

UC Merced

UC Merced Electronic Theses and Dissertations

Title

When Not Knowing Matters: Information and Morality in Economic Decisions

Permalink

<https://escholarship.org/uc/item/2kb0h3dk>

ISBN

9798197865311

Author

Hua, Tony

Publication Date

2026-05-07

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial License, available at <https://creativecommons.org/licenses/by-nc/4.0/>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, MERCED

**When Not Knowing Matters: Information and
Morality in Economic Decisions**

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

in

Economics

by

Tony Hua

Committee in charge:

Professor Robert Innes, Chair

Professor Zack Grossman

Professor Paul Smaldino

2026

The dissertation of Tony Hua is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

(Professor Zack Grossman)

(Professor Paul Smaldino)

(Professor Robert Innes, Chair)

University of California, Merced

2026

TABLE OF CONTENTS

	Signature Page	iii
	List of Figures	vii
	List of Tables	viii
	Acknowledgements	ix
	Curriculum Vitae	x
	Abstract	xvi
Chapter 1	Introduction	1
Chapter 2	Unwillingly Informed: the Prosocial Impact of Third-Party Informers	6
	2.1 Introduction	6
	2.2 Experimental Design	8
	2.2.1 Treatments	9
	2.2.2 Hypotheses	13
	2.2.3 Implementation	14
	2.3 Results	16
	2.3.1 Dictator Allocation Behavior in the Moral Wiggle-Room Game	16
	2.3.2 Dictator Information Choice in the Moral Wiggle-Room Game	20
	2.3.3 Dictator's decision to reward the informer	20
	2.3.4 Informer Behavior	21
	2.4 Conclusion	22
Chapter 3	Latent Norms: Prosocial Nudges Under Moral Ambiguity	25
	3.1 Introduction	25
	3.2 Hypotheses	27
	3.3 Experimental Design	28
	3.3.1 Moral Wiggle Room Games	28
	3.3.2 Dice Rolling Game	30
	3.3.3 Binary Choice Dictator Game	31

3.4	Results	32
3.4.1	Moral Wiggle Room Game	32
3.4.2	Dice Rolling Game	36
3.4.3	Binary Choice Dictator Game	37
3.5	A Bounded Rationality Model of Latent Norms	40
3.5.1	Preferences and Decision Problem	41
3.5.2	Optimal Attention and Action	41
3.5.3	Belief Timing: Commitment and Rationalization	42
3.5.4	Summary	43
3.6	Conclusion	44
Chapter 4	I Didn't Know Either: Social Information and Strategic Ignorance	46
4.1	Introduction	46
4.2	Experimental Design	47
4.2.1	Social Information Treatments	49
4.2.2	Hypotheses	50
4.3	Results	50
4.3.1	Data	50
4.3.2	Response to Social Information	51
4.4	Conclusion	56
Chapter 5	Conclusion	57
5.1	Information Avoidance and Social Asymmetry	57
5.2	From External Enforcement to Internal Activation	58
5.3	Internalized Shame and the Salience of Social Standards	59
5.4	A Unified View of Moral Regulation	59
5.5	Implications and Future Directions	62
5.6	Concluding Remarks	63
Appendix A		73
A.1	Summary Statistics and Demographics	73
A.2	Replication treatments	74
A.3	Treatment Diagram	77
A.4	Robustness check: removing age outliers	80
A.5	Informer Beliefs and Choice Reasoning	83
A.6	Hypothetical Informer treatments	85
A.7	Screenshots of Instructions	88

Appendix B	93
B.1 Demographics	93
B.2 Canonical Moral Wiggle Room Payoffs	97
B.3 Reported Dice Distributions	99
B.4 Framing of Belief Elicitation on Reporting Die Results	101
B.5 Experiment Interface for the MWRG	103
B.6 Experiment Interface for the DRG	116
B.7 Experiment Interface for the BCDG	126
Appendix C	136
C.1 Demographics	136
C.2 Analysis of Demographic Covariates and Result on <i>Ignorance Expectant</i> Dictators	138
C.3 Social Value Orientation (SVO) Moderation	140
C.4 Robustness Check Regression on Ignorance Norms	141
C.5 Sentiment and Linguistic Analysis of Subjects' Open Response	141
C.6 Moral Universalism Trust Survey	143
C.7 Conformity Score	146
C.8 Social Appropriateness Ratings of Ignorance	146
C.9 Experiment Interface for the MWRG	149

LIST OF FIGURES

Figure 1.1: Normative Evaluation of Willful Ignorance	3
Figure 2.1: Experiment Flow	11
Figure 2.2: <i>Info-First</i> Comparison	23
Figure 2.3: Bonus Withholding Rates	24
Figure 3.1: Ignorance Rates	33
Figure 3.2: Beliefs About Ignorance Rates	34
Figure 3.3: Dictator Behavior Across Treatments.	35
Figure 3.4: Distribution of Excess Payoffs	37
Figure 3.5: Belief Distribution in Dice Rolling Game	38
Figure 3.6: Prosocial Choice	39
Figure 4.1: Ignorance Rate and Beliefs	52
Figure 4.2: Ignorance Rates Across Norm Treatments	53
Figure 4.3: Ignorance Rates by Types	54
Figure 4.4: Distribution of Beliefs	55
Figure 5.1: Diagram of Moral Regulation	61
Figure A.1: Overview of Experimental Conditions	79
Figure A.2: The distribution of age by treatment	80
Figure B.1: Ignorance Rates By Treatments	98
Figure B.2: Reported Die Roll Distribution	100
Figure C.1: Distribution of participants' Universalism Trust scores.	145
Figure C.2: Perceived social appropriateness of information avoidance across norm treatments.	148

LIST OF TABLES

Table 2.1:	Payoff Table	9
Table 2.2:	Average Recipient Payoffs	17
Table 2.3:	Unwillingly-informed Switchers	18
Table 2.4:	Information Avoidance and Imposition	19
Table 3.1:	Summary of Experimental Designs and Treatment Conditions . . .	32
Table A.1:	Summary Demographics for Moral Wiggle Room Game	73
Table A.2:	Summary statistics for decision variables across <i>all</i> treatments . .	74
Table A.3:	The results of the replication treatments are comparable to those of Dana et al. (2007); Grossman (2014).	77
Table A.4:	Average recipient payment for dictators under the age of 55	81
Table A.5:	Main decisions for subjects under the age of 55	82
Table A.6:	Informer answering "yes" in the questionnaire, by treatment	83
Table A.7:	Informer Treatment	84
Table A.8:	Informer Bonus Treatment	84
Table B.1:	Summary Demographics for the Moral Wiggle Room Game	94
Table B.2:	Summary Demographics for the Dice Rolling Game	95
Table B.3:	Summary Demographics for the Binary Choice Dictator Game . . .	96
Table B.4:	Regression of Excess Payoffs On Framing	102
Table C.1:	Summary Demographics for the Moral Wiggle Room Game	137
Table C.2:	Demographics on Being a <i>Ignorance Expectant</i> Versus <i>Reveal</i> <i>Expectant</i> Dictator	139
Table C.3:	SVO moderation of ignorance in the Moral Wiggle Room Game . .	140
Table C.4:	Probit Regression on Ignorance Choice and Ignorance Environments	142
Table C.5:	Sentiment and Linguistic Theme Analysis by Expectant Type . . .	143
Table C.6:	Effect of Universalism Score on the Likelihood of Avoiding Information	144
Table C.7:	Effect of Conformity Score on the Likelihood of Following Norms or Being a <i>Ignorance Expectant</i> Dictator	146

ACKNOWLEDGEMENTS

To my dearest friends, who have stayed with me in ways words cannot fully express.

TONY HUA

(626)-315-0814 | [thua5\[at\]ucmerced.edu](mailto:thua5[at]ucmerced.edu)
<https://sites.google.com/view/tony-m-hua>
Citizenship: USA

EDUCATION

PhD <i>Economics</i> University of California, Merced	Aug. 2020 – May 2026 Merced, CA
BA <i>Major: Economics, Minor: Writing</i> University of California, Merced	Aug. 2010 – May 2015 Merced, CA
BS <i>Major: Applied Mathematics with Economics Emphasis</i> University of California, Merced	Aug. 2010 – May 2015 Merced, CA

RESEARCH INTERESTS

Economics and psychology, experimental economics, human-AI interaction, agent-based modeling

JOB MARKET PAPER

“Latent Norms: Prosocial Nudges Under Moral Ambiguity” (JMP)

People often look to social norms—what is considered socially appropriate—as a guide when making morally ambiguous decisions. Yet these norms are not always at the forefront of awareness, often remaining latent until brought to mind. I show that eliciting beliefs about others’ behavior can preemptively activate these *latent social norms*—internalized standards that are not always retrieved at the moment of choice. Across three incentivized experiments—a moral wiggle-room game, a dice-rolling task, and a binary-choice dictator game—I find that social information and the salience of norms shape behavior under moral ambiguity, but not when norms are already explicit. Simply eliciting beliefs about what others do *before* choice reduced selfish behavior even without providing new information: information avoidance fell by 18 percentage points, and cheating in excess earnings declined by 18%. In contrast, eliciting beliefs *after* choice produced the opposite pattern: participants rationalized selfish actions by inflating their stated expectations of others’ selfishness. When beliefs were elicited before choice, participants largely refrained from avoiding information. However, once they were shown that many others avoided information, they became more willing to do so themselves: a 70% norm increased avoidance by 10 percentage points, and even a 30% norm led to more avoidance than the control. By contrast, a 10% norm had no effect. In short, belief elicitation reduced selfish behavior, but this restraint weakened when ignorance appeared common. Together, these results suggest that belief elicitation can

mitigate selfish behavior by activating latent normative commitments at the point of decision.

PUBLICATIONS

“Unwillingly Informed: the Prosocial Impact of Third-Party Informers” (with Zack Grossman, Karine Nyborg, and Jo Thori Lind) *forthcoming at Journal of Environmental Economics and Management*

While people often avoid learning about inconvenient or undesirable consequences of their actions in order to behave selfishly, many social situations involve a third person who is in a position to impose this information. How does the presence of a potential informer affect information, behavior, and welfare in social decisions with moral wiggle-room? How prone are potential informers to impose this information and do information-avoidant decision-makers sanction informers that spoil their attempts to remain ignorant? We introduce a third-party informer into the moral wiggle-room game and find that dictators are undeterred from avoiding information. However, unwillingly-informed dictators revise their decisions to be less selfish and dictators are less likely to reward informers bringing bad news. Informers do not respond to the threat of retaliation, being highly likely to supply information. Interestingly, a subtle change in the choice interface - separating the dictator’s ignorance and allocation choices in two separate screens - causes a substantial reduction in the share of dictators choosing ignorance.

“Willful Ignorance in Social Decisions: Robust, Yet Contextually Sensitive” (with Zack Grossman), 2025, *Current Opinion in Psychology*, 2026, 67:102210

Although humans exhibit many prosocial behaviors, when the social benefits of their options are uncertain, surprisingly many avoid learning them before choosing, using ignorance as an excuse to dodge moral obligations and revert to selfish behavior. This kind of willful ignorance is robust in the sense that researchers have documented it using a wide array of methods, across diverse settings, and a time period spanning nearly two decades. At the same time, however, the degree to which it manifests is inconsistent across and within studies. Some of these inconsistencies stem from obvious factors, while the moderators driving others have yet to be identified or are poorly understood. This study synthesizes and organizes these contextual factors, providing recommendations for future research.

WORKING PAPERS

“AI as an Agent: Preferences for Truth and Drivers of Deception” (with Rob Innes) *under review at Experimental Economics*

This paper studies AI (ChatGPT5) behavior when acting as an agent for a human principal in two Sender-Receiver games: dictator and Gneezy (2005) type deception, both with payoff alternatives that stipulate either an equal or unequal (Sender benefiting) split of \$10. Treatments vary (1) objectives on the extent to which AI is to represent the Sender’s (v. Receiver’s) interests, (2) information about Sender preferences for honesty

and equality, (3) whether AI is deciding between honesty and dishonesty (equality and inequality) or, instead, advising the Sender, and (4) a welfare cost of dishonesty. Given neutral information about Senders' moral behavior, AI exhibits no innate preference for honesty or fairness. When told Senders predominantly prefer honesty, AI is more truthful; however, controlling for preferences over allocations, AI favors lies on net – the opposite of lie aversion. AI is more honest when there are welfare costs of dishonesty, more weight on Receiver interests in its posed objective, and when making a decision v. giving advice.

“I Didn’t Know Either: How Beliefs About Norms Shape Strategic Ignorance”

People often avoid information as a way to justify selfish behavior. However, such behavior often unfolds in social contexts, where expectations about others' behavior may shape moral decision-making. This study investigates how beliefs about descriptive norms influence strategic ignorance in a modified moral wiggle-room game. Participants first predicted how often others acquired information, then received randomly assigned feedback indicating high or low rates of ignorance before making their own decision as the dictator. Individuals were more likely to seek information when told that most others typically did so, and more likely to avoid information when told that ignorance was common. Norm-conforming behavior differed between *ignorance expectants*—those who expected ignorance and *reveal expectants*—those who expected information acquisition. *Ignorance expectants* adjusted only when norm cues strongly contradicted their prior beliefs, whereas *reveal expectants* exhibited modest but consistent responsiveness across norm environments. There is no evidence of ex-ante self-serving belief distortion, though limited behavioral change constrained opportunities to test for ex-post justification. These findings suggest that strategic ignorance responds to descriptive norms, but not through motivated belief distortion.

INVITED TALKS & PRESENTATIONS

Latent Norms: Prosocial Nudges Under Moral Ambiguity	November 2025
Behavioral & Experimental Economics	
Stanford-Caltech-UC Student Conference (BEESCUitS)	Santa Barbara, CA
Belief Elicitation as Behavioral Design: Forecasting Others' Actions Reduces Moral Wiggle Room	October 2025
UC Merced Brown Bag Series	Merced, CA
I Didn’t Know Either: How Beliefs About Norms Shape Strategic Ignorance	June 2025
Western Economic Association International (WEAI)	San Francisco, CA
[Poster] An Agent Based Model of Endogenous Group Identity Formation	May 2023
California Workshop on Evolutionary Social Science (CWESS)	Merced, CA

[Poster] Social Status and Labor Supply
All-California Labor Economics Conference (ACLEC)

September 2023
Santa Barbara, CA

REFEREE SERVICE

Experimental Economics

TEACHING EXPERIENCE

Game Theory	S25
Instructor of Record	UC Merced
Introduction to Economics	S21, S22
Teaching Assistant	UC Merced
Introduction to Business and Finance	F20
Teaching Assistant	UC Merced
Statistical Inference (equivalent to introductory statistics)	F21, F23, F24
Teaching Assistant	UC Merced
Intermediate Microeconomics	F22, S23, F25
Teaching Assistant	UC Merced
Game Programming (game design lecture)	F23, F24
Guest Lecturer	UC Merced
Undergraduate Research	F25
Guest Lecturer	UC Merced

ADDITIONAL EXPERIENCE

Graduate Student Mentor for the Undergraduate Research Opportunities Center (Summer 2024, Summer 2025)
Research Assistant for Professor Christian Fons-Rosen (Summer 2025)
Research Assistant for Professor Briana Ballis (Summer 2023)
Research Assistant for Professor Rob Innes (Summer 2022)
Research Assistant for Professor Zack Grossman (Summer 2021)

TRAININGS AND WORKSHOPS

Choice-Process Data Workshop at the Economic Science Association North American Meeting (November 2022)

Center for Behavioral Experiments Summer School (August 2022)
Workshop on Behavioral Game Theory, University of East Anglia (July 2022)
IFREE Workshop in Experimental Economics, Chapman University (June 2022)
Biopolitics Workshop, UC Merced (February 2022)
Centre for Experimental Social Sciences Nullfield oTree Course (June 2021)

ADDITIONAL GRADUATE LEVEL COURSEWORK

AI in Social Sciences (Neural Networks)
Modeling Social Behavior (Agent Based Modeling)
Political Cognition
Experimental Methods in Political Science
Game Theory in Political Science

SKILLS

Languages: English (Native)
Statistical Analysis Suites: R (proficient), STATA (proficient)
Simulation Suites: MATLAB (intermediate), NetLogo (proficient)
Programming: C# (advanced), Javascript (proficient), Python (proficient)
Frameworks: Unity (proficient), LIONESS (proficient), React JS (novice), oTree (proficient)
Document Creation: LaTeX, Google Apps Script, Filemaker

AWARDS AND FELLOWSHIPS

Graduate Dean's Dissertation Fellowship Spring 2026
Mercatus Center: Elinor Ostrom 2025-2026 Fellowship
UC Merced Economics Department: 2025 Research Grant
UC Merced Economics Department: 2024 Perry-Ellis Fellowship
UC Merced Economics Department: Professional Development Award
UC Merced: Grad Slam 2024 Finalist

INTERNAL SERVICE

Student Brown Bag Seminar Series Fall 2025
Cofounder and Organizer

2024 Deliberative Democracy Forum: Historically Marginalized Graduate and Professional Students
Participant

UC Merced 2022 Graduate Dean Search
Economics Graduate Student Representative

UC Merced GradEXCEL Peer Mentor for 2022 Cohort and 2024 Cohort

PROFESSIONAL EXPERIENCE

Game Designer 2023 – 2024
Goldhawk Interactive (Contractor) (remote)

Software Developer and Game Designer 2015 – 2019
Owner / Indie Developer Los Angeles, CA

Computer Science Classroom Instructor 2019 – 2020
9 Dots Community Learning Center Los Angeles, CA

REFERENCES

Zack Grossman

Associate Professor & Advisor
Economics
University of California, Merced
zgrossman@ucmerced.edu

Karine Nyborg

Professor & Co-author
Economics
University of Oslo
karine.nyborg@econ.uio.no

Rob Innes

Distinguished Professor & Co-author
Economics
University of California, Merced
rinnes@ucmerced.edu

Paul Smaldino

Professor
Cognitive and Information Sciences
University of California, Merced
psmaldino@ucmerced.edu

Elaine Denny

Assistant Professor
Political Science
University of California, Merced
edenny@ucmerced.edu

ABSTRACT OF THE DISSERTATION

When Not Knowing Matters: Information and Morality in Economic Decisions

by Tony Hua

Doctor of Philosophy in Economics

University of California Merced, 2026

Committee Chair: Professor Robert Innes

Individuals often face decisions in which the consequences of their actions for others are uncertain, and in many cases this uncertainty is actively maintained. A growing literature documents that individuals avoid information about the social consequences of their choices, a behavior commonly referred to as strategic ignorance. This dissertation argues that such ignorance is not merely an individual choice, but a socially embedded phenomenon shaped by interactions with others, internalized norms, and beliefs about the behavior of others. Using a series of experiments based on moral wiggle room environments, I examine how different features of the social context influence both the willingness to acquire information and the behavioral consequences of remaining ignorant. The results reveal a striking asymmetry: while many individuals avoid acquiring information themselves, others are readily willing to provide it. When information is imposed, a substantial share of decision-makers revise their choices in a more prosocial direction, demonstrating that ignorance—and its behavioral consequences—can be constrained by others. At the same time, prompting individuals to consider what constitutes appropriate behavior prior to making a decision increases prosocial actions, consistent with the activation of latent norms, whereas eliciting such beliefs after decisions instead reflects post-hoc justification. In addition, individuals respond to information about the behavior of others by aligning both their information acquisition and allocation choices with perceived norms. Taken together, these findings provide a unified account of moral decision-making in which behavior is shaped by a layered system of enforcement, ranging from external interventions to fully internalized regulation. The results highlight that the key determinant of behavior is not only what individuals know, but how the social environment shapes what they choose not to know.

Chapter 1

Introduction

Human decision-making often unfolds under uncertainty about the consequences of one's actions. In many cases, this uncertainty is not merely a constraint imposed by the environment but a condition that individuals actively maintain. People frequently avoid information about the social consequences of their choices—even when that information is readily available and costless to obtain (Momsen and Ohndorf, 2022). This pattern of behavior, commonly referred to as *strategic ignorance*, has been documented across a wide range of economic and social settings (Dana et al., 2007; Cain and Dana, 2012; Grossman and van der Weele, 2017; Exley and Kessler, 2023). Rather than confronting information that might compel them to act more prosocially, individuals sometimes choose not to know.

At first glance, such behavior appears puzzling. Standard economic models would predict that additional information weakly improves decision-making (Mas-Colell et al., 1995). Yet many experimental studies shows that individuals often decline opportunities to acquire information about how their actions affect others, particularly when that information could reveal a conflict between self-interest and social welfare (Dana et al., 2007; Grossman and van der Weele, 2017); this pattern persists even under double-blind conditions in which decisions are not observed by others (Larson and Capra, 2009). One interpretation is that ignorance serves as a psychological buffer: by remaining uninformed, individuals can behave in self-interested ways while preserving a positive self-image (Bénabou and Tirole, 2006, 2011; Festinger, 1957).

The standard experimental paradigm used to study this behavior is the moral wiggle-room game introduced by Dana et al. (2007). In this setting, a decision-maker chooses between options that affect both their own payoff and that of another individual, while the consequences for the other person are initially

uncertain. The decision-maker may choose to reveal this information before acting, but is not required to do so. This creates an environment in which remaining ignorant allows individuals to avoid learning whether their actions harm or benefit others. A large body of work shows that many individuals choose to remain uninformed in such settings (Feiler, 2014; Vu et al., 2023; Grossman and Hua, 2026), and that such behavior cannot be attributed solely to confusion or failed comprehension of the experiment (Exley and Kessler, 2023; van der Weele and Figueroa-Sisniega, 2026).

Figure 1.1 from Hua (2026) indicates how U.S. subjects recruited via the Prolific platform perceive behavior across scenarios in which individuals either possess or lack information about the consequences of their actions. Two patterns emerge. First, actions taken under ignorance are judged as less blameworthy than identical actions taken with full knowledge. Second, information avoidance is itself perceived as largely intentional. Taken together, these findings suggest that ignorance can function as a form of moral cover: it shifts responsibility away from the decision-maker even when the underlying action remains unchanged and the avoidance of information is perceived as intentional. These findings supplement and replicate related evidence from Krupka and Weber (2013) and Grossman and van der Weele (2017).¹

In this context, strategic ignorance can be viewed as a socially embedded behavior. When individuals face choices with uncertain social consequences, they may deliberately preserve that uncertainty in order to maintain flexibility in how their actions are interpreted. Ignorance allows individuals to exploit what has been termed *moral wiggle room*: the ability to act in self-interest while avoiding full acknowledgment of the harm caused (Dana et al., 2007; Fahrenwaldt et al., 2024).²

Thus, this dissertation advances a broader claim: *strategic ignorance is not simply an individual decision—it is shaped by the social environment*. Individuals do not merely avoid information because it is costly or inconvenient; they avoid it because

¹These normative evaluations are based on U.S. subjects recruited from Prolific and should therefore be interpreted with caution, particularly given concerns about the generalizability of findings from WEIRD populations (Henrich, 2020). This caveat is especially relevant because the moral weight placed on intentions varies substantially across societies. For example, Barrett et al. (2016) find cross-cultural variation in the extent to which intentions moderate moral judgment. Comparable cross-cultural evidence on whether ignorance mitigates blame remains limited. However, the behavioral signature of strategic ignorance—higher rates of selfish choice under ignorance than under full information—has been documented across 20 countries (Molho et al., 2025), suggesting that the behavioral phenomenon itself is not confined to a single cultural context, even if the underlying mechanisms may vary.

²The broader moral wiggle room phenomena includes any sort of excuse-seeking behavior that enables self-serving behavior. For other forms of moral wiggle room, see Exley (2016); Brütt et al. (2020); Falk et al. (2020); Exley and Kessler (2024); Vu and Molho (2025).

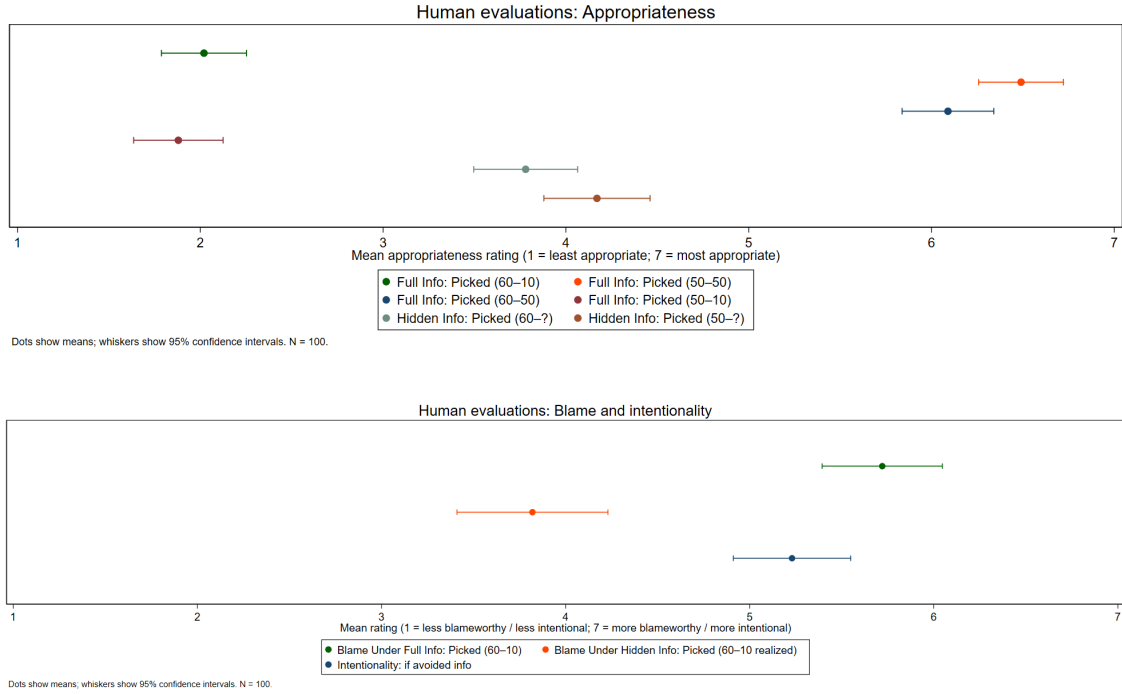


Figure 1.1: Normative Evaluation of Willful Ignorance

Notes: Human normative evaluations across scenarios. Payoffs correspond to the dictator and recipient, respectively. Full info denotes decisions made when the decision-maker observes the recipient’s payoff, while hidden info indicates that the recipient’s payoff is not observed. Dots represent means; whiskers denote 95% confidence intervals.

knowing would make them accountable for the consequences of their actions. At the same time, whether ignorance is possible, and how effective it is as a justification for behavior, depends on the social context in which decisions are made (Bartling et al., 2014; Momsen and Ohndorf, 2023b). Ignorance, in this sense, alters how actions are judged—by oneself and by others—and thereby changes the incentives governing behavior.

The central goal of this dissertation is to understand how different aspects of the social environment shape individuals’ willingness to remain ignorant and the consequences of that ignorance for behavior. Across three chapters, I examine how information avoidance is influenced by (i) the actions of others, (ii) the salience of internalized norms, and (iii) information about the behavior of others.

Chapter 2 studies how strategic ignorance interacts with the social environment through the actions of others. In many real-world settings, individuals are not alone

in their informational choices: others can provide, impose, or withhold information. Chapter 2 introduces a third-party “informer” into a standard moral wiggle-room environment, allowing another individual to reveal information that a decision-maker might prefer to avoid. The results show that while many individuals actively avoid information, others are willing to impose it, and doing so substantially increases prosocial behavior. These findings demonstrate that strategic ignorance is not solely an individual choice—it can be constrained by the actions of others.

Chapter 3 examines how internalized norms influence behavior when they are made salient. Social norms are often *latent*: individuals possess internal standards of appropriate behavior, but these standards do not always guide action unless they are brought to the forefront of decision-making. Chapter 3 uses belief elicitation as an experimental tool to study how the timing of belief formation affects behavior. When individuals are prompted to consider norms before making a decision, prosocial behavior increases. When beliefs are elicited after decisions, however, they instead reflect post-hoc rationalizations that justify self-interested choices. These results highlight an internal channel through which the social environment—via internalized norms—shapes behavior.

Chapter 4 examines how individuals respond to social information about the behavior of others. In many environments characterized by moral ambiguity, individuals look to others’ actions to infer what is appropriate, consistent with theories of descriptive norms and social learning (Cialdini et al., 1990; Fischbacher and Gächter, 2010). Chapter 4 studies how exposure to such information shapes both informational and allocation choices. The results show that individuals adjust their behavior in response to what they believe others do, aligning their actions more closely with perceived norms. This responsiveness highlights a comparative social channel through which strategic ignorance can be sustained or mitigated.

The chapters in this dissertation present a unified account of information avoidance as a socially embedded phenomenon. Ignorance is not merely the absence of information, but a strategic choice that interacts with social context, cognitive processes, and normative expectations. Individuals use ignorance to manage responsibility, others can intervene to limit it, internal norms can be activated or suppressed, and social information shapes how behavior aligns with those norms.

This perspective has broader implications beyond the experimental settings studied here. Many important economic and social decisions—from environmental consumption to ethical investment and public policy—are characterized by uncertainty about their consequences (Momsen and Ohndorf, 2020). In such contexts, the ability to remain ignorant can shape not only individual behavior but also collective outcomes. Understanding when and why individuals choose not to

know, and how this choice can be influenced, is therefore central to designing environments that promote socially desirable behavior.

By integrating insights from experimental economics, moral psychology, and behavioral theory, this dissertation contributes to a growing literature on the role of information in moral decision-making. It shows that the question is not only what individuals know, but also how the social environment shapes what they choose not to know—and why.

Chapter 2

Unwillingly Informed: the Prosocial Impact of Third-Party Informers

2.1 Introduction

Many social decisions involve uncertainty about their consequences, and individuals often avoid information about the impact of their actions in order to justify self-interested behavior (Dana et al., 2007; Grossman and van der Weele, 2017).¹ As discussed in Chapter 1, such strategic ignorance can serve as a means of managing moral responsibility. In many real-world settings, however, individuals do not make these decisions in isolation: others may provide or impose the very information one prefers to avoid.

Everyday interactions frequently expose individuals to information they might otherwise choose to ignore. When purchasing goods, a partner may encourage checking whether a product is ethically sourced. A friend might point out the environmental impact of a meal choice, or draw attention to the labor conditions behind a clothing purchase. Even when individuals actively avoid information, others may introduce it into the decision environment. These interactions suggest that the ability to remain ignorant—and thereby to avoid moral responsibility—may be limited by the presence and actions of others.

While prior research shows that individuals avoid information when doing so allows them to act more selfishly (Vu et al., 2023), less is known about how such behavior changes when others can impose information. If ignorance serves to

¹This chapter is based on material published in the *Journal of Environmental Economics and Management* (Grossman et al., 2026a).

reduce moral obligation, how do individuals respond when that ignorance is disrupted by others? Do individuals continue to avoid information in the presence of potential informers, and do informers anticipate that their actions may be unwelcome—particularly if imposing information may trigger social sanctions?²

To study these questions, we conduct an experiment based on a modified version of the moral wiggle-room game (Dana et al., 2007), which provides a simple environment in which individuals can choose whether to acquire information about the social consequences of their actions. We extend this framework by introducing a third party—an “informer”—who can reveal information that the decision-maker may prefer to avoid. If either the decision-maker or the informer reveals the information, the final choice is made under full information; otherwise, the decision-maker acts while remaining ignorant.

