayed effect. Under the Majority Treatment (MT), the delayed effect is consistently negative (approx. -1) across individual decisions, first proposals, and the final group decision, consistent with Hypothesis 2c. Crucially, the difference between the delayed effect for individuals ($\triangle_1^{MT} = -0.97$) and for the final group decision ($\triangle_3^{MT} = -1.07$) is not statistically significant (Matched-pair *t*-test: $t = -0.43$, $p > 0.1$), indicating that the majority rule fails to mitigate the discounting effect observed at the individual level. In sharp contrast, results from the Unanimity Treatment (UT) indicate a significant reduction in the group's sensitivity to delay. While individuals still display a negative delayed effect ($\triangle_1^{UT} = -1.04$), the effect for final group decisions is significantly less negative ($\triangle_3^{UT} = -0.71$). This shift toward greater resistance to discounting is marginally significant (Matched-pair *t*-test: $t = 1.70$, $p < 0.1$). Taken together, these findings suggest the unanimity rule—unlike the majority rule—mitigates the group's sensitivity to the erosion of value caused by delay, aligning with the prediction of Hypothesis 2d.

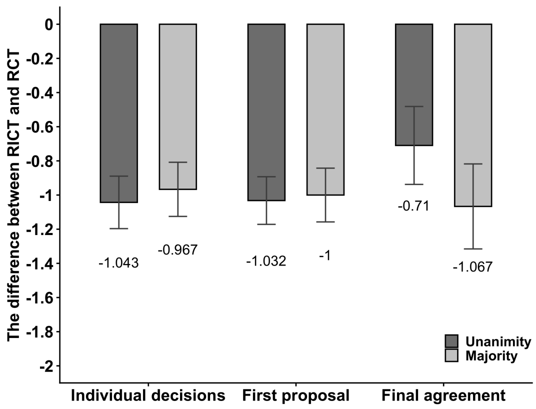


Figure 4.3: Differences between Risky-Intertemporal Choice and Risky Choice Tasks

I estimated a model in which the *Delayed Effect* is regressed on an indicator for the Majority Treatment (MT), an indicator for the Final Group Decision (FGD) stage, and their interaction, with standard errors clustered at the group level. Table 4.6 reports pooled panel regressions in Columns (1) and (2) and fixed effect estimates in Columns (3) and (4).

Table 4.6: Panel Estimates of the Delayed Effect ($D_{RIC} - D_{RC}$)

Specifications	Dependent variable: $D_{RIC} - D_{RC}$			
	All	All	MT	UT
	(1)	(2)	(3)	(4)
MT(=1)	0.032	0.008		
	(0.211)	(0.214)		
FGD(=1)	0.323*	0.323*	-0.067	0.323*
	(0.191)	(0.195)	(0.223)	(0.191)
MT×FGD	-0.389	-0.389		
	(0.295)	(0.289)		
Constant	-1.032***	-0.194	-1***	-1.032***
	(0.129)	(0.916)	(0.168)	(0.141)
Controls	No	Yes	No	No
Fixed effect	No	No	Yes	Yes
Observations	366	366	180	186
R-squared	0.0141	0.0141	0.0010	0.0299

Note: Standard errors clustered at the group level. MT is a dummy variable equal to 1 for the Majority Treatment and 0 for the Unanimity Treatment. FGD is a dummy variable equal to 1 for the stage of final group decisions and 0 for the preceding stages (i.e., individual decision or first proposal). Controls contain age, gender, education, and the Big Five personality traits.
$Significance levels indicated at * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Column (1) shows that the coefficient on the *FGD* dummy is positive and marginally significant ($\beta_{FGD} = 0.323$, $p < 0.1$), suggesting that groups tend to be less sensitive to the discounting effect than individuals under the unanimity treatment. This result maintains its magnitude and significance level after the inclusion of demographic controls and the

Big Five personality traits in Column (2). I then estimated the model separately for each treatment using fixed effects. In the Majority Treatment (Column (3)), the coefficient on *FGD* is small and statistically insignificant, consistent with my earlier finding that the majority rule does not mitigate the negative impact of delay. In the Unanimity Treatment (Column (4)), the coefficient on *FGD* is again positive and marginally significant, aligning with the pooled regression. These regressions reinforce the results from the simple matched-pair *t*-tests: the sensitivity to delay (or the erosion of risk-taking) appears to be mitigated in groups under the unanimity rule, but not under the majority rule.

4.5.3 The Effect of Individual Risk Attitudes on Group Decisions

Following the central analysis of treatment effects, this section conducts an exploratory analysis to investigate the underlying mechanisms—specifically, how individual risk attitudes shape final group decisions in both the RC and RIC tasks. I begin by comparing the distribution of risk attitudes across treatments, and then assessing how these attitudes relate to group outcomes under different decision rules. Finally, I examine whether the prevalence of a specific risk attitude within a group is correlated with its final decision, and whether this correlation varies by treatment.

Individual risk attitudes are classified into three categories: (i) risk averse ($D_1 < 5$), (ii) risk neutral ($D_1 \in \{5, 6\}$), and (iii) risk seeking ($D_1 > 6$), based on individual choices in the RC task. Table 4.7 presents the proportions of risk aversion (R.A.), risk neutrality (R.N.), and risk seeking (R.S.) under the majority (MT) and unanimity (UT) treatments. In MT, the shares of R.A., R.N., and R.S. are 53%, 41%, and 6%, respectively; in UT, the corresponding shares are 57%, 39%, and 4%. A Wilcoxon Mann-Whitney test reveals a statistically insignificant difference in the distribution of risk attitudes across treatments ($z = 0.621$, $p > 0.1$), confirming that random assignment yielded balanced groups in terms of baseline risk preferences.

Table 4.7: Distribution of Risk Preference Types

Treatments	Risk Aversion (R.A.)	Risk Neutrality (R.N.)	Risk Seeking (R.S.)
Majority Treatment (MT)	53%	41%	6%
Unanimity Treatment (UT)	57%	39%	4%
All	55%	40%	5%

Table 4.8 reports OLS regressions of final group decisions (D_3) on the majority treatment (MT), individual risk attitudes, interaction terms, and additional controls. In this analysis, the reference category comprises risk-seeking participants, while the other groups include risk-averse and risk-neutral individuals.

Table 4.8: OLS Estimates of Final Group's Decisions (D_3)

Specifications	Dependent variable: Final group's decisions			
	RC task (1)	RC task (2)	RIC task (3)	RIC task (4)
MT	0.833	0.947	0.75	0.715
	(0.747)	(0.784)	(0.982)	(0.989)
R.A.	-0.952*	-1.019*	-0.857**	-0.773**
	(0.543)	(0.583)	(0.346)	(0.384)
R.N.	-0.167	-0.209	-0.395	-0.344
	(0.561)	(0.605)	(0.367)	(0.415)
MT×R.A.	-0.528	-0.604	-0.912	-0.918
	(0.736)	(0.773)	(0.979)	(0.969)
MT×R.N.	-0.848	-0.993	-1.012	-0.915
	(0.745)	(0.785)	(1.065)	(1.051)
Constant	4.667***	4.962***	4***	4.700***
	(0.574)	(1.028)	(0.241)	(0.908)
Controls	No	Yes	No	Yes
Observations	183	183	183	183
R-squared	0.1290	0.1426	0.0610	0.0769
Wald test				
H_0: No significance in...				
...$\beta_{R.A.} = \beta_{R.N.}$	-0.785***	-0.81***	-0.462	-0.429
...$\beta_{R.A.} + \beta_{MT\times R.A.} = 0$	-1.48***	-1.623***	-1.769*	-1.691*
...$\beta_{R.N.} + \beta_{MT\times R.N.} = 0$	-1.015**	-1.202**	-1.407	-1.259
...$\beta_{R.A.} + \beta_{MT\times R.A.} = \beta_{R.N.} + \beta_{MT\times R.N.}$	-0.465**	-0.421**	-0.362	-0.432

***Note*:** Standard errors are clustered at the group level. The variables "*MT*", "*R.A.*", and "*R.S.*" are binary indicators. "MT" equals to 1 for the majority treatment and 0 for the unanimity treatment. "*R.A.*" equals to 1 if the participant is classified as risk averse, and 0 otherwise (i.e., risk neutral or risk seeking). "*R.S.*" equals to 1 if the participant is classified as risk seeking, and 0 otherwise (i.e., risk neutral or risk averse). Controls contain age, gender, education, and the Big Five personality traits.

[$]The reference category is *Risk-Seeking Participants*.

[$]Significance levels indicate at * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Columns (1) and (2) present results from the RC task. In Column (1), the coefficient on risk aversion (R.A.) is marginally significant and negative ($\beta_{R.A.} = -0.952$, $p < 0.1$), suggesting that, under the unanimity rule, risk-averse participants tend to be associated with fewer risky options in final group decisions relative to risk seekers. The coefficient on risk neutrality (R.N.) under the unanimity rule is statistically insignificant ($\beta_{R.N.} = -0.167$, $p > 0.1$), suggesting no meaningful difference from the baseline. Under the majority rule, the marginal effects of risk aversion and risk neutrality are given by $\beta_{R.A.} + \beta_{MT \times R.A.}$ and $\beta_{R.N.} + \beta_{MT \times R.N.}$, respectively. A Wald test reveals a significant difference between risk aversion and risk neutrality under the unanimity rule ($\beta_{R.A.} - \beta_{R.N.} = -0.785$, $p < 0.01$). Under the majority rule, the test provides suggestive evidence that risk aversion is associated with fewer risky options than risk neutrality ($\beta_{R.A.} + \beta_{MT \times R.A.} - \beta_{R.N.} - \beta_{MT \times R.N.} = -0.465$, $p < 0.05$). However, it is important to note that the interaction terms $\beta_{MT \times R.A.} = -0.528$ and $\beta_{MT \times R.N.} = -0.848$ do not reach statistical significance ($p > 0.1$). This indicates that, statistically, I cannot reject the null hypothesis that the effects of risk attitudes are identical across treatments. Thus, the data do not provide robust support for Hypothesis 3a regarding the dominance of risk aversion. Nevertheless, the qualitative patterns offer an exploratory insight consistent with the logic of Hypothesis 3b: under the majority rule, risk-neutral members appear to align with risk-averse members (both exhibiting negative coefficients), whereas under the unanimity rule, risk-neutral behavior remains indistinguishable from the risk-seeking baseline. Column (2), which includes additional controls, yields qualitatively similar patterns, suggesting that these observations are not driven by observable demographic shifts.

Columns (3) and (4) report regressions for the RIC task, which incorporates both risk and delayed payoffs. In Column (3), the coefficient on risk aversion remains significantly negative ($\beta_{R.A.} = -0.857$, $p < 0.05$), suggesting that risk-averse participants are associated with fewer risky-intertemporal choices under the unanimity rule. The coefficient on risk neutrality is again insignificant ($\beta_{R.N.} = -0.395$, $p > 0.1$). Under the majority rule, the marginal effect of risk aversion is negative and marginally significant ($\beta_{R.A.} + \beta_{MT \times R.A.} = -1.769$, $p < 0.1$), while the effect of risk neutrality remains statistically insignificant ($\beta_{R.N.} + \beta_{MT \times R.N.} = -1.407$, $p > 0.1$). This statistical insignificance suggests that risk-neutral members tend to behave similarly to the risk-seeking baseline. These results offer partial exploratory support for Hypothesis 3b in the RIC context. Risk-neutral members align with the risk-seeking baseline under the unanimity rule, as predicted, but do not exhibit a statistically significant shift toward risk aversion under the majority rule. This deviation suggests that the introduction of delay may alter the valuation process. Rather than signaling unstable preferences, it appears that the discounting of delayed rewards dampens the strategic differentiation between risk-neutral and risk-seeking members. The strong discounting effect likely overshadows the coalition incentives that were observed in the pure risk context.

Table 4.9 presents OLS regressions of final group decisions (D_3) on the number of risk-averse ($N_{R.A.}$) and risk-neutral ($N_{R.N.}$) participants, along with additional controls. These regressions examine how the composition of individual risk attitudes within a group relates to collective outcomes. Given that each group consists of three members ($0 \leq N_i \leq 3, i \in \{R.A., R.N., R.S.\}$) and $N_{R.A.} + N_{R.N.} + N_{R.S.} = 3$, only two attitude types are included in each specification, with risk-seeking participants as the reference category.

Columns (1)-(4) report regressions for the RC task under the majority (MT) and unanimity (UT) rules. Under MT, both $N_{R.A.}$ and $N_{R.N.}$ are significantly and negatively correlated with the number of risky options in final group decisions ($\beta_{N_{R.A.}} = -1.113$, $p < 0.01$; $\beta_{N_{R.N.}} = -0.634$, $p < 0.01$). A Wald test indicates that risk aversion exerts a significantly stronger (more negative) marginal effect than risk neutrality ($\beta_{N_{R.A.}} - \beta_{N_{R.N.}} = -0.479$, $p < 0.01$). These patterns remain qualitatively consistent after including additional controls (Column (2)).

Under UT, the baseline OLS specification (Column (3)) initially shows a marginally significant coefficient for $N_{R.N.}$ ($\beta_{N_{R.N.}} = -0.898$, $p < 0.1$). However, this fails to detect a significant effect when controls are added (Column (4): $\beta_{N_{R.N.}} = -0.806$, $p > 0.1$), and notably, robustness checks using an Ordered Probit model fail to detect a significant effect even in the baseline specification (see Table A.5 in Appendix A.3.1). This discrepancy suggests that the influence of risk-neutral members under UT is highly sensitive to model specification and likely not robust. Therefore, while risk aversion consistently drives safe choices under unanimity, the specific role of risk-neutral members remains statistically ambiguous in this context.

Table 4.9: OLS Estimates of Final Group's Decisions (D_3) by Tasks and Treatments

Specifications	Dependent variable: Final group's decisions							
	RC task				RIC task			
	MT	MT	UT	UT	MT	MT	UT	UT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$N_{R.A.}$	-1.113***	-1.173***	-1.319***	-1.285***	-1.225***	-1.222***	-0.915***	-0.935***
	(0.191)	(0.178)	(0.432)	(0.451)	(0.352)	(0.362)	(0.309)	(0.309)
$N_{R.N.}$	-0.634***	-0.699***	-0.898*	-0.806	-0.348	-0.281	-0.377	-0.336
	(0.189)	(0.178)	(0.502)	(0.520)	(0.297)	(0.289)	(0.311)	(0.310)
Constant	6.929***	8.023***	6.518***	7.591***	6.455***	6.574***	5.276***	6.055***
	(0.464)	(1.047)	(1.284)	(1.939)	(0.857)	(1.268)	(0.626)	(1.295)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	90	90	93	93	90	90	93	93
R-squared	0.3379	0.3700	0.1732	0.2240	0.4079	0.4319	0.1463	0.1725
Wald test								
H_0: No significant in...								
...$\beta_{N_{R.A.}} - \beta_{N_{R.N.}} = 0$	-0.479***	-0.474***	-0.421**	-0.479**	-0.877***	-0.941***	-0.538*	-0.599*

***Note*:** Standard errors are clustered at the group level. $N_{R.A.}$ and $N_{R.N.}$ denote the number of risk-averse and risk-neutral participants within each group, respectively. The baseline group is the number of risk seeking participants within each group. Controls contain age, gender, education, and the Big Five personality traits.

$The reference category is *Risk-Seeking Participants*.

$Significance levels indicate at * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Columns (5)-(8) report regressions for the RIC task. Across both decision rules, the marginal effect of $N_{R.A.}$ remains significantly negative, while the effect of $N_{R.N.}$ is statistically insignificant. This stands in partial contrast to the RC task, where risk-neutral members exerted a more distinct influence, particularly under the majority rule. Despite the overall insignificance of $N_{R.N.}$, Wald tests consistently reveal significant differences between the effects of $N_{R.A.}$ and $N_{R.N.}$. This indicates that even when delay is introduced, the presence of risk-averse members continues to exert a significantly stronger conservative pull compared to risk-neutral members. The results suggest that while the influence of risk-neutral members diminishes—likely because the discounting effect overshadows subtle strategic alignments—risk aversion remains a robust driver in shaping group decisions.

4.5.4 Coordination Mechanism

This section conducts an exploratory analysis of collective behavior to reveal the underlying mechanisms of consensus formation. The analysis proceeds in two stages. First, I examine coordination efficiency—specifically success and failure rates—to rule out the concern that the observed behavior is driven merely by a strategic avoidance of coordination failure. Having established that coordination is highly efficient across treatments, I then investigate how group members reach agreement, with particular attention to the mediating role of voting rules. The analysis tracks the decision-making process from initial proposals (D_2), through communication, to the final group decision.

Coordination Efficiency

To assess whether the observed choices were driven by a strategic avoidance of coordination failure (i.e., fear of obtaining zero payoff), I examined the frequency of failures and the number of rounds required to reach a decision.

The data indicate that coordination was highly efficient across all treatments. Coordination failure was entirely absent, with 100% of groups eventually reaching a final decision. Specifically, under the majority rule, 100% (30/30) of groups reached a decision in the first round in both the RC and RIC tasks. Similarly, under the unanimity rule, convergence was rapid and decisive: 94% (29/31) of groups reached agreement in the first round, with the remaining 6% (2/31) concluding in the second round.

The absence of prolonged bargaining or failure suggests that the greater risk aversion observed under the unanimity rule is unlikely to be a product of negotiation fatigue or fear of breakdown. Instead, this high rate of immediate agreement is likely facilitated by the communication mechanism, which enables members to exchange information and reduce strategic uncertainty prior to voting. By allowing participants to signal their preferences and constraints, communication reinforces the unanimity rule's function as a coordination device that clarifies the focal point. Participants appear to anticipate the constraint of full agreement and, aided by this pre-vote information exchange, strategically coordinate on the "safest common denominator"—typically the moderate or safe option—immediately at the initial stage. This interaction between communication and institutional constraints explains why groups exhibit cautious shifts without engaging in extended rounds of negotiation.

Mechanism Analysis

Proposal structures in D_2 are categorized into three types: TTS, AS, and NS,[3] as referenced earlier in the hypotheses. The subsequent exploratory analysis investigates how group members make choices to successfully coordinate and reach a collective decision when facing these three distinct proposal scenarios.

Figure 4.4 shows the distribution of AS, NS, and TTS across the RC and RIC tasks. In the RC task (Figures 4.4(a)-(b)), the shares under the majority rule (MT) are 3% (AS), 47% (NS), and 50% (TTS); under the unanimity rule (UT), the corresponding shares are 6%, 45%, and 48%. A Pearson's χ^2 test reveals no statistically significant difference ($\chi^2(2) = 0.951$, $p > 0.1$), indicating that the aggregate patterns of coordination are statistically indistinguishable across the rules in this context.

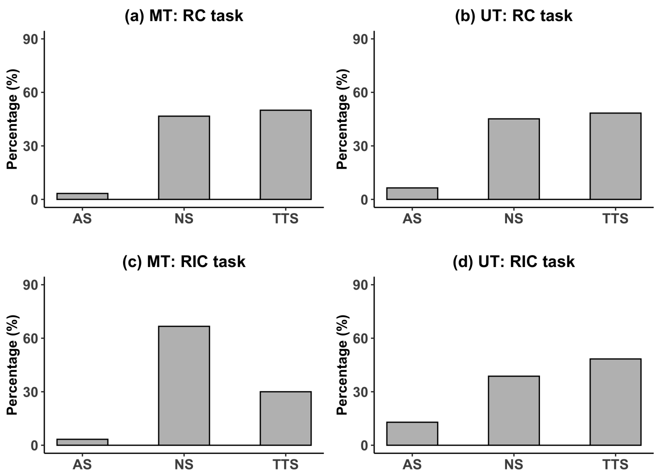


Figure 4.4: Distribution of AS, NS, and TTS

In the RIC task (Figures 4.4(c)-(d)), MT yields 3% (AS), 67% (NS), and 30% (TTS); UT yields 13%, 39%, and 48%, respectively. A statistically significant difference is detected (Pearson's χ^2 test: $\chi^2(2) = 17.216$, $p < 0.01$), driven by the higher prevalence of NS under the majority rule and a relatively higher share of coordination attempts (TTS) under the unanimity rule. This pattern suggests that institutional rules influence behavioral dynamics even at the first proposal stage: with delayed payoffs, participants under the majority rule appear more likely to submit highly heterogeneous proposals. This behavior is consistent with a strategic response where participants anticipate that a simple

[3] TTS indicates two out of three proposals are the same, AS indicates all proposals are the same, and NS indicates no proposals are the same. This classification reflects strategic variation across decision environments.

majority suffices, thereby reducing the immediate pressure to coordinate compared to the unanimity condition.

When facing the TTS or NS conditions, group members must engage in deliberation to identify a mutually acceptable proposal prior to voting. The final vote thus serves as a tractable and interpretable proxy for the group's coordinated outcome. While the discussion process remains a conceptual "black box", I treat the vote as a behavioral signal of consensus, allowing for analytical parsimony without sacrificing interpretive value.

Figure 4.5 depicts the distribution of voting outcomes under TTS and NS conditions across decision rules. Figures 4.5(a)-(b) show the outcome shares of the TTS condition for the RC and RIC tasks, respectively. Note that in the TTS condition (a two-to-one configuration), the final outcome presents two possibilities: selection of the majority-supported proposal or the minority-supported proposal. After communication, group members cast votes to determine the final decision.

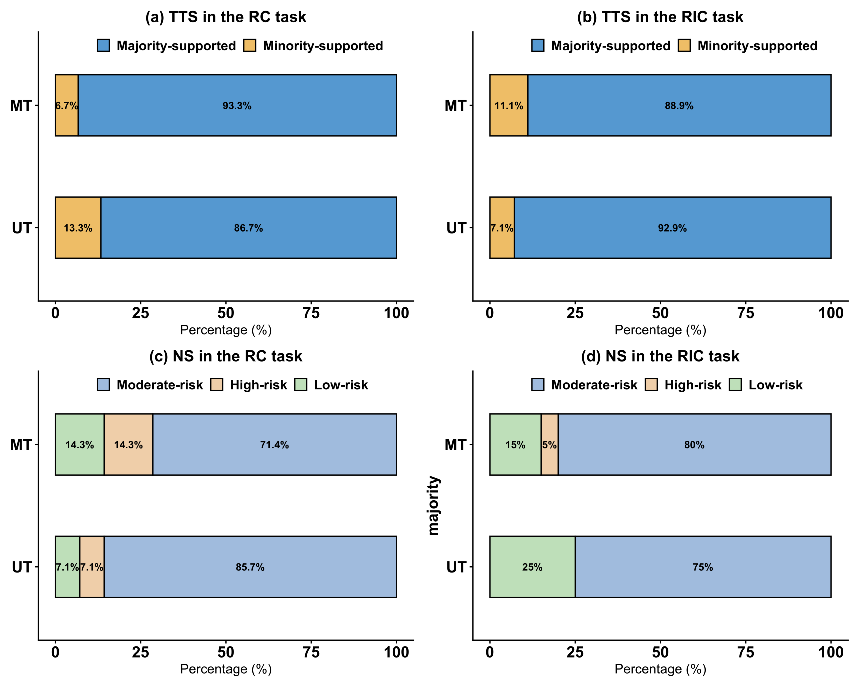


Figure 4.5: The Percentage of Proposals Chosen in Final Group's Decisions

TTS Condition: In the RC task, the majority-supported proposal was chosen in 93% of cases under the majority rule (MT) and 87% under the unanimity rule (UT). The corresponding percentages were 89% for MT and 93% for UT in the RIC task. A Fisher's exact test reveals no statistically significant difference between treatments in either task ($p > 0.1$). These results suggest that, regardless of the voting rule, groups consistently converge on the majority-supported proposal when initial preferences are partially aligned, aligning with Hypothesis 4a.

I next examine Hypotheses 4b and 4c to explore the motivations—normative conformity versus instrumental preferences (including the avoidance of risk or discounting)—

for choosing the majority-supported proposal. Specifically, I employ two-tailed t-tests to assess the difference in the risk levels between the majority-supported ($D_{2,MA}$) and minority-supported ($D_{2,MI}$) proposals, conditional on the majority-supported proposal being selected as the final group decision ($D_{3,MA}$). Table 4.10 summarizes the average number of risky choices for these proposals across treatments and tasks in the TTS condition.

Table 4.10: Average Number of Risky Choices by Proposal Support: Majority vs. Minority (TTS Condition)

Task	Decision Rule	$D_{2,MA}$	$D_{2,MI}$	$D_{2,MA}$ vs. $D_{2,MI}$	Two-tailed t-test (p-values)
RC Task (Pure Risk)	Majority Rule	4.14 (1.008)	4.57 (1.225)	$D_{2,MA} < D_{2,MI}$	0.117
	Unanimity Rule	4.69 (0.838)	4.92 (1.320)	$D_{2,MA} < D_{2,MI}$	0.255
RIC Task (Delayed Risk)	Majority Rule	3.88 (1.408)	2.88 (1.246)	$D_{2,MA} > D_{2,MI}$	0.052*
	Unanimity Rule	3.92 (1.093)	2.62 (1.938)	$D_{2,MA} > D_{2,MI}$	0.005***

Note: Standard deviations were clustered at the subject level. $D_{2,MA}$ and $D_{2,MI}$ represent average numbers of risky options in the majority-supported and the minority-supported proposal in the final group decision, respectively.
$^{\$}$Significance levels indicate at * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Table 4.10 shows that majority-supported proposals were nominally less risky than minority-supported proposals under both the majority (MT: 4.1 vs. 4.6; two-tailed t-test: $t = -1.21$, $p > 0.1$) and unanimity rules (UT: 4.7 vs. 4.9; two-tailed t-test: $t = -0.66$, $p > 0.1$), though neither difference was statistically significant. These results suggest that the selection of majority-backed proposals was unlikely driven by instrumental risk avoidance (Hypothesis 4c), as the majority options offered no significant safety advantage. Instead, the pattern is consistent with normative alignment (Hypothesis 4b).

In the RIC task, majority-supported proposals involved more risky-delayed choices than minority-supported proposals. This difference was marginally significant under the majority rule (MT: 3.9 vs. 2.9; two-tailed t-test: $t = 1.69$, $p < 0.1$) and significant under the unanimity rule (UT: 3.9 vs. 2.6; two-tailed t-test: $t = 2.71$, $p < 0.01$). These results indicate that group members favored majority-backed proposals even when they entailed greater risk and were subject to time discounting. This finding provides suggestive evidence for normative conformity: groups adhered to the majority principle even when doing so contradicted the instrumental incentives to select the safer and less discounted minority option. This observation reinforces the logic of Hypothesis 4b (the null).

NS Condition: Figures 4.5(c)-(d) show the distribution of final group decisions under the NS condition, categorized by proposal risk level (low, moderate, and high) across decision rules and tasks. Risk levels were defined relative to the proposal set within each group.

When facing the NS condition, in the RC task (Figure 4.5(c)), groups under the unanimity rule selected the moderate-risk proposal significantly more often (85.7%) than groups under the majority rule (71.4%). A marginally significant difference in moderate-risk selection (Fisher's exact test: $p < 0.1$) suggests that the unanimity rule tends to favor convergence on compromise options. This pattern is consistent with strategic coordination: the rule necessitates full agreement, likely positioning the moderate-risk proposal as the most viable middle ground to facilitate consensus. This offers partial support for Hypothesis 4d.

In the RIC task (Figure 4.5(d)), the preference for moderate-risk proposals remains stable across decision rules (80% for majority vs. 75% for unanimity). No significant difference was detected (Fisher's exact test: $p > 0.1$), which does not support Hypothesis 4d. Notably, the proportion of groups selecting the low-risk proposal under the unanimity rule (25%) is nominally higher than that under the majority rule (15%), although this difference implies only a descriptive trend rather than a statistical distinction. Despite the lack of significance, a potential mechanism warrants discussion: the introduction of delay may shift the focal point of compromise. Due to the discounting of future rewards, the subjective value of moderate-risk proposals is reduced. Consequently, members—especially those highly sensitive to discounting—might view the low-risk option as the more viable consensus point, effectively pulling the group outcome toward the safer end of the spectrum. This dynamic may attenuate the expected centralizing effect of unanimity in intertemporal contexts.

4.6 Conclusion and Discussion

This study examines how institutional decision rules (majority versus unanimity voting) shape collective choices in risky and intertemporal contexts. Using a modified Holt and Laury (2002) framework, I implement a task where safe options yield soon payoffs and risky options involve a four-week delay. In a lab experiment with 183 participants, central findings indicate that the cautious shift is rule-dependent: groups under the unanimity rule tend to exhibit greater risk aversion than individuals in the risky task. However, this institutional effect appears context-specific. While the unanimity rule seems to mitigate the erosion of value in intertemporal settings compared to the majority rule, the data do not support a robust divergence between the two rules in the presence of delay. This suggests that the strong cognitive bias of time discounting may attenuate the strategic differentiation usually induced by voting rules.

Exploratory analyses further suggest that individual risk preferences and coordination dynamics play a role in shaping these outcomes. Specifically, individual risk aversion is negatively associated with the number of risky choices, though the influence of risk neutrality appears sensitive to context and specification. Regarding coordination, the unanimity rule shows a tendency to favor convergence on moderate-risk proposals, whereas groups under the majority rule are more likely to select majority-supported proposals. These findings highlight how institutional rules mediate the aggregation of risk and intertemporal trade-offs, offering micro-foundational insights for financial decision-making under uncertainty.

From a policy perspective, these findings have salient implications for institutional design in settings where collective decisions are critical—such as corporate boards, regulatory committees, and financial institutions. While the majority rule facilitates faster resolutions, it risks marginalizing minority preferences and allowing short-term discounting to dominate. In contrast, the unanimity rule promotes inclusiveness and preserves the value of delayed prospects, albeit often at the expense of efficiency. This suggests that unanimity-based mechanisms may be particularly valuable in contexts requiring resistance to short-termism, ensuring that long-term strategic goals are not undervalued due to immediate pressures. Hybrid mechanisms combining elements of both rules may offer a balanced approach, particularly in contexts requiring both decisiveness and minority protection.

This study contributes to the literature on collective decision-making by delineating the boundary conditions of institutional influence. While prior work (Baker et al., 2008; Shupp and Williams, 2008) demonstrates that institutions condition risk-taking, my findings suggest that this conditioning power is not absolute. I identify a critical distinction: voting rules (particularly unanimity) are highly effective in mediating preference-based conflicts (such as differing risk appetites) but appear less effective in mitigating biases stemming from time discounting. This implies that the "cost of delay" acts as a dominant constraint that compresses the strategic space in which voting rules operate.

While this study focuses on equal-weight voting, future research should explore weighted voting systems, where asymmetries in voting power may introduce additional heterogeneity. The interaction between voting weights and decision rules in shaping group preferences warrants further investigation. Moreover, expanding the sample beyond university students to include more diverse populations, such as older adults or professionals, would enhance the external validity of the findings and offer deeper insights into real-world decision-making environments.

Chapter 5 Outlook

5.1 Summary: The Triple Dimensions of Context

This thesis begins with a central challenge: the classical economic assumption of stable and exogenous preferences fails to account for the complexity of real-world decision-making. It argues that choice context is not mere noise but a fundamental determinant of economic behavior. Through three interlinked experimental studies, the thesis develops a three-tier analytical framework—from individuals to groups—that systematically reveals the malleability of preferences, the moderating role of cultural norms, and the convergence effects of institutional design. Key findings include:

1. **Individual Level:** Social information (e.g., others' maximum donations) dynamically reshapes individuals' distributional preferences, inducing discontinuous shifts.
2. **Collective Action Level:** Cultural orientation (collectivism vs. individualism) significantly moderates the effectiveness of normative signals, giving rise to two distinct cooperation mechanisms: strategic imitation and endogenous compliance.
3. **Group Decision Level:** Formal rules (majority vs. unanimity) are not neutral in aggregating group preferences. They systematically influence outcomes in risk and intertemporal choices by amplifying or suppressing minority preferences—particularly those of the most cautious members.

While these three layers have been studied separately, the thesis's broader implication is that future research must move beyond this segmentation. A more ambitious agenda is needed: to investigate how informational, cultural, and institutional contexts interact dynamically within a unified framework. This chapter outlines an integrative research path aimed at bridging these dimensions to build a more context-sensitive model of economic decision-making.

5.2 Integrative Agenda I: The Cultural Sensitivity of Social Information

Chapter 2 demonstrates that social information—specifically, the "maximum giving" signal—can impact distributional preferences, though its underlying mechanism remains opaque. Chapter 3 shows that cultural norms significantly shape group behavior. A key unresolved question is whether the mechanism through which social information operates is itself culturally contingent.

The "maximum giving" signal used in Chapter 2 conveys achievement and high standards, which may be particularly salient in individualistic cultures such as Germany. In contrast, collectivist cultures like China may respond more strongly to signals emphasizing average behavior or common consensus (e.g., "80% of participants donated X"), which align with norms of conformity and harmony.

Future research could design a cross-cultural experiment combining the CES preferences elicitation method from Chapter 2 with the cultural manipulation from Chapter 3. This would allow a direct comparison of the relative effectiveness of different types of social information across cultural contexts:

- T1 (Individualistic framing): Participants are shown the highest donation amount.
- T2 (Collectivist framing): Participants are shown the average donation or the most common donation behavior.