We examine how the presence of a potential informer affects both information acquisition and allocation decisions, as well as how informers behave when they can impose information on others. In particular, we study whether individuals continue to avoid information when others can reveal it, whether informers choose to impose information even when it may be unwelcome, and how imposed information affects subsequent behavior.

Our design includes several treatments that vary the timing of information provision and the visibility of the informer. In the *Informer* treatment, decision-makers initially choose in isolation and only later learn that an informer may reveal information, allowing us to observe how externally imposed information affects subsequent choices. In the *Informer - Info First* treatment, the presence of the informer is known from the outset, enabling us to study whether the anticipation of information provision affects initial information choices. A corresponding control treatment, *Hidden Information - Info First*, isolates the effect of the informer’s presence. Finally, the *Informer - Bonus* treatment introduces a mechanism through which decision-makers can reward or punish informers, allowing us to explore whether the possibility of retaliation influences informer behavior.

Our results show that while many individuals actively avoid information, others are willing to impose it, and doing so substantially improves social outcomes. Informers frequently choose to reveal information, leading a substantial share of decision-makers to become unwillingly informed. Exposure to this information induces many individuals to revise their choices in a more prosocial direction,

²Previous evidence suggests that people self-select, such that altruistic types tend to pair up with information providers while selfish types avoid those who would impose information, implying strong preferences for information avoidance or information acquisition (Serra-Garcia and Szech, 2022; Soraperra et al., 2023; Mol et al., 2025).

increasing recipient payoffs. At the same time, the mere presence of a potential informer does not reduce information avoidance: individuals continue to avoid information even when they anticipate that it may be revealed by others.

These findings highlight a key role of the social environment in shaping strategic ignorance. While individuals may seek to avoid information in order to manage moral responsibility, others can intervene to limit this avoidance and thereby influence behavior. In this way, strategic ignorance is not solely an individual choice, but one that is shaped—and at times constrained—by the actions of others.

Our study contributes to the literature on strategic ignorance and prosocial behavior by showing how third parties affect both information acquisition and allocation decisions. While prior work has primarily focused on individual choices in isolation, we demonstrate that the presence of others can significantly alter the informational environment and its behavioral consequences. More broadly, the results underscore the importance of considering social interactions when studying information avoidance and its implications for economic and social outcomes.

2.2 Experimental Design

In this section, we describe the design of the experiment. Section 2.2.1 describes the basic structure and treatments of the experiment, Section 2.2.2 lays out the main hypotheses that we test, and Section 2.2.3 describes how we implemented the design. The experiment and analysis were preregistered in AEARCT-0011289. We preregistered a total of eight treatments, including two preparatory replication treatments, but only report four of them in the main body of this paper. The remaining four preregistered treatments are reported in A.6, including our justification for relegating them.

The basic experiment consists of two parts. Subjects first play the moral wiggle-room game (Part 1) and then answer an un-incentivized questionnaire (Part 2). In the moral wiggle-room game, a dictator unilaterally chooses between two options that determine her own payoff and that of another subject with whom she has been randomly and anonymously paired. While the dictator knows her own payoff for each option, the payoffs for the recipient are initially hidden. She knows that forgoing the higher payoff option could increase the recipient’s payoff, but doing so could also decrease it. However, the dictator can—but does not have to—click a button to costlessly reveal which is the case before choosing an allocation.

We used the same payoffs used by Dana et al. (2007) and featured in many

subsequent studies,³ featuring two equally likely payoff states, which are presented to subjects as tables, as shown in Table 2.1. In either state, choosing option A gets the dictator six experimental currency units (ECUs), while option B only yields 5.⁴ In the conflicting-interests state, choosing B brings the recipient’s payoff up from 1 to 5 ECU, while in the aligned-interest state it lowers their payoff from 5 to 1 ECU. Thus, the dictator initially is uncertain whether sacrificing their own payoff will help or harm the recipient, but she can resolve this uncertainty with the click of a button.

	Conflicting Interest		Aligned Interest	
Dictator chooses	Dictator receives	Recipient receives	Dictator receives	Recipient receives
A	6	1	6	5
B	5	5	5	1

Table 2.1: Payoff Table

Notes: The two payoff tables featured in the moral wiggle-room game (the conflicting and aligned states) matched those used by Dana et al. (2007) and were equally likely.

While our study is inspired by Lind et al. (2019), we follow the original design from Grossman and Hua (2026) but used other subjects as our recipients rather than an environmental charity. Before commencing our investigation, we ran a replication study, documented in A.2, which featured treatments corresponding to the *Baseline* and *Hidden Information* treatments from Dana et al. (2007). This study replicates the findings from Dana et al. (2007) in which substantial information avoidance reduced the recipient’s payoff, convincing us that our decision environment was suitable for answering questions about the informer’s impact.

2.2.1 Treatments

In this section, we describe our four main treatments. We ran two other preregistered treatments in which the informer is present, but unable to actually impose information. These were intended to help identify the mechanism by which the informer affects the dictator’s information choice, but given the absence of such an effect they yielded little additional insight. We have therefore relegated the

³For example, Grossman (2014); Grossman and van der Weele (2017)

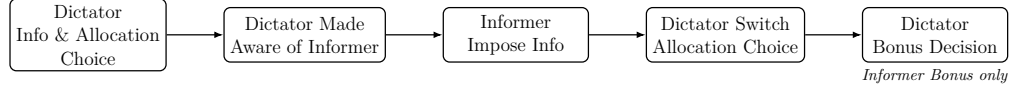
⁴The exchange rate was 1 ECU = 0.50 USD.

description of their design and the discussion of their results to A.6. A table summarizing the key features of each treatment is available in A.3, while instructions and screenshots are provided in A.7.

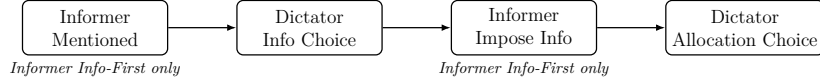
Informer The timeline of the *Informer* treatment is represented in Figure 2.1a. Dictators first play the standard moral-wiggle room game without any mention of a potential third-party, using a single decision screen for dictators' information and allocation choices. However, before the payoffs are realized, dictators are told about the existence and role of the informer. Without knowing the realized payoff state, the informer is asked the following question: "In the case that Person 1 did not reveal which table is being used for their pair, would you prefer to overturn their decision, making sure that they actually have this information when they make their choice between A and B?" Since the realized payoff state is unknown even to informers, their decisions of whether to provide information are independent of the content of the information. The dictator is then given a chance to revise her outcome choice. If she already self-informed or the informer chose to impose that information, she does so while seeing the recipient's true payoffs. If neither the dictator nor the informer chose to reveal the payoffs, the dictator remains ignorant while making this revised choice. Regardless of the information choices of the dictator and informer, the dictator is told about the informer's decision. We use this treatment to study, within subjects, the informer's impact on the chosen allocation.

Our decision to use a within-subjects design to examine the informer's impact on social outcomes warrants some justification. For the informer to have a nontrivial role, the dictator must be able to choose an outcome after being informed. This requires either revising an initial choice or splitting the decision screen in two, inserting the informer's information provision between the dictator's information and allocation choices. For the *Informer* treatment, we chose the former for three reasons. First, we were concerned that split decision screens would result in ignorance levels in the control treatment that were too low to allow us to meaningfully study the impact of the informer, as was the case in Lind et al. (2019). Second, since most studies using the moral wiggle-room game adhere to the original integrated interface, splitting the decision screen would make our findings less comparable to those of other studies. Finally, while the within-subjects design may bias dictators towards making the same initial and final choice, it has the advantage of allowing us to know the dictator's information preference in the absence of the informer and the within-subject impact of the informer on the chosen outcome. Furthermore, consistency bias would work against our hypotheses, rendering any results we do find even more interesting. Thus, we proceed by

comparing the dictator's interim outcome choice, made before she was aware of the informer, with her final choice.



(a) *Informer* and *Informer - Bonus*



(b) *Hidden Information - Info First* and *Informer - Info First*

Figure 2.1: Experiment Flow

Notes: Simplified experimental flows for the four main treatments.

Informer - Info First The *Informer* treatment cannot shed light on the impact of the informer's presence on dictators' own information choices, since dictators are unaware of the informer when choosing whether to seek information. For this reason we included the *Informer - Info First* treatment, in which the dictator is aware of the informer from the very start of the game. The timeline of this treatment is represented in Figure 2.1b.

As discussed above, disclosing the informer's presence from the start required us to divorce the information choice from the outcome choice and present them on separate screens, with the informer's choice implemented between them. We did this with the understanding that the separate-screen interface might substantially inhibit information avoidance, which is partly why we did not rely upon it to address our other research questions.

When making her information choice in this treatment, the dictator knows that the informer might impose the payoff information anyway. After the dictator's information choice, the informer chooses whether or not to do so, contingent on the dictator's avoidance. Finally, the dictator makes her allocation choice, either with or without knowledge of the recipient's payoffs, as determined by both her and the informer's decisions. We call this the *Informer - Info First* treatment both because the dictator learned about the informer before making her choice and because her information choice came first, separated from the outcome choice by the informer's turn.

Hidden Information - Info First The *Hidden Information - Info First* treatment serves as our baseline control condition for *Informer - Info First*. Its timeline is reflected in Figure 2.1b. This treatment replicates the classic hidden-information treatment in that there is no informer. However, the information and allocation decisions are presented sequentially on separate screens just like in the *Informer - Info First* treatment.⁵ This allows us to isolate the effect of the informer on information acquisition while holding constant other aspects of the interface.

Informer - Bonus We designed the *Informer - Bonus* treatment to study the dictator’s reciprocal response to being informed. It is similar to the *Informer* treatment, except that the dictator makes one additional choice following the final allocation choice. Similarly, the informer decides whether or not to impose information knowing that the dictator makes an additional bonus allocation choice in response to their decision. The timeline of this treatment is reflected in Figure 2.1a.

After having decided on revising her choice in response to the informer’s decision, the dictator is made aware of a subsequent decision: choosing whether or not to increase the informer’s payoff by 1 ECU. She does this knowing what the informer’s choice was, and the cost of providing the bonus is paid by the experimenters. Informers are aware of this when making their choice. Choosing to give this bonus could be regarded as a reward from the dictator or simply generosity, while withholding it could be regarded as a punishment. We examine this choice as a function of the dictator’s own information choice, the informer’s choice, and the content of the information imposed by the informer. We also compare the behavior of the informer in this treatment to their behavior in the *Informer* treatment to see if and how they react to the possibility of a reciprocal response from the dictator.

Two aspects of the design of this treatment warrant further discussion. First, the dictator’s experience in the *Informer - Bonus* treatment deviates from that in the *Informer* treatment only after the dictator has made her information and preliminary allocation choice. Hence we pool the data from these two single-screen treatments when analyzing the dictator’s preliminary allocation choices (before any switching). Second, we considered giving dictators the option to reduce informers’ default payoff instead of increasing it. Although this would have been formally equivalent to a bonus option combined with a correspondingly lower default payoff for informers, we opted

⁵The ‘Information First’ label for this treatment refers only to the fact that the dictator makes the information choice first, since there is no informer to tell her about.

against it, speculating that it might be perceived as psychologically harsher, making dictators hesitant to use the option. We doubt that this would have substantially changed our results, since (as discussed further below) most informers apparently did not expect their information to be unwelcome. Nevertheless, two competing norms for informer behavior might possibly be at play here—providing information and minding one’s own business (Fergusson et al., 2024)—and we cannot exclude the possibility that framing the sanctioning option as a bonus rather than a loss has subtly affected the salience of each of these two norms.

2.2.2 Hypotheses

In this section we present our preregistered hypotheses.⁶ Our first hypotheses are about how the informer’s presence changes the dictator’s behavior.⁷

Hypothesis 1 *Dictators behave more prosocially when there is a potential informer.*

We expected the overall outcomes to be more prosocial, leading to higher payoffs for the recipient when the informer is present. Part of why this might happen is simply because dictators end up with more information when it can be imposed by a third party because some dictators who avoid information will end up informed regardless. Lacking ignorance as an excuse to act selfishly, more of these will choose prosocially. We evaluated this hypothesis by comparing dictators’ interim outcome choices in the *Informer* and *Informer - Bonus* treatments, made before knowing of the informer’s existence, with their final choices. In particular, we expected unwillingly informed dictators to revise their decisions to behave more prosocially.

Another reason why dictators may behave more prosocially is because they are less likely to even try to avoid information when an informer is present. This could be the case if the introduction of the informer reinforces social obligations, makes the dictator feel more judged or scrutinized, or merely draw her own attention to the

⁶The order in which the hypotheses are presented here differs from that in which they appear in the preregistration document. However, their content remains the same. We adjusted the order merely to better align them with the presentation of our treatments and results.

⁷As mentioned above, we also conducted two preregistered *Hypothetical Informer* treatments, reported in A.6. These largely parallel the *Informer* and *Informer - Info First* treatments, but differ in that it is commonly known that the informer’s choice will not actually be implemented, and thus, the informer cannot actually influence the dictator’s information. The preregistered hypotheses relating to these treatments are derived from the question of whether the mere mention and description of a of a third-party potential informer might impact the dictator’s behavior, regardless of whether he can actually impose information upon her.

information choice, or if dictators dislike the idea of others imposing information on them. To investigate how the informer affects information acquisition, we compared the ignorance rates between the *Hidden Information - Info First* and *Informer - Info First*. Our second hypothesis posits that the possibility of having information imposed on the dictator makes her less likely to avoid it in the first place.

Hypothesis 2 *Dictators are less likely to avoid information when a potential informer is present.*

Turning our attention to the dictator’s response to the informer’s behavior, we have a third hypothesis. Based on the premise that some dictators choose without revealing, and thus presumably prefer ignorance, we expected that they would not like it when the informer imposes information. We tested the following hypothesis by comparing reward rates across the specified subgroups of dictators in the *Informer - Bonus* treatment.

Hypothesis 3 *Unwillingly informed dictators are less likely to reward the recipient than those who self-inform or are allowed to remain ignorant.*

Finally, informers may demure from imposing information either to respect the wishes of information avoiders, or to forestall a negative response. Our fourth hypothesis supposes the latter.

Hypothesis 4 *Informers are less likely to impose information in the Informer - Bonus treatment than they are in the Informer treatment.*

2.2.3 Implementation

We ran the experiments online from March to June 2024 recruiting English-speaking subjects in the United States above the age of 18 on the recruitment/payment platform Prolific. We configured the recruitment process to retrieve an even 50-50 split of male and female subjects and screened out subjects to ensure participation in only one session of the experiment. The recruitment website redirected subjects to the experiment website, which was created using the LIONESS web platform (Giamattei et al., 2020).

Subjects were told that the experiment consisted of two parts. The terminology of Part 1 and Part 2 was used solely to describe the sequence of decisions. Any subject who completed Part 1 were compensated accordingly based on their decisions in that part of the experiment. Part 1 consisted of the experimental design described above, while Part 2 was a questionnaire. A total of 577 Dictators and 394 Informers ($N = 971$) participated across our four treatments and fully completed Part 1.⁸ We paid

⁸Not counting any subjects who failed the comprehension checks or did not complete Part 1, a total of 1012 Dictators and 394 Informers ($N = 1406$) were recruited across all treatments, including

all subjects that completed Part 1 a fee of \$3, which corresponded to the minimum amount that any subject could earn. Subjects earning more than this were paid the difference via the bonus payment function on Prolific. We built a hidden 14.5-minute timer into the program, which was sufficient to allow virtually all subjects to complete Part 1 and for most subjects to complete Part 2. Subjects who completed both parts of the experiment early were shown the timer and were required to remain in the program until the timer had fully elapsed.⁹

Before Part 1, subjects answered a series of comprehension and attention checks. Any subjects who failed the checks were dropped out of the study. In the *Informer* and *Informer - Bonus* treatments, subjects were matched in two person lobbies, with one participant being assigned the role of the dictator and the other the informer. No subjects were assigned the role of recipient in this part. However, prior to starting Part 2, as an added bonus payment, subjects acting in the role of dictators were told they would be assigned as the recipient to the dictator choice of another subject with whom they had been anonymously matched. In this manner, each dictator from Part 1 played both roles, being matched with a recipient in Part 1 who was one of the other dictators. Importantly, this could not affect dictator behavior because they only learned about this extra payment after already having made their own decisions. Furthermore, Part 2 payments were independent of subjects' own behavior in Part 1 and there was no deception regarding earnings. Informers were paid a flat fee of 5 ECUs for both parts of the experiment.

The Part 2 questionnaire asked subjects to explain their decisions and describe their experience, along with a battery of questions on subjects' political party affiliation, concerns on the COVID-19 pandemic, concerns on climate change, an evaluation of the social appropriateness of various decisions that could be made by the dictator, and a self-assessment of their concern for self-image. Although Part 2 was unincentivized, subjects were required to remain in the study until the aforementioned timer had fully elapsed. In addition, we retrieved general, self-reported demographics information from the Prolific database. Finally, in the

those reported in the appendix. Observation totals were guided by the power analysis presented in our registered pre-analysis plan.

⁹After initial subjects complete the experiment, Prolific updates the advertised hourly compensation rate to reflect the average hourly rate as determined by the actual time it took the initial subjects to complete the experiment. If the actual duration deviates from the initial estimated duration by a large amount, this shows up as a markedly different advertised pay rate. This could affect the composition of subjects that self-select as participants because different treatments may take different amounts of time to complete, introducing a potential selection bias across treatments. We built the timer into the program to mitigate these risks and achieved the desired effect of stabilizing the advertised compensation rate between and within treatments.

Informer - Info First treatment, which was included to study dictators' information choice rather than allocation choices or informer behavior, the informer decisions were simulated using the distribution of informer decisions from the *Informer* treatment.¹⁰

2.3 Results

In Section 2.3.1, we present our analysis of dictators allocation choices and in Section 2.3.2 we examine their information choices. Section 2.3.3 presents our analysis of the dictator's decision whether or not to reward the informer, and our analysis of the informer's willingness to impose information is in Section 2.3.4. Appendix A.1 presents the detailed demographic statistics.

2.3.1 Dictator Allocation Behavior in the Moral Wiggle-Room Game

First, we look at the dictator's allocation choice, focusing on within-subject comparisons.

The *Informer* and *Informer - Bonus* treatments both feature an interim as well as a revised allocation choice. The only difference between these treatments was the bonus decision, which was not announced until after the interim allocation choice. Hence, interim choice data from these treatments can clearly be pooled. The first column of Table 2.2 shows the pooled data, while the second and third columns show the data from the *Informer* and *Informer - Bonus* treatments, respectively. The first row shows average interim payoff — i.e., what the recipient would have received in the absence of the informer — confirming that it is close to identical in the two treatments.

To study the effects of the informer, we next consider how dictators change their allocation decision between their interim and revised choice. The second row of Table 2.2 shows the average final payoff to the recipient, with the third row showing the difference between these two. As can be seen from the table, final payoffs are also very similar between the *Informer* and *Informer - Bonus* treatments. We thus proceed by pooling data on interim as well as final recipient payoffs from these two treatments.¹¹

¹⁰Effectively, the dictators in these treatments were randomly matched with one of the informers from the *Informer* treatment as a cost cutting measure.

¹¹Although the possibility of the bonus could affect the informer's behavior and, thus, the dictator's final information, we show below that informer behavior was close to identical across

Table 2.2: Average Recipient Payoffs

	Pooled	Bonus		Unwillingly informed	
		N	Y	N	Y
Interim	3.74 (1.86)	3.73 (1.87)	3.74 (1.86)	4.28 (1.54)	3.07 (2.00)
Final	4.30 (1.52)	4.37 (1.47)	4.24 (1.58)	4.28 (1.54)	4.33 (1.50)
Difference	0.57 [0.000]	0.63 [0.000]	0.49 [0.000]	0.00 [1.000]	1.25 [0.000]
<i>n</i>	367	189	178	201	166

Notes: Average recipient payoffs in the Informer and Informer - Bonus treatments. Standard errors shown in parenthesis below average payoffs. Square brackets show p-value for a t-test for difference in means between interim and final payoffs. The notation 0.000 means that the p-value is less than 0.001.

The average interim recipient payoff for the pooled treatments, shown in the first column, is 3.74, meaning that 68% of the recipients received a payoff of 5 instead of 1. After the informer is introduced, the recipient's average payoff climbs to 4.30, an increase of 15% and also corresponding to a 15 percentage point increase in the share of recipients getting the higher payoffs. The 0.57 difference between interim and final payoffs is highly significant ($t = 7.45$, $p < 0.001$).

To better understand the informer's impact on the recipient, we next zero in on when and how the dictator revised her initial choice. Of the 367 dictators in the *Informer* and *Informer - Bonus* treatments, only 56 of them (15%) revised their choice. This switching can be almost entirely attributed to dictators who had information imposed on them by the informer. We label these 166 dictators the unwillingly-informed in the fifth column. Virtually none of the 201 dictators, presented in the fourth column, switched: of the 127 who self-informed and the 74 who were allowed to remain ignorant, exactly one from each category changed their initial choice. We conclude from this that informers have a substantial impact on recipient outcomes through imposing information on those who would otherwise

the two treatments. Moreover, the Final payoff does not differ significantly between treatments for the two information statuses ($p = .67$, $p = .45$)

avoid it, many of whom subsequently choose more prosocial outcomes.

Table 2.3: Unwillingly-informed Switchers

Initial choice	Payoff state	
	Conflicting	Aligned
A	66% (42/64)	0% (0/67)
B	5% (1/19)	69% (11/16)

Notes: Percent of unwillingly-informed dictators who switched their initial allocation choice, by initial choice and payoff state. Number of subjects with each combination of initial choice and payoff state in parenthesis. A Pearson test for identical proportions in each cell yields $\chi^2(3) = 80.24$ ($p < 0.001$).

We dig into the switching behavior of the unwillingly-informed dictators in Table 2.3. It shows that of those who switched, their revised choice almost always benefited the recipient, even at their own cost. Of the 131 initially choosing the higher payoff for themselves (option A), only those who learned that their sacrifice would help the recipient ended up switching, with a full 66% of these 64 dictators doing so. Thus, the majority of selfish-acting information-avoiders were prompted to behave more prosocially by simply providing them the information they were avoiding.

Unsurprisingly, we also see this prosocial shift when it carries no personal cost. Consistent with the use of ignorance as an excuse to behave selfishly, only a minority of information-avoidant dictators choose the lower paying option, B. Of the 35 who did so and had the information imposed by the informer, we see a dramatic shift to the benefit of the recipient. Nineteen of these subjects learned that their sacrifice was helping the recipient and only one of these changed their mind about doing so. In contrast, 69% of the 16 who learned that increasing their own payoff would also help the recipient went ahead and made this change.

The shift in outcome choices due to the informer’s intervention has a major impact on the recipient’s payoffs. While the informer’s impact on the recipient’s payoffs is apparent in the paragraphs above, it stands out even more clearly among the recipients paired with unwillingly-informed dictators. Average payoffs for this group are shown in the fourth column of Table 2.2. Among this group, the introduction of the informer increases the average recipient payoffs from 3.07 to 4.33 ($t = 8.44$, $p < 0.001$), an increase of 41%. This corresponds to a jump in the number of recipients

getting the higher payoff from 52% to 83%. In contrast, there is no noticeable change in the recipient payoffs among the rest of the dictators, who either self-informed or were allowed to remain ignorant, as shown in the fifth column of the table.

Table 2.4: Information Avoidance and Imposition

	Total	Treatments			
		Inf.	Inf. Bonus	Hid. Info - IF	Inf. - IF
D avoids info	0.53 (0.50)	0.61 (0.49)	0.70 (0.46)	0.30 (0.46)	0.32 (0.47)
I imposes info	0.71 (0.45)	0.70 (0.46)	0.69 (0.46)		0.74 (0.44)
n (dictators)	578	189	178	105	105
n (informers)	394	202	192	0	0

Notes: Share of dictators avoiding information and of informers imposing it, by treatment. Standard errors in parentheses. The Hidden Info - IF treatment does not have any informers and Informer - IF uses simulated decisions based on informer decisions in Informer.

Finally, although not all of the informers chose to impose information, we consider the counterfactual scenario in which they did, meaning that *all* dictators ended up with full information, whether self-chosen or externally imposed. We extrapolate the switching behavior of the unwillingly-informed dictators, summarized in Table 2.3, to the 74 dictators who were allowed to remain ignorant. In that counterfactual, average recipient payoff is 4.51, corresponding to 88% receiving the higher payoff.¹² This constitutes an increase of 21% over the interim payoffs that arise when the dictator can remain ignorant at will.

The above analyses illustrate the impact that socially-provided information can have on dictator behavior and how much better the recipient may fare when another

¹²The counterfactual revised choices of the 240 info-avoiders are as follows. In the conflicting-interests payoff state, $(\frac{42}{64}) \cdot 92$ A -choosers switch to B and $(\frac{18}{19}) \cdot 22$ B -choosers stick with B , so 81.2 out of 114 (71%) counterfactually choose B . In the aligned-interests payoff state, $(\frac{67}{67}) \cdot 108$ of the A -choosers stick with A and $(\frac{11}{16}) \cdot 18$ of B -choosers switch to A , so 120.4 out of 126 (96%) counterfactually choose A . This leads to a counterfactual total of 201.6/240 (84%) of information-avoiders choosing the higher recipient payoff and $120 + 201.6 = 321.6$ out of 367 (88%) of dictators overall choosing the higher payoff. This corresponds to an average recipient payoff of 4.51.

person is present to limit the dictator’s information avoidance. Our findings offer very strong support for Hypothesis 1 and are summarized in the following result.

Result 1: Informers’ interventions make dictators behave more prosocially. This occurs via the channel of imposing information on avoidant dictators, who then act more prosocially.

2.3.2 Dictator Information Choice in the Moral Wiggle-Room Game

Next, we turn attention to the dictator’s information choice. While the informer clearly affected the dictator’s outcome choice through the provision of information, introducing the informer could have also helped the recipient by reducing the amount of information avoidance in the first place since self-informed dictators tend to choose more prosocially. Indeed, with ignorance rates in the *Informer* and *Informer - Bonus* treatments averaging 65%, there is room for a large effect. However, recall that to identify the impact of the informer without confounds, we need to compare across the two information-first treatments.

Panel (a) of Figure 2.2 shows the ignorance rates for these two treatments and reveals that the informer has little impact in this respect. The ignorance rate in the *Hidden Information - Info First* treatment is 0.30 and is almost identical to the 0.32 found in the *Informer - Info First* treatment ($Z = 0.34$, $p < 0.73$). The lack of support in our data for Hypothesis 2 is summarized in the following result and leads us to conclude that the informer’s impact is limited to the more direct channel of changing the dictator’s information.

Result 2: The known presence of a potential informer does not impact the dictator’s information acquisition choice.

However, we must note that these ignorance rates are drastically lower than those from the other treatments. We illustrate this by pooling the two *Info First* treatments together and displaying the resulting ignorance rate in panel (b) of the figure, next to the ignorance rate from the pooled *Informer* and *Informer - Bonus* treatments, which we collectively call the non-*Info-First* treatments. Changing the architecture by separating the dictator’s information and outcome choices in two separate screens lowered the average ignorance rate from 65% to 31% ($Z = 7.87$, $p < 0.001$).

2.3.3 Dictator’s decision to reward the informer

Hypothesis 3 posits that unwillingly-informed dictators are less likely to reward informers than those who self-inform or are allowed to remain ignorant. We explore

this hypothesis by examining that rate at which dictators *withheld* the bonus in the *Informer - Bonus* treatment. Figure 2.3 shows average bonus-withholding rate broken down by whether or not the informer changed the dictator’s information and by the content of that information. Panel (a) compares those who had information thrust upon them, whom we call the unwillingly informed, to the rest of the dictators, all of whom either self-informed or remained ignorant. Neither group was very prone to withhold the bonus, but the unwillingly informed were slightly more so. Ten percent of the 84 dictators in that group did so, compared to only 4% of the 94 remaining dictators. While the withholding rate among the unwillingly informed is more than double that among the others, both rates are quite low and the sample relatively small, so the difference in proportions is not statistically significant ($Z = 1.40$, $p < 0.16$; Fisher’s exact $p = 0.23$).

When we consider the *content* of the imposed information, however, the contrast is more stark. Forty-three of the unwillingly informed dictators learned that helping the recipient would lower their own payoff (conflicting-interests payoff state). We describe these as unwillingly-informed of bad news, and 14% of them withheld with the bonus. Contrasted with the 4% withholding rate among the remaining 135 dictators, we find that having bad news thrust upon them by the informer made the dictator between three and four times more likely to withhold the reward. This difference is significant in a statistical sense (at the 3% level for two-sided test of proportions and 4% level for Fisher’s exact test). Thus, our next result offers qualified support for Hypothesis 3.

Result 3: Dictators respond negatively to being unwillingly informed of bad news.

2.3.4 Informer Behavior

Were informers less likely to impose information when the dictator could respond? As shown in Table 2.4, they had a high propensity to supply information in general, with 69% and 71% doing so in the *Informer* and *Informer Bonus* treatments, respectively. A Fisher exact test fails to reject the null hypothesis that informers behave the same in both treatments, so we find no support for Hypothesis 4.

To understand the thinking and motives of informers, we asked them about their beliefs in the questionnaire. Interestingly, informers did not seem to foresee the substantial prevalence of willful ignorance; the vast majority of them (approximately 85%) thought that dictators had revealed the information themselves. In line with this, it turned out that only a small subset of informers in

the *Informer Bonus* treatment, about 8%, reported concerns about the bonus as a reason for their choice. Furthermore, of those who reported concerns about the bonus, a sizable minority believed that the bonus would have been withheld if they did *not* impose information. Given that roles were randomly assigned, these patterns suggest that informers generally do not expect dictators to remain ignorant. These miscalibrated beliefs, understandable because of the sensitivity of behavior to context, may drive the informer’s propensity to impose information despite the prevalence of willful ignorance. More detailed analysis of informer’s beliefs and the reasons they provided for their choices is provided in A.5.