This approach reframes social information not as a uniform input but as a culturally embedded signal. The central hypothesis is that social information must align with local normative expectations to maximize its behavioral impact. This has important implications for behavioral economics and nudge policy: there is no universally optimal nudge. Effective informational interventions require cultural tailoring.

5.3 Integrative Agenda II: The Institutional-Cultural Fit Hypothesis

Chapter 4 examines how majority and unanimity rules shape preference aggregation, while Chapter 3 highlights cultural differences in conflict resolution and norm compliance. This raises a deeper question: Is the effectiveness of institutional rules—such as voting mechanisms—contingent on the cultural context in which they operate?

I propose an "institutional-cultural fit" hypothesis. In individualistic cultures (e.g., the German sample), individuals may prioritize efficiency, decisiveness, and the expression of personal interests. The majority rule may thus be perceived as more efficient and democratic, even if it produces clear losers. In collectivist cultures (e.g., the Chinese sample), individuals may value group harmony, consensus, and the avoidance of overt conflict ("face culture"). The unanimity rule may be preferred, as it compels negotiation and consensus-building, minimizing public defeat of minority views—albeit at the cost of decision efficiency.

Future research could replicate the risk preference aggregation experiment from Chapter 4 within the cross-cultural framework of Chapter 3. Key outcome measures should include not only the group's final decision (e.g., patience or risk aversion), but also:

1. Time required to reach consensus.

2. Likelihood of decision failure (impasse).
3. Perceived procedural fairness and participant satisfaction.

This integrative approach moves beyond asking which rule is "better", and instead investigates which rule is more effective—or more resisted—within specific cultural contexts. The findings would offer practical guidance for corporate governance, international negotiations, and multicultural team management.

5.4 Integrative Agenda III: A Composite Experiment—The Dynamic System of Contextual Interaction

The ultimate goal of this paper is to understand the "triad of context". Future higher-level research should design a composite experiment that simultaneously manipulates all three dimensions—information, culture, and institutions—to observe their interaction effects.

Consider a simulated "corporate board" experiment with the following structure:

1. Institutional Dimension: Boards are randomly assigned to either the majority rule or the unanimity rule decision-making.
2. Cultural Dimension: The experiment is conducted in both Germany and China.
3. Informational Dimension: Boards must decide on the allocation between profit and corporate social responsibility (CSR), modeled as a public goods investment (similar to the dictator game in Chapter 2). Prior to voting, participants receive different types of industry-level social information:

    ◦ T1 (Competitive framing): "Your main competitor, Company A, invested a maximum of X million in CSR this quarter."
    ◦ T2 (Compliance framing): "The industry average CSR investment this quarter is Y million."

The design enables the examination of complex interaction effects. For instance, will a Chinese board operating under the unanimity rule (emphasizing harmony and consensus) respond most strongly to compliance-based information, resulting in the highest CSR investment? Conversely, will a German board under the majority rule (emphasizing individualism and efficiency) react most strongly to competitive signals? Such research treats social context as a dynamic system rather than a simple additive model of three independent variables.

5.5 Broader Implications: Algorithmic Governance in Digital Contexts

The analytical framework in this study offers critical insights into one of the most pressing challenges of our time—algorithmic governance. In the digital era, the context of choice is increasingly shaped by artificial intelligence (AI) systems. Recommendation engines, content feeds, and social media platforms have become primary sources of "social information", deeply influencing cultural norms and being embedded within digital institutional structures.

The thesis's discussion of "black boxes"—whether concerning how information reshapes preferences or how signals exert influence—becomes even more salient in the age of AI. Future research must address the following questions:

1. How do algorithmically generated social signals systematically—and potentially discriminatorily—reshape human preferences?
2. How can we design culturally aware AI systems that respect rather than erode diverse normative frameworks?
3. In algorithmically curated information environments, how do macro-level collective decision mechanisms (e.g., democratic voting) retain efficacy or become subject to manipulation?

In sum, this thesis begins with the premise that context matters and ultimately points toward a broader future. Understanding the interplay among information, culture, and institutions is not only a frontier in behavioral economics and sociology, but also a necessary foundation for designing fair, effective, and human-centered governance systems in an increasingly complex digital society.

References

Abadeer, A. S. (2015). Norms in collectivist versus individualist societies. In *Norms and Gender Discrimination in the Arab World*, pages 85–101. Springer.

Abdellaoui, M., Diecidue, E., and Öncüler, A. (2011). Risk preferences at different time periods: An experimental investigation. *Management Science*, 57(5):975–987.

Afriat, S. (1967). nthe construction of a utility function from expenditure datao. *International Economic Review*, 8:76–77.

Afriat, S. N. (1972). Efficiency estimation of production functions. *International economic review*, pages 568–598.

Afriat, S. N. (1987). Logic of choice and economic theory. *(No Title)*.

Agerström, J., Carlsson, R., Nicklasson, L., and Guntell, L. (2016). Using descriptive social norms to increase charitable giving: The power of local norms. *Journal of Economic Psychology*, 52:147–153.

Allcott, H. (2011). Social norms and energy conservation. *Journal of public Economics*, 95(9-10):1082–1095.

Allik, J. and Realo, A. (2004). Individualism-collectivism and social capital. *Journal of cross-cultural psychology*, 35(1):29–49.

Almås, I., Cappelen, A. W., Sørensen, E. Ø., and Tungodden, B. (2025). Fairness across the world.

Alt, J. E. and Shepsle, K. A. (1998). Rules, restrictions, constraints: Structure and process in the new institutional economics. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, 154(4):735–743.

Altemeyer, B. (1981). *Right-wing authoritarianism*. Univ. of Manitoba Press.

Amato, C., Gino, F., Montinari, N., and Sacco, P. (2020). Cheating, inequality aversion, and appealing to social norms. *Journal of Economic Behavior & Organization*, 179:767–776.

Ambrus, A., Greiner, B., and Pathak, P. A. (2015). How individual preferences are aggregated in groups: An experimental study. *Journal of public economics*, 129:1–13.

Anderhub, V., Güth, W., Gneezy, U., and Sonsino, D. (2001). On the interaction of risk and time preferences: An experimental study. *German Economic Review*, 2(3):239–253.

Anderson, L. R. and Stafford, S. L. (2009). Individual decision-making experiments with risk and intertemporal choice. *Journal of Risk and Uncertainty*, 38(1):51–72.

Andreoni, J. (1995). Cooperation in public-goods experiments: kindness or confusion? *The American Economic Review*, pages 891–904.

Andreoni, J. and Miller, J. (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2):737–753.

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

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Aquino, K. and Reed II, A. (2002). The self-importance of moral identity. *Journal of personality and social psychology*, 83(6):1423.

Arrow, K. J. (2020). *Social choice and individual values*. Barakaldo Books.

Asch, S. (1951). Effects of group pressure upon the modification and distortion of judgments.

Asheim, G. B., Claussen, C. A., and Nilssen, T. (2006). Majority voting leads to unanimity. *International Journal of Game Theory*, 35(1):91–110.

Bagaïni, A., Liu, Y., Kapoor, M., Son, G., Bürkner, P.-C., Tisdall, L., and Mata, R. (2025). A systematic review and meta-analyses of the temporal stability and convergent validity of risk preference measures. *Nature Human Behaviour*, 9(4):700–712.

Baharad, E., Ben-Yashar, R., and Nitzan, S. (2020). Variable competence and collective performance: unanimity versus simple majority rule. *Group Decision and Negotiation*, 29(1):157–167.

Baillon, A., Bleichrodt, H., Liu, N., and Wakker, P. P. (2016). Group decision rules and group rationality under risk. *Journal of Risk and Uncertainty*, 52(2):99–116.

Bainbridge, S. M. (2002). Why a board-group decisionmaking in corporate governance. *Vand. L. Rev.*, 55:1.

Baker, R. J., Laury, S. K., and Williams, A. W. (2008). Comparing small-group and individual behavior in lottery-choice experiments. *Southern Economic Journal*, 75(2):367–382.

Balliet, D., Wu, J., and De Dreu, C. K. (2014). Ingroup favoritism in cooperation: a meta-analysis. *Psychological bulletin*, 140(6):1556.

Banerjee, A. V. (1992). A simple model of herd behavior. *The quarterly journal of economics*, 107(3):797–817.

Banks, J. S. and Duggan, J. (2013). A bargaining model of collective choice. *Annals of Economics and Finance*, 14(3):1073–1110.

Bateman, I. and Munro, A. (2005). An experiment on risky choice amongst households. *The Economic Journal*, 115(502):C176–C189.

Baucells, M. and Heukamp, F. H. (2012). Probability and time trade-off. *Management Science*, 58(4):831–842.

Bedard, K. (2001). Human capital versus signaling models: university access and high school dropouts. *Journal of political economy*, 109(4):749–775.

Bereczkei, T., Birkas, B., and Kerekes, Z. (2010). Altruism towards strangers in need: costly signaling in an industrial society. *Evolution and Human Behavior*, 31(2):95–103.

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Bird, R. B. and Power, E. A. (2015). Prosocial signaling and cooperation among martu hunters. *Evolution and human behavior*, 36(5):389–397.

Black, D. (1948). On the rationale of group decision-making. *Journal of political economy*, 56(1):23–34.

Black, D. et al. (1958). The theory of committees and elections.

BliegeBird, R. and Smith, E. (2005). Signaling theory, strategic interaction, and symbolic capital. *Current anthropology*, 46(2):221–248.

Bornstein, G., Kugler, T., and Ziegelmeyer, A. (2004). Individual and group decisions in the centipede game: Are groups more "rational" players? *Journal of Experimental Social Psychology*, 40(5):599–605.

Bougheas, S., Nieboer, J., and Sefton, M. (2013). Risk-taking in social settings: Group and peer effects. *Journal of economic behavior & organization*, 92:273–283.

Bouton, L., Llorente-Saguer, A., and Malherbe, F. (2018). Get rid of unanimity rule: The superiority of majority rules with veto power. *Journal of Political Economy*, 126(1):107–149.

Bowles, S. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. *Journal of economic literature*, 36(1):75–111.

Brañas-Garza, P. (2007). Promoting helping behavior with framing in dictator games. *Journal of Economic Psychology*, 28(4):477–486.

Brunette, M., Cabantous, L., and Couture, S. (2015). Are individuals more risk and ambiguity averse in a group environment or alone? results from an experimental study. *Theory and Decision*, 78(3):357–376.

Buchan, N. R., Croson, R. T., and Dawes, R. M. (2002). Swift neighbors and persistent strangers: A cross-cultural investigation of trust and reciprocity in social exchange. *American journal of sociology*, 108(1):168–206.

Buchanan, J. M. and Tullock, G. (1965). *The calculus of consent: Logical foundations of constitutional democracy*, volume 100. University of Michigan press.

Burnstein, E. and Vinokur, A. (1977). Persuasive argumentation and social comparison as determinants of attitude polarization. *Journal of experimental social psychology*, 13(4):315–332.

Burton-Chellew, M. N. and West, S. A. (2021). Payoff-based learning best explains the rate of decline in cooperation across 237 public-goods games. *Nature human behaviour*, 5(10):1330–1338.

Cadsby, C. B., Hamaguchi, Y., Kawagoe, T., Maynes, E., and Song, F. (2007). Cross-national gender differences in behavior in a threshold public goods game: Japan versus canada. *Journal of Economic Psychology*, 28(2):242–260.

Camerer, C. F., Loewenstein, G., and Prelec, D. (2004). Neuroeconomics: Why economics needs brains. *The Scandinavian Journal of Economics*, 106(3):555–579.

Cappelen, A. W., Falch, R., Huang, Z., and Tungodden, B. (2025). Acceptance of inequality between children: Large-scale experimental evidence from china and norway. *The Economic Journal*, 135(667):999–1020.

Castore, C. H. and Murnighan, J. K. (1978). Determinants of support for group decisions. *Organizational Behavior and Human Performance*, 22(1):75–92.

Charness, G., Karni, E., and Levin, D. (2007). Individual and group decision making under risk: An experimental study of bayesian updating and violations of first-order stochastic dominance. *Journal of Risk and uncertainty*, 35(2):129–148.

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *The quarterly journal of economics*, 117(3):817–869.

Charness, G. and Sutter, M. (2012). Groups make better self-interested decisions. *Journal of economic perspectives*, 26(3):157–176.

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

Chen-Xia, X. J., Betancor, V., Rodríguez-Gómez, L., and Rodríguez-Pérez, A. (2023). Cultural variations in perceptions and reactions to social norm transgressions: a comparative study. *Frontiers in Psychology*, 14:1243955.

Chua, R. Y., Huang, K. G., and Jin, M. (2019). Mapping cultural tightness and its links to innovation, urbanization, and happiness across 31 provinces in china. *Proceedings of the National Academy of Sciences*, 116(14):6720–6725.

Chvaja, R., Řezníček, D., et al. (2019). An integrative framework of commitment displays in religious systems. *Human Ethology*, 34(1):41–52.

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology*, 58(6):1015.

Coles, P., Kushnir, A., and Niederle, M. (2013). Preference signaling in matching markets. *American Economic Journal: Microeconomics*, 5(2):99–134.

Compte, O. and Jehiel, P. (2017). On the optimal majority rule.

Congleton, R. D. (2004). The median voter model. In *The encyclopedia of public choice*, pages 707–712. Springer.

Corfman, K. P. and Lehmann, D. R. (1987). Models of cooperative group decision-making and relative influence: An experimental investigation of family purchase decisions. *Journal of Consumer Research*, 14(1):1–13.

Croson, R. and Buchan, N. (1999). Gender and culture: International experimental evidence from trust games. *American Economic Review*, 89(2):386–391.

Davis, J. H. (1973). Group decision and social interaction: A theory of social decision schemes.

Dawes, R. M., McTavish, J., and Shaklee, H. (1977). Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *Journal of personality and social psychology*, 35(1):1.

De Dreu, C. K., Weingart, L. R., and Kwon, S. (2000). Influence of social motives on integrative negotiation: a meta-analytic review and test of two theories. *Journal of personality and social psychology*, 78(5):889.

de Haan, T., Offerman, T., and Sloof, R. (2011). Noisy signaling: theory and experiment. *Games and Economic Behavior*, 73(2):402–428.

Del Carpio, M. B., Alpizar, F., and Ferraro, P. J. (2022). Time and risk preferences of individuals, married couples and unrelated pairs. *Journal of Behavioral and Experimental Economics*, 97:101794.

Denant-Boemont, L., Diecidue, E., and l'Haridon, O. (2017). Patience and time consistency in collective decisions. *Experimental economics*, 20(1):181–208.

Deutsch, M. and Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology*, 51(3):629.

Dougherty, K., Pitts, B., Moeller, J., and Ragan, R. (2014). An experimental study of the efficiency of unanimity rule and majority rule. *Public Choice*, 158(3):359–382.

Dougherty, K. L. and Edward, J. (2005). A nonequilibrium analysis of unanimity rule, majority rule, and pareto. *Economic Inquiry*, 43(4):855–864.

Doyle, J. R. (2013). Survey of time preference, delay discounting models. *Judgment and Decision making*, 8(2):116–135.

Duffy, J. and Kornienko, T. (2010). Does competition affect giving? *Journal of Economic Behavior & Organization*, 74(1-2):82–103.

Duggan, J. and Martinelli, C. (2001). A bayesian model of voting in juries. *Games and Economic Behavior*, 37(2):259–294.

Dvorak, F. and Fischbacher, U. (2024). Social learning with intrinsic preferences. *arXiv preprint arXiv:2402.18452*.

Eckel, C. C. and Grossman, P. J. (1996). Altruism in anonymous dictator games. *Games and economic behavior*, 16(2):181–191.

Engle-Warnick, J. and Slonim, R. L. (2006). Learning to trust in indefinitely repeated games. *Games and Economic Behavior*, 54(1):95–114.

Epper, T. F. and Fehr-Duda, H. (2024). Risk in time: The intertwined nature of risk taking and time discounting. *Journal of the European Economic Association*, 22(1):310–354.

Erev, I. and Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American economic review*, pages 848–881.

Eriksson, L. (2019). Social norms as signals. *Social Theory and Practice*, pages 579–599.

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

Feddersen, T. and Pesendorfer, W. (1998). Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting. *American Political science review*, 92(1):23–35.

Fehr, E. and Charness, G. (2025). Social preferences: fundamental characteristics and economic consequences. *Journal of Economic Literature*, 63(2):440–514.

Fehr, E. and Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and human behavior*, 25(2):63–87.

Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.

Festinger, L. (1954). A theory of social comparison processes. *Human relations*, 7(2):117–140.

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

Fischbacher, U. and Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3):525–547.

Fisman, R., Jakiela, P., and Kariv, S. (2015a). How did distributional preferences change during the great recession? *Journal of Public Economics*, 128:84–95.

Fisman, R., Jakiela, P., Kariv, S., and Markovits, D. (2015b). The distributional preferences of an elite. *Science*, 349(6254):aab0096.

Fisman, R., Kariv, S., and Markovits, D. (2007). Individual preferences for giving. *American Economic Review*, 97(5):1858–1876.

Fleck, R. K. and Hanssen, F. A. (2013). Judicial review as a constraint on tyranny of the majority. *The Journal of Law, Economics, & Organization*, 29(2):303–331.

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

Frey, B. S. and Meier, S. (2004). Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment. *American economic review*, 94(5):1717–1722.

Fukuyama, F. (1996). *Trust: The social virtues and the creation of prosperity*. Simon and Schuster.

Gächter, S., Gerhards, L., and Nosenzo, D. (2017). The importance of peers for compliance with norms of fair sharing. *European Economic Review*, 97:72–86.

Gächter, S. and Herrmann, B. (2009). Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1518):791–806.

Gächter, S., Herrmann, B., and Thöni, C. (2010). Culture and cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1553):2651–2661.

Gächter, S. and Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature*, 531(7595):496–499.

Gelfand, M. (2019). *Rule makers, rule breakers: Tight and loose cultures and the secret signals that direct our lives*. Scribner.

Gelfand, M. J. (2012). Culture's constraints: International differences in the strength of social norms. *Current Directions in Psychological Science*, 21(6):420–424.

Gelfand, M. J., Nishii, L. H., and Raver, J. L. (2006). On the nature and importance of cultural tightness-looseness. *Journal of applied psychology*, 91(6):1225.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliach, A., Ang, S., Arnadottir, J., et al. (2011). Differences between tight and loose cultures: A 33-nation study. *science*, 332(6033):1100–1104.

Gheorghiu, M. A., Vignoles, V. L., and Smith, P. B. (2009). Beyond the united states and japan: Testing yamagishi's emancipation theory of trust across 31 nations. *Social Psychology Quarterly*, 72(4):365–383.

Giletta, M., Choukas-Bradley, S., Maes, M., Linthicum, K. P., Card, N. A., and Prinstein, M. J. (2021). A meta-analysis of longitudinal peer influence effects in childhood and adolescence. *Psychological bulletin*, 147(7):719.

Gillet, J., Schram, A., and Sonnemans, J. (2009). The tragedy of the commons revisited: The importance of group decision-making. *Journal of public Economics*, 93(5-6):785–797.

Gintis, H. (2003). The hitchhiker's guide to altruism: Gene-culture coevolution, and the internalization of norms. *Journal of theoretical biology*, 220(4):407–418.

Gintis, H., Smith, E. A., and Bowles, S. (2001). Costly signaling and cooperation. *Journal of theoretical biology*, 213(1):103–119.

Giuliani, F., Cannito, L., Gigliotti, G., Rosa, A., Pietroni, D., and Palumbo, R. (2023). The joint effect of framing and defaults on choice behavior. *Psychological research*, 87(4):1114–1128.

Glätzle-Rützler, D., Lergetporer, P., and Sutter, M. (2021). Collective intertemporal decisions and heterogeneity in groups. *Games and Economic Behavior*, 130:131–147.

Glynia, N., Thum, M. P., and Xefteris, D. (2025). Unanimity vs. majority: Proposing under incomplete information. *Majority: Proposing under Incomplete Information (September 17, 2025)*.

Goeschl, T., Kettner, S. E., Lohse, J., and Schwieren, C. (2018). From social information to social norms: Evidence from two experiments on donation behaviour. *Games*, 9(4):91.

Goldstein, N. J., Cialdini, R. B., and Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of consumer Research*, 35(3):472–482.

Gonzalez-Jimenez, V. and Müller, W. (2025). Are groups more rational than individuals? evidence from revealed preference tests. *Evidence from Revealed Preference Tests (September 05, 2025)*.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., and Ditto, P. H. (2011). Mapping the moral domain. *Journal of personality and social psychology*, 101(2):366.

Green, L. and Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological bulletin*, 130(5):769.

Greif, A. and Kingston, C. (2011). Institutions: rules or equilibria? In *Political economy of institutions, democracy and voting*, pages 13–43. Springer.

Griskevicius, V., Tybur, J. M., Sundie, J. M., Cialdini, R. B., Miller, G. F., and Kenrick, D. T. (2007). Blatant benevolence and conspicuous consumption: when romantic motives elicit strategic costly signals. *Journal of personality and social psychology*, 93(1):85.

Gross, J. and De Dreu, C. K. (2021). Rule following mitigates collaborative cheating and facilitates the spreading of honesty within groups. *Personality and Social Psychology Bulletin*, 47(3):395–409.

Guinier, L. and Carter, S. L. (1994). The tyranny of the majority: Fundamental fairness in representative democracy. *(No Title)*.

Gunnthorsdottir, A., Thorsteinsson, P., and Olafsson, S. P. (2023). Reciprocity or community? different cultural pathways to cooperation and welfare. *Cross-Cultural Research*, 57(4):391–428.

Guttman, J. M. (1998). Unanimity and majority rule: the calculus of consent reconsidered. *European Journal of Political Economy*, 14(2):189–207.

Häge, F. M. (2008). *Decision-Making in the Council of the European Union: The Role of Committees*. Optima Grafische Communicatie.

Haley, K. J. and Fessler, D. M. (2005). Nobody's watching?: Subtle cues affect generosity in an anonymous economic game. *Evolution and Human behavior*, 26(3):245–256.

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

Haslam, S. A., Jetten, J., Postmes, T., and Haslam, C. (2009). Social identity, health and well-being: An emerging agenda for applied psychology. *Applied Psychology-an International Review-Psychologie Appliquee-Revue Internationale*, 58(1):1–23.

Hastie, R. and Kameda, T. (2005). The robust beauty of majority rules in group decisions. *Psychological review*, 112(2):494.

Hausman, D. M. (2011). *Preference, value, choice, and welfare*. Cambridge University Press.

Hawkes, K. (1991). Showing off: tests of an hypothesis about men's foraging goals. *Ethology and sociobiology*, 12(1):29–54.

He, H., Martinsson, P., and Sutter, M. (2012). Group decision making under risk: An experiment with student couples. *Economics Letters*, 117(3):691–693.

Heine, S. J. and Ruby, M. B. (2010). Cultural psychology. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(2):254–266.

Heisenberg, D. (2005). The institution of 'consensus' in the european union: Formal versus informal decision-making in the council. *European Journal of Political Research*, 44(1):65–90.

Heller, S. B. and Kessler, J. B. (2024). Information frictions and skill signaling in the youth labor market. *American Economic Journal: Economic Policy*, 16(4):1–33.

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., et al. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *science*, 327(5972):1480–1484.

Herrmann, B., Thoni, C., and Gachter, S. (2008). Antisocial punishment across societies. *Science*, 319(5868):1362–1367.

Heywood, J. S. and Wei, X. (2004). Education and signaling: evidence from a highly competitive labor market. *Education Economics*, 12(1):1–16.

Hill, K. and Hurtado, A. M. (2017). *Ache life history: The ecology and demography of a foraging people*. Routledge.

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

Horwitz, M. J. (1966). Tocqueville and the tyranny of the majority. *The review of politics*, 28(3):293–307.

Huff, L. and Kelley, L. (2003). Levels of organizational trust in individualist versus collectivist societies: A seven-nation study. *Organization science*, 14(1):81–90.

Inukai, K., Shimodaira, Y., and Shiozawa, K. (2022). Empirical properties of an extended ces utility function in representing distributional preferences. *ISER DP*, (1199).

Ioannou, C. A. and Romero, J. (2014). A generalized approach to belief learning in repeated games. *Games and Economic Behavior*, 87:178–203.

Isaac, R. M. and Walker, J. M. (1988). Communication and free-riding behavior: The voluntary contribution mechanism. *Economic inquiry*, 26(4):585–608.

Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of personality and social psychology*, 50(6):1141.

Jackson, M. O. and Yariv, L. (2015). Collective dynamic choice: the necessity of time inconsistency. *American Economic Journal: Microeconomics*, 7(4):150–178.

Jetten, J., Postmes, T., and McAuliffe, B. J. (2002). 'we're all individuals': Group norms of individualism and collectivism, levels of identification and identity threat. *European journal of social psychology*, 32(2):189–207.

Johnson, R. C. (1962). A study of children's moral judgments. *Child Development*, pages 327–354.

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

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.

Kahneman, D. and Tversky, A. (1984). Choices, values, and frames. *American psychologist*, 39(4):341.

Kameda, T. and Davis, J. H. (1990). The function of the reference point in individual and group risk decision making. *Organizational behavior and human decision processes*, 46(1):55–76.

Karpowitz, C. F., Mendelberg, T., and Shaker, L. (2012). Gender inequality in deliberative participation. *American Political Science Review*, 106(3):533–547.

Kim, J., Lim, W., and Schweighofer-Kodritsch, S. (2020). Bargaining and time preferences: An experimental study.

Kimbrough, E. O., Krupka, E. L., Kumar, R., Murray, J. M., Ramalingam, A., Sánchez-Franco, S., Sarmiento, O. L., Kee, F., and Hunter, R. F. (2024). On the stability of norms and norm-following propensity: a cross-cultural panel study with adolescents. *Experimental Economics*, 27(2):351–378.

Kimbrough, E. O. and Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3):608–638.

Kimbrough, E. O. and Vostroknutov, A. (2018). A portable method of eliciting respect for social norms. *Economics Letters*, 168:147–150.

Koch, B. J. and Koch, P. T. (2007). Collectivism, individualism, and outgroup cooperation in a segmented china. *Asia Pacific Journal of Management*, 24(2):207–225.

Kocher, M. G. and Sutter, M. (2005). The decision maker matters: Individual versus group behaviour in experimental beauty-contest games. *The Economic Journal*, 115(500):200–223.

Kogan, N. and Wallach, M. A. (1967). *Risk taking as a function of the situation, the person, and the group*. Holt, Rinehart & Winston.

Kohlberg, L. (1981). The philosophy of moral development: Moral stages and the idea of justice.

König, T. and Junge, D. (2008). Veto player theory and consensus behaviour. In *Unveiling the Council of the European Union: Games Governments Play in Brussels*, pages 81–98. Springer.

Krick, E. (2017). The myth of effective veto power under the rule of consensus. dynamics and democratic legitimacy of collective decision-making by "tacit consent". *Négociations*, 27(1):109–128.

Kruger, D. J. and Kruger, J. S. (2018). What do economically costly signals signal?: A life history framework for interpreting conspicuous consumption. *Evolutionary Psychological Science*, 4(4):420–427.

Krupka, E. and Weber, R. A. (2009). The focusing and informational effects of norms on pro-social behavior. *Journal of Economic psychology*, 30(3):307–320.

Kübler, D., Müller, W., and Normann, H.-T. (2008). Job-market signaling and screening: An experimental comparison. *Games and Economic Behavior*, 64(1):219–236.

Kugler, T., Kausel, E. E., and Kocher, M. G. (2012). Are groups more rational than individuals? a review of interactive decision making in groups. *Wiley interdisciplinary reviews: Cognitive science*, 3(4):471–482.

Laughlin, P. R. and Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellective tasks. *Journal of Experimental Social Psychology*, 22(3):177–189.

Ledyard, J. O. et al. (1994). *Public goods: A survey of experimental research*. Division of the Humanities and Social Sciences, California Inst. of Technology.

Leung, K. (1997). Negotiation and reward allocations across cultures.

Levitan, L. C. and Verhulst, B. (2016). Conformity in groups: The effects of others' views on expressed attitudes and attitude change. *Political behavior*, 38(2):277–315.

Lewis, A., Carrera, S., Cullis, J., and Jones, P. (2009). Individual, cognitive and cultural differences in tax compliance: Uk and italy compared. *Journal of Economic Psychology*, 30(3):431–445.

List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political economy*, 115(3):482–493.

Liu, C.-H., Tsai, S.-F., Chen, T.-C., and Cheng, H.-W. (2024). The signaling role of feedback in the repeated public goods game: Experimental evidence from the laboratory. *Plos one*, 19(2):e0299196.

Loewenstein, G., Angner, E., et al. (2003). Predicting and indulging changing preferences. *Time and decision: Economic and psychological perspectives on intertemporal choice*, 12:351–391.

Macé, A. and Treibich, R. (2024). Repeated majority voting.

Maciejovsky, B., Sutter, M., Budescu, D. V., and Bernau, P. (2013). Teams make you smarter: How exposure to teams improves individual decisions in probability and reasoning tasks. *Management Science*, 59(6):1255–1270.

Maggi, G. and Morelli, M. (2006). Self-enforcing voting in international organizations. *American Economic Review*, 96(4):1137–1158.

Maletz, D. J. (2002). Tocqueville's tyranny of the majority reconsidered. *The Journal of Politics*, 64(3):741–763.

Markus, H. and Kitayama, S. (1991). Culture and the self: implications for cognition, emotion, and motivation. *Psychological Review*, 98:224–253.

Martin, R. and Randal, J. (2008). How is donation behaviour affected by the donations of others? *Journal of Economic Behavior & Organization*, 67(1):228–238.

Masclet, D., Colombier, N., Denant-Boemont, L., and Loheac, Y. (2009). Group and individual risk preferences: A lottery-choice experiment with self-employed and salaried workers. *Journal of Economic Behavior & Organization*, 70(3):470–484.

Masclet, D., Noussair, C., Tucker, S., and Villeval, M.-C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review*, 93(1):366–380.

McAndrew, F. T. (2021). Costly signaling theory. In *Encyclopedia of evolutionary psychological science*, pages 1525–1532. Springer.

McAuliffe, B. J., Jetten, J., Hornsey, M. J., and Hogg, M. A. (2003). Individualist and collectivist norms: When it's ok to go your own way. *European journal of social psychology*, 33(1):57–70.

Miller, C. E. (1985). Group decision making under majority and unanimity decision rules. *Social Psychology Quarterly*, pages 51–61.

Mizrahi, D., Laufer, I., and Zuckerman, I. (2020). Collectivism-individualism: Strategic behavior in tacit coordination games. *PLoS One*, 15(2):e0226929.

Mondak, J. J. (2010). *Personality and the foundations of political behavior*. Cambridge University Press.

Morone, A., Nuzzo, S., and Temerario, T. (2017). Decision process, preferences over risk and consensus rule: a group experiment. *Preferences Over Risk and Consensus Rule: A Group Experiment (May 18, 2017)*.

Morris, M. W., Hong, Y.-y., Chiu, C.-y., and Liu, Z. (2015). Normology: Integrating insights about social norms to understand cultural dynamics. *Organizational behavior and human decision processes*, 129:1–13.

Morton, R. B. (1987). A group majority voting model of public good provision. *Social Choice and Welfare*, 4(2):117–131.

Moscovici, S. and Zavalloni, M. (1969). The group as a polarizer of attitudes. *Journal of personality and social psychology*, 12(2):125.

Mu, Y., Kitayama, S., Han, S., and Gelfand, M. J. (2015). How culture gets embrained: Cultural differences in event-related potentials of social norm violations. *Proceedings of the National Academy of Sciences*, 112(50):15348–15353.

Murphy, J. H. and Banerjee, S. (2015). A caveat for the application of the critical cost efficiency index in induced budget experiments. *Experimental Economics*, 18(3):356–365.

Nehring, K., Pivato, M., and Puppe, C. (2016). Unanimity overruled: Majority voting and the burden of history. *Journal of Theoretical Politics*, 28(4):552–597.

Nemeth, C. (1977). Interactions between jurors as a function of majority vs. unanimity decision rules. *Journal of Applied Social Psychology*, 7(1):38–56.

North, D. C. (1990). Institutions, institutional change and economic performance. *Cambridge University*.

Noussair, C. and Wu, P. (2006). Risk tolerance in the present and the future: An experimental study. *Managerial and Decision Economics*, 27(6):401–412.

Nunner-Winkler, G. (2007). Development of moral motivation from childhood to early adulthood. *Journal of Moral education*, 36(4):399–414.

O'Garra, T., Capraro, V., and Kujal, P. (2019). The effect of social information in the dictator game with a taking option. *arXiv preprint arXiv:1907.13471*.

Oh, S. H. (2013). Do collectivists conform more than individualists? cross-cultural differences in compliance and internalization. *Social Behavior and Personality: an international journal*, 41(6):981–994.

Ohbuchi, K.-I., Fukushima, O., and Tedeschi, J. T. (1999). Cultural values in conflict management: Goal orientation, goal attainment, and tactical decision. *Journal of cross-cultural psychology*, 30(1):51–71.

Ohtsubo, Y., Miller, C. E., Hayashi, N., and Masuchi, A. (2004). Effects of group decision rules on decisions involving continuous alternatives: The unanimity rule and extreme decisions in mock civil juries. *Journal of Experimental Social Psychology*, 40(3):320–331.

Oosterbeek, H., Sloof, R., and Van De Kuilen, G. (2004). Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental economics*, 7(2):171–188.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge university press.