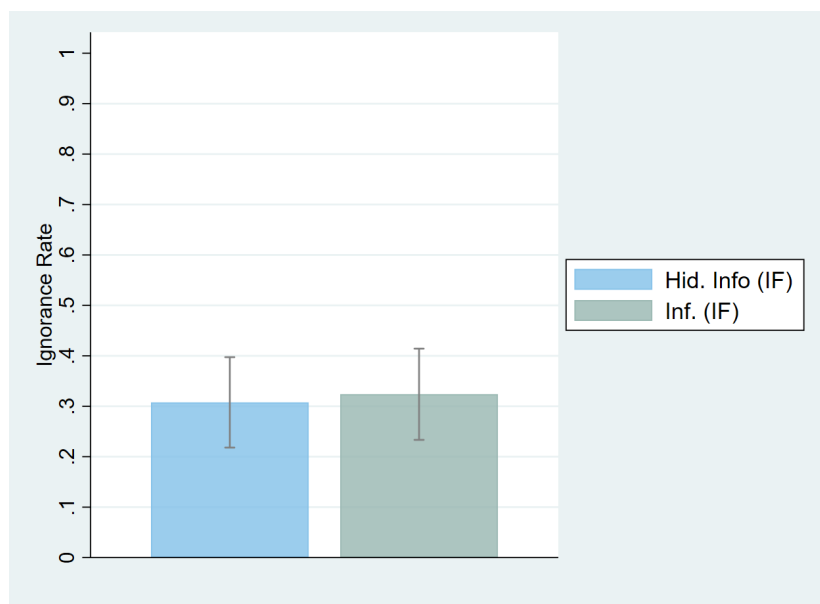
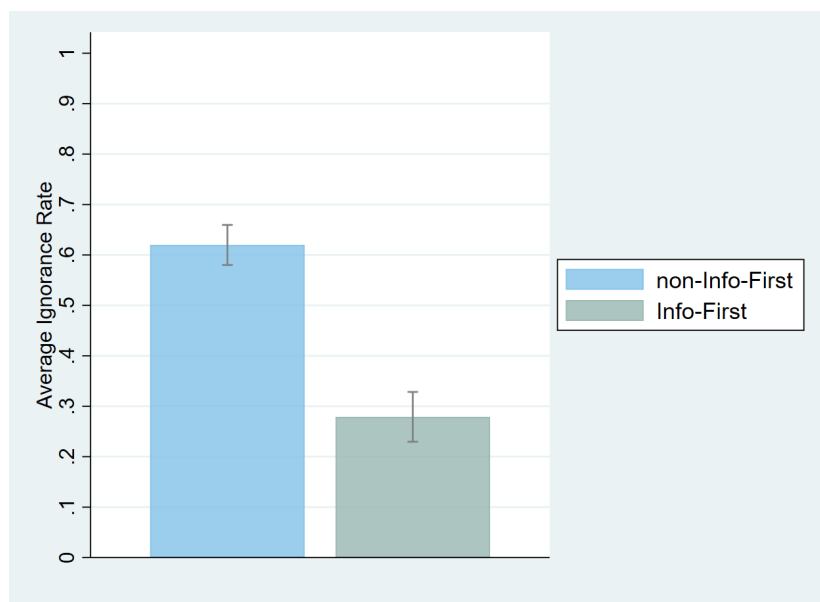
Result 4: Informers appear to be insensitive to the threat of punishment.

2.4 Conclusion

This chapter examines how the presence of others shapes strategic ignorance by introducing a third party who can impose information. While many individuals actively avoid information about the consequences of their actions (Dana et al., 2007; Grossman and van der Weele, 2017), we find that others are frequently willing to provide that information, even when it may be unwelcome. When information is imposed, a substantial share of individuals revise their choices in a more prosocial direction, leading to meaningful improvements in social outcomes.

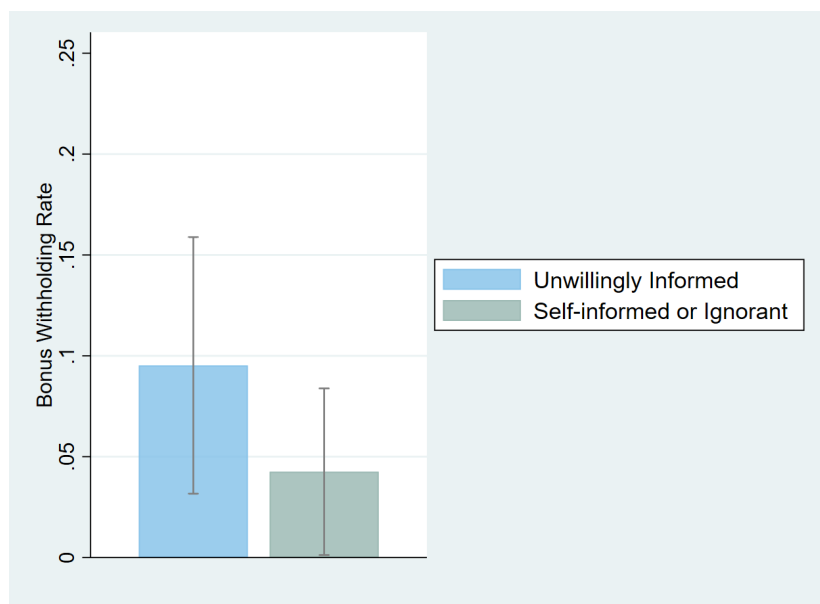
These findings demonstrate that strategic ignorance is not solely an individual choice. Although individuals may seek to avoid information in order to manage moral responsibility, the social environment can constrain this behavior. The presence of others introduces a channel through which information can be provided endogenously, limiting the effectiveness of ignorance as a justification for self-interested action. The results highlight an external channel through which the social environment shapes behavior: others can intervene to reduce information avoidance and promote more prosocial outcomes.

At the same time, we find that individuals continue to avoid information even when they anticipate that it may be revealed by others. This suggests that the preference for ignorance remains robust, but that its consequences depend on the surrounding social context. Yet, despite many studies on the moral wiggle room game, under what circumstances ignorance persists remains an open question (Grossman and Hua, 2026).

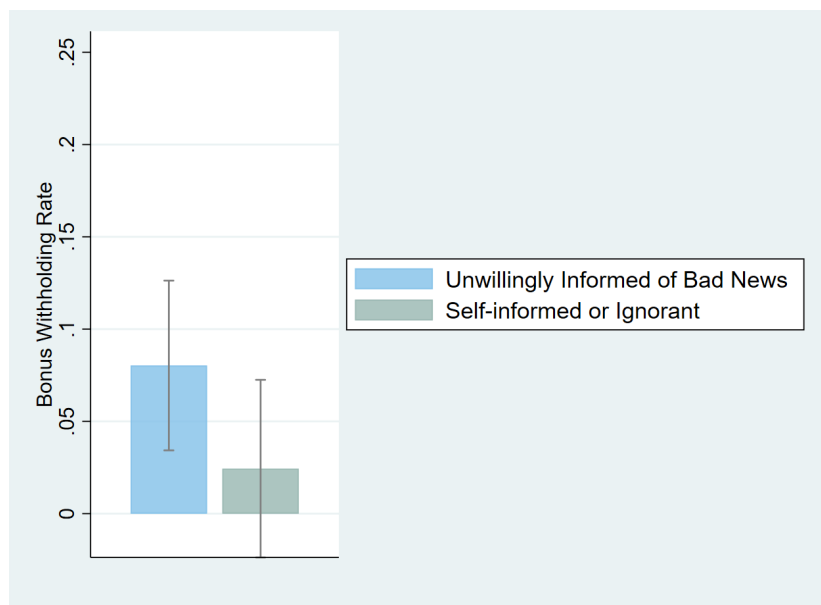
Figure 2.2: *Info-First* Comparison(a) Ignorance rates for the Hidden Information (*Info First*) and Informer (*Info First*) treatments.(b) Average ignorance rates in the non-*Info-First* and *Info-First* treatments.

Notes: Comparison of ignorance rates across the two *Info First* treatments and between those treatments and the non-*Info-First* treatments. Bars indicate 95% confidence intervals.

Figure 2.3: Bonus Withholding Rates



(a) Self-informed and ignorant versus unwillingly informed.



(b) Self-informed, ignorant, and unwillingly informed of good news versus unwillingly informed of bad news.

Notes: Bonus withholding rates, dictators in the Informer - Bonus treatment. Bars indicate 95% confidence intervals.

Chapter 3

Latent Norms: Prosocial Nudges Under Moral Ambiguity

3.1 Introduction

Individuals often face decisions in which the consequences of their actions for others are uncertain.¹ As discussed in Chapter 2, individuals may exploit this uncertainty by avoiding information, thereby preserving flexibility in how their actions are interpreted and reducing perceived responsibility (Dana et al., 2007; Grossman and van der Weele, 2017). While Chapter 2 shows that the ability to remain ignorant can be constrained by others, an important question remains: why are people so willing to impose information on others when they themselves may not choose to acquire information? The imposition of information on others serves as a mechanism for the enforcement of social norms (Akerlof, 1980; Bicchieri, 2006; Krupka and Weber, 2013). Yet, as evident from Chapter 2, norms do not appear to be universally applied. This pattern reveals an important asymmetry. Individuals may recognize the normative value of information when deciding whether others should know, yet fail to apply the same standard when deciding whether to know themselves.

One possibility is that norms are not absent, but latent: they may guide behavior only when features of the decision environment make them salient. Under this interpretation, individuals may possess internal standards of appropriate behavior that do not necessarily guide action unless activated or brought to mind.²

¹A version of this chapter is circulated as a working paper (Hua, 2025b).

²For instance, Bhattacharya and Dugar (2022) find that reminding vendors in fish markets about

This view is consistent with theories of norm salience and goal activation, which suggest that contextual cues can make particular standards more behaviorally relevant (Cialdini et al., 1990; Bargh et al., 2001; Bicchieri, 2006). It is also consistent with evidence that making moral standards salient can reduce dishonest or antinormative behavior (Thaler and Sunstein, 2008; Mazar et al., 2008). This suggests that behavior may depend not only on the availability of external information, but also on the salience of internal beliefs about what constitutes appropriate conduct.

This mechanism is related to, but distinct from, the literature on intuition and deliberation in economic games. Meta-analytic evidence suggests that quick judgment, intuition may increase cooperation in social dilemmas, though subsequent work emphasizes heterogeneity and smaller effects (Rand, 2016; Kvarven et al., 2020). In dictator-game settings, however, the evidence is less consistent with a simple intuition-promotes-altruism account (Fromell et al., 2020). The present chapter examines how the timing of social reflection matters: whether prompting beliefs before choice activates latent norms, and whether eliciting beliefs after choice instead facilitates rationalization.

In this chapter, I examine how eliciting beliefs about what others do affects behavior in environments characterized by moral wiggle room. Building on work showing that social customs and norms can shape behavior (Akerlof, 1980; Bicchieri, 2006; Krupka and Weber, 2013), and that behavior in economic games may reflect preferences for following norms rather than strict preferences over allocations (Kimbrough and Vostroknutov, 2016), I study whether prompting individuals to state their beliefs about what others do before making a decision increases the salience of prosocial norms, thereby constraining self-interested action. This contrasts with standard social-preference accounts, which model prosocial behavior as arising from concerns over payoff distributions, reciprocity, or efficiency (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002). Conversely, when beliefs are elicited after decisions, they may instead reflect post-hoc rationalizations that justify behavior (Kunda, 1990; Bénabou and Tirole, 2011).

To investigate this, I conduct a series of experiments in which the timing of belief elicitation is varied relative to the decision. When beliefs are elicited prior to choice, individuals are required to consider what they perceive others do before making their own decision. When beliefs are elicited after the decision, they cannot influence the choice directly and instead capture individuals' interpretations of their own behavior.

The results show that the timing of belief elicitation has a substantial impact on superstition can promote honesty.

behavior. Eliciting beliefs before the decision increases prosocial behavior, consistent with the activation of latent norms. In contrast, eliciting beliefs after the decision leads to responses that align with individuals' chosen actions, suggesting that beliefs may serve as self-serving justification. These findings indicate that norms are often latent and can be activated or suppressed depending on how the decision environment is structured.

This chapter highlights an internal channel through which the social environment shapes behavior. Even in the absence of external information or intervention, prompting individuals to reflect on norms can influence their choices. In this sense, the social environment operates not only through others' actions, but also through the activation of internalized standards of behavior.

In effect, this chapter asks if our prescription of whether or not others should seek information can be self-prescribed. Meanwhile, the next chapter turns to a complementary mechanism, examining how individuals respond to social information about the behavior of others once such norms become active.

3.2 Hypotheses

I present a series of hypotheses that support a theory of latent norms. First, in accordance with theories of cognitive dissonance (Festinger, 1957; Rabin, 1994; Konow, 2000; Matthey and Regner, 2011), I posit that people strive for consistency between what they believe and how they behave, and that second, such consistency is usually pursued *after* the fact rather than anticipated in advance—precisely what we would expect if prosocial norms remain latent and are only activated when brought to mind. Second, in line with a literature on self-awareness (Diener and Wallbom, 1976), prompting beliefs about others' behaviors before acting forces these expressed beliefs to become salient. By doing so, these beliefs nudge behavior towards alignment with the prosocial norms they have already acknowledged. Third, when they act before beliefs are elicited, their actions become salient, and they subsequently rationalize their beliefs to fit their behavior (Festinger, 1957). These dynamics motivate the hypotheses below:

Hypothesis 1: Norm activation through belief elicitation.

When beliefs are elicited before a choice, they activate latent social norms and constrain selfishness. By making norms salient at the moment of decision, belief elicitation increases adherence to prosocial standards.

Hypothesis 2: Post-hoc rationalization.

When beliefs are elicited after a choice, they adjust to justify behavior. In this case, individuals rationalize selfishness by assuming that others behave selfishly as well.

Together, these hypotheses capture the asymmetry in how beliefs and actions interact: beliefs constrain behavior when formed beforehand, but serve as self-serving rationalizations when formed afterwards.

3.3 Experimental Design

The study investigates three paradigms in which individuals can justify or resist self-interested behavior: (i) avoiding information about consequences for others, (ii) misreporting unverifiable outcomes, and (iii) making distributional choices under clear fairness norms. To study information avoidance, I implement two versions of the moral wiggle room game (MWRG) (Dana et al., 2007). To study dishonest reporting, I use the dice rolling game (DRG) (Fischbacher and Föllmi-Heusi, 2013). Finally, to establish boundary conditions where fairness norms are already unambiguous (Andreoni and Bernheim, 2009), I include a binary choice dictator game (BCDG). Across these paradigms, treatments vary in the timing of belief elicitation, allowing me to examine whether eliciting beliefs before action primes and activates norms, thereby constraining selfishness, or whether eliciting them after action instead facilitates post hoc rationalization.

The study was pre-registered in AEARCT-0016348.³ The experimental interfaces were programmed using LIONESS Lab (Giamattei et al., 2020). The sample was restricted to English-speaking subjects in the United States from the age of 19 to 80 with at least 10 completed studies on their Prolific profile and an approval rating of 99–100%. Recruitment was set to retrieve an even split of male and female subjects. To mitigate risks of sample imbalance from varying subject availability, randomization was conducted within each experiment: for each experiment, all treatment arms were fielded concurrently and participants entering that experiment were randomly assigned. Detailed demographics are available in Appendix B.1.

3.3.1 Moral Wiggle Room Games

The moral wiggle room game (MWRG) builds on Dana et al. (2007), in which dictators choose between two allocations of experimental currency units (ECUs)

³Link to pre-registration: <https://doi.org/10.1257/rct.16348-2.6>

for themselves and another participant, with the recipient’s payoff hidden unless the dictator actively chooses to reveal it. This setup creates scope for strategic ignorance, allowing dictators to avoid information that would make selfishness more transparent.⁴

Timing of belief elicitation. This version of the MWRG isolates the causal role of belief elicitation timing. Dictators face a payoff structure in which *Option A* yields 65 ECUs for the dictator and either 0 or 50 for the recipient, while *Option B* yields 50 ECUs for the dictator and the complementary payoff for the recipient.⁵ Subjects were asked:

“In another study, participants were asked to participate as Person 1. What percentage of people acting as Person 1 (the dictator) do you believe revealed the payoff tables?”

The key treatment variation concerns when beliefs about others’ information-seeking behavior are elicited. In the *Belief First* condition, subjects report their beliefs prior to making their decision, whereas in the *Belief After* condition, beliefs are elicited only after the decision has been made.⁶

This design tests whether eliciting beliefs prior to choice makes perceived social norms cognitively salient, thereby activating internalized standards that constrain selfish behavior (Hypothesis 1). In contrast, when beliefs are elicited after the decision, they cannot influence the choice directly and instead reflect post hoc rationalizations that align with behavior (Hypothesis 2).

Full Info and Self/Self Conditions. To distinguish norm activation from alternative explanations such as confusion, complexity aversion, or baseline reluctance to acquire information (Oprea, 2020), I include two benchmark conditions: *Full Info* and *Self/Self*. The *Full Info* condition removes the opportunity for strategic ignorance by revealing the payoff table before the

⁴In the MWRG, the DRG, and the BCDG, payments were converted at a rate of 40 ECUs = 1 USD.

⁵Compared to typical moral wiggle room games, payoffs were adjusted to 65 vs. 50 for the dictator to heighten selfish temptation, and 0 vs. 50 for the recipient to sharpen the trade-off. As shown in Appendix B.2, which reports a pretrial using the canonical payoff ratios from Dana et al. (2007), dictators behaved in a way consistent with the results reported here.

⁶Two additional, pre-registered benchmark treatments, *Full Info* (recipient payoff always revealed) and *Self/Self* (dictator’s choice only affected own payoffs), were simultaneously conducted with the *Belief First* and *Belief After* treatments.

allocation decision. Because the consequences for the recipient are known, this condition provides a benchmark for prosocial behavior when ignorance cannot serve as moral cover.

The *Self/Self* condition preserves the information-acquisition structure of the hidden-information game but removes the moral stakes. In this condition, both possible payoff consequences accrue only to the decision-maker, so revealing the payoff table remains useful for understanding the decision problem but has no implications for another participant. Non-revelation in this condition therefore captures baseline information avoidance due to confusion, inattention, or complexity rather than the desire to avoid morally relevant information (Exley and Kessler, 2023; Oprea, 2024). Comparing the hidden-information treatments to *Self/Self* helps isolate the portion of ignorance that is specifically tied to moral consequences, while comparing allocation behavior to *Full Info* shows whether belief elicitation moves choices toward the benchmark observed when consequences are known.

3.3.2 Dice Rolling Game

In the DRG, participants are asked to privately roll a six-sided die and are paid in ECUs according to the number they report, with certain outcomes yielding larger payoffs. Because actual rolls are unobservable, this design creates scope for dishonest inflation under plausible deniability. Treatments vary only in the timing of belief elicitation: participants state their beliefs about others’ reporting behavior either before their own report (*Belief First*) or afterward (*Belief After*). To measure dishonesty, I aggregate payoffs in the die-roll tasks by calculating *excess payoffs*—defined as the difference between participants’ average reported earnings and the truthful benchmark implied by a uniform distribution of rolls (Fischbacher and Föllmi-Heusi, 2013). Because actual rolls are unobserved, this measure identifies dishonesty at the treatment level rather than classifying any individual report as dishonest. Before rolling the die, subjects were asked:

“In a similar study, participants were also asked to report their die roll. What percentage of people do you believe accurately (inaccurately) reported their number?”⁷

⁷The experiment varied the framing of the belief elicitation question between positive valence (“accurately”) and negative valence (“inaccurately”). Results are consistent across framings, though effects were larger in the negative framing. Appendix B.4 reports differences in behavior between framing.

As with the MWRG, belief elicitation are not incentivized.⁸ Participants were told that one of their seven reported die rolls would be realized for payment. As with the setup with the MWRG, this design tests whether eliciting beliefs in the *Belief First* condition activates latent norms and constrains dishonesty (Hypothesis 1), whereas eliciting beliefs in the *Belief After* condition produces self-serving rationalizations of misreporting (Hypothesis 2). The DRG thus complements the MWRG by extending the analysis from information avoidance to dishonest reporting, demonstrating the generalizability of belief elicitation effects across domains of moral ambiguity.

3.3.3 Binary Choice Dictator Game

An important question is whether eliciting beliefs about what others do serves as a general prosocial nudge, or whether its effects arise only within contexts of moral ambiguity. The BCDG provides a benchmark case where fairness norms are unambiguous, and there is little scope for moral wiggle room. Following the full information condition from Dana et al. (2007), dictators choose between two allocations: Option B, an egalitarian split (e.g., 50–50), or Option A, a selfish allocation (e.g., 65–0). Because the trade-off is explicit, fairness considerations should already be salient without additional prompting. Subjects in this experiment were asked:

“In another study, participants were asked to participate as Person 1. What percentage of people acting in the role of Person 1 do you believe picked option A (B)?”

The key treatment again varies whether beliefs about others’ behavior are elicited prior to the dictator’s decision (*Belief First*) or after the dictator’s decision (*Belief After*). As in the DRG described above, beliefs were elicited under two frames: either about the proportion of others who chose Option A or the proportion who chose Option B.

⁸This design choice is particularly important in the DRG: incentivizing beliefs about others’ honesty may lead subjects to suspect that the experimenter has independent knowledge of how much cheating occurs. Such a perception could itself suppress dishonest reporting—if participants infer that their own misreporting might be indirectly observed—thereby confounding the mechanism of interest. By keeping belief elicitation unincentivized, the present study avoids this potential confound and provides a cleaner causal test of the norm-activation mechanism.

Table 3.1: Summary of Experimental Designs and Treatment Conditions

Paradigm		Treatments	Purpose
MWRG: Moral wiggle game Timing of elicitation	Moral room with	<i>Belief First, Belief After, Full Info, Self/Self</i>	Tests whether eliciting beliefs before choice activates latent norms and constrains selfishness (H1), or whether eliciting beliefs after choice produces post hoc rationalizations (H2). <i>Full Info</i> and <i>Self/Self</i> provides boundaries on prosocial behavior and complexity aversion.
DRG: Dice rolling game	Dice	<i>Belief First, Belief After</i>	Extends analysis to dishonest misreporting: does eliciting beliefs before constrain dishonesty, and after produce rationalizations?
BCDG: Binary choice dictator game	Binary dictator	<i>Belief First, Belief After</i>	Establishes a boundary case: with unambiguous fairness norms, tests whether elicitation acts as a generalized prosocial nudge (null or negative effect expected).

3.4 Results

To structure the results, I proceed in three parts. I first analyze the Moral Wiggle Room Game (MWRG) in Section 3.4.1, which provides the main test. I then turn to the Dice Rolling Game (DRG) in Section 3.4.2, which extends the analysis to dishonest reporting under private information and provides an additional test of the hypotheses. Finally, I examine the Binary Choice Dictator Game (BCDG) in Section 3.4.3, a setting without moral ambiguity, to establish a boundary condition: when fairness norms are unambiguous, belief elicitation has little effect.

3.4.1 Moral Wiggle Room Game

To begin, I examine behavioral responses in the MWRG, where the timing of belief elicitation was experimentally manipulated. Figure 3.1 reports information avoidance across the *Belief First* and *Belief After* conditions. Eliciting beliefs before choice substantially reduced ignorance: dictators avoided information 18

percentage points less in the *Belief First* condition, consistent with Hypothesis 1 (test of proportions; $p < 0.001$).

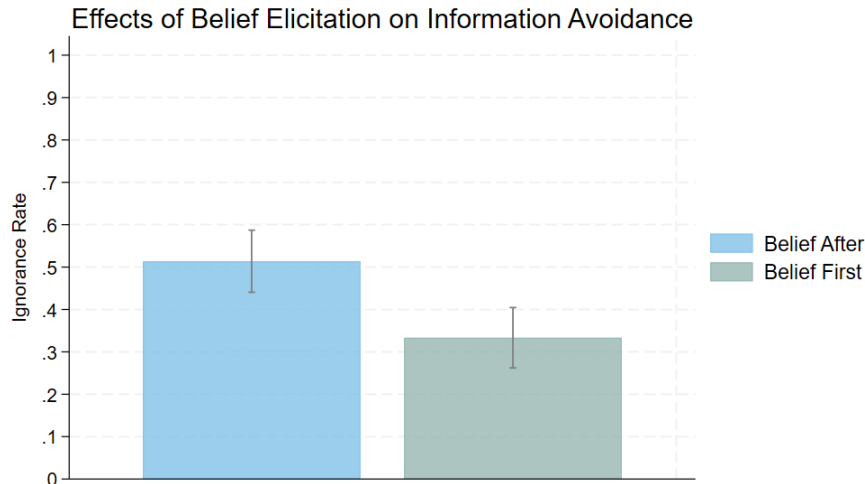


Figure 3.1: Ignorance Rates

Notes: Information avoidance rates among dictators across belief elicitation timing. Error bars denote 95 percentage confidence intervals.

In the *Belief After* condition of the MWRG, beliefs were elicited only after dictators made their allocation decisions. Figure 3.2 reports the belief distributions across *Belief First* and *Belief After*. The distributions differed significantly (Kolmogorov–Smirnov; $D = 0.1611$, $p = 0.021$), consistent with Hypothesis 2. In other words, eliciting beliefs after choice allows beliefs to serve as self-serving justifications, whereas eliciting beliefs before choice activates norms and constrains selfish behavior.

Together, these results support the hypothesis that eliciting beliefs prior to choice activates latent norms and reduces selfish behavior. To better understand the sources of ignorance, the *Self/Self* condition provides a benchmark for the role of complexity aversion (Exley and Kessler, 2023). In this condition, where subjects choose payoffs only for themselves, all participants are incentivized to reveal the payoff tables. Nevertheless, ignorance remains at about 18 percentage points, reflecting an aversion to revealing information even in the absence of self-image concerns. To contextualize this effect, Figure 3.3b shows that in the Conflicting Interest Game, *Belief First* nearly restores prosocial choice to the level observed

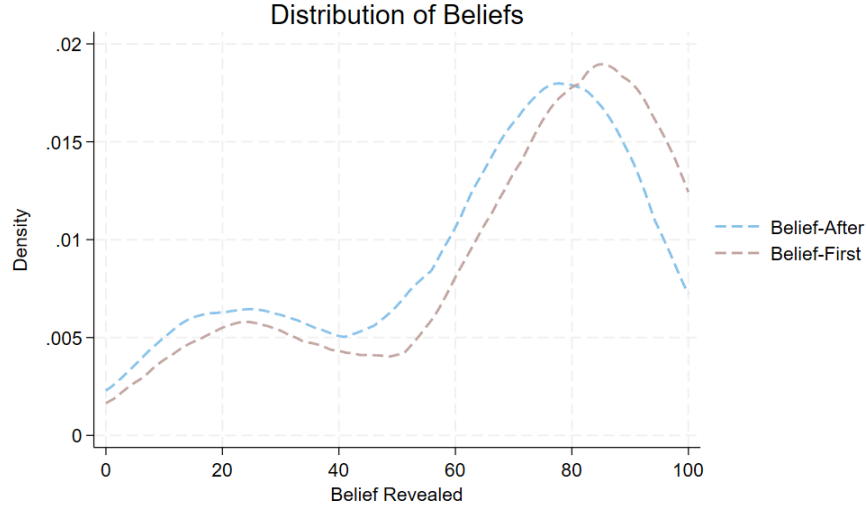


Figure 3.2: Beliefs About Ignorance Rates

under *Full Information*. Specifically, participants chose the equitable 50–50 split over the 65–0 split at rates closely matching the *Full Info* benchmark, demonstrating a near-complete recovery of prosocial behavior.

Ignorance from complexity aversion:

$$\frac{0.177(\text{Self.Self})}{0.514(\text{Belief.After})} = 34.4\%$$

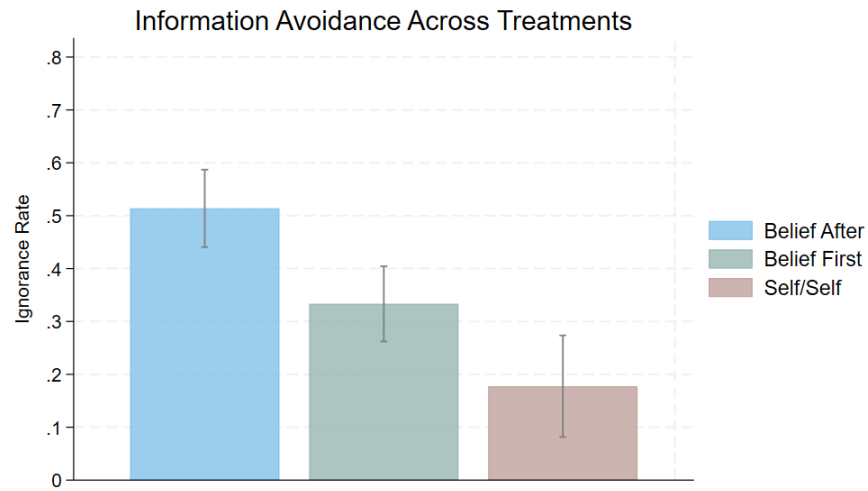
Ignorance recovered from social image:

$$\frac{0.514 - 0.333(\text{Belief.After} - \text{Belief.First})}{0.514(\text{Belief.After})} = 35.2\%$$

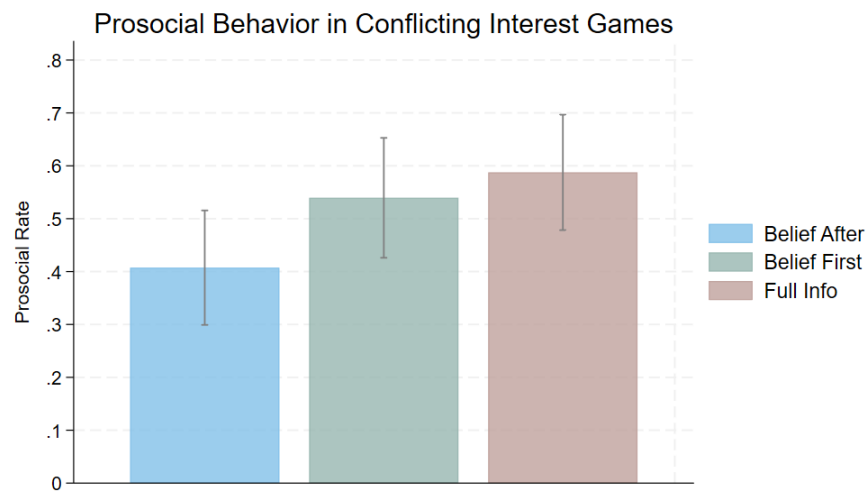
Ignorance from selfishness:

$$\frac{0.333 - 0.177(\text{Belief.First} - \text{Self.Self})}{0.514(\text{Belief.After})} = 30\%$$

The decomposition is similar to Exley and Kessler (2023), where self-image explained 40 to 50% of ignorance, although slightly higher here, likely due to payoff differences. Decomposing overall ignorance, I estimate that 34.4% arises from complexity aversion, 35.2% from social image concerns (mitigated by belief elicitation), and the remaining 30% from selfish preferences. These shares indicate



(a) Ignorance rates across treatments.



(b) Prosocial rates across Conflicting Interest Games.

Figure 3.3: Dictator Behavior Across Treatments.

Notes: Bars indicate 95% confidence intervals.

how belief elicitation activates social image concerns and, in turn, drives the near-complete recovery of prosocial choice.