Oyserman, D., Coon, H. M., and Kemmelmeier, M. (2002). Rethinking individualism and collectivism: evaluation of theoretical assumptions and meta-analyses. *Psychological bulletin*, 128(1):3.

O'Garra, T. and Sisco, M. R. (2020). The effect of anchors and social information on behaviour. *Plos one*, 15(4):e0231203.

Parks, C. D. and Vu, A. D. (1994). Social dilemma behavior of individuals from highly individualist and collectivist cultures. *Journal of Conflict Resolution*, 38(4):708–718.

Pattanaik, P. K. and Xu, Y. (2015). *Rationality and context-dependent preferences*. Springer.

Patterson, S. C. and Caldeira, G. A. (1988). Party voting in the united states congress. *British Journal of Political Science*, 18(1):111–131.

Poole, K. T. and Daniels, R. S. (1985). Ideology, party, and voting in the us congress, 1959–1980. *American Political Science Review*, 79(2):373–399.

Posner, E. (2009). *Law and Social Norms*. Harvard University Press.

Posner, E. A. (2002). The signaling model of social norms: Further thoughts. *U. Rich. L. Rev.*, 36:465.

Potters, J. and Xu, Y. (2020). Social information and selfishness. *Journal of Economic Behavior & Organization*, 177:327–340.

Potz, M. (2022). Costly commitments "under his eye": Reconceptualizing the costly signaling theory of religion. *Journal of the American Academy of Religion*, 90(3):599–617.

Pruitt, D. G. (2013). *Negotiation behavior*. Academic press.

Rammstedt, B. and John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of research in Personality*, 41(1):203–212.

Reinstein, D. and Riener, G. (2012). Reputation and influence in charitable giving: an experiment. *Theory and decision*, 72(2):221–243.

Roberts, G. (1998). Competitive altruism: from reciprocity to the handicap principle. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1394):427–431.

Roccas, S., Sagiv, L., Schwartz, S. H., and Knafo, A. (2002). The big five personality factors and personal values. *Personality and social psychology bulletin*, 28(6):789–801.

Rockenbach, B., Sadrieh, A., and Mathauschek, B. (2007). Teams take the better risks. *Journal of Economic Behavior & Organization*, 63(3):412–422.

Romme, A. G. L. (2004). Unanimity rule and organizational decision making: A simulation model. *Organization Science*, 15(6):704–718.

Rozin, P. and Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and social psychology review*, 5(4):296–320.

Saad, G. (2007). *The evolutionary bases of consumption*. Psychology Press.

Samek, A., Cowell, J. M., Cappelen, A. W., Cheng, Y., Contreras-Ibáñez, C., Gomez-Sicard, N., Gonzalez-Gadea, M. L., Huepe, D., Ibáñez, A., Lee, K., et al. (2020). The development of social comparisons and sharing behavior across 12 countries. *Journal of Experimental Child Psychology*, 192:104778.

Samuelson, P. A. (1938a). The empirical implications of utility analysis. *Econometrica, Journal of the Econometric Society*, pages 344–356.

Samuelson, P. A. (1938b). A note on the pure theory of consumer's behaviour: an addendum. *Economica*, 5(19):353.

Samuelson, W. and Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of risk and uncertainty*, 1(1):7–59.

Saracevic, S., Schlegelmilch, B. B., and Wu, T. (2022). How normative appeals influence pro-environmental behavior: The role of individualism and collectivism. *Journal of Cleaner Production*, 344:131086.

Sartori, G. (2016). The theory of democracy revisited. In *Democracy: A Reader*, pages 192–196. Columbia University Press.

Schildberg-Hörisch, H. (2018). Are risk preferences stable? *Journal of Economic Perspectives*, 32(2):135–154.

Schultz, W., Dietz, T., and Stern, P. (2002). *Knowledge, information and household recycling: Examining the knowledge-deficit model.* National Academy of Sciences Washington, DC.

Shang, J. and Croson, R. (2009). A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *The economic journal*, 119(540):1422–1439.

Shapiro, J. (2010). Discounting for you, me and we. Technical report, Working Paper 39, Institute for Financial Management and Research.

Sherif, M. (1936). The psychology of social norms.

Shupp, R. S. and Williams, A. W. (2008). Risk preference differentials of small groups and individuals. *The Economic Journal*, 118(525):258–283.

Sia, C.-L., Tan, B. C., and Wei, K.-K. (2002). Group polarization and computer-mediated communication: Effects of communication cues, social presence, and anonymity. *Information Systems Research*, 13(1):70–90.

Sliwka, D. (2007). Trust as a signal of a social norm and the hidden costs of incentive schemes. *American Economic Review*, 97(3):999–1012.

Smith, E. A. (2004). Why do good hunters have higher reproductive success? *Human Nature*, 15(4):343–364.

Smith, E. A. and Bird, R. L. B. (2000). Turtle hunting and tombstone opening: Public generosity as costly signaling. *Evolution and human behavior*, 21(4):245–261.

Soler, M. (2012). Costly signaling, ritual and cooperation: Evidence from candomblé, an afro-brazilian religion. *Evolution and Human Behavior*, 33(4):346–356.

Somasundaram, J. and Eli, V. (2022). Risk and time preferences interaction: An experimental measurement. *Journal of Risk and Uncertainty*, 65(2):215–238.

Sosis, R. (2003). Why aren't we all hutterites? costly signaling theory and religious behavior. *Human nature*, 14(2):91–127.

Spadaro, G., Graf, C., Jin, S., Arai, S., Inoue, Y., Lieberman, E., Rinderu, M. I., Yuan, M., Van Lissa, C. J., and Balliet, D. (2022). Cross-cultural variation in cooperation: A meta-analysis. *Journal of Personality and Social Psychology*, 123(5):1024.

Spence, A. M. (1974a). Market signaling: Informational transfer in hiring and related screening processes. *(No Title)*.

Spence, M. (1974b). Competitive and optimal responses to signals: An analysis of efficiency and distribution. *Journal of Economic theory*, 7(3):296–332.

Spence, M. (1976). Informational aspects of market structure: An introduction. *The Quarterly Journal of Economics*, pages 591–597.

Spence, M. (2002). Signaling in retrospect and the informational structure of markets. *American economic review*, 92(3):434–459.

Stasson, M. F., Kameda, T., Parks, C. D., Zimmerman, S. K., and Davis, J. H. (1991). Effects of assigned group consensus requirement on group problem solving and group members' learning. *Social Psychology Quarterly*, pages 25–35.

Stewart, A. J. and Plotkin, J. B. (2014). Collapse of cooperation in evolving games. *Proceedings of the National Academy of Sciences*, 111(49):17558–17563.

Stigler, G. J. and Becker, G. S. (1977). De gustibus non est disputandum. *The american economic review*, 67(2):76–90.

Strand, J. R. and Rapkin, D. P. (2011). Weighted voting in the united nations security council: A simulation. *Simulation & Gaming*, 42(6):772–802.

Sunstein, C. (2002). The law of group polarization. *Journal of political philosophy*.

Sutter, M. (2009). Individual behavior and group membership: Comment. *American Economic Review*, 99(5):2247–2257.

Szolnoki, A. and Perc, M. (2014). Costly hide and seek pays: unexpected consequences of deceit in a social dilemma. *New Journal of Physics*, 16(11):113003.

Thaler, R. H. and Sunstein, C. R. (2021). *Nudge: The final edition*. Penguin.

Thomadsen, R., Rooderkerk, R. P., Amir, O., Arora, N., Bollinger, B., Hansen, K., John, L., Liu, W., Sela, A., Singh, V., et al. (2018). How context affects choice. *Customer Needs and Solutions*, 5(1):3–14.

Tindale, R. S. and Kameda, T. (2000). 'social sharedness' as a unifying theme for information processing in groups. *Group Processes & Intergroup Relations*, 3(2):123–140.

Ting-Toomey, S. (1988). Intercultural conflict styles: A face negotiation theory', in yy kim.

Triandis, H. C. (1995). Individualism & collectivism.

Triandis, H. C. (2001). Individualism-collectivism and personality. *Journal of personality*, 69(6):907–924.

Triandis, H. C. (2018). *Individualism and collectivism*. Routledge.

Tropman, J. E. (2013). *Effective meetings: Improving group decision making*, volume 17. Sage Publications.

Tsebelis, G. (1995). Decision making in political systems: Veto players in presidentialism, parliamentarism, multicameralism and multipartyism. *British journal of political science*, 25(3):289–325.

Tsebelis, G. (2002). *Veto players: How political institutions work*. Princeton University Press.

Tsebelis, G. (2013). Bridging qualified majority and unanimity decisionmaking in the eu. *Journal of European Public Policy*, 20(8):1083–1103.

Tversky, A. and Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics*, 106(4):1039–1061.

Tversky, A. and Simonson, I. (1993). Context-dependent preferences. *Management science*, 39(10):1179–1189.

Tyler, T. R. (2006). *Why people obey the law*. Princeton university press.

Udah, E. B. and Ayara, N. (2014). Institutions, governance structure and economic performance nexus in nigeria. *Journal of economics and sustainable development*, 5(3).

Uz, I. (2015). The index of cultural tightness and looseness among 68 countries. *Journal of Cross-Cultural Psychology*, 46(3):319–335.

Varian, H. R. (1982). The nonparametric approach to demand analysis. *Econometrica: Journal of the Econometric Society*, pages 945–973.

Velden, F. S. T., Beersma, B., and De Dreu, C. K. (2007). Majority and minority influence in group negotiation: the moderating effects of social motivation and decision rules. *Journal of Applied Psychology*, 92(1):259.

Vinokur, A. and Burstein, E. (1974). Effects of partially shared persuasive arguments on group-induced shifts: A group-problem-solving approach. *Journal of Personality and Social Psychology*, 29(3):305.

Viscusi, W. K., Phillips, O. R., and Kroll, S. (2011). Risky investment decisions: How are individuals influenced by their groups? *Journal of Risk and Uncertainty*, 43(2):81–106.

Vishkin, A., Kitayama, S., Berg, M., Diener, E., Gross-Manos, D., Ben-Arieh, A., and Tamir, M. (2023). Adherence to emotion norms is greater in individualist cultures than in collectivist cultures. *Journal of Personality and Social Psychology*, 124:1256–1276.

Wabara, K. (2021). Does unanimity protect minority preferences? evidence from financial risk-taking in an award-winning game show. *Evidence from Financial Risk-Taking in an Award-Winning Game Show (October 11, 2021)*.

Waldman, M. (2016). The dual avenues of labor market signaling. *Labour Economics*, 41:120–134.

Wallensteen, P. and Johansson, P. (2015). 2 the un security council: Decisions and actions. *The UN Security Council in the 21st century*, pages 27–54.

Wang, Q., Liu, L., and Chen, X. (2020). Evolutionary dynamics of cooperation in the public goods game with individual disguise and peer punishment. *Dynamic Games and Applications*, 10(3):764–782.

Wang, X. (2021). Costly participation and the evolution of cooperation in the repeated public goods game. *Dynamic Games and Applications*, 11(1):161–183.

Warntjen, A. (2010). Between bargaining and deliberation: decision-making in the council of the european union. *Journal of European Public Policy*, 17(5):665–679.

Weber, T. O., Schulz, J. F., Beranek, B., Lambarraa-Lehnhardt, F., and Gächter, S. (2023). The behavioral mechanisms of voluntary cooperation across culturally diverse societies: Evidence from the us, the uk, morocco, and turkey. *Journal of Economic Behavior & Organization*, 215:134–152.

Wei, Z., Zhao, Z., and Zheng, Y. (2019). Following the majority: Social influence in trusting behavior. *Frontiers in Neuroscience*, 13:89.

Weiß, M., Hein, G., and Hewig, J. (2021). Between joy and sympathy: Smiling and sad recipient faces increase prosocial behavior in the dictator game. *International journal of environmental research and public health*, 18(11):6172.

Williams, S. and Taormina, R. J. (1993). Unanimous versus majority influences on group polarization in business decision making. *The Journal of social psychology*, 133(2):199–205.

Wilson, M. and Daly, M. (1985). Competitiveness, risk taking, and violence: The young male syndrome. *Ethology and sociobiology*, 6(1):59–73.

Wong, R. Y.-m. and Hong, Y.-y. (2005). Dynamic influences of culture on cooperation in the prisoner's dilemma. *Psychological science*, 16(6):429–434.

Wright, J. (1999). Altruism as a signal: Zahavi's alternative to kin selection and reciprocity. *Journal of Avian Biology*, pages 108–115.

Wright, S. C., Taylor, D. M., and Moghaddam, F. M. (1990). Responding to membership in a disadvantaged group: From acceptance to collective protest. *Journal of personality and social psychology*, 58(6):994.

Wu, J., Zhang, J., and Wang, C. (2023). Student performance, peer effects, and friend networks: Evidence from a randomized peer intervention. *American Economic Journal: Economic Policy*, 15(1):510–542.

Xiao, W. S. (2021). The role of collectivism–individualism in attitudes toward compliance and psychological responses during the covid-19 pandemic. *Frontiers in Psychology*, 12:600826.

Xu, Q., He, S., Li, Z., Duan, R., and Li, P. (2023). Voluntary or reluctant? social influence in charitable giving: an erp study. *Social Cognitive and Affective Neuroscience*, 18(1):nsad010.

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and social Psychology*, 51(1):110.

Yamagishi, T., Cook, K. S., and Watabe, M. (1998). Uncertainty, trust, and commitment formation in the united states and japan. *American journal of sociology*, 104(1):AJSv104p165–194.

Yamagishi, T. and Yamagishi, M. (1994). Trust and commitment in the united states and japan. *Motivation and emotion*, 18(2):129–166.

Zahavi, A. (1977). The cost of honesty (further remarks on the handicap principle). *Journal of theoretical Biology*, 67(3):603–605.

Zhang, J. and Casari, M. (2012). How groups reach agreement in risky choices: an experiment. *Economic Inquiry*, 50(2):502–515.

Zhou, J., Yang, J., and Faye, B. (2025). Examining the impact of total rewards on proactivity among chinese knowledge employees: the moderating role of vertical and horizontal collectivist orientations. *Frontiers in Psychology*, 16:1487464.

Appendix

A.1 The Measurement of Changes in Distributional Preferences

A.1.1 CCEI Calculations

We evaluate the internal consistency of individual choices using the Generalized Axiom of Revealed Preference (GARP), which enables the identification and exclusion of decision-making patterns that violate basic rationality. Under GARP, if allocation π_i is indirectly revealed preferred to π_j, then π_j must not be strictly directly revealed preferred to π_i—that is, π_i cannot lie strictly within the budget set when π_j is chosen.

To measure the degree of GARP consistency, we calculate the Critical Cost Efficiency Index (CCEI), first introduced by Afriat (1967, 1972). The CCEI ranges from 0 to 1, with a value of 1 indicating perfect consistency with utility maximization and lower values reflecting greater violations. Following Afriat (1987) and Murphy and Banerjee (2015), we calculate the CCEI as follows.

For two bundles x_i and x_j,

$$D_{ij} = \frac{p_i x_j}{p_i x_i} - 1$$

where p_i is the price combination when the bundle x_i is chosen. Let $d_{ij} = \max\left\{D_{ij}, D_{ji}\right\}$, the cross cost efficiency index is defined as follows:

$$e_{ij} = 1 - \max\left\{0, -d_{ij}\right\}$$

Then, we have the critical cost efficiency index as follows:

$$e^* = \min_{i \neq j}\left\{e_{ij}\right\}$$

In our experiment, average values of CCEI for BT are 0.933 in Stage I, 0.953 in Stage II, and 0.897 across both stages. For MIT, the corresponding values are 0.936, 0.943, and 0.892, respectively (see Table A.1). No statistically significant differences emerge between the treatments in Stage I (Wilcoxon Mann–Whitney test: z=-0.046, p=0.963) or Stage II (Wilcoxon Mann–Whitney test: z=0.311, p=0.756).

Figure A.1 shows the distribution of CCEI values for BT and MIT in Stage I (*Rounds 1-10*), Stage II (*Rounds 11-20*), and the combined data (*Rounds 1-20*). In both treatments, approximately 80% of participants achieve CCEI values above 0.9 in each stage. When all 20 rounds are considered, this proportion declines to around 70%, likely reflecting the increased probability of decision errors over a larger choice set, which lowers overall consistency with utility maximization.

We classify participants as behaviorally consistent with GARP—and thus rational—if their CCEI exceeds 0.8. By this criterion, 87 participants in BT (88.8%) and 109 in MIT (90.8%) are deemed rational. These findings indicate that most subjects made decisions broadly aligned with utility maximization. For robustness, all analyses are conducted on both the full sample (all observations) and a restricted sample of participants with CCEI$\geq$0.8.[1]

Table A.1: Average Critical Cost Efficiency Indices (CCEI) by Treatment and Stage

Treatments	All subjects			CCEI$\geq$0.8		
	Stage I	Stage II	Stage I+Stage II	Stage I	Stage II	Stage I+Stage II
BT	0.933	0.953	0.897	0.956	0.977	0.937
	(0.091)	(0.111)	(0.132)	(0.027)	(0.031)	(0.034)
MIT	0.936	0.943	0.892	0.952	0.976	0.923
	(0.083)	(0.128)	(0.126)	(0.029)	(0.036)	(0.052)

Note: Standard deviations are reported in parentheses. Stage I and Stage II correspond to *Rounds 1-10* and *Rounds 11-20*, respectively.

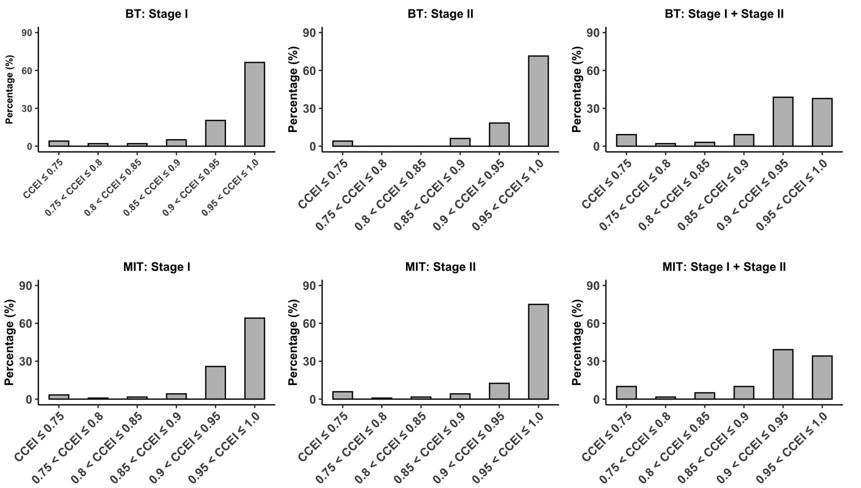


Figure A.1: Distribution of CCEIs: Separate Stages vs. Pooled Sample

A.1.2 Changed Distributional Preferences

Social information regarding maximum giving alters some participants' allocation behavior in Stage II. Figures A.2–A.6 plot individual-level indifference curves, derived from estimated $\hat{\alpha}$ and $\hat{\rho}$, illustrating preference shifts between stages. Preferences are classified as Perfect Substitutes, Leontief, Selfish, or Cobb-Douglas. Notably, curves in Stage

[1] The full sample includes 98 and 120 participants in BT and MIT, respectively. Excluding those with CCEI<0.8 yields rational subsamples of 87 and 109 participants.

II often differ substantially from those in Stage I. For example, Subject MIT01 exhibits a substitute preference in Stage I but becomes entirely selfish in Stage II. Similarly, Subject MIT11 shifts from a substitute preference ($\hat{\rho} = 0.867$) to complete selfishness ($\hat{\alpha} = 1$) after receiving social information.

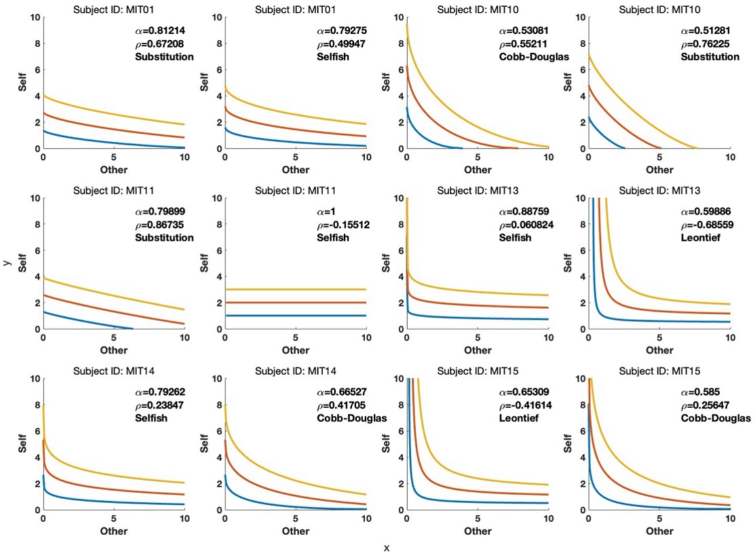


Figure A.2: Indifference Curves of Changed Preferences in BT across Stages

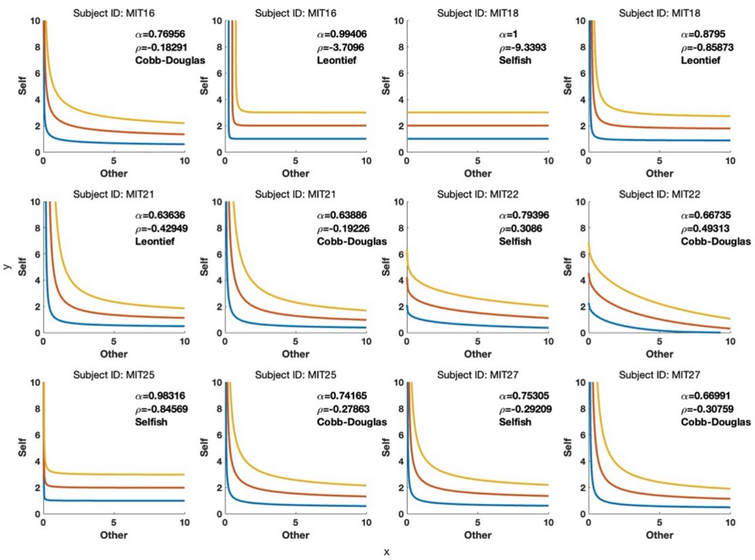


Figure A.3: Indifference Curves of Changed Preferences in MIT across Stages

A.1.3 Experimental Instructions

Welcome, and thank you for participating in this experiment. The session includes several tasks (games) and a short survey. The entire experiment will take about 40-50 min-

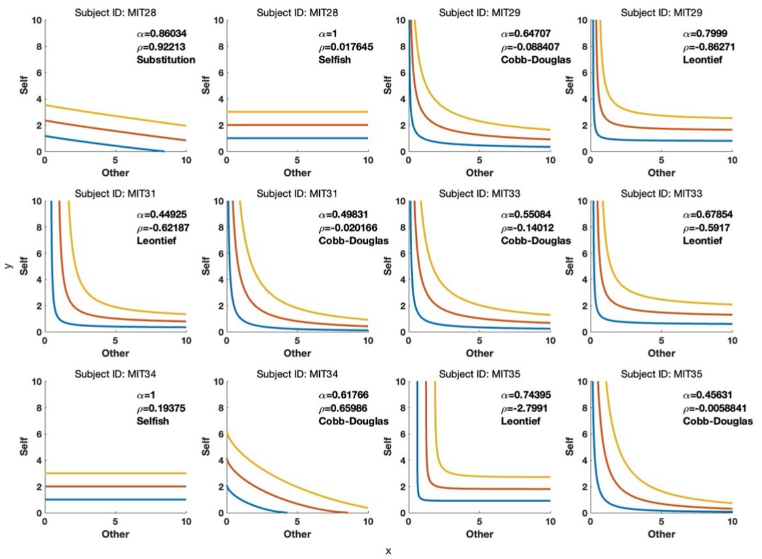


Figure A.4: Indifference Curves of Changed Preferences in MIT across Stages

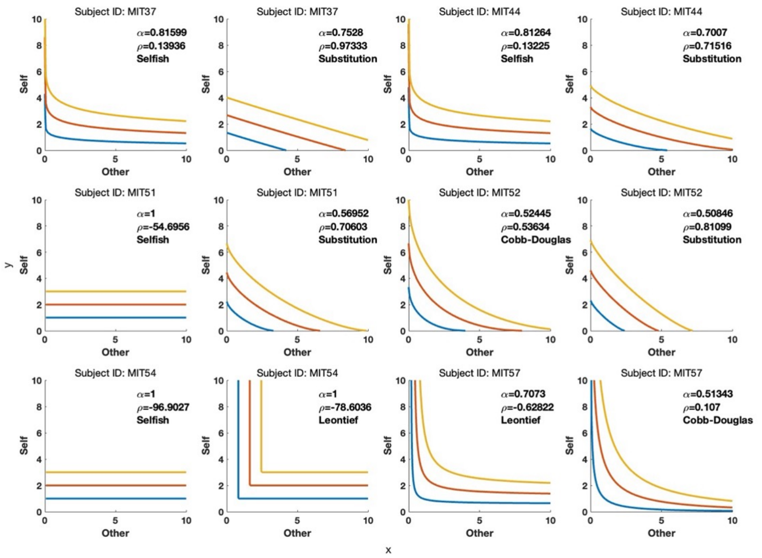


Figure A.5: Indifference Curves of Changed Preferences in MIT across Stages

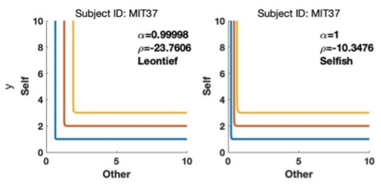


Figure A.6: Indifference Curves of Changed Preferences in MIT across Stages utes, and the survey will take less than 2 minutes.[2]

Now you have 15 minutes to read the instructions and complete the exercise. After that, the experimenter will read the instructions aloud to help you better understand the experimental context.

Silence and Conduct

(i) Please do not talk to other participants during the experiment, especially while the experimenter is reading the instructions.
(ii) Turn off your mobile phone.
(iii) If you have any questions, do not ask them aloud. Instead, raise your hand and the experimenter will assist you individually.

Payoffs

(i) You will receive a monetary payoff based on your performance in the experiment.
(ii) Each experimental currency point (token) is worth 0.3 RMB (€0.037).
(iii) Your total payoff will include the earnings from the tasks plus a fixed show-up fee of 10 RMB (€1.22).

Anonymity

(i) Your identity will remain completely anonymous throughout the experiment.
(ii) You will not know who the other participants are, and they will not know who you are.

Data Protection

In accordance with EU data protection regulations and Chinese law, all data and decisions will be handled anonymously and will not be disclosed.

If you agree, please click "Next" to start the experiment.

Instructions for Baseline Treatment

The experiment consists of 20 decision rounds and will take approximately 40 minutes. All participants will make a series of decisions with varying budget lines.

[2] The original instructions were administered in Chinese, and this document presents the English translation.

In each round, you will make a decision that affects you and the other participant. A budget line will appear on the screen (see Figure A.7), representing two dimensions: *Self* (your payoff) and *Other* (the other participant's payoff). You will choose a point on this line by clicking, which determines the allocation of tokens between you and the other participant. The selected point corresponds to a pair (x, y), where x is the amount of tokens for the *Other*, and y is the amount of tokens for *Self*. After making a decision, click "*Confirmation*" to enter the next round.

After all 20 rounds are completed, all participants will be randomly assigned to groups of two, and the computer randomly assigns roles: one participant becomes *Red*, the other *Blue*. If you are assigned the *Red* role, one of your 20 decisions will be randomly selected to determine the final payment for both you and the other participant (plus a fixed show-up fee). If the other participant is assigned *Red*, one of his/her decisions will be used instead.

At the end of the experiment, you will complete a short questionnaire. It does not collect personal information such as name or identity.

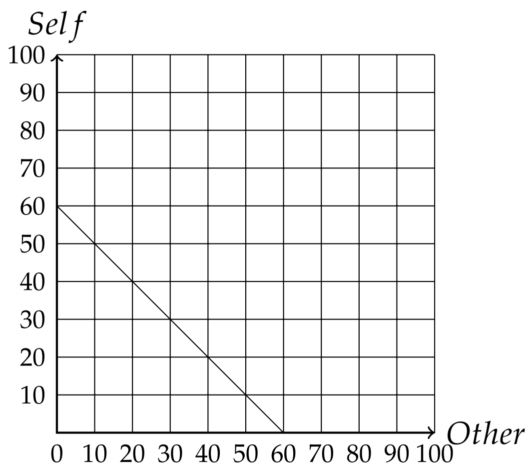


Figure A.7: The Budget Constraint: A Formal Example

Instructions for Maximum Information Treatment

The experiment consists of 20 decision rounds and will take approximately 40 minutes. All participants will make a series of decisions with varying budget lines.

In each round, you will make a decision that affects you and the other participant. A budget line will appear on the screen (see Figure A.7), representing two dimensions: *Self* (your payoff) and *Other* (the other participant's payoff). You will choose a point on this line by clicking, which determines how tokens are allocated between you and the other participant. The selected point corresponds to a pair (x, y), where x is the amount of tokens for the *Other*, and y is the amount of tokens for *Self*. After making a decision, click "*Confirmation*" to enter the next round.

After all 20 rounds are completed, all participants will be randomly assigned to groups of two, and the computer randomly assigns roles: one participant becomes *Red*, the other *Blue*. If you are assigned the *Red* role, one of your 20 decisions will be randomly selected to determine the final payment for both you and the other participant (plus a fixed show-up fee). If the other participant is assigned *Red*, one of his/her decisions will be used instead.

At the end of the experiment, you will complete a short questionnaire. It does not collect personal information such as name or identity.

After participants complete Stage I (Round 1-10), this information will be displayed on the screen before they begin Stage II.

> **New Instructions:** You have completed 10 rounds of decisions. Starting from the next stage, before each round, you will be shown the amount given by the participant who allocated the most to the other participant in the previous round. *For example, before making your decision in Round i+1, you will see the highest allocation made in Round i (for rounds* $10 \leq i \leq 19$*)*

Exercise

Please complete the exercise. If you have any questions, please raise your hand. The experimenters will come to you and answer your questions privately.

(1) If you are made the role *Blue*, the other participant who is matched with you is assigned the role __________ (*Red* or *Blue*); your role is __________ (*Red* or *Blue*) in the next round.
(2) If you are assigned the role of *Red*, your decisions will determine the payments for yourself and the other participant? (Yes or No)
(3) As for budget line 1 (Figure A.8), you get __________ tokens and the other participant gets __________ tokens.

(4) As for the budget line 2 (Figure A.9), you get __________ tokens and the other participant gets __________ tokens.

(5) As for the budget line 3 (Figure A.10), you get __________ tokens and the other participant gets __________ tokens.

(6) As for the budget line 4 (Figure A.11), you get __________ tokens and the other participant gets __________ tokens.

(7) As for the budget line 5 (Figure A.12), you get __________ tokens and the other participant gets __________ tokens.

(8) As for the budget line 6 (Figure A.13), you get __________ tokens and the other participant gets __________ tokens.