3.4.2 Dice Rolling Game

Having shown in the MWRG that belief elicitation constrains ignorance, I now test whether this extends to dishonest reporting under private information via the dice rolling game (Fischbacher and Föllmi-Heusi, 2013). I analyze dishonesty by calculating excess payoffs, defined as the difference between the average of the participants' reported earnings across seven rolls and the truthful benchmark of 25 ECUs implied by the uniform distribution of a six sided dice.⁹ Under truth-telling, reported outcomes should yield a symmetric distribution centered at zero excess. Figure 3.4 shows that cheating is evident in both the *Belief First* and *Belief After* conditions, with participants overstating high-paying outcomes and underreporting sixes. Average payoffs declined from 30.41 in *Belief After* to 29.44 in *Belief First*. Relative to the truthful benchmark, this corresponds to a reduction in excess payoffs from 5.41 to 4.44, roughly 18%. A two-sided t -test confirms that the difference in mean excess payoffs is statistically significant ($t = 2.31$, $p = 0.021$). Because excess payoffs are non-normally distributed, I also compare the two distributions using a Wilcoxon rank-sum test, which similarly rejects equality across treatments ($z = 2.40$, $p = 0.016$). Together, these results support Hypothesis 1: belief elicitation curbs cheating by activating honesty norms.

Next, I examine the distribution of beliefs. Figure 3.5 displays the distribution of beliefs about accurate reporting. Subjects in the *Belief After* condition were more likely to believe that others had been less honest. These results parallel the moral wiggle room game: after exploiting the opportunity to cheat, participants appear to rationalize their behavior by lowering their expectations of others. This provides further support to Hypothesis 2.

Although the modal belief in both the *Belief First* and *Belief After* conditions was that most others cheated, eliciting beliefs before choice still reduced dishonest reporting. This suggests that merely being aware of others' dishonesty does not increase one's own tendency to cheat. Instead, making norms salient appears to shift individuals away from baseline self-interested behavior without amplifying selfishness.

⁹Appendix B.3 shows the distribution of die rolls between the single roll and six roll reports.

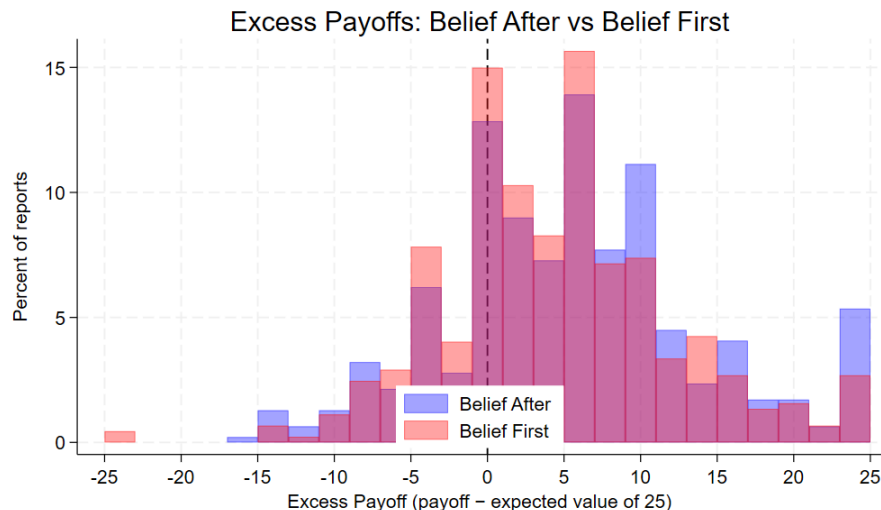


Figure 3.4: Distribution of Excess Payoffs

Notes: Distribution of excess payoffs between treatment conditions. Note that a truthful baseline would result in a symmetrical distribution.

3.4.3 Binary Choice Dictator Game

In the Binary Choice Dictator Game, where dictators chose between a selfish allocation (65–0) and a fair split (50–50), the effects of belief elicitation were modest, with statistically insignificant differences across framings—all within 7–8 percentage points (Figure 3.6). Notably, when participants were prompted to predict the selfish option, prosocial choice declined slightly relative to baseline—suggesting that elicitation in this frame may have reduced the salience of the 50–50 fairness norm. Since equal division is widely recognized as the prevailing prosocial standard in dictator games (Andreoni and Bernheim, 2009), these results imply that belief elicitation can sometimes crowd out norm salience rather than reinforce it, consistent with the latent norms account but indicating that framing matters for whether elicitation strengthens or weakens prosocial behavior.

The inclusion of the BCDG serves a specific purpose: to test whether the effects of belief elicitation hinge on the presence of moral wiggle room itself. In the MWRG and DRG, ambiguity created space for self-serving rationalizations, and belief elicitation curtailed such behavior by activating otherwise latent norms. By contrast, in the BCDG, the fairness norm was already highly salient under full information, leaving little room for ambiguity. Here, asking participants to forecast selfishness may have

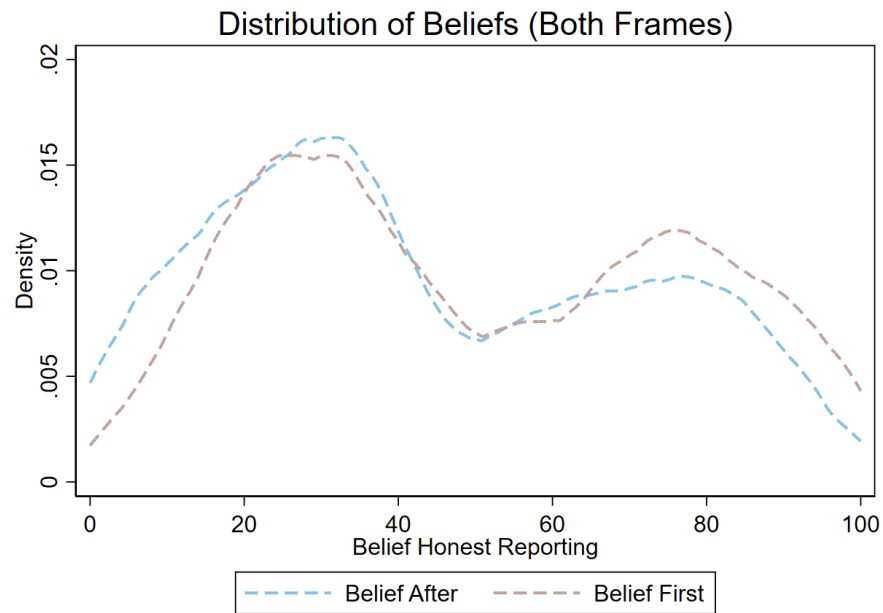


Figure 3.5: Belief Distribution in Dice Rolling Game

Notes: Beliefs about honesty in the Dice Rolling Game. A Kolmogorov–Smirnov tests confirm that the distributions differ significantly between the *Belief First* and *Belief After* conditions ($D = 0.0949$, $p = 0.033$).

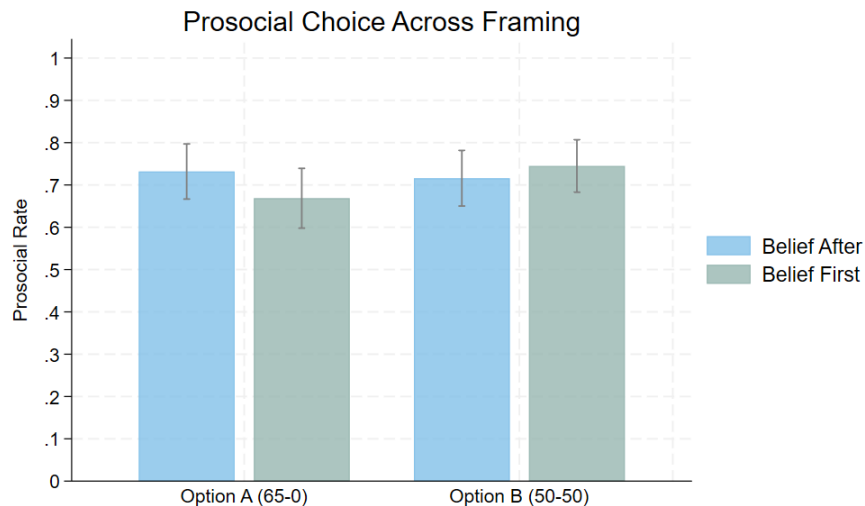


Figure 3.6: Prosocial Choice

Notes: Prosocial choice rates (50–50 over 65–0) under different framings of belief elicitation: “What proportion of subjects chose Option A?” versus “What proportion of subjects chose Option B?” Results are shown separately for the Option A and Option B frames. Error bars denote 95 percentage confidence intervals. Tests of proportions for Option A and Option B: $p = 0.528$ and $p = 0.196$, respectively.

inadvertently shifted attention away from the fairness norm, leading to a modest, but statistically insignificant crowding-out effect with a 6 pp drop in prosocial choice (test of proportions; $p = 0.196$). The results suggest that belief elicitation functions as a salience intervention whose impact depends on baseline norm visibility: when norms are latent, elicitation activates them and increases prosociality; when norms are already salient, certain frames of elicitation may reduce their force.

The null effect in the BCDG is also consistent with the broader mixed evidence on deliberation and altruistic giving in dictator games (Fromell et al., 2020). When the fairness tradeoff is explicit, the relevant norm may already be salient, leaving little additional role for belief elicitation as a norm-activation device.

3.5 A Bounded Rationality Model of Latent Norms

Drawing on theories of bounded rationality and sparse attention (Gabaix, 2014, 2019), I develop a model to illustrate how belief elicitation can activate otherwise latent norms. Motivated by a rational perspective (Dana and van der Weele, 2026), the model assumes that decision-makers have limited attention and therefore do not fully attend to every component of their utility function at all times. Instead, they allocate attention across competing considerations, including material payoffs and social or normative concerns.

In morally ambiguous settings, material incentives may be especially salient at the moment of choice. When the selfish option is immediately available, decision-makers may place disproportionate attention on private material gains while giving less weight to social consequences or internalized standards of appropriate behavior. Belief elicitation before choice changes this attentional environment. By requiring decision-makers to consider what others do or what behavior is socially appropriate before making their own decision, belief elicitation shifts attention toward the social component of the decision problem. As a result, norms that would otherwise remain latent become decision-relevant.

The model below formalizes this mechanism. It shows how the timing of belief elicitation can affect behavior even without changing payoffs, information, or external incentives: belief elicitation before choice increases attention to normative considerations, whereas belief elicitation after choice leaves the initial allocation decision governed more heavily by material incentives.

Agents choose an action $a \in [0, 1]$, where $a = 0$ indicates a fully selfish action and $a = 1$ a fully prosocial action.¹⁰ Material payoffs $\pi(a)$ are decreasing in a : $\pi'(a) < 0$. Let $D \in [0, 1]$ denote the perceived share of others taking prosocial actions (descriptive norm). Cognitive limits determine whether the agent attends to social utility: $A \in [0, 1]$, where $A = 0$ corresponds to latent norms (ignored in choice) and $A = 1$ to fully active norms (fully incorporated). Intermediate A captures partial attention.

¹⁰The continuous action space is a reduced-form way of representing the degree of prosociality in the choice. The binary moral wiggle-room game is nested as the special case $a \in \{0, 1\}$ as in Grossman and van der Weele (2017), but allowing a to vary continuously makes the comparative statics transparent by accommodating evidence that subjects respond to the “price” of fairness: the private payoff cost of choosing the fair or prosocial option rather than the selfish option (van der Weele and Figueroa-Sisniega, 2026).

3.5.1 Preferences and Decision Problem

When norms are attended to, prosocial actions yield social utility

$$S(a; D) = s(a) [\alpha + \beta D], \quad \alpha, \beta \geq 0, \quad (3.1)$$

where $s(a)$ is increasing in a with $s(0) = 0$ and $s'(a) > 0$. The intercept α absorbs any baseline personal norms or intrinsic moral cost, including injunctive norms that are treated as relatively fixed in this setting (Nyborg, 2011; Bašić and Verrina, 2024; Shani and Zeelenberg, 2026).¹¹ The term D captures the idea that the force of this norm may depend on how many others are believed to recognize or follow it. Thus, even if only those who follow a norm are expected to enforce it, a higher perceived share of prosocial actors increases the expected normative pressure associated with selfish behavior.

Attention incurs cognitive cost $C(A)$, with $C(0) = 0$, $C'(A) > 0$ and $C''(A) \geq 0$. $\pi(a)$ denotes material benefits from action. Decision utility is therefore

$$U(a, A; D) = \pi(a) + A \cdot S(a; D) - C(A). \quad (3.2)$$

The agent chooses (a, A) to maximize U .

In most environments, A is naturally low because paying attention to norms requires cognitive effort and yields no immediate material benefit. When social considerations are not made salient—such as when norms are implicit or unobserved—the marginal value of attention is small relative to its cost $C'(A)$. As a result, agents typically operate with *latent norms*, focusing on material payoffs unless an external cue (e.g., norm information, belief elicitation, or an informer) raises the salience of $S(a; D)$ and makes attention worthwhile.

3.5.2 Optimal Attention and Action

At an interior optimum (a^*, A^*) , the first-order conditions are

$$\pi'(a^*) + A^* S_a(a^*; D) = 0 \quad (3.3)$$

$$S(a^*; D) - C'(A^*) = 0 \quad (3.4)$$

¹¹Because selfish actions are privately payoff-maximizing, beliefs about others' behavior may provide a particularly important reference point for other-regarding choices. This is consistent with evidence that individuals often have preferences for norm-following rather than fixed preferences over allocations (Kimbrough and Vostroknutov, 2016). Eliciting beliefs about others' behavior may therefore make the relevant descriptive standard more salient, increasing the weight placed on prosocial action.

where $S_a(a; D) \equiv \frac{\partial S(a; D)}{\partial a}$. Hence a^* is increasing in attention A^* and in the descriptive norm D , while A^* increases in D (and in α). When $C(A) = \kappa A$ is linear, the solution for A typically lies at the corners $\{0, 1\}$.

The prosocial choice is continuous:

$$a^*(D) = a(A^*(D)), \quad \frac{\partial a^*}{\partial D} > 0. \quad (3.5)$$

3.5.3 Belief Timing: Commitment and Rationalization

In many settings, agents are asked to state their beliefs about D before or after choosing a . The timing matters. If beliefs are elicited *before* action, the agent forms a belief D knowing that it will shape the socially appropriate benchmark $S(a; D)$. Stating a belief therefore creates a commitment: choosing a selfish action after claiming that most others behave prosocially is psychologically costly. As a result, *Belief-First* subjects attenuate selfish actions and behave more prosocially.

If beliefs are instead elicited *after* the action, the direction of causality reverses. The agent first picks a to maximize material payoffs with potentially latent norms, and then forms a belief D that justifies this choice. Selfish actions are rationalized by selecting a lower D (“others are selfish too”), reducing any dissonance between norms and behavior.¹² Hence belief-after subjects report lower prosocial beliefs and choose more selfish actions than belief-first subjects.

When beliefs are elicited first, D is interpreted as being close to subjects’ natural or honest beliefs about social norms. Stating beliefs merely makes norms psychologically salient, increasing attention A and shifting behavior toward prosociality.

Full attention is not a requirement for the model: even when norms are only partially active, belief-first creates a constraint, while belief-after supports rationalization.

Formal Timing of Belief-First and Belief-After

The commitment and rationalization mechanisms can be expressed formally as two different sequential maximization problems. The Belief First (BF) and Belief After (BA) cases correspond to different sequential maximization problems. Let D_0 denote a prior belief and $C_D(D; D_0)$ a convex cost of adjusting beliefs.¹³ Let $\hat{a}(D)$

¹²Rabin (1994) argues that cognitive dissonance creates an incentive to normalize immoral behavior as being socially acceptable.

¹³Because decision-makers have some general, if imprecise, sense of their social environment, reported beliefs cannot be adjusted without constraint. Convexity captures the assumption that

be the normatively endorsed action (e.g. $\hat{a}(D) = 1$ when $s(a) = a$), and let $\Psi(a, D)$ capture the psychological cost of deviating from this benchmark, for example

$$\Psi(a, D) = \frac{\psi}{2} [a - \hat{a}(D)]^2, \quad \psi \geq 0.$$

Belief First (BF). When beliefs are elicited before action, the agent solves a two-stage problem. In the first stage,

$$D^{BF} \in \arg \max_D \left\{ -C_D(D; D_0) + \max_{a,A} [\pi(a) + AS(a; D) - C(A) - \Psi(a, D)] \right\}. \quad (3.6)$$

In the second stage, given D^{BF} ,

$$(a^{BF}, A^{BF}) \in \arg \max_{a,A} [\pi(a) + AS(a; D^{BF}) - C(A) - \Psi(a, D^{BF})]. \quad (3.7)$$

The term $\Psi(a, D)$ links the belief chosen in the first stage to the subsequent action, generating a commitment effect.

Belief After (BA). When beliefs are elicited after action, the sequence is reversed. In the first stage, the agent chooses (a, A) given the prior D_0 :

$$(a^{BA}, A^{BA}) \in \arg \max_{a,A} [\pi(a) + AS(a; D_0) - C(A)]. \quad (3.8)$$

In the second stage, after observing own action a^{BA} , the agent selects a belief D to rationalize this choice:

$$D^{BA} \in \arg \max_D [S(a^{BA}; D) - C_D(D; D_0)]. \quad (3.9)$$

Because D is chosen after a , beliefs adjust to support the realized behavior rather than constrain it ex ante, capturing rationalization.

3.5.4 Summary

In this model, an agent chooses both a prosocial action a and the amount of attention A devoted to social considerations. Material incentives are captured by $\pi(a)$, while social utility $S(a; D)$ increases in both the action and the perceived

small deviations from prior beliefs are relatively easy to rationalize, while increasingly extreme deviations from one's prior intuition become disproportionately costly.

descriptive norm D . The parameter α describes a baseline pull toward prosociality, whereas β measures how strongly behavior responds to the perceived behavior of others. Cognitive effort is costly through $C(A)$, so norms can be *latent*: when attention is low, choices reflect only material payoffs; when attention is high, norms are incorporated. The limited-attention assumption is natural in morally ambiguous decisions because subjects face multiple competing considerations, including own payoffs, others' payoffs, and social standards.

Information about norms or forced exposure to such information effectively raises A , making otherwise latent norms behaviorally relevant. Belief elicitation before action creates a commitment effect, while belief elicitation after action leads to rationalization.

3.6 Conclusion

This chapter examines how the salience of internalized norms influences behavior in environments characterized by moral ambiguity. By varying the timing of belief elicitation, I show that prompting individuals to consider what others would do before making a decision increases prosocial actions, consistent with the activation of social norms. In contrast, when beliefs are elicited after decisions, they tend to align with individuals' chosen actions, suggesting a role for post-hoc justification (Bénabou and Tirole, 2011; Kunda, 1990).

One possible interpretation of this mechanism is that belief elicitation activates a form of internalized norm sensitivity: concern about violating perceived social standards even in the absence of external observation or sanction. This channel is related to, but distinct from, guilt, which is typically tied to responsibility for harm or wrongdoing, and from explicit shame, which depend more directly on observability and evaluation by others (Battigalli and Dufwenberg, 2007; Tangney et al., 2007; Bénabou and Tirole, 2006; Bicchieri et al., 2023; Molho et al., 2025). In the present setting, belief elicitation may make perceived standards of appropriate behavior more salient before action, increasing the psychological cost of violating those standards and thereby constraining self-serving behavior. Such interpretation could explain why strategic ignorance fails to materialize when reciprocity norms are active (van der Weele et al., 2014).

The results indicate that social norms may often be latent. Individuals may possess internal standards of appropriate behavior, but these standards do not necessarily guide action unless they are made salient at the moment of choice (Gabaix, 2014, 2019). Belief elicitation provides a simple mechanism for activating these norms, thereby influencing behavior even in the absence of external

enforcement.

At the same time, the results highlight a dual role for beliefs. While they can guide behavior when elicited *ex ante*, they can also serve to rationalize behavior *ex post*, emphasizing the importance of the timing and context in which beliefs are formed and reported. Notably, the results indicate a prosocial nudge, with no significant evidence that people hold latent anti-social norms within the context of the moral wiggle room game and the dice rolling game.¹⁴

These findings reveal an internal channel through which the social environment shapes behavior. Even without external intervention, the structure of the decision environment can activate or suppress norms, influencing the extent to which individuals act in a prosocial manner.

¹⁴Fergusson et al. (2024) shows that certain norms are inherently anti-social, such as a norm against snitching even when doing so would achieve socially desirable outcomes.

Chapter 4

I Didn't Know Either: Social Information and Strategic Ignorance

4.1 Introduction

In many social environments, individuals do not make decisions in isolation.¹ Rather, they observe or infer the behavior of others and use this information to guide their own actions. As discussed in Chapter 1, strategic ignorance allows individuals to manage moral responsibility under uncertainty (Dana et al., 2007; Grossman and van der Weele, 2017). Chapters 2 and 3 show how this behavior is shaped by external interventions and the salience of internal norms. This chapter examines a complementary mechanism: how individuals respond to information about what others do?

When the appropriate course of action is ambiguous, individuals often rely on descriptive norms—beliefs about the behavior of others—to inform their choices (Cialdini et al., 1990; Bicchieri, 2006). Observing that others behave selfishly may legitimize similar behavior, while observing prosocial actions may reinforce norm compliance. Experimental evidence shows that individuals condition their behavior on their beliefs about others' actions, adjusting their choices accordingly (Fischbacher and Gächter, 2010). The evidence suggests that people not only respond to price incentives, but that they also respond to social incentives—adjusting their own behavior in response to descriptive norms or what others do (Allcott, 2011; Hallsworth et al., 2017), and in response to what others believe is socially appropriate (Schultz et al., 2007).

¹A version of this chapter is circulated as a working paper (Hua, 2025a)

In this chapter, I study how exposure to information about others’ behavior affects both information acquisition and allocation decisions. In particular, I examine whether individuals are more or less likely to avoid information when they believe that others do so, and whether behavior converges toward perceived norms.

To investigate this, I conduct experiments in which participants receive information about the behavior of others in similar decision environments. This information varies across treatments, allowing me to identify how individuals adjust their choices in response to different normative signals. As a secondary line of inquiry, following an established line of research on motivated social beliefs (Oprea and Yuksel, 2022; d’Adda et al., 2024), this experiment tests whether or not people may preemptively form self-serving beliefs in anticipation of exploiting ignorance for their own gains.

The results show that individuals are responsive to social information. When participants believe that others behave selfishly, they are more likely to avoid information and act in their own self-interest, in line with evidence of a “boomerang” effect in which social information about others behavior can motivate self-serving behaviors Schultz et al. (2007). Conversely, when they are exposed to more prosocial behavior, their own actions shift in that direction. These findings suggest that internal norms are sensitive to social information. However, contrary to this prediction, I find no evidence that people pre-emptively form self-serving beliefs in anticipation of exploiting ignorance. This null result is informative: it suggests that pre-choice belief reports are not merely anticipatory rationalizations. Instead, the evidence is more consistent with Chapter 3’s interpretation that making beliefs about social norms salient attenuates moral wiggle room, while self-serving belief distortion is more likely to arise *ex post* (Dimant et al., 2020; Exley and Kessler, 2024).

This chapter highlights a social channel through which strategic ignorance operates. Information about others’ behavior influences both how individuals interpret norms and how they act under uncertainty. Together with the previous chapters, the results demonstrate that information avoidance is a context-dependent behavior shaped by multiple dimensions of the social environment.

4.2 Experimental Design

This chapter examines how individuals respond to social information about the behavior of others. While previous chapters focus on externally imposed information and the activation of internal norms, this chapter studies how

descriptive norms—information about what others do—shape both information acquisition and allocation decisions.

The experimental design builds on the moral wiggle-room framework (Dana et al., 2007). As in previous chapters, participants make decisions that affect both their own payoff and that of another individual, with the option to remain ignorant about the consequences of their actions.

Norm information. The experiment is designed to study how beliefs about norms influence behavior when exposed to information about descriptive norms. For this reason, beliefs are *always* elicited prior to choice in this version. After reporting their beliefs about others’ information-seeking behavior, dictators are randomly assigned to one of three *Norm* treatments that describe the percentage of subjects who remained ignorant: *70% Ignorance*, *30% Ignorance*, *10% Ignorance*. A *No Norm* condition in which no social information serves as a control group. The ignorance rates were retrieved from other studies using the moral wiggle room game and thus reflect real ignorance rates.² Payoffs followed from Dana et al. (2007) with amounts of 6 vs 5 for the dictator, and 1 vs 5 for the recipient. These conditions allow me to test whether behavior is responsive to descriptive norms, while the *No Norm* condition provides a benchmark for behavior absent normative cues. Specifically, prior to making a decision as the dictator, subjects are asked:

“In another study, participants were asked to participate as Person 1. What percentage of people acting as Person 1 (the dictator) do you believe revealed the payoff tables?”

In addition, subjects are randomly assigned to a *Known* or *Unknown* condition: in the *Known* condition, they are informed ex ante that they will play the MWRG, while in the *Unknown* condition, they report their beliefs first before being informed that they will participate as the dictator.

This design enables two tests. First, it tests whether individuals strategically distort beliefs in anticipation of selfishness by comparing belief reports between the *Known* and *Unknown* conditions. If anticipating one’s own morally relevant choice induces self-serving belief distortion, subjects in the *Known* condition should report lower expected reveal rates, or equivalently higher expected ignorance rates, than

²The 70% Ignorance and 30% Ignorance environments were retrieved from Grossman et al. (2026b), and the 10% Ignorance environment was retrieved from Lind et al. (2019). Here, differences in the subject pools and subtle changes to the interfaces resulted in different ignorance rates. Subjects were told to predict behavior from a “similar” but not exactly the same studies. Thus, no deception was employed in this experiment.

subjects in the *Unknown* condition. Second, the design tests whether behavior is responsive to descriptive norms when these are made salient, by comparing choices across the *70% Ignorance*, *30% Ignorance*, and *10% Ignorance* treatments. The payoff structure follows the typical payoff structure from moral wiggle room games Dana et al. (2007); Vu et al. (2023).

4.2.1 Social Information Treatments

To study the impact of social information, we introduce treatments in which participants are provided with information about the behavior of others in similar decision environments. This information varies across treatments and is presented prior to the participant’s decision.

In particular, the design consists of a 2×4 between-subjects structure that varies both the information participants receive about others and the context in which beliefs are formed.

First, participants are asked to predict the proportion of other individuals who chose to acquire information in a similar decision environment. This belief elicitation stage captures participants’ prior expectations about descriptive norms.

Second, participants are randomly assigned to one of four norm conditions. In the *No Norm* condition, no feedback is provided. In the remaining conditions, participants receive information about the behavior of others, indicating that either 10%, 30%, or 70% of individuals chose to avoid information in a comparable setting. This variation provides exogenous shifts in perceived norms, allowing us to identify how social information influences behavior.

Finally, the design includes a framing manipulation. In the *Known* condition, participants are aware from the outset that they will make the decision themselves. In the *Unknown* condition, participants initially reason about another individual’s decision before learning that they will take on that role, enabling a test on whether the anticipation of making a morally relevant decision affects elicited beliefs.

After receiving this information, participants make their own decision in the moral wiggle-room game, choosing whether to acquire information and selecting an allocation. This structure allows us to study how both prior beliefs and externally provided social information shape information avoidance and subsequent behavior. In particular, participants are informed about the proportion of previous participants who chose to acquire information.

4.2.2 Hypotheses

I focus on two conceptual hypotheses that describe how individuals respond to social information.

Hypothesis 1: Norm responsiveness. Individuals are responsive to descriptive norms. When provided with information about the behavior of others, they adjust their behavior in the direction of those norms, with stronger responses when their prior beliefs differ from observed behavior.

Hypothesis 2: Ex-ante self-serving belief distortion. Individuals who anticipate exploiting moral wiggle room may pre-emptively distort their beliefs about others' behavior in a self-serving direction. If beliefs are used to justify selfish behavior before the decision is made, then individuals who know they will later make the dictator decision should report lower expected reveal rates—equivalently, higher expected ignorance rates—than individuals who report beliefs before learning that they will make the decision themselves.

These hypotheses imply that social information can influence both information acquisition and allocation decisions. I examine these channels in the results.

4.3 Results

This section begins by outlining the procedures for participant recruitment and data collection. It then describes the data cleaning process, including exclusions and adjustments made to ensure data quality. Next, the analysis verifies whether participants' behavior aligns with theoretical expectations before systematically presenting the key findings. The results are structured to first examine overall patterns, followed by tests of the study's main hypotheses, and concluding with exploratory analyses of unexpected findings.

4.3.1 Data

The study was pre-registered under AEARCT-0014367,³ and the experimental interface was programmed using LIONESS (Giamattei et al. (2020)). Participants were recruited via Prolific in January and February of 2025. The sample was restricted to English-speaking subjects in the United States from the age of 19 to 80 with at least 10 completed studies on their Prolific profile with an approval rating of 99-100%. Recruitment was set to retrieve an even split of male and female subjects. To mitigate risks of sample imbalance from varying subject availability,

³Link to pre-registration: <https://doi.org/10.1257/rct.14367-1.2>

all treatment arms ran simultaneously with subjects being randomly assigned to one of eight treatments in a 2×4 factorial design. A hidden timer (15–16 minutes) regulated study duration, and participants received a base payment of \$2.50–\$2.70 for completing the study.⁴ Additional earnings from decision-making tasks were provided through bonus payments at the study’s conclusion at a rate of 2 ECUs = \$1.

Two participants who revealed the payoff tables but did not make an allocation decision were excluded from the analysis, resulting in a final sample of 1,020. Additionally, 107 participants did not complete the questionnaire before the timer expired. Partial demographic data were recovered for 62 of them using Prolific user records.⁵ The sample was qualitatively balanced across treatment arms, consisting primarily of adults averaging between 37–39 years old, with most identifying as White and only a small proportion currently enrolled as students. A full breakdown of the demographics are available in Appendix C.1.

To explore how these beliefs correlate with actual behavior, Figure 4.1 plots average ignorance rates across bins of belief percentages. A clear negative trend emerges, with participants who believed more of their peers revealed information were themselves more likely to seek information, suggesting a behavioral anchoring effect consistent with norm sensitivity.

4.3.2 Response to Social Information

I begin by examining how participants respond to information about others’ behavior. Figure 4.2 shows the relationship between the behavior of others and participants’ own choices. Participants exposed to higher levels of prosocial behavior are more likely to choose the prosocial option themselves, while those exposed to more selfish behavior tend to act more selfishly.

Figure 4.2 shows the ignorance rates across the four information treatments. Dictators were 8–12 percentage points more likely to avoid information in the high ignorance (70%) environment compared to the low ignorance (30% and 10%) environments respectively. This corresponds to a relative reduction in ignorance behavior of approximately 20–30%, suggesting that even subtle social information cues can meaningfully shift decision-making in morally ambiguous contexts.

⁴In the second half of data collection, the timer was extended from 15 to 16 minutes to allow more time for the final questionnaire. The base payment was adjusted accordingly to maintain a consistent advertised rate on Prolific.

⁵For the remaining 45 participants, a technical error—mostly occurring in the first data collection batch—prevented data retrieval.