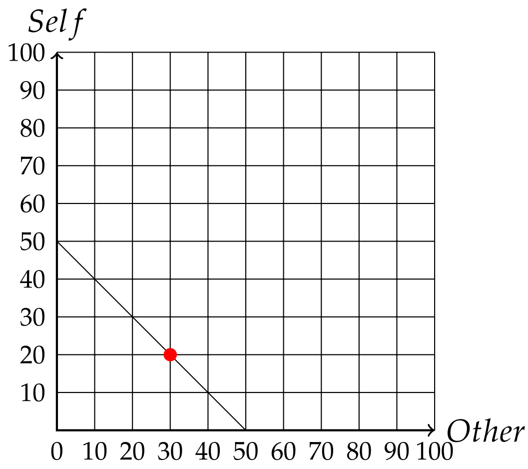


Figure A.8: Allocation Decision in Budget 1

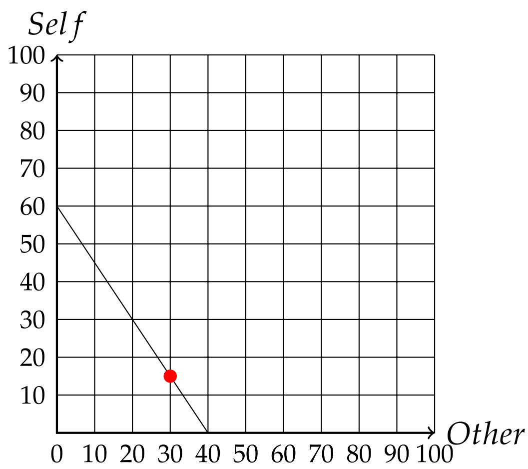


Figure A.9: Allocation Decision in Budget 2

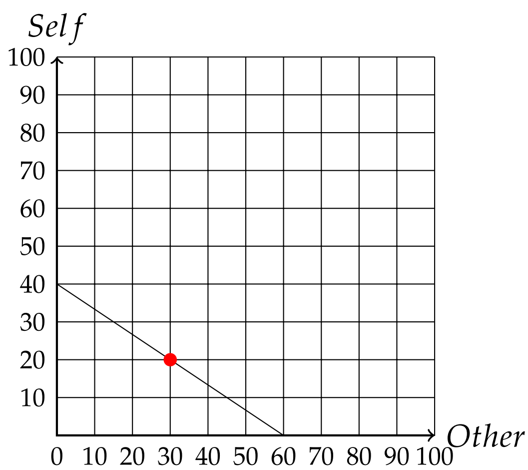


Figure A.10: Allocation Decision in Budget 3

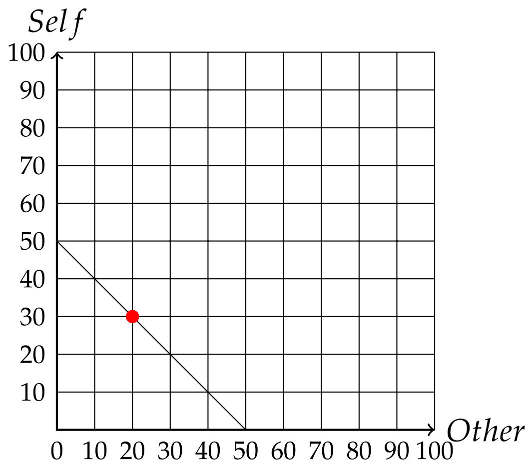


Figure A.11: Allocation Decision in Budget 4

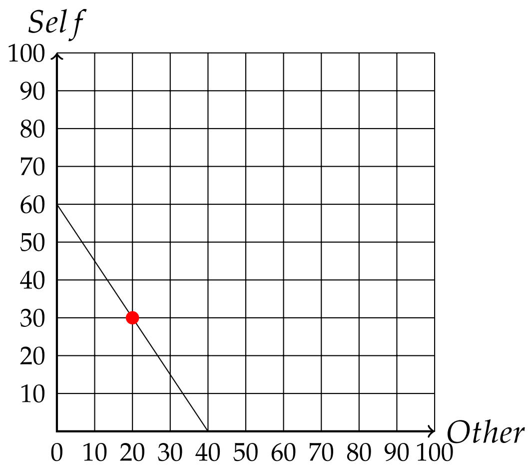


Figure A.12: Allocation Decision in Budget 5

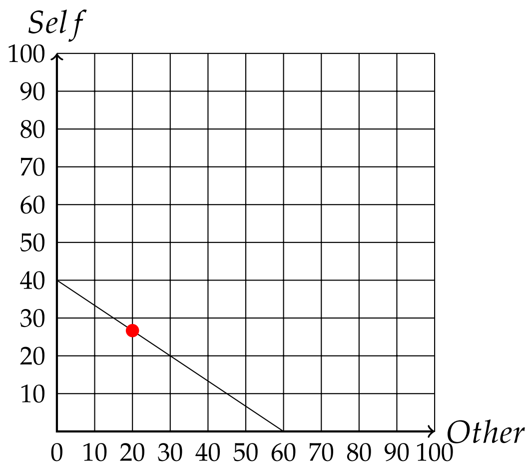


Figure A.13: Allocation Decision in Budget 6

A.2 Strategic Misrepresentation vs. Intrinsic Compliance: How Culture Shapes Normative Signaling and Cooperation in a Repeated Public Goods Game

Welcome, and thank you for participating in this experiment. The session includes several tasks (games) and a short survey. The entire experiment will take approximately 40-50 minutes, and the survey will take less than two minutes.

Silence and Conduct

(i) Communication is strictly prohibited between participants, especially while the experimenter is reading the instructions.
(ii) Turn off your mobile phone.
(iii) If you have any questions, do not ask them aloud. Instead, raise your hand and the experimenter will assist you individually.

Payoffs

(i) You will receive a monetary payoff based on your performance in the experiment.
(ii) Each experimental currency point (token) is worth €0.07.[3]
(iii) Your total payoff will comprise earnings from the tasks plus a fixed show-up fee of €3.00.[4]

Anonymity

(i) Your identity will remain completely anonymous throughout the experiment.
(ii) You will not know who the other participants are, and they will not know who you are.

Data Protection

In accordance with EU data protection regulations and Chinese law, all data and decisions will be handled anonymously and will not be disclosed.

[3] The experimental exchange rate of one token is equal to €0.07 for German participants and 0.1 RMB (€0.012) for Chinese participants.
[4] German participants received a show-up fee of €3.00, and Chinese participants received a show-up fee of 15 RMB (€1.8).

A.2.1 Task I: Dice Game

In this part, you will play a dice game where your profit depends on the result of your die rolls. Each participant may earn a different amount, as the outcomes will vary.

On the table in front of you, there are a die and a cup. Your task is to roll the die twice using the cup.

1. **Roll the die twice** using the cup. *Note: Do not show your roll outcomes to other participants.*
2. After rolling twice, please enter the **first roll result** and its **corresponding profit** on the screen.

You will see a table showing the profit for each number in Table A.2. *For example: if your first roll is 1, you earn 5 tokens. If it's 6, you earn 0 tokens.*

Table A.2: Numbers and Profits for the Dice Game

Numbers	1	2	3	4	5	6
Profits	5	10	15	20	25	0

A.2.2 Task II: Trash Game

In this task, you are given 100 balls, which represent pieces of trash—such as plastic bottles, used paper, and similar items. Your job is to **distribute these balls between Box A and Box B**.

Important Notes:

(i) The total number of balls placed in Box A and Box B **must add up to 100**.
(ii) **Box A** represents a **trash can**.
(iii) **Box B** represents the **ground**.

Placing balls in Box A means you are properly disposing of trash. Placing balls in Box B means you are littering by throwing trash on the ground. The commonly accepted social norm is: "***People should throw all trash balls into the trash can (Box A).***"

Task 2 relates to Task 3. Your decision in **Task 2** will determine how you are grouped in **Task 3**. In other words, the way you distribute trash balls between Box A and Box B will affect which participants you are matched with in the next task. You will be placed in a group with three other participants whose number of trash balls placed in **Box A** is **similar to yours**. *For example: if you put 80 balls into Box A, you will be grouped with others who also placed around 80 balls into Box A.*

Note: *This paragraph explains the relationship between tasks and appears* ***only*** *in the instructions for the* ***signaling treatment****. It is* ***not included*** *in the instructions for the* ***baseline group*** *or the* ***internalization of social norms treatment****.*[5]

Each ball placed in Box A earns you **0.25 tokens**, while each ball placed in Box B earns you **0.4 tokens**. Let ***X*** be the number of balls you place in Box A. Then, ***(100-X)*** balls are placed in Box B. Your profit is calculated as:

$$Profit = 0.25 \cdot X + 0.4 \cdot (100 - X)$$

Your final payoff will be based on your total profit, converted at an exchange rate of **€0.07 per token**:

$$Payoff = 0.07 \times Profit = 0.07 \times [0.25 \cdot X + 0.4 \cdot (100 - X)]$$

Exercise

To ensure clear understanding of the game, please answer the following questions.

You are given an endowment of **100 trash balls**. (i) How many tokens will you earn for placing one trash ball in the trash can (Box A)?

(A) 1 token (B) 0.5 tokens (C) 0.25 tokens (D) 0.4 tokens

(ii) How many tokens will you earn for placing one trash on the ground (Box B)?

(A) 1 token (B) 0.5 tokens (C) 0.25 tokens (D) 0.4 tokens

You are given an endowment of **100 trash balls**. (i) How many tokens will you get from putting all 100 trash balls in the trash can (Box A)?

(A) 10 tokens (B) 20 tokens (C) 25 tokens (D) 40 tokens

(ii) How many tokens will you get from throwing 100 trash balls to the ground (Box B)?

(A) 10 tokens (B) 20 tokens (C) 25 tokens (D) 40 tokens

[5] The instructions for the internalization of social norms treatment are identical to those of the baseline group. In both treatments, participants are **not informed** about how group assignments are made, so they share the same set of instructions.

(iii) How many tokens will you get in total from putting 70 trash balls in the trash can (Box A) and throwing 30 trash balls to the ground (Box B)?

(A) 14.5 tokens (B) 20.5 tokens (C) 29.5 tokens

(iv) Putting trash balls in the trash can is consistent with pro-morality (following social norms).

(A) True (B) False

A.2.3 Task III: Repeated Public Goods Game

In this task, you will be anonymously assigned to a group of four participants.[6] The task consists of 10 rounds, and your group composition will remain the same throughout. Your identity will not be revealed at any point, so other group members will not know who you are.

In each round:

- You will receive **20 tokens**.
- You must decide how to allocate these tokens between your **private account** and **public account**.

The decision procedure is as follows:

Step 1. Make your investment in the public account for *Round i*.
Step 2. Review the results for this round, including:
- Your investment
- Total investment by other group members
- Profit from your private account
- Profit from the public account
- Total profit for the round

Step 3. Proceed to the next round (*Round i+1*).

For example, if you invest X tokens in the public account (where $0 \le X \le 20$*), and your group members invest Y, M, and N tokens respectively, your profit for that round is calculated as:*

$$Profit = 20 - X + 0.4 \cdot (X + Y + M + N)$$

where $20 - X$ is your profit from the private account (tokens you kept), and $0.4 \cdot (X + Y + M + N)$ is your share of the public account return.

Your final payoff in this task is based on the **average profit across all 10 rounds**, converted

[6] The description in ST is: "*you will be assigned to a group of four participants according to the principle outlined in Task 2*"

at an exchange rate of **€0.07 per token**:

$$Payoff = 0.07 \times \frac{1}{10} \times \sum_{i=1}^{10} Profit_i$$

Exercise

You may invest any amount between **0 and 20 tokens** in the public account. If you invest ***X*** tokens, and the other group members invest ***Y***, ***M***, and ***N*** tokens respectively, your profit for the round will be: $20 - X + 0.4 \cdot (X + Y + M + N)$.

Scenario 1: When your investment was 0 tokens, and the other members invested 20, 20, and 20 tokens, respectively, your profits in this round would be _____ tokens, and the profits of the other members would be _____, _____, and _____ tokens, respectively.

Scenario 2: When your investment was 20 tokens, and the other members invested 15, 15, and 15 tokens, respectively, your total profits in this round would be _____ tokens, and the profits of the other members would be _____, _____, and _____ tokens, respectively.

Scenario 3: When your investment was 20 tokens, and the other members invested 0, 0, and 0 tokens, respectively, your total profits in this round would be _____ tokens, and the profits of the other members would be _____, _____, and _____ tokens, respectively.

A.3 Group Decisions Rules and Intertemporal Risk Choices

A.3.1 Robustness Checks using Ordered Probit Models

This appendix presents robustness checks for the main regression analyses reported in Chapter 4. Given that the dependent variable (number of risky choices) is bounded and ordinal (integers from 0 to 10), I re-estimated the baseline specifications using Ordered Probit models.

Table A.3 corresponds to the analysis of decision changes (Table 4.5 in Chapter 4). Table A.4 and Table A.5 correspond to the analyses of risk attitudes and delayed effects, respectively. Across all specifications, the signs and statistical significance of the key coefficients remain qualitatively identical to the OLS estimates reported in Chapter 4. This confirms that my findings are not driven by the choice of the linear probability model.

Table A.3: Ordered Probit Estimates of Changes in Risky Choices ($D_3 - D_1$)

Specifications	Dependent variable: $D_3 - D_1$			
	RC task		RIC task	
	(1)	(2)	(3)	(4)
MT(=1)	0.262*	0.269*	0.009	0.026
	(0.148)	(0.151)	(0.156)	(0.159)
Male(=1)		0.007		-0.260*
		(0.159)		(0.143)
Age		-0.003		0.025**
		(0.012)		(0.011)
Education		-0.025		-0.043
		(0.086)		(0.103)
Extraversion		-0.091		0.154**
		(0.069)		(0.078)
Agreeableness		-0.018		-0.102
		(0.097)		(0.104)
Conscientiousness		0.005		-0.090
		(0.084)		(0.085)
Neuroticism		0.138		0.028
		(0.072)		(0.067)
Observations	183	183	183	183
R-squared	0.0045	0.0152	0.0000	0.0193

Note: Standard errors are clustered at the group level. The term $D_3 - D_1$ denotes the difference in the number of risky-soon (or risky-delayed) choices between the final group decision and the corresponding individual decision. The variable "*MT*" is a binary indicator equal to 1 for the majority treatment and 0 for the unanimity treatment. Male is a dummy variable equal to 1 for male participants and 0 for female participants.
$^{\$}$Significance levels indicate at * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.4 presents the Ordered Probit estimates corresponding to the analysis of risk attitudes in Chapter 4 (Table 4.8). Consistent with the OLS results, the coefficient for Risk Aversion (R.A.) remains consistently negative and significant across most specifications, confirming that risk-averse members exert a strong influence on group decisions.

While there are minor variations in the significance levels of specific Wald tests compared to the linear specifications—particularly in the RIC task—the overall pattern remains robust: risk aversion acts as a dominant constraint under both rules, whereas risk neutrality shows a similar but weak effect in the RC task but is not significantly correlated with group decisions in the RIC task.

Table A.4: Ordered Probit Estimates of Final Group's Decisions (D_3)

Specifications	Dependent variable: $D_3 - D_1$			
	RC task		RIC task	
	(1)	(2)	(3)	(4)
MT	0.859	0.997	0.900	0.924
	(0.812)	(0.842)	(1.014)	(0.990)
R.A.	-0.969*	-1.048*	-0.686**	-0.621**
	(0.513)	(0.543)	(0.288)	(0.299)
R.N.	-0.165	-0.207	-0.268	-0.214
	(0.569)	(0.602)	(0.313)	(0.332)
MT×R.A.	-0.559	-0.651	-1.020	-1.065
	(0.784)	(0.812)	(0.988)	(0.961)
MT×R.N.	-0.867	-1.043	-1.083	-1.079
	(0.812)	(0.844)	(1.036)	(0.997)
Male		0.193		-0.184
		(0.172)		(0.155)
Age		-0.007		-0.005
		(0.011)		(0.012)
Education		-0.084		-0.039
		(0.079)		(0.081)
Extraversion		0.044		0.017
		(0.061)		(0.081)
Agreeableness		-0.061		-0.036
		(0.101)		(0.101)
Conscientiousness		0.001		0.063
		(0.087)		(0.102)
Neuroticism		0.052		-0.032
		(0.097)		(0.084)
Observations	183	183	183	183
R-squared	0.0458	0.0514	0.0256	0.0302
Wald test				
H_0: No significance in...				
...$\beta_{R.A.} = \beta_{R.N.}$	-0.804***	-0.841***	-0.418*	-0.407*
...$\beta_{R.A.} + \beta_{MT\times R.A.} = 0$	-1.528***	-1.699***	-1.706*	-1.686*
...$\beta_{R.N.} + \beta_{MT\times R.N.} = 0$	-1.032**	-1.25**	-1.351	-1.293
...$\beta_{R.A.} + \beta_{MT\times R.A.} = \beta_{R.N.} + \beta_{MT\times R.N.}$	-0.496**	-0.449**	-0.355	-0.393*

Note: Standard errors are clustered at the group level. The term $D_3 - D_1$ denotes the difference in the number of risky-soon (or risky-delayed) choices between the final group decision and the corresponding individual decision. The variable "*MT*" is a binary indicator equal to 1 for the majority treatment and 0 for the unanimity treatment. Male is a dummy variable equal to 1 for male participants and 0 for female participants.
$^{\$}$Significance levels indicate at * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.5: Ordered Probit Estimates of Final Group's Decisions (D_3) by Tasks and Treatments

Specifications	Dependent variable: Final group's decisions							
	RC task				RIC task			
	MT	MT	UT	UT	MT	MT	UT	UT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$N_{R.A.}$	-1.807***	-1.956***	-1.546**	-1.555**	-1.453***	-1.496	-0.823***	-0.859***
	(0.553)	(0.531)	(0.701)	(0.697)	(0.373)	(0.367)	(0.285)	(0.286)
$N_{R.N.}$	-1.151**	-1.298***	-1.132	-1.089	-0.422	-0.339	-0.269	-0.240
	(0.526)	(0.492)	(0.732)	(0.719)	(0.338)	(0.317)	(0.293)	(0.282)
Male		0.030		-0.403*		0.431*		-0.056
		(0.325)		(0.223)		(0.014)		(0.228)
Age		-0.001		-0.018		-0.002		0.011
		(0.017)		(0.017)		(0.015)		(0.013)
Education		-0.017		0.003		-0.246		-0.142
		(0.131)		(0.121)		(0.178)		(0.134)
Extraversion		0.035		0.021		-0.0003		-0.037
		(0.075)		(0.117)		(0.114)		(0.111)
Agreeableness		-0.270*		-0.036		-0.027		-0.101
		(0.139)		(0.161)		(0.109)		(0.103)
Conscientiousness		-0.130		0.084		0.024		0.081
		(0.125)		(0.122)		(0.148)		(0.154)
Neuroticism		0.027		-0.163		0.141		0.051
		(0.191)		(0.119)		(0.121)		(0.108)
Observations	90	90	93	93	90	90	93	93
R-squared	0.1626	0.1803	0.0886	0.1050	0.1693	0.1861	0.0629	0.0714
Wald test								
H_0: No significant in...								
...$\beta_{N_{R.A.}} - \beta_{N_{R.N.}} = 0$	-0.656***	-0.658***	-0.414**	-0.466**	-1.031***	-1.157***	-0.554*	-0.619**

Note: Standard errors are clustered at the group level. $N_{R.A.}$ and $N_{R.N.}$ denote the number of risk-averse and risk-neutral participants within each group, respectively. The baseline group is the number of risk seeking participants within each group. Controls contain age, gender, education, and Big Five personality traits.