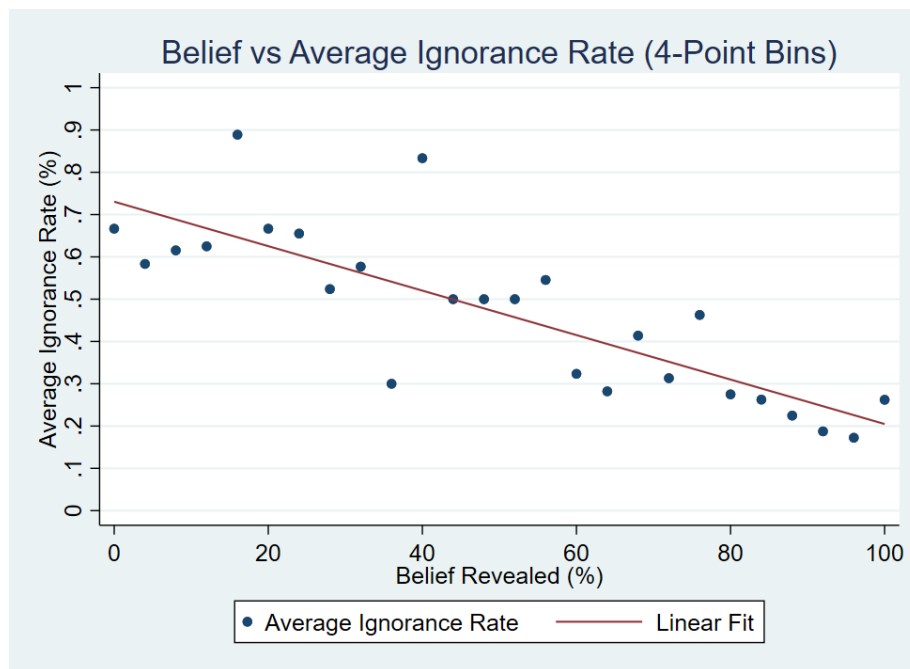


Figure 4.1: Ignorance Rate and Beliefs

Notes: This figure provides a descriptive overview of the relationship between participants' beliefs and their own information choices. Participants are grouped into four-percentage-point bins based on their stated belief about the share of dictators who revealed the payoff table. For each bin, the figure plots the average ignorance rate among participants in that bin.

A nonparametric Cuzick trend test confirms this pattern ($z = 2.69, p = 0.0071$), indicating a statistically significant monotonic decline in ignorance rates from high to low ignorance norms. Appendix C.4 shows this result is robust to demographic controls, such as age, sex, race, and income. This indicates that the decision to avoid information is influenced by the prevailing social norms.

These results indicate that individuals adjust their behavior in the direction of observed norms, consistent with Hypothesis 1.

The results suggest that individuals adjust their behavior based on how much their prior beliefs deviate from observed norms. This effect is particularly pronounced among *ignorance expectant* dictators—those who initially believed that most others would choose to remain ignorant. Although they represent a minority of the sample ($N = 217$), these individuals responded strongly when exposed to a

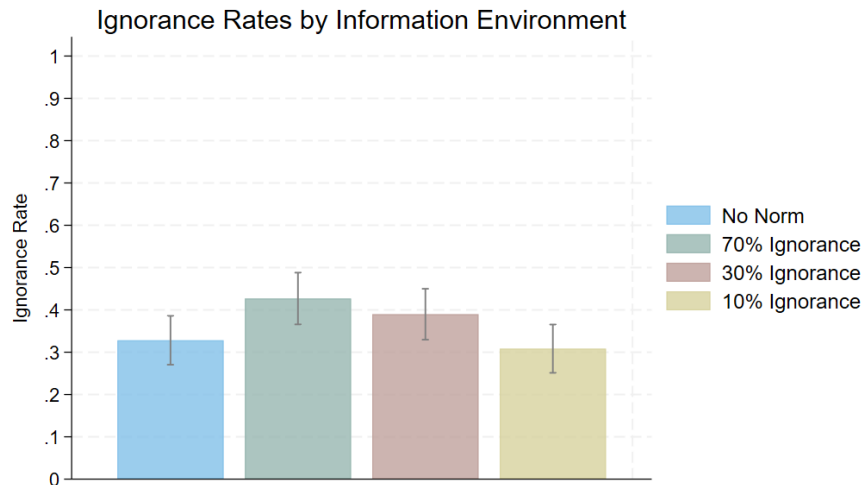


Figure 4.2: Ignorance Rates Across Norm Treatments

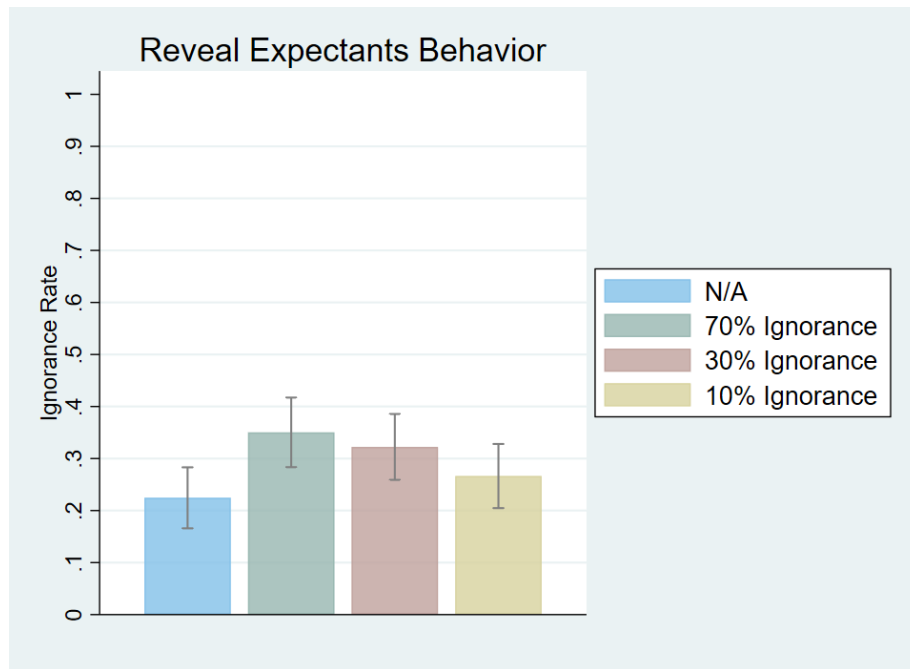
Notes: Bars indicate 95% confidence intervals.

10% Ignorance environment.⁶ Specifically, for *ignorance expectant* dictators, exposure to the 10% Ignorance environment decreased the likelihood of avoiding information by approximately 20 percentage points compared to the pooled sample of other categories. A two-sided difference of means test confirms this effect is statistically significant ($t = 2.76$, $p < 0.001$). This provides evidence for *Hypothesis 1 - Norm Responsiveness*—the idea that individuals whose prior beliefs about ignorance deviate from observed norms adjust their behavior accordingly.

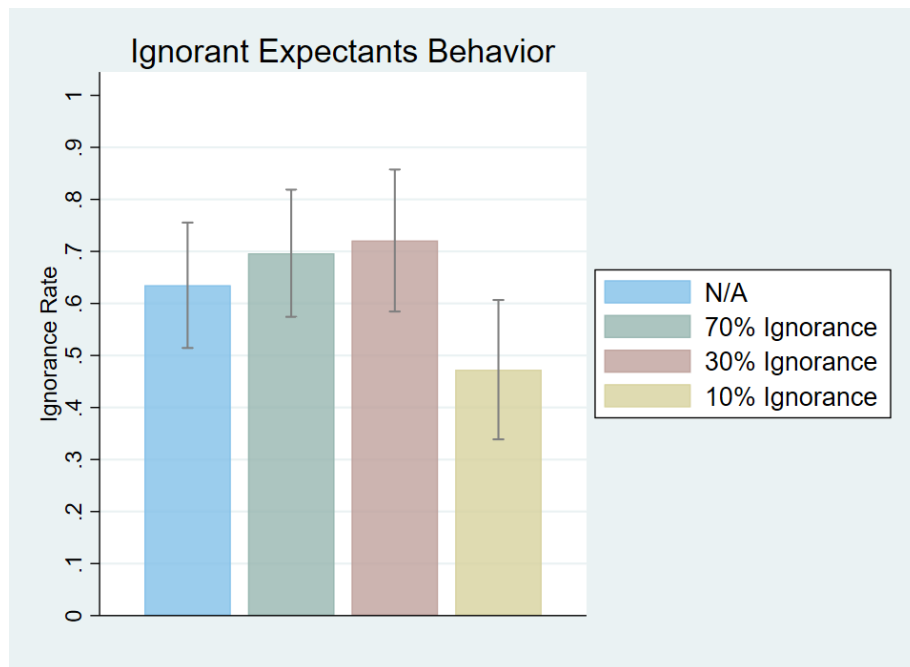
Compared to *reveal expectant* dictators, *ignorance expectant* dictators were far less likely to reveal payoffs but are significantly more likely to follow norms in the 10% Ignorance environment.

Next, I examine whether subjects preemptively form self-serving beliefs about social norms in anticipation of a selfish choice. Figure 4.4 shows no statistically detectable difference in the distribution of beliefs between subjects in the *Known* and *Unknown* conditions. Thus, the data do not support *Hypothesis 2*: anticipating one's own dictator decision does not appear to induce ex-ante self-serving belief distortion.

⁶In Appendix C.7, I check whether *ignorance expectant* dictators were more likely to exhibit norm-following characteristics by regressing the likelihood of being a *ignorance expectant* dictator on the subject's conformity score (Mehrabian and Steff (1995)).



(a) Ignorance rates of subjects who predicted an ignorance rate below 50%. N = 803.



(b) Ignorance rates of subjects who predicted an ignorance rate at or above 50%. N = 217.

Figure 4.3: Ignorance Rates by Types

Notes: Information avoidance rates among reveal expectant and ignorance expectant dictators across norm treatments. Bars indicate 95% confidence intervals.

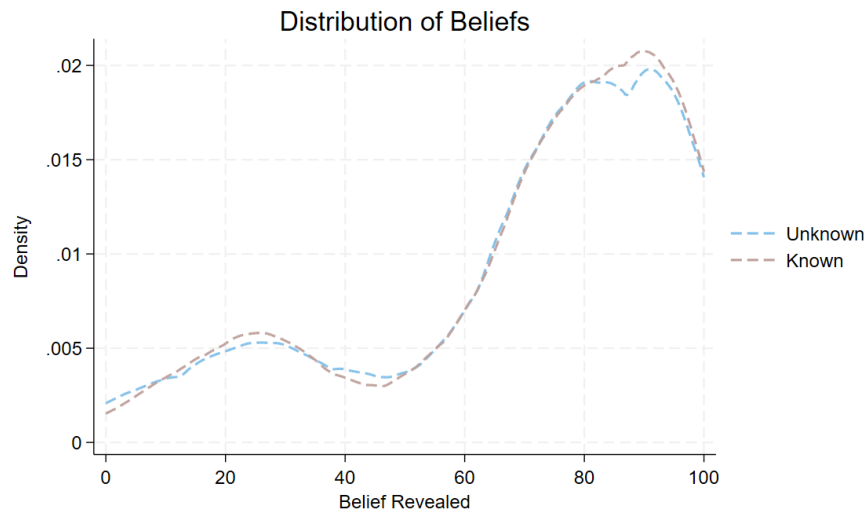


Figure 4.4: Distribution of Beliefs

Notes: Distribution of participants' beliefs about the percentage of dictators who revealed the payoff tables.

A potential concern with the design is that subjects in the *Unknown* condition may nevertheless anticipate that they will later participate in the same game whose behavior they are asked to predict. If so, their belief reports may not reflect ex ante expectations about others' behavior. Instead, subjects who anticipate their own later information choice might form or report beliefs strategically, using pessimistic expectations about others' willingness to reveal information to justify their own future ignorance.

Any anticipatory rationalization would likely attenuate rather than generate the effects from eliciting beliefs first. If some subjects form pessimistic beliefs in advance to justify later ignorance, this would make belief elicitation before choice less effective at activating prosocial norms. Yet in Chapter 3, behavior in the *Belief First* condition resembles behavior in the *Full Info* condition, where the consequences for others are known and ignorance cannot serve as moral cover. This suggests that belief elicitation before choice largely restores prosocial behavior rather than generating the pessimistic, self-serving beliefs characteristic of post-choice rationalization.

4.4 Conclusion

This chapter examines how social information about the behavior of others shapes strategic ignorance. We show that individuals adjust both their information acquisition and allocation decisions in response to perceived norms. When participants are exposed to more self-interested behavior, they are more likely to avoid information and act in their own interest. Conversely, exposure to more prosocial behavior leads to increased information acquisition and more prosocial choices.

These findings demonstrate that strategic ignorance is not only influenced by internal motivations or external constraints, but also by social comparison. Individuals use the behavior of others as a reference point for their own actions, particularly in environments where the appropriate course of action is ambiguous. In such settings, descriptive norms provide a guide for behavior, shaping both how individuals interpret their choices and how they act upon them.

The results also highlight that information avoidance itself is socially responsive. The decision to remain ignorant is not fixed, but varies with beliefs about what others do. When ignorance is perceived to be common, individuals are more likely to adopt it as a strategy; when others are seen to acquire information and act prosocially, this behavior becomes more prevalent.

At the same time, I find no evidence that individuals distort beliefs in a self-serving direction before making their own decision. This null result suggests against deliberate anticipatory rationalization, and instead supports the interpretation that beliefs elicited before action can make latent norms behaviorally salient.

Taken together, these findings identify a comparative social channel through which the social environment shapes behavior. While Chapter 2 shows that others can constrain ignorance by imposing information, and Chapter 3 demonstrates that internalized norms can be activated to guide behavior, this chapter shows that individuals also look to others when deciding how to act. Strategic ignorance is therefore best understood as a context-dependent behavior, shaped by multiple dimensions of the social environment.

Chapter 5

Conclusion

This dissertation examines how individuals manage information in morally relevant decisions and how this process is shaped by the social environment. Across three chapters, a consistent pattern emerges: individuals often avoid information that would make the consequences of their actions more salient, yet they are simultaneously responsive to social forces that constrain such avoidance. Taken together, the findings show that strategic ignorance is not merely an individual decision, but a socially embedded behavior shaped by external intervention, internalized norms, and beliefs about others.

5.1 Information Avoidance and Social Asymmetry

A central empirical regularity that emerges across the dissertation is an asymmetry between the willingness to acquire information and the willingness to provide it. In Chapter 2, many decision-makers actively avoid information about the consequences of their actions, even when that information is costless. At the same time, a substantial share of third parties—informers—are willing to impose that very information on others.

This asymmetry is striking. The same individuals who prefer not to know are embedded in a social environment in which others are willing to ensure that they do. When placed in the role of the informer, participants frequently choose to reveal information, thereby increasing prosocial behavior and improving outcomes for others. Yet when in the role of decision-maker, they often avoid acquiring that same information.

This pattern suggests that behavior depends critically on whether individuals are

thinking about their own actions or about the actions of others. When individuals make decisions for themselves, they may avoid information to preserve flexibility in how their behavior is interpreted. When considering others, however, they appear more willing to enforce norms by ensuring that relevant information is brought to light. In this sense, the social environment introduces a tension between private incentives to remain ignorant and collective tendencies to promote accountability.

5.2 From External Enforcement to Internal Activation

While Chapter 2 highlights an external channel through which information avoidance is constrained, Chapter 3 demonstrates that similar behavioral effects can arise even in the absence of third-party intervention. Simply prompting individuals to think about what others do—through belief elicitation—substantially reduces information avoidance and increases prosocial behavior.

The key insight is that making beliefs about others salient changes how individuals evaluate their own actions. When beliefs are elicited prior to decision-making, individuals are prompted to consider what others typically do or what is socially appropriate. This shifts the frame of the decision from a private choice to one that is implicitly evaluated against a social standard. As a result, individuals behave more prosocially, even though no new information about payoffs is introduced and no external enforcement is present.

This mechanism mirrors the effect of informers in Chapter 2, but operates entirely through internal channels. In both cases, behavior changes because individuals can no longer rely on ignorance as a justification. In Chapter 2, this is because information is imposed externally. In Chapter 3, it is because perceived social norms are made cognitively salient.

Whether or not this phenomenon generalizes to other conditions remains an open question. Prior evidence suggests that strategic ignorance is only strategic when individuals want to avoid confronting an undesirable situation (Spiekermann and Weiss, 2016; Hochman et al., 2026), rather than when they are already motivated to engage (Felgendreher, 2018; Balafoutas et al., 2021; Moyal and Schurr, 2022; Momsen and Ohndorf, 2023a). A latent norm interpretation implies that such a norm already exists within the individual. Thus, while a theory of latent norms may explain why moral wiggle room is contextually sensitive, it does not explain why such underlying beliefs about norms exists in the first place.

5.3 Internalized Shame and the Salience of Social Standards

One interpretation of this internal channel is that belief elicitation activates what may be described as *internalized shame*: a sensitivity to violating perceived social standards even in the absence of external observation. Unlike guilt, which depends on known harm, or social shame, which depends on being observed and evaluated by others (Tangney et al., 2007; Bicchieri et al., 2023), internalized shame operates through individuals' beliefs about what others view as appropriate behavior. Thus, many latent norms may be primarily socially motivated, serving as higher order beliefs about social standards that one might not personally hold. Because higher order beliefs are more distant from one's intrinsic motivations, they are manifested as internalized shame, relying on perceived social considerations.

When individuals are prompted to consider what others do, they implicitly bring these social standards to mind. Acting in violation of those standards may then carry a psychological or emotional cost, even in private settings. In this way, belief elicitation can substitute for external enforcement by activating an internalized form of social evaluation.

Importantly, this mechanism does not rely on observability or reputational concerns. As shown in Chapter 3, behavior changes even when actions remain private and no external consequences are introduced. Instead, it is sufficient to make perceived social expectations salient at the moment of choice.

The relationship between different forms of norm enforcement is illustrated in Figure 5.1, which presents a layered framework ranging from formal sanctions and deterrence to socially enforced shame and, ultimately, to fully internalized regulation. At the more internalized level, behavior is guided not by external consequences but by individuals' own perceptions of social expectations. The results of this dissertation suggest that such internal mechanisms can play a central role in shaping behavior, particularly in environments characterized by moral ambiguity.

5.4 A Unified View of Moral Regulation

Taken together, the findings across chapters point to a unified view of moral regulation in which behavior is shaped by multiple, interacting channels:

1. **External enforcement (Chapter 2):** Others can constrain information avoidance by imposing information, thereby increasing accountability.

2. **Internal activation (Chapter 3):** Making social norms salient activates internalized standards that guide behavior, even in private settings.
3. **Social information (Chapter 4):** Beliefs about what others do shape both information acquisition and allocation decisions, influencing how norms are interpreted and applied.

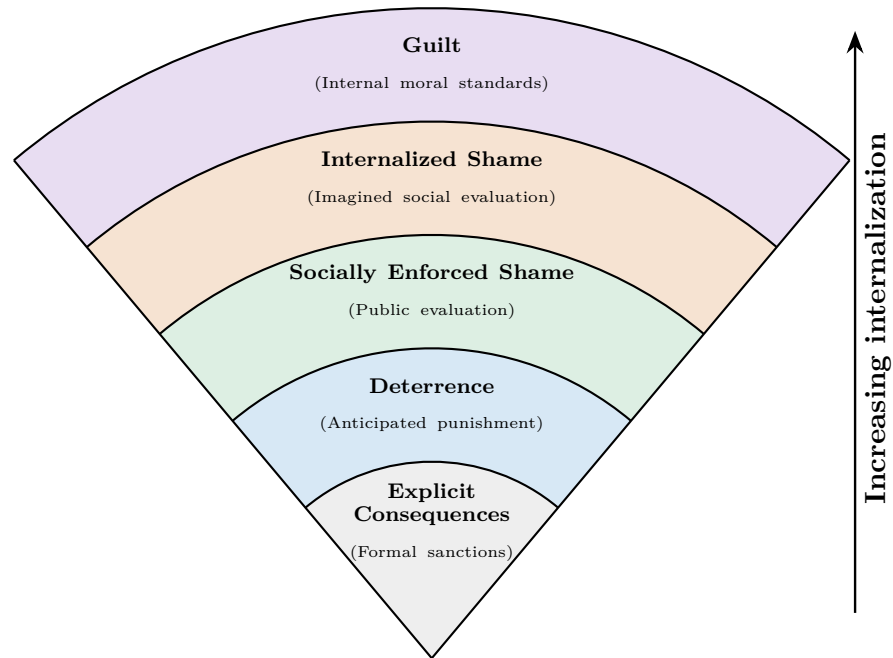


Figure 5.1: Diagram of Moral Regulation

Notes: Layers of norm enforcement, ranging from external to internal mechanisms. Enforcement can arise through formal sanctions and deterrence, through social evaluation (socially enforced shame), or through internalized responses to perceived social expectations (internalized shame and guilt). Moving toward more internalized layers, behavior is increasingly regulated without the need for external monitoring or punishment.

These channels are complementary. External interventions can directly alter the information environment, while internal and social mechanisms determine how individuals interpret and respond to that environment. Importantly, the effectiveness of each channel depends on context. In some settings, external enforcement may be necessary to overcome strong incentives for ignorance. In others, subtle changes in how decisions are framed may be sufficient to activate internalized norms.

5.5 Implications and Future Directions

The results have broader implications for understanding behavior in real-world settings where the consequences of actions are uncertain or obscured. In domains such as ethical consumption, environmental decision-making, and organizational behavior, individuals often face opportunities to remain uninformed about the social impact of their choices. The findings of this dissertation suggest that such ignorance is not fixed, but can be shaped by both external interventions and internal cognitive processes.

Policies that increase transparency or enable third-party information provision may reduce harmful behavior by limiting the scope for strategic ignorance. At the same time, interventions that make social norms salient—such as prompting individuals to reflect on what others do or what is considered appropriate—may achieve similar effects at lower cost by activating internalized forms of regulation.

Several directions for future research emerge from this work. First, understanding how beliefs about norms evolve over time, and how those beliefs interact with behavior, is critical for explaining the persistence or erosion of cooperation. This question is especially important because the absence of information may itself shape adaptive behavior: evolutionary models suggest that payoff ignorance can alter which strategies survive in repeated social environments (Friedman et al., 2015; Huck et al., 2017). Second, further work is needed to distinguish more precisely among psychological mechanisms such as guilt, internalized shame, social-image concerns, and norm-following, and to identify the conditions under which each becomes dominant. Third, future research should examine how these mechanisms vary across cultural and institutional contexts. Cross-cultural evidence shows substantial variation in social preferences and fairness norms across societies (Henrich et al., 2001, 2005), raising the possibility that strategic ignorance, norm activation, and information imposition may operate differently where responsibility, intention, and moral obligation are understood differently. Finally, future research should examine how institutional design can

leverage both external and internal channels of enforcement to promote socially desirable behavior in complex environments. This is particularly important given mixed evidence on the effectiveness of information campaigns, which suggests that simply providing information may not be sufficient when individuals have incentives to avoid, reinterpret, or discount it (Thunström et al., 2026).

5.6 Concluding Remarks

This dissertation shows that the question is not only what individuals know, but also how the social environment shapes what they choose to know—and how they act on that knowledge. Strategic ignorance is not simply a failure of information acquisition, but a behavior that reflects the interplay between private incentives and social expectations.

When individuals are left alone, they may choose not to know. But when others intervene, or when social norms are made salient, this ignorance becomes more difficult to sustain. In this way, moral behavior is shaped not only by information itself, but by the social meaning attached to it—and by the extent to which individuals are made to confront that meaning, either by others or by themselves.

Bibliography

- Akerlof, G. A. (1980). A theory of social custom, of which unemployment may be one consequence. *The Quarterly Journal of Economics*, 94(4):749–775.
- Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9-10):1082–1095.
- Andreoni, J. and Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636.
- Balafoutas, L., Sandakov, F., and Zhuravleva, T. (2021). No moral wiggle room in an experimental corruption game. *Frontiers in Psychology*, 12.
- Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., and Trötschel, R. (2001). The automated will: Nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology*, 81(6):1014–1027.
- Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M. T., Fitzpatrick, S., Gurven, M., Henrich, J., Kanovsky, M., Kushnick, G., Pisor, A., Scelza, B. A., Stich, S., von Rueden, C., Zhao, W., and Laurence, S. (2016). Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences*, 113(17):4688–4693.
- Bartling, B., Engl, F., and Weber, R. A. (2014). Does willful ignorance deflect punishment? – an experimental study. *European Economic Review*, 70:512–524.
- Battigalli, P. and Dufwenberg, M. (2007). Guilt in games. *American Economic Review*, 97(2):170–176.

- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., and Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4):323–370.
- Bašić, I. and Verrina, L. (2024). Personal norms — and not only social norms — shape economic behavior. *Journal of Public Economics*, 234:104954.
- Bhattacharya, H. and Dugar, S. (2022). Business norm versus norm-nudge as a contract-enforcing mechanism: Evidence from a real marketplace. *European Economic Review*, 144:104078.
- Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press, Cambridge, UK.
- Bicchieri, C., Dimant, E., and Sonderegger, S. (2023). It’s not a lie if you believe the norm does not apply: Conditional norm-following and belief distortion. *Games and Economic Behavior*, 138:321–354.
- Bolton, G. E. and Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1):166–193.
- Brütt, K., Schram, A., and Sonnemans, J. (2020). Endogenous group formation and responsibility diffusion: An experimental study. *Games and Economic Behavior*, 121:1–31.
- Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678.
- Bénabou, R. and Tirole, J. (2011). Identity, morals, and taboos: Beliefs as assets *. *The Quarterly Journal of Economics*, 126(2):805–855.
- Cain, D. and Dana, J. (2012). Paying people to look at the consequences of their actions. Working paper.
- Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869.
- 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):1015–1026.
- Dana, J. and van der Weele, J. J. (2026). Willful ignorance: A perspective from economics. *Current Opinion in Psychology*, 68:102210.

- Dana, J., Weber, R., and Kuang, X. J. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1):67–80.
- Diener, E. and Wallbom, M. (1976). Effects of self-awareness on antinormative behavior. *Journal of Research in Personality*, 10(1):107–111.
- Dimant, E., van Kleef, G. A., and Shalvi, S. (2020). Requiem for a nudge: Framing effects in nudging honesty. *Journal of Economic Behavior & Organization*, 172:247–266.
- Duch, M. L., Grossmann, M. R. P., and Lauer, T. (2020). z-tree unleashed: A novel client-integrating architecture for conducting z-tree experiments over the internet. *Journal of Behavioral and Experimental Finance*, 28:Article 100400.
- d’Adda, G., Gao, Y., Golman, R., and Tavoni, M. (2024). Strategic information avoidance, belief manipulation and the effectiveness of green nudges. *Ecological Economics*, 222:108191.
- Enke, B., Fisman, R., Freitas, L. M., and Sun, S. (2024). Universalism and political representation: Evidence from the field. *American Economic Review: Insights*, 6(2):214–29.
- Enke, B., Rodríguez-Padilla, R., and Zimmermann, F. (2022). Moral universalism: Measurement and economic relevance. *Management Science*, 68(5):3590–3603.
- Exley, C. L. (2016). Excusing selfishness in charitable giving: the role of risk. *Review of Economic Studies*, 83(2):587–628.
- Exley, C. L. and Kessler, J. B. (2023). Information Avoidance and Image Concerns. *The Economic Journal*, 133(656):3153–3168.
- Exley, C. L. and Kessler, J. B. (2024). Motivated errors. *American Economic Review*, 114(4):961–87.
- Fahrenwaldt, A., tho Pesch, F., Fiedler, S., and Baumert, A. (2024). What’s moral wiggle room? a theory specification. *Judgment and Decision Making*, 19:e17, 1–29.
- Falk, A., Neuber, T., and Szech, N. (2020). Diffusion of being pivotal and immoral outcomes. *The Review of Economic Studies*, 87(5):2205–2229.
- Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

- Feiler, L. (2014). Testing models of information avoidance with binary choice dictator games. *Journal of Economic Psychology*, 45:253–267.
- Felgendreher, S. (2018). Do consumers choose to stay ignorant? the role of information in the purchase of ethically certified products. Technical Report 717, Working Paper in Economics.
- Fergusson, L., Guerra, J.-A., and Robinson, J. A. (2024). Anti-social norms. NBER Working Paper 32717, National Bureau of Economic Research.
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press, Stanford, CA.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10:171–178.
- 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.
- Fischbacher, U. and Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, 100(1):541–56.
- Friedman, D., Huck, S., Oprea, R., and Weidenholzer, S. (2015). From imitation to collusion: Long-run learning in a low-information environment. *Journal of Economic Theory*, 155:185–205.
- Fromell, H., Nosenzo, D., and Owens, T. (2020). Altruism, fast and slow? evidence from a meta-analysis and a new experiment. *Experimental Economics*, 23:979–1001.
- Gabaix, X. (2014). A sparsity-based model of bounded rationality. *Quarterly Journal of Economics*, 129(4):1661–1710.
- Gabaix, X. (2019). Behavioral inattention. *Handbook of Behavioral Economics: Applications and Foundations 1*, 2:261–343.
- Giamattei, M., Yahosseini, K. S., Gächter, S., and Molleman, L. (2020). Lioness lab: a free web-based platform for conducting interactive experiments online. *Journal of the Economic Science Association*, 6:95–111.
- Grossman, Z. (2014). Strategic ignorance and the robustness of social preferences. *Management Science*, 60(11):2659–2665.

- Grossman, Z. and Hua, T. (2026). Willful ignorance in social decisions: Robust, yet contextually sensitive. *Current Opinion in Psychology*, 67:102211.
- Grossman, Z., Hua, T., Lind, J. T., and Nyborg, K. (2026a). Unwillingly informed: the prosocial impact of third-party informers. *Journal of Environmental Economics and Management*, 138:103330.
- Grossman, Z., Hua, T., Lind, J. T., and Nyborg, K. (2026b). Unwillingly informed: the prosocial impact of third-party informers. *Journal of Environmental Economics and Management*, 138:103330.
- Grossman, Z. and van der Weele, J. (2017). Self-image and willful ignorance in social decisions. *Journal of the European Economic Association*, 15(1):173–217.
- Hallsworth, M., List, J. A., Metcalfe, R. D., and Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148:14–31.
- Henrich, J. (2020). *The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous*. Farrar, Straus and Giroux, New York.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., and McElreath, R. (2001). In search of homo economicus: Behavioral experiments in 15 small-scale societies. *American Economic Review*, 91(2):73–78.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N. S., Hill, K., Gil-White, F., Gurven, M., Marlowe, F. W., Patton, J. Q., and Tracer, D. (2005). “economic man” in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, 28(6):795–815.
- Hochman, G., Kalagy, T., Malul, S., and Yosef, R. (2026). Choosing not to know: The emotional and sociocultural architecture of pension willful ignorance. *Current Opinion in Psychology*, 67:102181.
- Hua, T. (2025a). I didn’t know either: How beliefs about norms shape strategic ignorance. Technical Report 5390751, SSRN.
- Hua, T. (2025b). Latent norms: Prosocial nudges under moral ambiguity.

- Hua, T. (2026). When does moral wiggle room arise? implications from large language models. Working paper.
- Huck, S., Leutgeb, J., and Oprea, R. (2017). Payoff information hampers the evolution of cooperation. *Nature Communications*, 8:15147.
- Kimbrough, E. O. and Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3):608–638.
- Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. *American Economic Review*, 90(4):1072–1091.
- Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3):480–498.
- Kvarven, A., Strømland, E., Wollbrant, C., Andersson, D., Johannesson, M., Tinghög, G., Västfjäll, D., Myrseth, K. O. R., Mørkbak, M. R., and Hauge, K. E. (2020). The intuitive cooperation hypothesis revisited: A meta-analytic examination of effect size and between-study heterogeneity. *Journal of the Economic Science Association*, 6(1):26–42.
- Larson, T. and Capra, C. M. (2009). Exploiting moral wiggle room: Illusory preference for fairness? a comment. *Judgment and Decision Making*, 4(6):467–474.
- Lind, J. T., Nyborg, K., and Pauls, A. (2019). Save the planet or close your eyes? testing strategic ignorance in a charity context. *Ecological Economics*, 161:9–19.
- Mas-Colell, A., Whinston, M. D., and Green, J. R. (1995). *Microeconomic Theory*. Oxford University Press, New York.
- Matthey, A. and Regner, T. (2011). Do i really want to know? a cognitive dissonance-based explanation of other-regarding behavior. *Games*, 2(1):114–135.
- Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45(6):633–644.

- Mehrabian, A. and Steffl, C. A. (1995). Basic temperament components of loneliness, shyness, and conformity. *Social Behavior and Personality: An International Journal*, 23(3):253–263.
- Mol, J. M., Soraperra, I., and van der Weele, J. J. (2025). Spoiling the party: Experimental evidence on the willingness to transmit inconvenient ethical information. *Experimental Economics*, 28(1):181–199.
- Molho, C., Soraperra, I., Schultz, J., and Shalvi, S. (2025). Guilt- and shame-driven prosociality across societies. *Nature Human Behavior*.
- Momsen, K. and Ohndorf, M. (2020). When do people exploit moral wiggle room? an experimental analysis of information avoidance in a market setup. *Ecological Economics*, 169:106479.
- Momsen, K. and Ohndorf, M. (2022). Information avoidance, selective exposure, and fake (?) news: Theory and experimental evidence on green consumption. *Journal of Economic Psychology*, 88:102457.
- Momsen, K. and Ohndorf, M. (2023a). Expressive voting versus information avoidance: experimental evidence in the context of climate change mitigation. *Public Choice*, 194:45–74.
- Momsen, K. and Ohndorf, M. (2023b). Information avoidance: Self-image concerns, inattention, and ideology. *Journal of Economic Behavior & Organization*, 211:386–400.
- Moyal, A. and Schurr, A. (2022). The effect of deliberate ignorance and choice procedure on pro-environmental decisions. *Ecological Economics*, 200:107512.
- Nyborg, K. (2011). I don’t want to hear about it: Rational ignorance among duty-oriented consumers. *Journal of Economic Behavior & Organization*, 79(3):263–274.
- Oprea, R. (2020). What makes a rule complex? *American Economic Review*, 110(12):3913–51.
- Oprea, R. (2024). Decisions under risk are decisions under complexity. *American Economic Review*, 114(12):3789–3811.
- Oprea, R. and Yuksel, S. (2022). Social exchange of motivated beliefs. *Journal of the European Economic Association*, 20(2):667–699.

- Rabin, M. (1994). Cognitive dissonance and social change. *Journal of Economic Behavior & Organization*, 23(2):177–194.
- Rand, D. G. (2016). Cooperation, fast and slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychological Science*, 27(9):1192–1206.
- Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5):429–434.
- Schwartz, S. H. (1977). Normative influences on altruism. *Advances in Experimental Social Psychology*, 10:221–279.
- Serra-Garcia, M. and Szech, N. (2022). The (in)elasticity of moral ignorance. *Management Science*, 68(7):4755–5555.
- Shani, Y. and Zeelenberg, M. (2026). The pain of suspecting and the comforts of knowing the worst. *Current Opinion in Psychology*, 67:102208.
- Soraperra, I., van der Weele, J., Villeval, M. C., and Shalvi, S. (2023). The social construction of ignorance: Experimental evidence. *Games and Economic Behavior*, 138:197–213.
- Spiekermann, K. and Weiss, A. (2016). Objective and subjective compliance: A norm-based explanation of ‘moral wiggle room’. *Games and Economic Behavior*, 96:170–183.
- Tangney, J. P., Stuewig, J., and Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology*, 58:345–372.
- Thaler, R. H. and Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press, New Haven, CT.
- Thunström, L., van ’t Veld, K., and Shogren, J. F. (2026). Interventions that reduce willful ignorance of policy-relevant information. *Current Opinion in Psychology*, 67:102209.
- van der Weele, J. J. and Figueroa-Sisniega, C. (2026). Inconvenient truths: A note on information avoidance and the price of fairness. *European Economic Review*, 181:105188.

- van der Weele, J. J., Kulisa, J., Kosfeld, M., and Friebe, G. (2014). Resisting moral wiggle room: how robust is reciprocal behavior? *American Economic Journal: Microeconomics*, 6(3):256–264.
- Vu, L. and Molho, C. (2025). Reluctant altruism: Underlying mechanisms and global variations. *Current Opinion in Psychology*, 66:102107.
- Vu, L., Soraperra, I., Leib, M., van der Weele, J., and Shalvi, S. (2023). Ignorance by choice: A meta-analytic review of the underlying motives of willful ignorance and its consequences. *Psychological Bulletin*, 149(9-10):611–635.

Appendix A

A.1 Summary Statistics and Demographics

In this appendix, we present the demographic data that was automatically collected by Prolific and the additional data we collected using the questionnaire embedded in our experiment. Table A.1 provides summary statistics on all available demographic variables. As Prolific routinely expires transitory variables, many subjects did not have their current employment status available.

	Base. n=117	Hid. Info n=109	Hyp. Inf. n=99	Inf. n=391	Inf. Bonus n=370	Hid. Info (IF) n=105	Hyp. Inf. (IF) n=110	Inf. (IF) n=105	Total n=1,406
Age	37.14 (11.91)	34.33 (14.05)	35.48 (10.72)	37.86 (12.67)	35.13 (11.90)	35.06 (11.51)	38.14 (13.35)	36.28 (12.29)	36.33 (12.39)
Sex									
Female	55 (54.5%)	49 (45.0%)	50 (50.5%)	198 (50.6%)	189 (51.1%)	51 (48.6%)	55 (50.0%)	50 (47.6%)	697 (50.1%)
Male	46 (45.5%)	60 (55.0%)	49 (49.5%)	189 (48.3%)	175 (47.3%)	53 (50.5%)	55 (50.0%)	55 (52.4%)	682 (49.1%)
Not Consented	0 (0.0%)	0 (0.0%)	0 (0.0%)	4 (1.0%)	6 (1.6%)	1 (1.0%)	0 (0.0%)	0 (0.0%)	11 (0.8%)
Ethnicity									
White	75 (74.3%)	72 (66.1%)	68 (68.7%)	243 (62.1%)	246 (66.5%)	70 (66.7%)	81 (73.6%)	75 (71.4%)	930 (66.9%)
Black	6 (5.9%)	7 (6.4%)	11 (11.1%)	46 (11.8%)	33 (8.9%)	11 (10.5%)	5 (4.5%)	5 (4.8%)	124 (8.9%)
Asian	8 (7.9%)	18 (16.5%)	9 (9.1%)	40 (10.2%)	41 (11.1%)	6 (5.7%)	10 (9.1%)	11 (10.5%)	143 (10.3%)
Mixed	3 (3.0%)	11 (10.1%)	7 (7.1%)	38 (9.7%)	27 (7.3%)	9 (8.6%)	11 (10.0%)	7 (6.7%)	113 (8.1%)
Other	5 (5.0%)	1 (0.9%)	3 (3.0%)	15 (3.8%)	12 (3.2%)	5 (4.8%)	2 (1.8%)	5 (4.8%)	48 (3.5%)
Not Consented	4 (4.0%)	0 (0.0%)	1 (1.0%)	9 (2.3%)	11 (3.0%)	4 (3.8%)	1 (0.9%)	2 (1.9%)	32 (2.3%)
Student Status									
Not Student	47 (88.7%)	45 (76.3%)	54 (73.0%)	229 (83.6%)	199 (72.4%)	59 (77.6%)	66 (84.6%)	69 (92.0%)	768 (79.7%)
Current Student	6 (11.3%)	14 (23.7%)	20 (27.0%)	41 (15.0%)	70 (25.5%)	16 (21.1%)	12 (15.4%)	6 (8.0%)	185 (19.2%)
Not Consented	0 (0.0%)	0 (0.0%)	0 (0.0%)	4 (1.5%)	6 (2.2%)	1 (1.3%)	0 (0.0%)	0 (0.0%)	11 (1.1%)
Political Party									
Democrat	39 (33.3%)	42 (38.5%)	38 (38.4%)	135 (34.5%)	130 (35.1%)	39 (37.1%)	34 (30.9%)	35 (33.3%)	492 (35.0%)
Republican	10 (8.5%)	11 (10.1%)	13 (13.1%)	44 (11.3%)	61 (16.5%)	17 (16.2%)	19 (17.3%)	19 (18.1%)	194 (13.8%)
Independent/Other	68 (58.1%)	56 (51.4%)	48 (48.5%)	194 (49.6%)	162 (43.8%)	49 (46.7%)	56 (50.9%)	51 (48.6%)	684 (48.6%)
Not Consented	0 (0.0%)	0 (0.0%)	0 (0.0%)	18 (4.6%)	17 (4.6%)	0 (0.0%)	1 (0.9%)	0 (0.0%)	36 (2.6%)
Employment Status									
Full-Time	24 (20.5%)	24 (22.0%)	22 (22.2%)	116 (29.7%)	107 (28.9%)	34 (32.4%)	30 (27.3%)	40 (38.1%)	397 (28.2%)
Part-Time	9 (7.7%)	15 (13.8%)	7 (7.1%)	44 (11.3%)	39 (10.5%)	10 (9.5%)	11 (10.0%)	6 (5.7%)	141 (10.0%)
Not Employed	13 (11.1%)	13 (11.9%)	12 (12.1%)	51 (13.0%)	56 (15.1%)	9 (8.6%)	17 (15.5%)	14 (13.3%)	185 (13.2%)
Other	1 (0.9%)	5 (4.6%)	6 (6.1%)	19 (4.9%)	15 (4.1%)	5 (4.8%)	5 (4.5%)	3 (2.9%)	59 (4.2%)
Not Consented	70 (59.8%)	52 (47.7%)	52 (52.5%)	161 (41.2%)	153 (41.4%)	47 (44.8%)	47 (42.7%)	42 (40.0%)	624 (44.4%)

Notes: Parentheses indicate standard deviation or percentage based on category.

Table A.1: Summary Demographics for Moral Wiggle Room Game

	Base. n=117	Hid. Info n=109	Hyp. Inf. n=99	Inf. n=391	Inf. Bonus n=370	Hid. Info (IF) n=105	Hyp. Inf. (IF) n=110	Inf. (IF) n=105	Total n=1,406
Dict. Ignorance Rate	. (.)	0.55 (0.50)	0.58 (0.50)	0.61 (0.49)	0.70 (0.46)	0.30 (0.46)	0.21 (0.41)	0.32 (0.47)	0.50 (0.50)
Inf. Supplied Info	. (.)	. (.)	0.75 (0.44)	0.71 (0.46)	0.69 (0.46)	. (.)	0.72 (0.45)	0.74 (0.44)	0.72 (0.45)
Recip. Mean Payoff	4.01 (0.00)	3.53 (0.00)	3.87 (0.00)	4.37 (0.00)	4.24 (0.00)	4.09 (0.00)	4.24 (0.00)	4.43 (0.00)	4.18 (0.24)

Notes: Parentheses indicate standard deviation.

Table A.2: Summary statistics for decision variables across *all* treatments

A.2 Replication treatments

Lind et al. (2019) did not find much information avoidance, even *without* an informer present, which made it difficult to identify the impact of the third-party informer. To proceed with our investigation, we first needed to identify a decision environment in which the kind of information avoidance documented by Dana et al. (2007) would prevail. In this appendix, we describe the design of a preregistered replication experiment that we conducted before the *Informer* treatments reported in the main body of this paper. We explain the steps we took during the pilot process to arrive at the ultimate design and then we report the results of the replication treatments. Because we found non-trivial levels of information avoidance and our results were otherwise comparable of Dana et al. (2007) and similar replications, we concluded that the decision environment was suitable for investigating the impact of the informer and proceeded with our main treatments.

Pilots and design Because having to actively choose ignorance can reduce participants’ propensity to do so (see Grossman (2014)), we favored layout and defaults that avoided presenting the information decision as an active choice. We also avoided highlighting the information choice more than absolutely necessary to avoid provoking extra scrutiny or reflection upon the information choice. We also aimed for brief and simple instructions rather than elaborate explanations, making it possible to ‘close one’s eyes’ for those who may want to do so.

We initially preregistered our study in April 2022.¹ In that initial plan, we pointed out that since our research questions can only be explored in a context where strategic ignorance does prevail, our study would proceed in two stages: First, we would explore in Stage 1 whether there is indeed strategic ignorance in our context, using the *Baseline* and *Hidden Information* treatments (corresponding to the *Baseline* and *Hidden Information* treatments in Dana et al. (2007)). For sufficient strategic ignorance to be considered present, we required that some

¹See <https://www.socialscienceregistry.org/versions/141944/docs/version/document>

dictators must not reveal information in *Hidden Information*, and that the fraction of dictators choosing in a manner inconsistent with a preference for the fair outcome—meaning either avoiding information or acting selfishly conditional on knowingly being in the conflicting-interests payoff state—must be significantly larger than the share of dictators choosing selfishly in the *Baseline* treatment. If strategic ignorance turned out to be sufficiently prevalent to allow exploration of our primary research questions, then we would proceed to Stage 2. If not, the design would be modified and tested again, aiming to trigger strategic ignorance, before proceeding to Stage 2.

In the spring and summer of 2022, we collected Stage 1 data according to the previous preregistered plan. The experiments were programmed using zTree unleashed (Fischbacher, 2007; Duch et al., 2020) and were run using the recruitment/payment platform Prolific. Using a total sample of 232 subjects and hence 116 pairs (55 in *Baseline* and 61 in *Hidden Information*), we found that although some dictators did choose not to reveal information in the *Hidden Information* treatment, and a larger share of dictators facing the conflicting-interest payoffs chose the selfish option in the *Hidden Information* treatment than in the *Baseline* treatment, the latter difference was not statistically significant ($p = 0.15$). Consequently, according to our initial plan, we revised our Stage 1 design and procedures before proceeding to Stage 2.

We registered the updated version of our experiment on the Social Science Registry.² It featured four main changes aimed at favoring higher levels of information avoidance. First, we introduced attention checks and timers to reduce noise from inattentive subjects (possibly also bots). Second, we adapted the experimental design to further favor passive information avoidance by having the information and allocation choices made on the same screen, as in Dana et al. (2007). In the previous Stage 1 pilot attempt, the information choice was made before the allocation option choice, on a separate screen; thus, to stay ignorant the dictator had to actively choose to proceed without asking for information (a feature copied from Lind et al. (2019)). This anticipated the need in the *Informer* treatments to leave room for an informer to provide information to an information-avoidant dictator.

Third, we changed our choice of programming tool. Combining subject interaction in z-Tree Unleashed and the Prolific platform turned out to be more complicated than expected, introducing impractical experimental routines and long time lags between recruitment and experiment. This may have contributed to a major attrition problem – losing about 50 percent of subjects between the

²See <https://www.socialscienceregistry.org/trials/11289>

recruitment stage and the experiment itself. Since attrition might not be random with respect to our main variables of interest, the high attrition rate could be one reason for the apparently low or missing prevalence of strategic ignorance. Instead, we used LIONESS (Giamattei et al., 2020) instead of z-tree Unleashed (Fischbacher, 2007; Duch et al., 2020).

Fourth, instead of exploring the effects of dictators' option to punish informers (reducing the informer's payoff by 2 dollars) we chose to explore the effects of dictators' option to reward informers by a bonus of 0.50 dollars. This was due to our belief that the previous sanction opportunity may have been too harsh, possibly causing Dictators not to use it at all. We did not conduct any *Informer* treatments in Stage 1, so this design change was not motivated by empirical results in Stage 1. Only the new Stage 1 data, collected in the Spring of 2023, is included in our analysis.

Results of the replication treatments We ask two main questions of the replication-treatment data. First, is there a sufficient amount of information avoidance to allow us to investigate the impact of the informer in subsequent treatments? Second, does the behavior in this sample resemble that found in Dana et al. (2007) and in other replications, such as that of (Grossman, 2014)? We had two hundred and twenty-six subjects participate as dictators in our replication treatments, and their choices provide clear 'yes' answers to both questions. Based upon these findings we decided to proceed with Stage 2 of our preregistered experiment and conduct the treatments with the informer.

Of the 109 dictators in the *Hidden Information* treatment, 60 chose not to reveal the payoffs, amounting to an ignorance rate of 55%. This leaves plenty of room for an intervention such as introducing an informer to have a measurable impact on information avoidance. Furthermore, this level of information avoidance is sufficient to negatively affect social outcomes. In the *Hidden Information* treatment, 36 (62%) out of the 58 recipients who were paired with dictators facing the conflicting-interests payoffs ended up with the lower payoff. However, only 29 (25%) out of the 117 recipients in the *Baseline* treatment received the lower payoff. One additional fact worth mentioning is that the average recipient payoff in the *Hidden Information* treatment was 3.53, which is very similar to the average interim recipient payoffs (3.73 and 3.74) recorded in the *Informer* and *Informer - Bonus* treatments, respectively.

We turn to Table A.3 to compare the results of our replication treatment with both the original results from Dana et al. (2007) and those of the Grossman (2014) replication (*Baseline* and *Default NR* treatments. It fully reproduces Table 2 from Grossman (2014) in the second and third columns, for comparison with the results

Table A.3: The results of the replication treatments are comparable to those of Dana et al. (2007); Grossman (2014).

	Replication	Grossman (2014)	DWK
Baseline			
<i>n</i>	117	26	19
Inconsistent with pref for (5,5)	25	35	26
Hidden Info			
<i>n</i>	109	39	32
Inconsistent with pref for (5,5)	60	60	53
Ignorance	55	45	44
(5,5) given revealed and in CI game (<i>N</i>)	79 (24)	54 (13)	75 (8)
\$-maximizing given ignorant (<i>N</i>)	88 (60)	89 (19)	86 (14)

*Note: This table fully reproduces Table 2 from (Grossman, 2014) and adds the results from our replication experiment, which are displayed in the first column. Apart from numbers showing sample size (*n*), all numbers are percentages.*

from our replication experiment, which are displayed in the first column. We report the percentage of subjects exhibiting five behaviors: selfish choice in the *Baseline* treatment, then—in the *Hidden Information* treatment—choosing in a way inconsistent with a desire to ensure a higher payoff for the recipient, not revealing the payoffs, choosing unselfishly conditional on revealing and seeing the conflicting-interests payoffs, and choosing the higher payoff conditional on remaining ignorant. The values of these measures in our data closely track those of DWK and Grossman (2014), with a sample more than four times as large.³

A.3 Treatment Diagram

This appendix provides an overview of the key features of all experimental conditions. The table below summarizes the design elements that distinguish each treatment. The treatments are listed in rows, with the main treatments that are discussed in the text indicated in bold. The remaining rows correspond to the treatments reported in the appendices. To reflect the status of *Baseline* and *Hidden Information* as replication treatments, they appear above the main treatments.

³The one exception to this is the fact that the giving rate while knowingly facing the conflicting-interests payoffs is lower in Grossman (2014) than in the other two experiments.

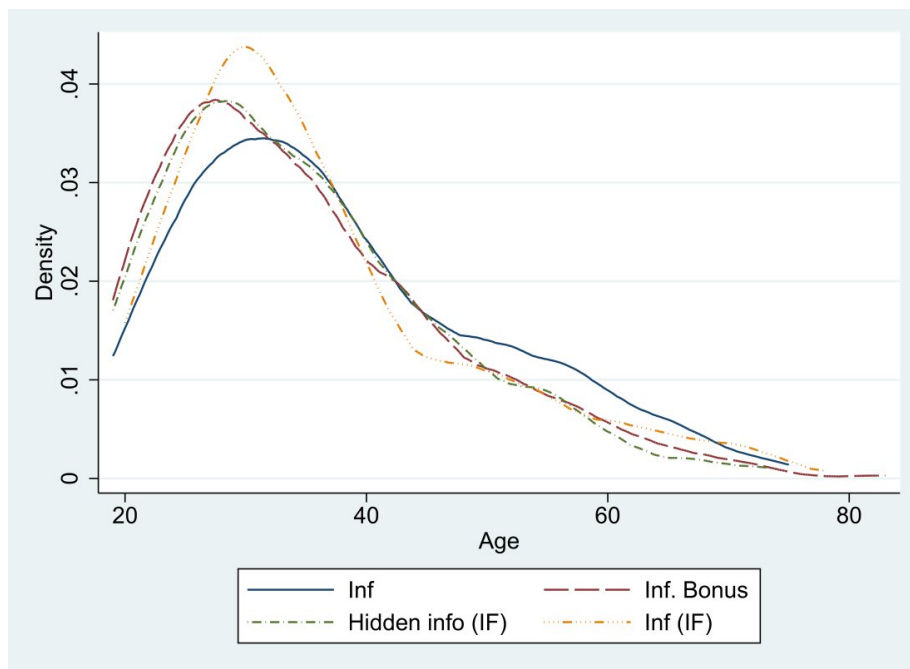
Key features include whether an informer (real or hypothetical) is present at all, at what point the dictator is made aware of the informer (if present), whether or not the informer's choice could actually be implemented, if the dictator is able to provide a bonus to the informer, and if the experimental condition has a split screen setup, separating the information decision from the allocation decision.

Treatment	Informer Involved?	Dictator Aware of Informer During Info Choice?	Dictator May Revise Choice	Informer Choice Implemented?	Dictator Bonus Option?	Split Screen?
Baseline	–	–	–	–	–	–
Hidden Information	–	–	–	–	–	–
Informer	✓	✗	✓	✓	–	–
Informer Info First	✓	✓	–	✓	–	✓
Hidden Info Info First	–	–	–	–	–	✓
Informer Bonus	✓	✗	✓	✓	✓	–
Hypothetical Informer	✓	✗	✓	✗	–	–
Hypothetical Informer Info First	✓	✓	–	✗	–	✓

Figure A.1: Overview of Experimental Conditions

Notes: Overview of the eight experimental conditions and key design features. Each row corresponds to one condition, and columns indicate whether specific informational or institutional elements are present at different stages of the decision process. A ✓ indicates that the feature is present, ✗ denotes feature is not present, and – indicates that the feature is not relevant for that condition. The main treatments of interest are shown in bold.

Figure A.2: The distribution of age by treatment



A.4 Robustness check: removing age outliers

We noted in Section 2.3 that our data featured some degree of age imbalance across treatments. In particular, the average age in the *Informer* and *Informer - Info First* treatments was slightly higher than in the other two treatments, to the point that we can reject the hypothesis of equality of age across treatments. Here we argue that this difference is driven by outliers on the high end of the age range and that this age imbalance does not drive our main results.

Figure A.2 shows the age distribution by treatment. All treatments peak close to age 30, although the *Informer* and *Informer - Info First* treatments peak slightly higher than the other two treatments. Most notable, though, is the fact that the *Informer* treatment has a thicker upper tail. We proceed by re-examining our main results while excluding the upper tails of all distributions and focusing only on subjects below age 55. We find that all of our main results hold up for this restricted sample and conclude that our results are not driven by age imbalance across treatments.

Table A.5 replicates Table 2.2 from the main text for subjects below the age of 55. In the two treatments featured in the table, this excludes 39 out of 367

Table A.4: Average recipient payment for dictators under the age of 55

	Pooled	Bonus		Unwillingly informed	
		N	Y	N	Y
Interim	3.76 (1.85)	3.80 (1.84)	3.71 (1.88)	4.34 (1.49)	3.03 (2.01)
Final	4.32 (1.51)	4.41 (1.42)	4.22 (1.59)	4.32 (1.51)	4.32 (1.51)
Difference	0.56 [0.000]	0.61 [0.000]	0.51 [0.000]	-0.02 [0.319]	1.29 [0.000]
n	328	170	158	182	146

Notes: This table replicates Table 2.2 for subjects below the age of 55.

subjects. The table shows average recipient payoffs both interim and final, by treatment and broken down by whether or not the corresponding dictator was unwillingly informed. Focusing on the last two columns, we see that when the dictator is unwillingly informed of the payoff state, the recipient's average payoff increases from 3.03 to 4.32 ECU. However, there is no corresponding increase for the recipients of dictators who were not unwillingly informed.

Table A.5 replicates Table 2.4 from the main text for subjects below the age of 55. It summarizes the information choices of the dictator and informer across all four treatments. In the info-first treatments, the dictator chooses ignorance 29 and 34 percent of the time, respectively. As with the full sample, there is no evidence that the presence of the informer reduces the dictator's propensity to avoid information. Furthermore, as in the full sample, there is a large gap between the ignorance rates in the info-first treatments and the other treatments. Finally, the overall rate at which the informer chose to impose info is 71%, which exactly matches the rate observed in the full sample. Furthermore, there is little difference across treatments in this behavior.

Table A.5: Main decisions for subjects under the age of 55

	Total	Treatments			
		Inf.	Inf. Bonus	Hid. Info (IF)	Inf. (IF)
D avoids info	0.53 (0.50)	0.59 (0.49)	0.71 (0.46)	0.29 (0.46)	0.34 (0.48)
I imposes info	0.71 (0.46)	0.70 (0.46)	0.68 (0.47)		0.76 (0.43)
N (dictators)	518	170	158	96	94
N (informers)	422	170	158	0	94

Notes: This table replicate Table 2.4 for subjects below the age of 55.

A.5 Informer Beliefs and Choice Reasoning

Once informers had made their main decision, we asked informers two questions about their beliefs. The first question was whether they believed that the dictator had revealed the payoff table and the second was whether they thought that the dictator would want the informer to provide that information. We did not include analysis of subjects' responses to these questions in the main text because informer behavior did not differ much between treatments and the analysis adds little insight. However, we report the analysis here for completeness and the raw data is available upon request.

Table A.6 summarizes these results by treatment. The vast majority of informers in the *Informer* and *Informer-Bonus* treatments believed that dictators had revealed the table ahead of time (between 84% to 87%), notably higher than the percentage of dictators who had sought information. This may suggest that people hold a different standard for culpability when not tempted by moral wiggle room, or that informers, who knew dictator's decision problem but had not themselves faced the single decision screen decision context, either did not expect exactly such a decision context or did not realize its effect.

Question	Informer	Informer - Bonus
Do you think they chose to find out the table?	168/193 (87%)	155/185 (84%)
If Person 1 did not reveal which table was actually being used, did you think that they wanted you to reveal the table for them?	78/194 (40%)	90/185 (48%)

Table A.6: Informer answering "yes" in the questionnaire, by treatment

We also had two undergraduate research assistants code the informer responses according to the following categories: a concern for the bonus being withheld (only in *Informer Bonus*), procedural concern for informed decision making, a desire to help the recipient, respecting the dictator's choice, believing that the dictator wanted information, and a catchall other category which includes no responses. For each response to each question, each coder assigned a value to an indicator variable for the response belonging to each category.

The results reported below are the average value across all respondents for each category and across both undergraduate coders. The category with by far the most representation among informers' responses is "More information is better," which is consistent with large numbers of informers choosing to impose the information. However, the second most prevalent category is "Respected dictator's choice," which may indicate a sensitivity to the fact that the information may have been

unwelcome. Only 8% of responding informers were concerned about the bonus being withheld. Furthermore, a minority of those informers believed that they would be more likely to receive the bonus if they chose to impose the information. This low share is natural, given that informers did not foresee the high share of ignorant dictators; apparently, they did not realize that information may be unwanted to the extent that dictators avoided it. Overall, informer responses are consistent with the finding of Mol et al. (2025) that informers' decisions are largely driven by their own information preferences, with some awareness of dictator preferences.

Reason	Informer
More information is better	0.43
Wanted to help recipient	0.09
Respected dictator's choice	0.20
Believed dictator wanted information	0.09
Other	0.19

Table A.7: Informer Treatment

Reason	Informer
Worried about bonus	0.08
More information is better	0.37
Wanted to help recipient	0.09
Respected dictator's choice	0.19
Believed dictator wanted information	0.12
Other	0.14

Table A.8: Informer Bonus Treatment

A.6 Hypothetical Informer treatments

Acknowledging the fact that the mere presence of a third party may affect the dictator’s information choice, we planned, preregistered, and ran two treatments in which the third party’s presence and role was announced, but their decision to inform was not implemented. We call these the *Hypothetical Informer* treatments and we ran both a regular and an *Info First* version. This appendix presents the design of these treatments and the results. Given that in our main treatments we found that the informer’s impact came only through his imposition of information and that the informer did not affect the dictator’s information choice, it is no surprise that we found no treatment effects in the *Hypothetical Informer* treatments. For the very same reason, we relegated discussion of these treatments to this appendix. In both treatments, the informer decisions were simulated using the distribution of informer decisions from the *Informer* treatment.

Hypothetical Informer

The *Hypothetical Informer* treatment is identical to *Hidden Information* in the sense that the dictator freely chooses her own information without any interference. Just like in the *Informer* treatment, however, after the dictator makes her initial choice she is told about the existence of another participant with a third role. This role and its description to the dictator exactly matches that of the *Informer* treatment in just about every respect. Specifically, the informer is told about the dictator’s decision problem and is asked if they prefer to provide the information about the recipient’s payoffs to the dictator, should the dictator choose not to seek it herself. The informer’s decision is shared with the dictator after the fact. However, the key difference is that the informer’s decision will never actually be implemented. Thus, the dictator has a chance to revise her decision after learning about the informer, their role, and whether or not they chose to inform. But dictators who chose not to inform themselves stay ignorant, regardless of what the informer chose. All of this is known to both the informer and the dictator. We hypothesized that merely introducing the presence of this third party and describing their *hypothetical* role as an informer could sway the dictator’s behavior, much in the way that subtle changes in choice architecture have been noted to sway dictator’s information choices, ostensibly by increasing the salience of their information choice and the sense of scrutiny surrounding it.

We used the *Hypothetical Informer* treatment to evaluate any changes to dictator’s outcome choices stemming from the mere introduction of a potential informer. Just as with the regular informer treatments, we compare within subjects the initial outcome choice to the revised outcome choice.

Hypothetical Informer - Info First

The *Hypothetical Informer - Info First* treatment parallels the *Hidden Information - Info First* in that it separates the dictator's information choice from the outcome choice and disposes with the revised outcome choice. However, as in the *Hypothetical Informer* treatment, the informer's decision is never implemented and this is common knowledge. Just as with the *Info First* treatments reported in the main text, we compare between subjects the ignorance rates in this treatment to that of the *Hidden Info - Info First* control treatment.

Screenshots of the hypothetical informer treatments are presented below.