$The reference category is *Risk-Seeking Participants*.

$Significance levels indicate at * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

A.3.2 Instructions

Welcome!
Welcome to the experiment, and thank you for participating. Please remain silent and do not communicate with other participants during the session.

You will make a series of decisions throughout the experiment. Your final earnings depend on your performance. Each experimental point is worth €0.50. In addition to your earnings, you will receive a fixed show-up fee of €3.00. At the end of the experiment, you will be asked to complete a brief survey and questionnaire, which will take less than five minutes.

This study investigates decision-making and offers you the opportunity to earn money based on your choices. Please follow these instructions carefully:

1. Turn off your mobile phone.

2. Do not engage in unrelated activities such as browsing the internet, playing games, or reading. Violations will result in exclusion from the experiment and forfeiture of payment.

3. If you have a question, raise your hand. The experimenters will assist you privately. Please do not ask questions aloud.

4. All data will be collected anonymously and will not be linked to your identity.

If you agree to participate, please click "**Next**" to begin. On the next page, you will see the experimental instructions and begin making your decisions.

Instructions for Part A: Individual Risky Choice Task

You will participate in a four-part experiment: Part A, Part B, Part C, and Part D. Your final payoff will be based on one randomly selected part, and within that part, one of your decisions will be randomly chosen to determine your earnings.

You can collect your payment at our office (AWI 00.012) either tomorrow or four weeks later, depending on your choice. The exact pickup date will be sent to you via email.

If you agree to participate, please click "**Next**". On the following page, you will see the instructions for Part A and begin making your decisions.

Individual Decision Task You will make choices between two options in 10 decision-making scenarios. Each scenario presents:

i) ***Option A:*** Receive 8.0 points with 100% certainty **tomorrow**.

ii) ***Option B:*** Receive 12.0 points with probability *p*, or 4.0 points with probability *1-p*, also **tomorrow**.

*Example: In Scenario 1, Option B offers a 10% chance of earning 12.0 points and a 90% chance of earning 4.0 points **tomorrow**. If you choose Option A, you will receive 8.0 points with certainty **tomorrow**.*

Please note: your earnings from this part will only be paid if it is randomly selected as the payoff-relevant part.

After completing all 10 choices, click "**Next**" to continue with the experiment.

Table A.6: Payoff for the Risky Choice (RC) Task

No.	Option A (***Tomorrow***)	Or...	Option B (***Tomorrow***)
1	8.0 points with 100%	A ○ ○ B	12.0 points with 10%, otherwise 4.0 points
2	8.0 points with 100%	A ○ ○ B	12.0 points with 20%, otherwise 4.0 points
3	8.0 points with 100%	A ○ ○ B	12.0 points with 30%, otherwise 4.0 points
4	8.0 points with 100%	A ○ ○ B	12.0 points with 40%, otherwise 4.0 points
5	8.0 points with 100%	A ○ ○ B	12.0 points with 50%, otherwise 4.0 points
6	8.0 points with 100%	A ○ ○ B	12.0 points with 60%, otherwise 4.0 points
7	8.0 points with 100%	A ○ ○ B	12.0 points with 70%, otherwise 4.0 points
8	8.0 points with 100%	A ○ ○ B	12.0 points with 80%, otherwise 4.0 points
9	8.0 points with 100%	A ○ ○ B	12.0 points with 90%, otherwise 4.0 points
10	8.0 points with 100%	A ○ ○ B	12.0 points with 100%, otherwise 4.0 points

Instructions for Part B: Collective Risky Choice Task

In this part of the experiment, you will be randomly assigned to a team of three participants (Participant 1, Participant 2, and Participant 3) and will make decisions together. You will receive the same list of 10 decision-making scenarios as Part A. However, the team's final decision will be determined by the team's voting outcome.

Step 1: Each team member submits an individual proposal by completing the choice list, identical to that as in Part A.

Step 2: You and your teammates will communicate and coordinate via computer chat.

Step 3: The final decision will be made by a majority vote, requiring agreement from **at least two out of three** members. If no agreement is reached in the first round, all members may revise their proposals, continue discussion, and vote again. If no decision is reached after five rounds of voting and communication, the team will receive no payoff for Part B.

In the unanimity treatment, Step 3 is modified as follows:

> Step 3: The final decision must be reached by a unanimous vote of all three members. If no agreement is reached in the first round, all members may revise their proposals, continue discussion, and vote again. If no decision is reached after five rounds, the team will receive no payoff for Part B.

Please note that the profit from this part will be paid out individually to each team member only if this part is randomly selected as the payoff-relevant one. After reading the instructions, click "**Next**" to proceed.

In this part of the experiment, you are randomly assigned to a group of three participants. You are Participant 3. ***Your task is to submit a proposal for the group by completing the following choice list.***

You will face 10 decision-making scenarios, each offering two options:

Option A: Receive 8.0 points with 100% certainty, paid out **tomorrow**.

Option B: Receive 12.0 points with probability p, or 4.0 points with probability $1 - p$, also paid out **tomorrow**.

Example: In Scenario 1, Option B gives each team member a 10% chance of earning 12.0 points and a 90% chance of earning 4.0 points ***tomorrow****. If you choose Option A, each member receives 8.0 points with certainty* ***tomorrow****.*

Once you have completed all choices, click "**Next**" to continue with the experiment.

Coordination: Communication and Voting

You are Participant 1. On this page, you will see three proposals for the decision-making list. The colored list on the left is your own proposal from Round 1; the other two were submitted by your teammates.

Take a moment to review and compare the proposals. You may communicate with your teammates via computer chat to discuss the options and the voting. When ready, cast your vote for one of the three proposals.

The final decision will be made by a **majority vote**, meaning **at least two out of three** teammates must agree on the same proposal.

In the unanimity treatment, voting instructions are as follows:

> Take a moment to review and compare the proposals. You may communicate with your teammates via computer chat to discuss the options and voting. When ready, cast your vote for one of the three proposals. The final decision will be made by a **unanimous vote**, meaning **all three teammates** must agree on the same proposal.

If no agreement is reached in this round, all team members will proceed to revise and resubmit their proposals. You will then enter a new round of discussion and voting. If no decision is reached after five rounds, the team will receive no payoff for Part B.

After voting, click "**Next**" to continue the experiment.

Table A.7: An Example for the RC Task of the Group in Part B

No.	**Option A** (*Tomorrow*)	P1	P2	P3 (me)	**Option B** (*Tomorrow*)
1.	8.0 points with 100%.	A	A	A	12.0 points with 10%, otherwise 4.0 points.
2.	8.0 points with 100%.	A	A	A	12.0 points with 20%, otherwise 4.0 points.
3.	8.0 points with 100%.	A	A	A	12.0 points with 30%, otherwise 4.0 points.
4.	8.0 points with 100%.	A	A	A	12.0 points with 40%, otherwise 4.0 points.
5.	8.0 points with 100%.	A	B	A	12.0 points with 50%, otherwise 4.0 points.
6.	8.0 points with 100%.	B	B	A	12.0 points with 60%, otherwise 4.0 points.
7.	8.0 points with 100%.	B	B	A	12.0 points with 70%, otherwise 4.0 points.
8.	8.0 points with 100%.	B	B	B	12.0 points with 80%, otherwise 4.0 points.
9.	8.0 points with 100%.	B	B	B	12.0 points with 90%, otherwise 4.0 points.
10.	8.0 points with 100%.	B	B	B	12.0 points with 100%, otherwise 4.0 points.

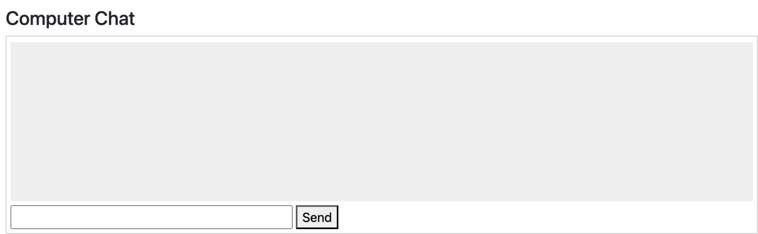


Figure A.14: Computer Chatbox in Part B

Instructions for Part C: Individual Risky-Intertemporal Choice Task

This is a choice list in which you will make a choice between two options in each of 10 scenarios.

In every scenario, you must choose between:

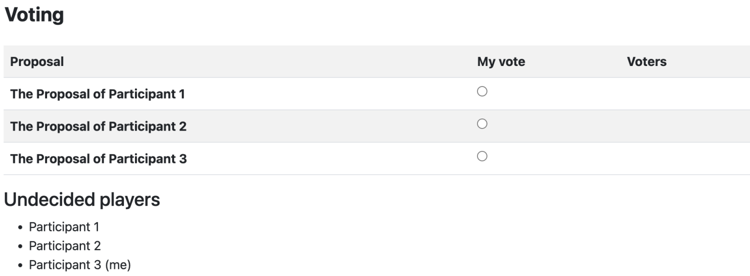


Figure A.15: Cast a Vote in Part B

Option A: Receive 8.0 points with 100% certainty, paid out **tomorrow**.

Option B: Receive 12.0 points with probability p, or 4.0 points with probability $1 - p$, paid out **four weeks later**.

Example: In Scenario 1: if you choose Option B, you will have a 10% chance of earning 12.0 points and a 90% chance of earning 4.0 points ***four weeks later****. If you choose Option A, you will receive 8.0 points* ***tomorrow*** *with certainty.*

Please note: the profit from this part will only be paid out if it is randomly selected as the payoff-relevant part.

After completing all 10 choices, click "**Next**" to continue with the experiment.

Table A.8: Payoff for the Risky-Intertemporal Choice (RIC) Task

No.	Option A (*Tomorrow*)	Or...	Option B (*Four Weeks Later*)
1	8.0 points with 100%	A ○ ○ B	12.0 points with 10%, otherwise 4.0 points
2	8.0 points with 100%	A ○ ○ B	12.0 points with 20%, otherwise 4.0 points
3	8.0 points with 100%	A ○ ○ B	12.0 points with 30%, otherwise 4.0 points
4	8.0 points with 100%	A ○ ○ B	12.0 points with 40%, otherwise 4.0 points
5	8.0 points with 100%	A ○ ○ B	12.0 points with 50%, otherwise 4.0 points
6	8.0 points with 100%	A ○ ○ B	12.0 points with 60%, otherwise 4.0 points
7	8.0 points with 100%	A ○ ○ B	12.0 points with 70%, otherwise 4.0 points
8	8.0 points with 100%	A ○ ○ B	12.0 points with 80%, otherwise 4.0 points
9	8.0 points with 100%	A ○ ○ B	12.0 points with 90%, otherwise 4.0 points
10	8.0 points with 100%	A ○ ○ B	12.0 points with 100%, otherwise 4.0 points

Instructions for Part D: Collective Risky-Intertemporal Choice Task

In this part of the experiment, you will work in a team of three participants (Participant 1, Participant 2, and Participant 3) to make joint decisions. You will receive the same list of 10 decision-making scenarios as in Part C, but this time, the final decision will be determined by a group vote.

Step 1: Each team member submits a proposal by completing the choice list, identical to Part C.

Step 2: You and your teammates may communicate and coordinate via computer chat.

Step 3: The final decision will be made by a **majority vote**, requiring agreement from **at least two out of three** team members. If no agreement is reached in the first round, all members may revise their proposals, continue discussion, and vote again. If no decision is reached after five rounds, the team will receive no payoff for Part D.

In the unanimity treatment, Step 3 is modified as follows:

Step 3: The final decision must be reached by a **unanimous vote of all three** team members. If no agreement is reached in the first round, all members may revise their proposals, continue discussion, and vote again. If no decision is reached after five rounds, the team will receive no payoff for Part D.

Please note that the profit from this part will be paid individually to each team member only if this part is randomly selected as the payoff-relevant one. After reading the instructions, click "**Next**" to proceed.

In this part of the experiment, you are one of three participants assigned to a group. You are Participant 2. ***Your task is to submit a proposal for the group by completing the following choice list.***

You will face 10 decision-making scenarios, each offering two options:

Option A: Receive 8.0 points with 100% certainty, paid out **tomorrow**.

Option B: Receive 12.0 points with probability p, or 4.0 points with probability $1 - p$, paid out **four weeks later** to each team member.

Example: In Scenario 1, if you choose Option B, each team member will have a 10% chance of receiving 12.0 points and a 90% chance of receiving 4.0 points ***four weeks later****. If you choose Option A, each member will receive 8.0 points* ***tomorrow*** *with certainty.*

Once you have completed all choices, click "**Next**" to continue with the experiment.

Coordination: Communication and Voting

You are Participant 1. On this page, you will see three proposals for the decision-making list. The choice list on the left is your own proposal from Round 1; the other two were submitted by your teammates.

Take time to review and compare your proposal with those of your teammates. You may communicate with them via computer chat to discuss the proposals and the voting. When ready, cast your vote for one of the three proposals.

The final decision will be made by a **majority vote**, meaning **at least two out of three** team members must agree on the same proposal.

In the unanimity treatment, voting instructions are as follows:

Take time to review and compare your proposal with those of your teammates. You may communicate with them via computer chat to discuss the proposals and the voting. When ready, cast your vote for one of the three proposals. The final decision will be made by a **unanimous vote**, meaning all three members must agree on the same proposal.

If no agreement is reached in this round, all team members will proceed to revise and resubmit their proposals. You will then enter a new round of discussion and voting. If no decision is reached after five rounds, the team will receive no payoff for Part D.

After voting, click "**Next**" to continue with the experiment.

Table A.9: An Example of the RIC Task for Groups in Part D

No.	**Option A (Tomorrow)**	P1	P2	P3 (me)	**Option B (Four weeks later)**
1.	8.0 points with 100%.	A	A	A	12.0 points with 10%, otherwise 4.0 points.
2.	8.0 points with 100%.	A	A	A	12.0 points with 20%, otherwise 4.0 points.
3.	8.0 points with 100%.	A	A	A	12.0 points with 30%, otherwise 4.0 points.
4.	8.0 points with 100%.	A	A	A	12.0 points with 40%, otherwise 4.0 points.
5.	8.0 points with 100%.	A	A	A	12.0 points with 50%, otherwise 4.0 points.
6.	8.0 points with 100%.	A	B	A	12.0 points with 60%, otherwise 4.0 points.
7.	8.0 points with 100%.	B	B	A	12.0 points with 70%, otherwise 4.0 points.
8.	8.0 points with 100%.	B	B	B	12.0 points with 80%, otherwise 4.0 points.
9.	8.0 points with 100%.	B	B	B	12.0 points with 90%, otherwise 4.0 points.
10.	8.0 points with 100%.	B	B	B	12.0 points with 100%, otherwise 4.0 points.

Figure A.16: Computer Chatbox in Part D

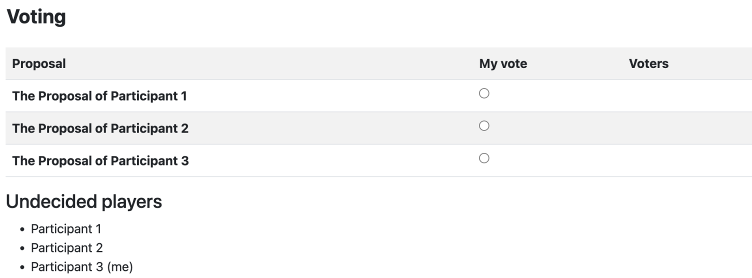


Figure A.17: Cast a Vote in Part D

Instructions for the Big Five Personality Traits

Please evaluate the following statements by completing the sentence: "I see myself as someone who ..." This section will take only a few minutes to complete.

Table A.10: The Big Five Personality Traits

	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree Strongly
... is reserved	○	○	○	○	○
... is generally trusting	○	○	○	○	○
... tends to be lazy	○	○	○	○	○
... is relaxed, handles stress	○	○	○	○	○
... has few artistic interests	○	○	○	○	○
... is outgoing, sociable	○	○	○	○	○
... tends to find fault with others	○	○	○	○	○
... does a thorough job	○	○	○	○	○
... gets nervous easily	○	○	○	○	○
... has an active imagination	○	○	○	○	○