Dictator Decision with Hypothetical Informer

Dictator decision screen - present on Hypothetical Informer

Which role will you have? You have been assigned the role of Person 1, the one who DOES choose between A and B.

Before you make your decision, here is some new information: a third participant, Person 3, has also been matched with your pair. Person 3 was given a description of the choice you faced and asked the following question: "In the case that Person 1 did not reveal which table is being used for their pair, would you prefer to overturn their decision, making sure that they actually have this information when they make their choice between A and B?" Person 3's payoff is independent of their choice.

The question is a purely hypothetical one, however; Person 3 is not actually able to provide any information to you. Thus, you will get to know which table is used only if you requested this information yourself. Person 3's payoff is independent of their choice.

Below you will choose whether or not to reveal the payoff table. After that, you will be informed about Person 3's decision. If you requested the information, you will then be told which table is actually being used. We will then ask you to choose between options A and B.

Please click "Continue" to move on and make your decision. If you wish to reveal which payoff table is being used, click "Reveal" first.

Option:	Person 1 Receives	Person 2 Receives
A	6	?
B	5	?

Reveal

Continue

Dictator Information with Hypothetical Informer (Info-First)

Dictator instructions - present on Hypothetical Informer - Info-First

Although you just made your decision, here is some new information: a third participant, Person 3, has also been matched with your pair. Person 3 was given a description of the choice you faced and asked the following question: "In the case that Person 1 did not reveal which table is being used for their pair, would you prefer to overturn their decision, making sure that they actually have this information when they make their choice between A and B?"

The question is a purely hypothetical one, however; Person 3 is not actually able to provide any information to you. Thus, you will get to know which table is used only if you requested this information yourself. Person 3's payoff is independent of their choice.

On the next screen, you will be informed about Person 3's decision. If you requested the information or if Person 3 decided to make sure that you have it, you will be told which table is actually being used.

We will then ask you to choose between options A and B even if you already did so on the previous decision screen. Your previous choice will then be filled out by default, and you can revise or keep that choice as you see fit. Payoffs to both you and Person 2 will be based on your final choices on the next screen.

Continue

Results of the Hypothetical Informers treatments We ran the *Hypothetical Informer* treatment with 98 dictator participants and the *Hypothetical Informer -*

Info First Treatment with 110 dictator participants.

First, we examine the extent to which a hypothetical informer affects information avoidance. Recall from Table 2.4 that the ignorance rates in the *Hidden Information - Info First* and *Informer - Info First* treatments were 0.30 and 0.32, respectively. The *Hypothetical Informer - Info First* treatment featured a similarly low ignorance rate with only 23 (21%) out of 110 dictators avoiding information. Although the non-*Info First* treatments are not designed for comparing the information choice, we also note that the 58% of the 98 dictators in the *Hypothetical Informer* treatment remained ignorant, which is very close to the 55% in the *Hidden Information* treatment and the 61% in the *Informer* treatment. We conclude from this that the hypothetical informer has little impact on information choice, which is unsurprising given the lack of an impact from the actual informer.

Next, we consider the impact of the hypothetical informer on recipient payoffs. The average interim recipient payoff in the *Hypothetical Informer* treatment was 3.90, slightly above the 3.74 we found in the treatments with an effectual informer. The introduction of the hypothetical informer had only a trivial impact. Only one of the 98 dictators changed their decision after learning about the hypothetical informer, which resulted in an average final payoff of 3.94 for the recipient.⁴ We conclude that a hypothetical informer does not have a significant impact on the dictator's choices nor on social outcomes. Dictators only respond when information is actually imposed.

Result: Hypothetical informers do not affect ignorance rates, nor do they lead to any impact on recipient payoffs.

⁴This one dictator chose to reveal and was in the conflicting-interests payoff state. They switched their choice from A to B, increasing the recipient's payoff.

A.7 Screenshots of Instructions

General Instructions 1

First page of instructions - dictators and informers - present on every treatment

Thank you for participating in this experiment. It has two parts. First you will get instructions for Part 1 and complete it. Then you will get instructions for Part 2, which is entirely separate from Part 1, and complete it. Your earnings from the two parts will be added together to determine your final payment. All payment amounts will be stated in terms of Experimental Currency Units (ECUs), which will be converted to US Dollars at the end of the experiment at a rate of 2 ECUs = \$1.

Please do not leave your computer and please keep your attention on this browser tab until the experiment has been completed. Be aware that there will be timers and attention checks on subsequent parts of this experiment.

Part 1 instructions: You will be randomly paired with another Prolific participant. Everyone will remain anonymous to each other. One of you, Person 1, will choose between two options, A and B, that will determine how much each of you gets paid. The other person, Person 2, cannot do anything to impact this decision.

We will use a payoff table to show you the payment amounts that go with each option. Below is an example. The payoff table shows that if Person 1 chooses B, for example, then they get 3 ECUs and Person 2 receives 4 ECUs. You can see this by looking at the bottom row. This is just an example, though. When it is time for the actual choice, we will show you a payoff table with different numbers in it.

Example Payoff Table		
Person 1 Chooses	Person 1 Receives	Person 2 Receives
A	1	2
B	3	4

To make sure you understand how people's payments depend on which option is chosen, we will ask you a question below.

Question: In this example, if Person 1 chooses A then what payment does Person 2 receive?

button appears in 17 seconds

Remaining time: 01:59

General Instructions 2

Second page of instructions - dictators and informers - present on every treatment

So what is the actual payoff table for your pair? What are the actual payment amounts that go with Person 1's two options?

There are two possible payoff tables shown below. Your pair will use one of them, but Person 1 won't be told at first which one is being used. All they will know is that either one could be the actual payoff table, with each being equally likely.

However, Person 1 can click a button before making their choice and reveal which payoff table is actually being used. This is optional and not required. They could also just go ahead and choose one of two options, A or B. Person 2 will not be told whether or not Person 1 revealed the payoff table.

Please confirm your understanding by answering the questions below. You will not be able to participate if you do not answer these questions correctly. Immediately after you pass these questions, you will be paired with another participant. This may involve a short wait. Once you are paired, the experiment will continue with an attention check, so please pay attention to your screen.

One Possible Payoff Table		
Person 1 Chooses	Person 1 Receives	Person 2 Receives
A	6	1
B	5	5

Another Possible Payoff Table		
Person 1 Chooses	Person 1 Receives	Person 2 Receives
A	6	5
B	5	1

Question 1: According to these two payoff tables, if you are Person 1 and you choose B, what will you receive?

Question 2: If you are Person 1 and all you know is that one of these two payoff tables is actually being used, what can you say about what will Person 2 receive if you choose B?

button appears in 12 seconds

Remaining time: 02:54

Attention Check

Attention check - dictators and informers - present on every treatment

This is an attention check. To participate, you must answer it correctly.

The question you are being asked now is very simple, when asked for a number, you must select '4828'.

Please choose a number.

4828
15
900
-1

Click here to continue

Attempts left to answer the control questions: 1

Remaining time: 00:43

Informer Instructions

Informer instructions - informers only - present on Informer and Informer Bonus treatments

In addition to Person 1 and Person 2, there is a third role in this study, Person 3. **Instead of being Person 1 or 2 and making the decision you just read about, you have been assigned the role of Person 3.**

Recall that Person 1 will not know which payoff table is being used, but they can reveal that information. While Person 1 is making their decision, **all you have to do is answer the following question:**

"In the case that Person 1 did not reveal which table is being used for their pair, would you prefer to overturn their decision, making sure that they actually have this information when they make their choice between A and B?"

This is the only thing that you have to do for Part 1 of this experiment at you will be paid **3 ECU regardless of how you respond**. You will receive this additional payment as a bonus payment on Prolific.

On the next page you will make your choice.

When you are done making your decision, Part 1 of the experiment will be over and we will give you instructions for Part 2.

Click here to continue

Dictator Decision (Same Screen)

Dictator instructions - dictators only - present on all Single Screen treatments

Which role will you have? You have been assigned the role of Person 1, the one who DOES choose between A and B.

To make your choice, click the corresponding button below. If you wish to reveal which payoff table is being used, click "Reveal".

When you are done making your decision, Part 1 of the experiment is over. We will tell you what your Part 1 payment will be and then give you instructions for Part 2.

Option:	Person 1 Receives	Person 2 Receives
A	6	?
B	5	?

Please select one of the options below, then click OK.

A
B
Reveal
Click here to continue

Dictator Decisions 1 (Info First)

Dictator instructions - dictators only - present on Informer and Informer Bonus treatments

Which role will you have? You have been assigned the role of Person 1, the one who DOES choose between A and B.

Before you make your decision, here is some new information: a third participant, Person 3, has also been matched with your pair. Person 3 was given a description of the choice you faced and asked the following question: "In the case that Person 1 did not reveal which table is being used for their pair, would you prefer to overturn their decision, making sure that they actually have this information when they make their choice between A and B?" Person 3's payoff is independent of their choice.

Below you will choose whether or not to reveal the payoff table. After that, you will be informed about Person 3's decision. If you requested the information or if Person 3 decided to make sure that you have it, you will then be told which table is actually being used. We will then ask you to choose between options A and B.

Please click "Continue" to move on and make your decision. If you wish to reveal which payoff table is being used, click "Reveal" first.

Option:	Person 1 Receives	Person 2 Receives
A	6	?
B	5	?

Reveal
Continue

Informer Decisions

Informer instructions - informers only - present on Informer and Informer Bonus treatments

In the case that Person 1 did not reveal which table is being used for their pair, would you prefer to overturn their decision, making sure that they actually have this information when they make their choice between A and B?

Yes
No
Continue

Dictator Decision 2 (Info First)

Informer instructions - dictators only - present on all Single Screen treatments

The actual payoff table being used is shown below.

Option:	Person 1 Receives	Person 2 Receives
A	6	5
B	5	1

Please select one of the options, then click to continue.

A
B
Click here to continue

Dictator Informed

Dictator informed - present on Informer and Informer Bonus treatments

Although you just made your decision, here is some new information: a third participant, Person 3, has also been matched with your pair. Person 3 was given a description of the choice you faced and asked the following question: "In the case that Person 1 did not reveal which table is being used for their pair, would you prefer to overturn their decision, making sure that they actually have this information when they make their choice between A and B?" Person 3's payoff is independent of their choice.

On the next screen, you will be informed about Person 3's decision. If you requested the information or if Person 3 decided to make sure that you have it, you will be told which table is actually being used.

We will then ask you to choose between options A and B even if you already did so on the previous decision screen. Your previous choice will then be filled out by default, and you can revise or keep that choice as you see fit. Payoffs to both you and Person 2 will be based on your final choices on the next screen.

Continue

Appendix **B**

B.1 Demographics

Tables B.1, C.1, B.2, and B.3 report detailed demographic characteristics of participants across all treatment arms in the Moral Wiggle Room Game (MWRG), the Dice Rolling Game (DRG), and the Binary Choice Dictator Game respectively. The variables include age, gender, student status, race and ethnicity, household income, education level, and political party affiliation.

Table B.1: Summary Demographics for the Moral Wiggle Room Game

	Treatment Arms									Total
	Unk. N/A n=132	Unk. 70% Ig. n=119	Unk. 30% Ig. n=119	Unk. 10% Ig. n=134	Kwn. N/A n=127	Kwn. 70% Ig. n=134	Kwn. 30% Ig. n=135	Kwn. 10% Ig. n=120	n=1,020	
age	38.30 (12.10)	38.27 (12.48)	38.80 (11.10)	39.02 (11.44)	39.11 (13.67)	37.12 (12.73)	37.33 (12.51)	38.13 (11.60)	38.25 (12.22)	
female	0.55 (0.50)	0.51 (0.50)	0.58 (0.50)	0.51 (0.50)	0.52 (0.50)	0.42 (0.50)	0.46 (0.50)	0.44 (0.50)	0.50 (0.50)	
student	0.17 (0.38)	0.17 (0.37)	0.06 (0.24)	0.10 (0.30)	0.12 (0.33)	0.16 (0.37)	0.12 (0.33)	0.12 (0.33)	0.13 (0.33)	
race										
White	94 (74.6%)	80 (73.4%)	77 (69.4%)	82 (66.1%)	89 (71.2%)	90 (70.9%)	94 (74.6%)	88 (76.5%)	694 (72.1%)	
Black	12 (9.5%)	12 (11.0%)	13 (11.7%)	18 (14.5%)	10 (8.0%)	21 (16.5%)	13 (10.3%)	11 (9.6%)	110 (11.4%)	
Hispanic	8 (6.3%)	5 (4.6%)	10 (9.0%)	9 (7.3%)	6 (4.8%)	8 (6.3%)	7 (5.6%)	2 (1.7%)	55 (5.7%)	
East Asian	7 (5.6%)	4 (3.7%)	3 (2.7%)	4 (3.2%)	8 (6.4%)	4 (3.1%)	3 (2.4%)	5 (4.3%)	38 (3.9%)	
Southeast Asian	3 (2.4%)	5 (4.6%)	4 (3.6%)	4 (3.2%)	8 (6.4%)	3 (2.4%)	3 (2.4%)	4 (3.5%)	34 (3.5%)	
MENA	1 (0.8%)	1 (0.9%)	0 (0.0%)	2 (1.6%)	1 (0.8%)	0 (0.0%)	1 (0.8%)	0 (0.0%)	6 (0.6%)	
Other Race	1 (0.8%)	2 (1.8%)	4 (3.6%)	5 (4.0%)	3 (2.4%)	1 (0.8%)	5 (4.0%)	5 (4.3%)	26 (2.7%)	
income										
\$0	6 (4.9%)	4 (3.7%)	1 (0.9%)	4 (3.6%)	6 (5.1%)	2 (1.7%)	5 (4.3%)	2 (1.8%)	30 (3.3%)	
less than \$20,000	23 (18.9%)	23 (21.5%)	15 (14.0%)	21 (18.8%)	27 (23.1%)	30 (26.1%)	26 (22.2%)	30 (27.5%)	195 (21.5%)	
less than \$40,000	14 (11.5%)	10 (9.3%)	23 (21.5%)	20 (17.9%)	19 (16.2%)	19 (16.5%)	27 (23.1%)	16 (14.7%)	148 (16.3%)	
less than \$60,000	23 (18.9%)	17 (15.9%)	18 (16.8%)	20 (17.9%)	25 (21.4%)	23 (20.0%)	22 (18.8%)	16 (14.7%)	164 (18.1%)	
less than \$80,000	23 (18.9%)	22 (20.6%)	18 (16.8%)	14 (12.5%)	19 (16.2%)	15 (13.0%)	16 (13.7%)	18 (16.5%)	145 (16.0%)	
less than \$100,000	11 (9.0%)	12 (11.2%)	11 (10.3%)	11 (9.8%)	7 (6.0%)	8 (7.0%)	9 (7.7%)	11 (10.1%)	80 (8.8%)	
greater than \$100,000	22 (18.0%)	19 (17.8%)	21 (19.6%)	22 (19.6%)	14 (12.0%)	18 (15.7%)	12 (10.3%)	16 (14.7%)	144 (15.9%)	
education										
Less than high school	0 (0.0%)	0 (0.0%)	1 (0.9%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.8%)	0 (0.0%)	2 (0.2%)	
High school graduate	41 (33.6%)	45 (42.5%)	40 (37.7%)	40 (35.7%)	50 (42.7%)	47 (40.9%)	55 (46.6%)	37 (33.9%)	355 (39.2%)	
Bachelors	61 (50.0%)	41 (38.7%)	43 (40.6%)	53 (47.3%)	45 (38.5%)	56 (48.7%)	48 (40.7%)	52 (47.7%)	399 (44.1%)	
Masters	16 (13.1%)	18 (17.0%)	22 (20.8%)	19 (17.0%)	22 (18.8%)	8 (7.0%)	11 (9.3%)	17 (15.6%)	133 (14.7%)	
Doctorate	4 (3.3%)	2 (1.9%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	4 (3.5%)	3 (2.5%)	3 (2.8%)	16 (1.8%)	
political_party										
Republican	34 (27.9%)	22 (20.8%)	24 (22.9%)	26 (23.2%)	24 (20.7%)	26 (22.6%)	27 (22.9%)	27 (24.8%)	210 (23.3%)	
Independent	41 (33.6%)	42 (39.6%)	41 (39.0%)	39 (34.8%)	40 (34.5%)	42 (36.5%)	42 (35.6%)	29 (26.6%)	316 (35.0%)	
Democrat	47 (38.5%)	42 (39.6%)	40 (38.1%)	47 (42.0%)	52 (44.8%)	47 (40.9%)	49 (41.5%)	53 (48.6%)	377 (41.7%)	
religion										
Christianity	58 (47.9%)	48 (45.3%)	54 (51.4%)	52 (46.4%)	50 (43.5%)	57 (49.6%)	51 (43.2%)	50 (45.9%)	420 (46.6%)	
Islam	1 (0.8%)	2 (1.9%)	1 (1.0%)	3 (2.7%)	1 (0.9%)	0 (0.0%)	2 (1.7%)	0 (0.0%)	10 (1.1%)	
Judaism	0 (0.0%)	4 (3.8%)	2 (1.9%)	5 (4.5%)	3 (2.6%)	2 (1.7%)	2 (1.7%)	1 (0.9%)	19 (2.1%)	
Hinduism	0 (0.0%)	0 (0.0%)	1 (1.0%)	0 (0.0%)	2 (1.7%)	0 (0.0%)	1 (0.8%)	1 (0.9%)	5 (0.6%)	
Buddhism	2 (1.7%)	1 (0.9%)	0 (0.0%)	1 (0.9%)	5 (4.3%)	0 (0.0%)	1 (0.8%)	1 (0.9%)	11 (1.2%)	
Agnosticism	20 (16.5%)	23 (21.7%)	20 (19.0%)	20 (17.9%)	22 (19.1%)	27 (23.5%)	29 (24.6%)	24 (22.0%)	185 (20.5%)	
Atheism	21 (17.4%)	20 (18.9%)	15 (14.3%)	14 (12.5%)	16 (13.9%)	21 (18.3%)	17 (14.4%)	20 (18.3%)	144 (16.0%)	
Other Religion	19 (15.7%)	8 (7.5%)	12 (11.4%)	17 (15.2%)	16 (13.9%)	8 (7.0%)	15 (12.7%)	12 (11.0%)	107 (11.9%)	

Notes: This table presents detailed demographic characteristics of participants across all treatment conditions. Variables include age, gender, student status, race/ethnicity, income, education, and political affiliation. Each cell shows the proportion or mean value for the group, with standard deviations in parentheses where applicable.

Table B.2: Summary Demographics for the Dice Rolling Game

	Treatment Arm				
	BA - Inaccurate n=232	BA - Accurate n=235	BF - Inaccurate n=213	BF - Accurate n=234	Total n=914
age	42.68 (12.56)	41.43 (13.67)	42.24 (13.47)	40.64 (12.26)	41.73 (12.99)
female	0.48 (0.50)	0.50 (0.50)	0.51 (0.50)	0.50 (0.50)	0.50 (0.50)
student	0.10 (0.29)	0.11 (0.31)	0.08 (0.27)	0.11 (0.31)	0.10 (0.30)
race					
White	172 (74.8%)	162 (70.1%)	156 (74.3%)	160 (69.3%)	650 (72.1%)
Black	23 (10.0%)	34 (14.7%)	24 (11.4%)	27 (11.7%)	108 (12.0%)
Hispanic	12 (5.2%)	9 (3.9%)	20 (9.5%)	21 (9.1%)	62 (6.9%)
East Asian	8 (3.5%)	5 (2.2%)	4 (1.9%)	5 (2.2%)	22 (2.4%)
Southeast Asian	6 (2.6%)	11 (4.8%)	2 (1.0%)	7 (3.0%)	26 (2.9%)
MENA	1 (0.4%)	1 (0.4%)	0 (0.0%)	0 (0.0%)	2 (0.2%)
Other Race	8 (3.5%)	9 (3.9%)	4 (1.9%)	11 (4.8%)	32 (3.5%)
income					
\$0	3 (1.3%)	6 (2.6%)	3 (1.4%)	3 (1.3%)	15 (1.7%)
less than \$20,000	37 (16.1%)	38 (16.5%)	32 (15.3%)	42 (18.2%)	149 (16.5%)
less than \$40,000	33 (14.3%)	41 (17.7%)	36 (17.2%)	44 (19.0%)	154 (17.1%)
less than \$60,000	46 (20.0%)	53 (22.9%)	48 (23.0%)	51 (22.1%)	198 (22.0%)
less than \$80,000	40 (17.4%)	28 (12.1%)	36 (17.2%)	35 (15.2%)	139 (15.4%)
less than \$100,000	21 (9.1%)	28 (12.1%)	20 (9.6%)	13 (5.6%)	82 (9.1%)
greater than \$100,000	50 (21.7%)	37 (16.0%)	34 (16.3%)	43 (18.6%)	164 (18.2%)
education					
Less than high school	1 (0.4%)	2 (0.9%)	2 (1.0%)	0 (0.0%)	5 (0.6%)
High school graduate	60 (26.1%)	72 (31.2%)	73 (34.9%)	85 (36.8%)	290 (32.2%)
Bachelors	120 (52.2%)	96 (41.6%)	92 (44.0%)	90 (39.0%)	398 (44.2%)
Masters	43 (18.7%)	53 (22.9%)	35 (16.7%)	46 (19.9%)	177 (19.6%)
Doctorate	6 (2.6%)	8 (3.5%)	7 (3.3%)	10 (4.3%)	31 (3.4%)
political_party					
Republican	72 (31.3%)	68 (29.4%)	55 (26.3%)	51 (22.1%)	246 (27.3%)
Independent	60 (26.1%)	62 (26.8%)	67 (32.1%)	76 (32.9%)	265 (29.4%)
Democrat	98 (42.6%)	101 (43.7%)	87 (41.6%)	104 (45.0%)	390 (43.3%)

Notes: This table presents detailed demographic characteristics of participants across all dice rolling game treatment conditions. Variables include age, gender, student status, race/ethnicity, income, education, and political affiliation. Each cell shows the proportion or mean value for the group, with standard deviations in parentheses where applicable.

Table B.3: Summary Demographics for the Binary Choice Dictator Game

	Treatment Arm				Total n=726
	BA - (65-0) n=179	BA - (50-50) n=183	BF - (65-0) n=172	BF - (50-50) n=192	
age	43.45 (13.45)	41.51 (12.46)	42.79 (12.79)	40.71 (12.79)	42.09 (12.89)
female	0.52 (0.50)	0.50 (0.50)	0.47 (0.50)	0.50 (0.50)	0.50 (0.50)
student	0.11 (0.32)	0.16 (0.36)	0.10 (0.30)	0.10 (0.30)	0.12 (0.32)
race					
White	128 (77.1%)	120 (69.0%)	114 (69.5%)	129 (73.7%)	491 (72.3%)
Black	14 (8.4%)	23 (13.2%)	17 (10.4%)	17 (9.7%)	71 (10.5%)
Hispanic	10 (6.0%)	10 (5.7%)	8 (4.9%)	14 (8.0%)	42 (6.2%)
East Asian	9 (5.4%)	10 (5.7%)	11 (6.7%)	5 (2.9%)	35 (5.2%)
Southeast Asian	1 (0.6%)	8 (4.6%)	9 (5.5%)	4 (2.3%)	22 (3.2%)
MENA	0 (0.0%)	0 (0.0%)	1 (0.6%)	2 (1.1%)	3 (0.4%)
Other Race	4 (2.4%)	3 (1.7%)	4 (2.4%)	4 (2.3%)	15 (2.2%)
income					
\$0	5 (3.0%)	3 (1.7%)	6 (3.7%)	3 (1.7%)	17 (2.5%)
less than \$20,000	23 (13.9%)	32 (18.4%)	26 (15.9%)	30 (17.1%)	111 (16.3%)
less than \$40,000	30 (18.1%)	30 (17.2%)	30 (18.3%)	26 (14.9%)	116 (17.1%)
less than \$60,000	32 (19.3%)	32 (18.4%)	26 (15.9%)	37 (21.1%)	127 (18.7%)
less than \$80,000	20 (12.0%)	30 (17.2%)	25 (15.2%)	23 (13.1%)	98 (14.4%)
less than \$100,000	23 (13.9%)	11 (6.3%)	16 (9.8%)	19 (10.9%)	69 (10.2%)
greater than \$100,000	33 (19.9%)	36 (20.7%)	35 (21.3%)	37 (21.1%)	141 (20.8%)
education					
Less than high school	1 (0.6%)	0 (0.0%)	1 (0.6%)	0 (0.0%)	2 (0.3%)
High school graduate	60 (36.1%)	61 (35.1%)	59 (36.0%)	63 (36.0%)	243 (35.8%)
Bachelors	66 (39.8%)	83 (47.7%)	69 (42.1%)	80 (45.7%)	298 (43.9%)
Masters	34 (20.5%)	28 (16.1%)	29 (17.7%)	25 (14.3%)	116 (17.1%)
Doctorate	5 (3.0%)	2 (1.1%)	6 (3.7%)	7 (4.0%)	20 (2.9%)
political_party					
Republican	55 (33.1%)	56 (32.2%)	43 (26.2%)	52 (29.7%)	206 (30.3%)
Independent	43 (25.9%)	63 (36.2%)	49 (29.9%)	54 (30.9%)	209 (30.8%)
Democrat	68 (41.0%)	55 (31.6%)	72 (43.9%)	69 (39.4%)	264 (38.9%)

Notes: This table presents detailed demographic characteristics of participants across all binary choice dictator game treatment conditions. Variables include age, gender, student status, race/ethnicity, income, education, and political affiliation. Each cell shows the proportion or mean value for the group, with standard deviations in parentheses where applicable.

B.2 Canonical Moral Wiggle Room Payoffs

In an initial set of pilot studies, I implemented the canonical payoffs of the moral wiggle room game as introduced by Dana et al. (2007). These payoffs led most subjects to reveal the payoff tables. Since the reveal rate may already approximate the true prosocial rate in a full-information Conflicting Interest Game, the degree of ignorance attributable to social image concerns—rather than pure selfishness—was likely limited. As such, the canonical design may represent an upper bound on the effects of belief elicitation, given that it already maximizes the salience of social image concerns even without an intervention.

The pilot design also included a *Placebo* condition, in which participants were asked to estimate an unrelated factual statistic (e.g., average daily coffee consumption) before making their choice. This served to test whether simply answering any pre-task question might shift behavior. In practice, reveal rates in the *Placebo* were indistinguishable from the *Belief After* condition, confirming that the placebo prompt itself had no behavioral effect. By contrast, *Belief First* elicitation significantly increased the likelihood of revealing information relative to both *Belief After* and *Placebo*. Across the canonical-payoff pilot, belief elicitation raised prosocial choice by about 8.5 percentage points—a smaller effect than in the main design but directionally consistent with the mechanism of norm activation.

Figure B.1 illustrates these results. Taken together, the pilot and main experiments suggest that the effectiveness of belief elicitation is moderated by the salience of social image concerns embedded in the payoff structure, but the mechanism of norm activation is robust across designs. The inclusion of a *Full Information* and *Self/Self* treatment in the modified design allowed for a more careful examination of the effects of belief elicitation beyond this canonical setting.

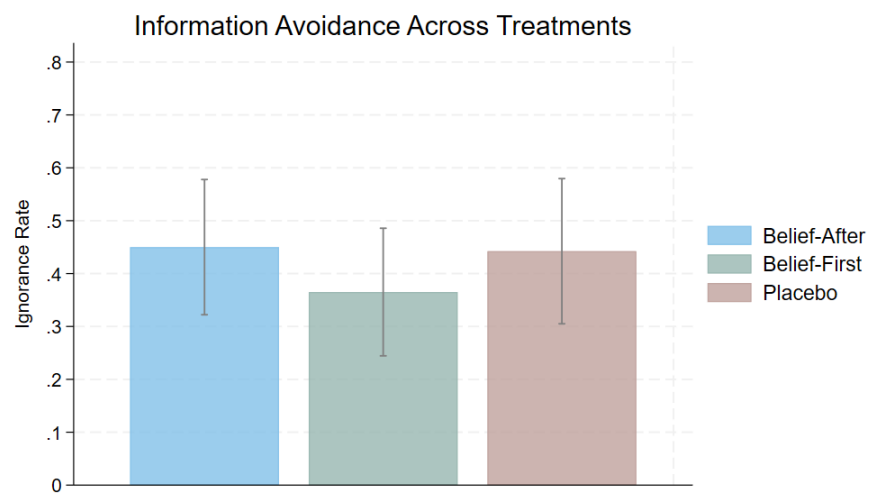


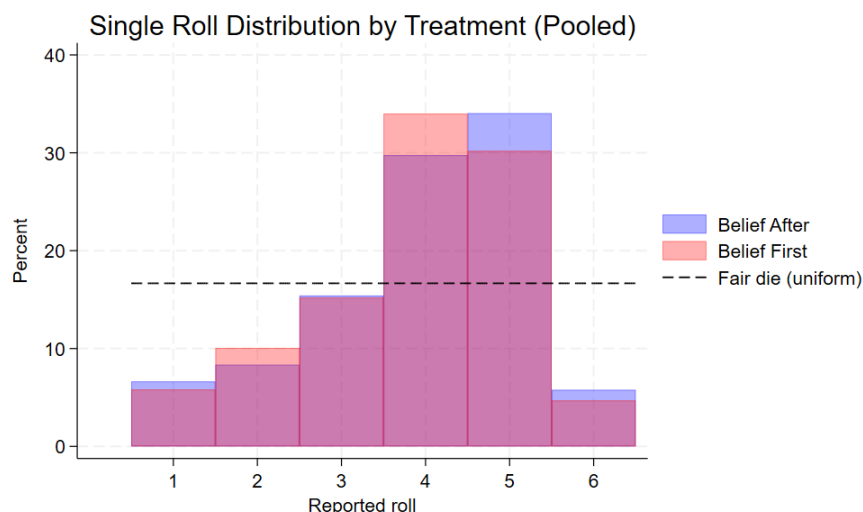
Figure B.1: Ignorance Rates By Treatments

Notes: Distribution of excess payoffs between treatment conditions. Note that a truthful baseline would result in a symmetrical distribution.

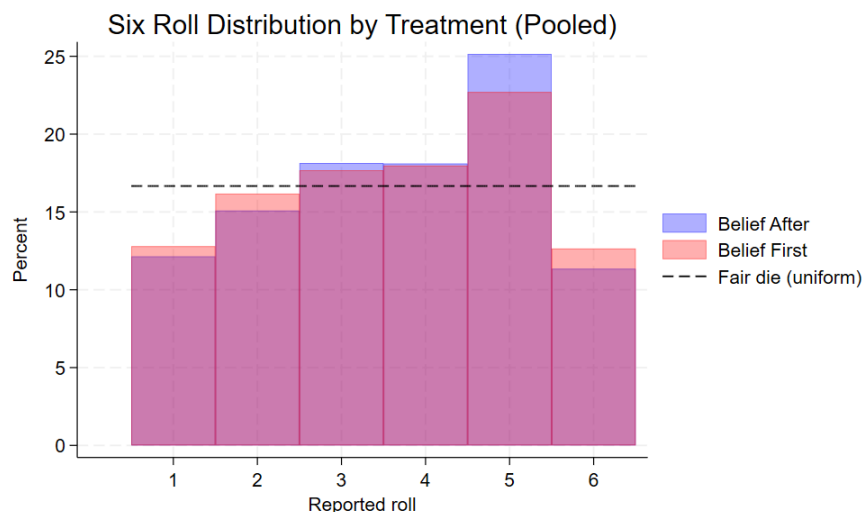
B.3 Reported Dice Distributions

This appendix reports the distributions of reported dice outcomes compared to the uniform benchmark. Figures B.2a and B.2b present the pooled distributions for the single-roll and six-roll tasks. Consistent with prior work (Fischbacher and Föllmi-Heusi, 2013), reports deviate from uniformity through underreporting of six and inflation of higher-paying outcomes, especially fours and fives. The *Belief First* treatment curbs this tendency more clearly in the six-roll condition. Lying once provides ample moral wiggle room—participants can justify a single dishonest report as a harmless deviation—but sustaining misreporting across six outcomes demands repeated justification, which appears more difficult when beliefs about others’ behavior are made salient in advance. In this sense, eliciting beliefs not only activates latent honesty norms but also constrains the use of moral wiggle room when dishonesty requires consistency over multiple opportunities. *Accurate* and *Inaccurate* framings show qualitatively similar distributions, with effects generally weaker under the *Accurate* framing.

Overall, the distributions confirm the same pattern captured by excess payoffs: dishonesty is present in both single and multiple roll settings, but belief elicitation—especially when placed before reporting—reduces the extent of cheating, with the strongest effects emerging when lying requires repeated misreporting rather than a one-off deviation.



(a) Reported distribution of single rolls, pooled across accuracy framings. Dashed line denotes the uniform benchmark (fair die).



(b) Reported distribution of six rolls, pooled across accuracy framings. Dashed line denotes the uniform benchmark (fair die).

Figure B.2: Reported Die Roll Distribution

Notes: Reported dice distributions under pooled accuracy framings. While dishonesty is evident in both settings through underreporting of six and inflation of higher outcomes, the Belief First treatment reduces cheating more strongly in the six-roll condition, suggesting that repeated misreporting makes moral wiggle room harder to sustain. Full distributions by accuracy framing are available in the replication package.

B.4 Framing of Belief Elicitation on Reporting Die Results

Within the DRG, I further varied the framing of the belief-elicitation task. Participants were asked either:

“In another study, participants were asked to participate as Person 1. What percentage of people in that study do you think **accurately** reported their number?”

or

“In another study, participants were asked to participate as Person 1. What percentage of people in that study do you think **inaccurately** reported their number?”

Table B.4 show that the *Inaccurate* frame produced a stronger reduction in dishonest reporting than the *Accurate* frame, suggesting that negatively valenced framing may heighten norm salience more effectively. This asymmetry likely reflects that “inaccuracy” brings to mind a *norm violation*, which is more salient than “accuracy” as a baseline expectation (Schwartz, 1977; Baumeister et al., 2001).

Table B.4: Regression of Excess Payoffs On Framing

	(1)	(2)	(3)
	(mean) excess	(mean) excess	(mean) excess
Inaccurate Frame	-1.468** (0.638)	-1.429** (0.653)	-1.233* (0.632)
Accurate Frame	-0.659 (0.649)	-0.529 (0.648)	-0.334 (0.623)
Demographics Control	No	Yes	Yes
SVO Control	No	No	Yes
Observations	914	900	900

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.5 Experiment Interface for the MWRG

This appendix contains screenshots of the experimental interface for the MWRG. When appropriate, alternative versions of the same screen corresponding to different treatment arms are provided. For the *Known* condition, when the subject is aware that they will later participate as the dictator, the interface refers to the dictator as “you.” In the *Unknown* condition, before making the allocation decision, the interface refers to the dictator as “Person 1.”

Page 1 Instructions

All Treatments

Thank you for participating in this study. Depending on the decisions you make, you can earn additional bonus payments. All payment amounts will be stated in terms of Experimental Currency Units (ECUs) converted at a rate of 2 ECUs = \$1.

We will provide you with 5 ECUs (\$2.50) for the completion of the study. Additional bonus ECUs earned will be provided within 72 hours after the study has been completed. You may withdraw from the study at any time. However, if you end the study before completion, you will forfeit all payment, including any bonuses earned.

Please do not leave your computer, and keep your attention on this browser tab until the study has been completed. There will be timers and attention checks. If you remain idle for an extended amount of time, your participation in the study will be terminated.

Continue

Page 2 Instructions

Unknown treatment

In studies like this one, we typically present people with a scenario in which one person chooses between two options that determine their own payoff and that of another person. We are going to describe one of these scenarios to you.

Afterwards, you will make some predictions about what you think other people have done in that scenario.

Here is an example to help you understand what is going on. After you answer some questions about it to confirm your understanding, we will describe the actual scenario to you.

There are two people: Person 1 and Person 2. Person 1 picked between options A and B to decide how much they and a second individual, Person 2, will receive.

Person 2 does not make any decisions.

In this example, the payoff table below shows that if Person 1 chooses B, then they get 3 ECUs and Person 2 receives 4 ECUs.

Example Payoff Table		
Person 1 Chooses	Person 1 Receives	Person 2 Receives
A	1	2
B	3	4

Question 1: In this example, if Person 1 chooses A, then how much does Person 1 receive?

Question 2: In this example, if Person 1 chooses A, then how much does Person 2 receive?

button appears in 13 seconds

Page 2 Instructions

Known treatment

In studies like this one, we typically present people with a scenario in which one person chooses between two options that determine their own payoff and that of another person. We are going to describe one of these scenarios to you.

Afterwards, you will make some predictions about what you think other people have done in that scenario. Then you will face that scenario yourself and make your own decisions.

Here is an example to help you understand what is going on. After you answer some questions about it to confirm your understanding, we will describe the actual scenario to you.

There are two people: Person 1 and Person 2. Person 1 picked between options A and B to decide how much they and a second individual, Person 2, will receive.

Person 2 does not make any decisions.

In this example, the payoff table below shows that if Person 1 chooses B, then they get 3 ECUs and Person 2 receives 4 ECUs.

Example Payoff Table		
You Choose	You Receive	Person 2 Receives
A	1	2
B	3	4

Question 1: In this example, if you as Person 1 choose A, then what payment do you receive?

Question 2: In this example, if you as Person 1 choose A, then what payment does Person 2 receive?

button appears in 13 seconds

Page 3 Instructions

Unknown treatment

Now we are going to describe to you the actual scenario that other participants faced. At the bottom of this page, you will answer some questions to confirm your understanding.

In the actual scenario, Person 1 will see the Hidden Payoff Table which hides Person 2's payoffs from Person 1. However, Person 1 may **optionally** reveal the Hidden Payoff Table to determine if it is one of **two possible** versions of the payoff tables (shown below): Payoff Table 1 or Payoff Table 2, with an **equal chance** of it being either.

The computer will choose the actual payoff table ahead of time. Revealing the payoff table only shows which version is being used. **Person 2 will not know if Person 1 chooses to reveal** the Hidden Payoff Table.

According to these two payoff tables, Person 1 will always receive 6 ECUs if they pick option A and 5 ECUs if they pick option B. Person 2 will receive either 1 or 5 ECUs depending on the actual payoff table.

Hidden Payoff Table		
Person 1 Choose	Person 1 Receive	Person 2 Receives
A	6	???
B	5	???

Payoff Table 1		
Person 1 Chooses	Person 1 Receives	Person 2 Receives
A	6	1
B	5	5

Payoff Table 2		
Person 1 Chooses	Person 1 Receives	Person 2 Receives
A	6	5
B	5	1

Several people in the other study acting as Person 1 made a decision whether or not to reveal the Hidden Payoff Table. We will ask you to **predict the percentage of those people acting as Person 1 who chose to reveal the table**.

Question 3 : If Person 1 did not reveal the Hidden Payoffs Table and chooses B, what will Person 2 receive?

Person 2 will always get 6
Person 2 will always get 5
Person 2 will always get 1
Person 2 will get either 5 or 1
Person 2 will get either 6 or 5

Question 4: If Person 1 chooses A, what will Person 1 receive?

Person 1 will always get 6
Person 1 will always get 5
Person 1 will always get 1
Person 1 will get either 5 or 1
Person 1 will get either 6 or 5

button appears in 14 seconds

Page 3 Instructions

Known treatment

Now we are going to describe to you the actual scenario that other participants faced and that you will later participate in. At the bottom of this page, you will answer some questions to confirm your understanding.

In the actual scenario, you will see the Hidden Payoff Table which hides Person 2's payoffs from you. However, you may **optionally** reveal the Hidden Payoff Table to determine if it is one of **two possible** versions of the payoff tables (shown below): Payoff Table 1 or Payoff Table 2, with an **equal chance** of it being either.

The computer will choose the actual payoff table ahead of time. Revealing the payoff table only shows which version is being used. **Person 2 will not know if you choose to reveal** the Hidden Payoff Table.

According to these two payoff tables, you will always receive 6 ECUs if you pick option A and 5 ECUs if you pick option B. Person 2 will receive either 1 or 5 ECUs depending on the actual payoff table.

Hidden Payoff Table		
You Choose	You Receive	Person 2 Receives
A	6	???
B	5	???

Payoff Table 1		
You Choose	You Receive	Person 2 Receives
A	6	1
B	5	5

Payoff Table 2		
You Choose	You Receive	Person 2 Receives
A	6	5
B	5	1

Several people in the other study acting as Person 1 made a decision whether or not to reveal the Hidden Payoff Table. We will ask you to **predict the percentage of those people acting as Person 1 who chose to reveal the table**. After your prediction, you will act as Person 1 in a similar experiment.

Question 3: If you did not reveal the Hidden Payoffs Table and choose A, what will Person 2 receive?

Person 2 will always get 6
Person 2 will always get 5
Person 2 will always get 1
Person 2 will get either 5 or 1
Person 2 will get either 6 or 5

Question 4: If you choose A, what will you receive?

You will always get 6
You will always get 5
You will always get 1
You will get either 5 or 1
You will get either 6 or 5

button appears in 13 seconds

Page 4 Instructions

Unknown treatment

We want to give you a chance to familiarize yourself with the choice interface as it was presented to the other participants to help you understand it before you make your prediction.

Person 2's payoffs are hidden by "???" but Person 1 can choose to reveal the payoffs before choosing between option A and option B. **Choosing to reveal is free.**

Please take a moment to imagine what Person 1 would do in this scenario. The interface below is just a mock-up. The buttons do not do anything if clicked.

Hidden Payoff Table		
Person 1 Chooses	Person 1 Receives	Person 2 Receives
A	6	???
B	5	???

A

B

Reveal Payoff Table

button appears in 8 seconds

Page 4 Instructions

Known treatment

We want to give you a chance to familiarize yourself with the choice interface as it was presented to the other participants to help you understand it before you make your prediction.

Person 2's payoffs are hidden by "???" but Person 1 can choose to reveal the payoffs before choosing between option A and option B. **Choosing to reveal is free.**

Please take a moment to imagine what you would do in this scenario. The interface below is just a mock-up. The buttons do not do anything if clicked.

Hidden Payoff Table		
You Choose	You Receive	Person 2 Receives
A	6	???
B	5	???

A

B

Reveal Payoff Table

button appears in 7 seconds

Page 5 Instructions

Unknown treatment

Hidden Payoff Table		
Person 1 Chooses	Person 1 Receives	Person 2 Receives
A	6	???
B	5	???

This is the Hidden Payoffs Table that participants acting in the role of Person 1 sees when making their decision. Person 2's payoffs are hidden by "???" but **Person 1 can choose to reveal the payoffs before choosing between option A and option B. Choosing to reveal is free.**

What percentage of people acting in the role of Person 1 do you believe revealed the payoff tables?

If your prediction is more accurate than the majority of the other participants, you will earn an additional ECU.

Use the slider and plus or minus buttons below to make your prediction.

No one revealed		Everyone revealed
-	Percentage of people who revealed: %	+

button appears in 13 seconds

Page 5 Instructions

Known treatment

Hidden Payoff Table		
Person 1 Chooses	Person 1 Receives	Person 2 Receives
A	6	???
B	5	???

This is the Hidden Payoffs Table that participants acting in the role of Person 1 sees when making their decision. Person 2's payoffs are hidden by "???" but **Person 1 can choose to reveal the payoffs before choosing between option A and option B. Choosing to reveal is free.**

Before you participate as Person 1, what percentage of people acting in the role of Person 1 do you believe revealed the payoff tables?

If your prediction is more accurate than the majority of the other participants, you will earn an additional ECU.

Use the slider and plus or minus buttons below to make your prediction.

No one revealed		Everyone revealed
-	Percentage of people who revealed: %	+

button appears in 18 seconds

Page 6 Instructions

Unknown treatment

Now we will ask you to act in the role of Person 1. Another Prolific participant will be matched and assigned the role of Person 2. You will earn additional ECUs based on your decision. Recall that these are the two possible payoff tables, with both being equally likely.

Payoff Table 1		
You Choose	You Receive	Person 2 Receives
A	6	1
B	5	5

Payoff Table 2		
You Choose	You Receive	Person 2 Receives
A	6	5
B	5	1

You will not know which payoff table is being used, but you can choose to reveal which payoff table is being used. We will ask you to make your decision on the next page.

Continue

Page 6 Instructions

Known treatment

Now it is time for you to make your decision as Person 1. Another Prolific participant will be matched and assigned the role of Person 2. You will earn additional ECUs based on your decision. Recall that these are the two possible payoff tables, with both being equally likely.

Payoff Table 1		
You Choose	You Receive	Person 2 Receives
A	6	1
B	5	5

Payoff Table 2		
You Choose	You Receive	Person 2 Receives
A	6	5
B	5	1

You will not know which payoff table is being used, but you can choose to reveal which payoff table is being used. We will ask you to make your decision on the next page.

Continue

Page 7 Instructions

30% Ignorance treatment - Overestimated Ignorance

A majority: 70% or 7 out of 10 people chose to reveal the payoff tables.

You predicted that 53% of people revealed the payoff tables.

You **underestimated** the amount of people who had revealed the payoff tables. People were less likely to keep Person 2's payoffs hidden than you had predicted.

You are **Person 1**. If you choose "Reveal Payoff Table", the next page will reveal the payoffs Person 2 will receive when you choose between A and B.

You Choose	You Receive	Person 2 Receives
A	6	???
B	5	???

A
B
Reveal Payoff Table

Continue

Page 7 Instructions

70% Ignorance treatment - Underestimated Ignorance

A minority: 30% or 3 out of 10 people chose to reveal the payoff tables.

You predicted that 51% of people revealed the payoff tables.

You **overestimated** the amount of people who had revealed the payoff tables. People were more likely to keep Person 2's payoffs hidden than you had predicted.

You are **Person 1**. If you choose "Reveal Payoff Table", the next page will reveal the payoffs Person 2 will receive when you choose between A and B.

You Choose	You Receive	Person 2 Receives
A	6	???
B	5	???

A
B
Reveal Payoff Table

Continue

Page 7 Instructions

No Norm treatment

You predicted that 24% of people revealed the payoff tables.

You are **Person 1**. If you choose "Reveal Payoff Table", the next page will reveal the payoffs Person 2 will receive when you choose between A and B.

You Choose	You Receive	Person 2 Receives
A	6	???
B	5	???

A
B
Reveal Payoff Table

Continue

Page 7 Instructions

Revealed payoffs

Actual payoff table shown:

You choose	You Receive	Person 2 Receives
A	6	5
B	5	1

Please select one of the options, then click to continue.

A
B

Continue

Page 8 Instructions

Appropriateness Rating

Please respond to the questions below. When you reach the bottom, click the button to move to the next page. When you reach the end of the allotted time, you will automatically be redirected back to Prolific.

How socially appropriate do you think **other people** believed it was to **reveal** the payoff table, in order to show Person 2's payoffs? If your response matches the most common response of the other participants in this study, you will receive 1 extra ECU.

Socially Inappropriate ☐ ☐ ☐ ☐ Socially Appropriate

How socially appropriate do you think **other people** believed it was to **not reveal** the payoff table, in order to show Person 2's payoffs? If your response matches the most common response of the other participants in this study, you will receive 1 extra ECU.

Socially Inappropriate ☐ ☐ ☐ ☐ Socially Appropriate

How socially appropriate do you **personally** believe it is to **reveal** the payoff table, in order to show Person 2's payoffs?

Socially Inappropriate ☐ ☐ ☐ ☐ Socially Appropriate

How socially appropriate do you **personally** believe it is to **not reveal** the payoff table, leaving Person 2's payoffs hidden?

Socially Inappropriate ☐ ☐ ☐ ☐ Socially Appropriate

button appears in 6 seconds

Remaining time: 11:00

Page 9 Instructions

Free Response Explanation

Please respond to the questions below. When you reach the bottom, click the button to move to the next page. When you reach the end of the allotted time, you will automatically be redirected back to Prolific.

1. Briefly explain your decision when acting as Person 1.

remaining characters 300

2. Briefly explain how you made your prediction regarding the percentage of people acting as Person 1 who revealed the payoff tables in the other study.

remaining characters 300

button appears in 17 seconds

Remaining time: 10:30

Page 10 Instructions

The Conformity Scale

Please respond to the questions below. When you reach the bottom, click the button to move to the next page. When you reach the end of the allotted time, you will automatically be redirected back to Prolific.

I often rely on, and act upon, the advice of others.

Very strong disagreement ☐ ☐ ☐ ☒ ☐ ☐ ☐ ☐ Very strong agreement

I would like to be the last one to change my opinion in a heated argument on a controversial topic.

Very strong disagreement ☐ ☐ ☐ ☒ ☐ ☐ ☐ ☐ Very strong agreement

Generally, I'd rather give in and go along for the sake of peace than struggle to have my way.

Very strong disagreement ☐ ☐ ☐ ☒ ☐ ☐ ☐ ☐ Very strong agreement

I tend to follow family tradition in making political decisions.

Very strong disagreement ☐ ☐ ☐ ☒ ☐ ☐ ☐ ☐ Very strong agreement

Basically, my friends are the ones who decide what we do together.

Very strong disagreement ☐ ☐ ☐ ☒ ☐ ☐ ☐ ☐ Very strong agreement

A charismatic and eloquent speaker can easily influence and change my ideas.

Very strong disagreement ☐ ☐ ☐ ☒ ☐ ☐ ☐ ☐ Very strong agreement

I am more independent than conforming in my ways.

Very strong disagreement ☐ ☐ ☐ ☒ ☐ ☐ ☐ ☐ Very strong agreement

If someone is very persuasive, I tend to change my opinion and go along with them.

Very strong disagreement ☐ ☐ ☐ ☒ ☐ ☐ ☐ ☐ Very strong agreement

I don't give in to others easily.

Very strong disagreement ☐ ☐ ☐ ☒ ☐ ☐ ☐ ☐ Very strong agreement

I tend to rely on others when I have to make an important decision quickly.

Very strong disagreement ☐ ☐ ☐ ☒ ☐ ☐ ☐ ☐ Very strong agreement

I prefer to make my own way in life rather than find a group I can follow.

Very strong disagreement ☐ ☐ ☐ ☒ ☐ ☐ ☐ ☐ Very strong agreement

Continue

Page 11 Instructions

Demographics Survey

Please respond to the questions below. When you reach the bottom, click the button to move to the next page. When you reach the end of the allotted time, you will automatically be redirected back to Prolific.

1. What is your assigned sex at birth?

Male ☐

Female ☐

Other ☐

2. How many years old are you?

3. What is your race, ethnic, or ancestral background?

White ☐

Black or African-American ☐

Hispanic or Latino/Latina/Latinx ☐

East Asian ☐

Southeast Asian ☐

Middle Eastern or North African ☐

Other ☐

4. What is your annual income before taxes?

\$0

\$20,000 or less

\$40,000 or less

\$60,000 or less

\$80,000 or less

\$100,000 or less

Greater than \$100,000

5. What is your employment status?

Full-Time

Part-Time

Self-Employed

Unemployed

Retired

Other

6. What is your highest educational attainment?

Less than high school

High school graduate

Bachelor's degree

Master's or professional degree

Doctoral degree

7. Are you currently a student?

Yes

No

8. In terms of your political leanings, how would you rate yourself?

Extremely liberal

Extremely conservative

9. With what political party, if any, do you sympathize?

Democrat

Independent/none

Republican

10. With which religious affiliation do you most identify with?

Christianity

Islam

Judaism

Hinduism

Buddhism

Agnosticism

Atheism

Other

11. How often do you go to church/temple or participate in any religious activities each month?

0

1

2

3

4

5

6+

Continue

Page 12 Instructions

End Screen

Thank you for participating in this study! As an **additional bonus**, we will assign you as Person 2 (the recipient) for the decision made by a different Person 1. This bonus will be provided to you within 72 hours of the completion of this study.

Please keep this browser page open. Once the timer has elapsed, you will automatically be redirected to Prolific for completion and payment. **If you close the browser before the timer elapses, you will not be paid.**

Additional survey questions are on the following pages.

Remaining time: 05:56

Page 13 Instructions

Optional Moral Universalism Trust Scale

Please keep this browser page open. Once the timer has elapsed, you will automatically be redirected to Prolific for completion and payment. **If you close the browser before the timer elapses, you will not be paid.**

There are various ways in which people can cheat on others or take advantage of them. How much do you trust that different people will not cheat on you or take advantage of you? **In each row below, how would you split 100 "trust points" between a randomly-selected person who lives in the United States, and the individual displayed on the right (who is part of a particular social group)?** The closer you drag the slider to one individual, the more you trust that individual, relative to the other individual. **Please assume all of the individuals below live in the United States.**

Randomly-Selected Person M who lives in the US	Trust this person  Trust this person	Someone who shares your religious beliefs (e.g. a fellow Christian)
Randomly-Selected Person N who lives in the US	Trust this person  Trust this person	Someone of your same race/ethnicity (e.g. a fellow Hispanic person)
Randomly-Selected Person O who lives in the US	Trust this person  Trust this person	A member of one of your past current organizations (local church, leisure club or association, etc.)
Randomly-Selected Person P who lives in the US	Trust this person  Trust this person	Someone who lives in your local neighborhood
Randomly-Selected Person Q who lives in the US	Trust this person  Trust this person	A former or current colleague at work or school

Remaining time: 05:35

Page 14 Instructions

Optional Questions

Please keep this browser page open. Once the timer has elapsed, you will automatically be redirected to Prolific for completion and payment. **If you close the browser before the timer elapses, you will not be paid.**

Do you donate to charity?

☐	Frequently
☐	Sometimes
☐	Rarely
☐	Never

Do you play video games?

☐	Frequently
☐	Sometimes
☐	Rarely
☐	Never
☐	Continue

Remaining time: 05:03

B.6 Experiment Interface for the DRG

This appendix contains screenshots of the experimental interface for the DRG.

Page 1 Instructions

All Treatments

Thank you for participating in this study. All payment amounts will be stated in Experimental Currency Units (ECUs) which are converted at a rate of 40 ECUs = \$1.

You will be paid a minimum of 40 ECUs for completing this study, but depending on the decisions you make, you can earn additional bonus payments to be provided within 72 hours after the study has been completed. You may withdraw from the study at any time. However, if you end the study before completion, you will forfeit all payment, including any bonuses earned.

Please do not leave your computer, and keep your attention on this browser tab until the study has been completed. There will be timers and attention checks. If you remain idle for an extended amount of time, your participation in the study will be terminated.

Continue

Page 2 Instructions

All Treatments

In this study, we will ask you to perform a series of tasks that will determine your bonus payment. We will randomly pick your decision from one of the following two tasks.

In the first task, we will ask you to roll a 6-sided die and then report the value to us. Your payoff will be determined by the Payoff Table below. Note that a reported 6 will yield 0 ECUs.

Payoff Table

Die Roll						
Payoff	10 ECUs	20 ECUs	30 ECUs	40 ECUs	50 ECUs	0 ECUs

For an example, if you report 3, then you will receive 30 ECUs.

To test your understanding, please answer the following comprehension questions.

Question 1: If you reported a 2, how many ECU's will you receive?

Question 2: If you reported a 5, how many ECU's will you receive?

Question 3: If you reported a 6, how many ECU's will you receive?

Continue

Attempts left to answer the control questions: 3

Page 3 Instructions

Accurate Frame only

But before you roll the die, we would like to ask you to make a prediction. In a similar study, participants were also asked to report their die roll.

In another study, participants were asked to participate as Person 1. What percentage of people acting in the role of Person 1 do you believe revealed the payoff tables?

What percentage of people in that study do you think inaccurately reported their number? Recall that participants were paid a bonus based on the number they reported.

Use the slider and plus or minus buttons below to make your prediction.



button appears in 11 seconds

Page 3 Instructions

Inaccurate Frame only

But before you roll the die, we would like to ask you to make a prediction. In a similar study, participants were also asked to report their die roll.

What percentage of people in that study do you think accurately reported their number? Recall that participants were paid a bonus based on the number they reported.

Use the slider and plus or minus buttons below to make your prediction.

No one accurate

 Everyone accurate

Percentage of people who **accurately reported** %

button appears in 17 seconds

Page 4 Instructions

Accurate Frame only

Payoff Table						
Die Roll						
Payoff	10 ECUs	20 ECUs	30 ECUs	40 ECUs	50 ECUs	0 ECUs

Please roll a **6-sided die with values between 1 to 6**. You may use a physical die, a virtual die on your browser, or a dice-rolling app on your phone.

Enter the number you rolled below. Payment is determined by the Payoff Table above. Recall that a **reported 6 will yield 0 ECUs**.

Continue

Page 4 Instructions

Inaccurate Frame only

Payoff Table

Die Roll						
Payoff	10 ECUs	20 ECUs	30 ECUs	40 ECUs	50 ECUs	0 ECUs

You predicted that 84% of people inaccurately reported their die rolls.

Please roll a 6-sided die with values between 1 to 6. You may use a physical die, a virtual die on your browser, or a dice-rolling app on your phone.

Enter the number you rolled below. Payment is determined by the Payoff Table above. Recall that a reported 6 will yield 0 ECUs.

Continue

Page 5 Instructions

Accurate Frame only

Payoff Table

Die Roll						
Payoff	10 ECUs	20 ECUs	30 ECUs	40 ECUs	50 ECUs	0 ECUs

This is the second task. Now we will ask you to roll the same die six times. Please remember the numbers you rolled. We will randomly select from your previous report and one of the six reports below and pay you according to the Payoff Table above. Recall that a reported 6 will yield 0 ECUs.

You predicted that 73% of people accurately reported their die rolls.

Enter the **first** number below.

Enter the **second** number below.

Enter the **third** number below.

Enter the **fourth** number below.

Enter the **fifth** number below.

Enter the **sixth** number below.

button appears in 10 seconds

Page 5 Instructions

Belief After - Inaccurate Frame only

Payoff Table

Die Roll						
Payoff	10 ECUs	20 ECUs	30 ECUs	40 ECUs	50 ECUs	0 ECUs

This is the second task. Now we will ask you to roll the same **die six times**. Please remember the numbers you rolled. We will **randomly** select from your previous report and one of the six reports below and pay you according to the Payoff Table above. Recall that a **reported 6 will yield 0 ECUs**.

You predicted that 84% of people inaccurately reported their die rolls.

Enter the **first** number below.

Enter the **second** number below.

Enter the **third** number below.

Enter the **fourth** number below.

Enter the **fifth** number below.

Enter the **sixth** number below.

button appears in 4 seconds

Page 3 Instructions

Belief After treatment only

Payoff Table

Die Roll						
Payoff	10 ECUs	20 ECUs	30 ECUs	40 ECUs	50 ECUs	0 ECUs

Please roll a **6-sided die with values between 1 to 6**. You may use a physical die, a virtual die on your browser, or a dice-rolling app on your phone.

Enter the number you rolled below. Payment is determined by the Payoff Table above. Recall that a **reported 6 will yield 0 ECUs**.

Continue

Page 4 Instructions

Belief After treatment only

Payoff Table						
Die Roll						
Payoff	10 ECUs	20 ECUs	30 ECUs	40 ECUs	50 ECUs	0 ECUs

This is the second task. Now we will ask you to roll the same **die six times**. Please remember the numbers you rolled. We will **randomly** select from your previous report and one of the six reports below and pay you according to the Payoff Table above. Recall that a **reported 6 will yield 0 ECUs**.

Enter the **first** number below.

Enter the **second** number below.

Enter the **third** number below.

Enter the **fourth** number below.

Enter the **fifth** number below.

Enter the **sixth** number below.

button appears in 11 seconds

Page 5 Instructions

Accurate Frame only

Now, we would like to ask you to make a prediction. In a similar study, participants were also asked to report their die roll.

What percentage of people in that study do you think accurately reported their number? Recall that participants were paid a bonus based on the number they reported.

Use the slider and plus or minus buttons below to make your prediction.



button appears in 13 seconds

Page 5 Instructions

Inaccurate Frame only

Now, we would like to ask you to make a prediction. In a similar study, participants were also asked to report their die roll.

What percentage of people in that study do you think inaccurately reported their number? Recall that participants were paid a bonus based on the number they reported.

Use the slider and plus or minus buttons below to make your prediction.

No one inaccurate

 Everyone inaccurate

-
 Percentage of people who **inaccurately reported** %
 +

button appears in 17 seconds

Page 6 Instructions

All treatment conditions

Please respond to the questions below. When you reach the bottom, click the button to move to the next page.

How socially appropriate do you think **other people** believed it was to **accurately report** the dice roll? If your response matches the most common response of the other participants in this study, you will receive an additional 5 ECUs.

Socially Inappropriate

 Socially Appropriate

How socially appropriate do you think **other people** believed it was to **inaccurately report** the dice roll? If your response matches the most common response of the other participants in this study, you will receive an additional 5 ECUs.

Socially Inappropriate

 Socially Appropriate

How socially appropriate do you **personally** believe it is to **accurately report** the die role?

Socially Inappropriate

 Socially Appropriate

How socially appropriate do you **personally** believe it is to **inaccurately report** the die roll?

Socially Inappropriate

 Socially Appropriate

button appears in 7 seconds

Remaining time: 09:59

Page 7 Instructions

All treatment conditions

Please respond to the questions below. When you reach the bottom, click the button to move to the next page.

1. Briefly explain how you made your prediction regarding the percentage of people who **inaccurately reported** their dice rolls.

remaining characters 300

2. Have you ever participated in a study that asked you to do the same task as this one?

Yes
No

button appears in 10 seconds

Remaining time: 09:52

Page 8 Instructions

All Treatments

Please respond to the questions below. When you reach the bottom, click the button to move to the next page.

1. What is your assigned sex at birth?

Male
Female
Other

2. How many years old are you?

3. What is your race, ethnic, or ancestral background?

White
Black or African-American
Hispanic or Latino/Latina/Latinx
East Asian
Southeast Asian
Middle Eastern or North African
Other

4. What is your annual income before taxes?

\$0
\$20,000 or less
\$40,000 or less
\$60,000 or less
\$80,000 or less
\$100,000 or less
Greater than \$100,000

5. What is your employment status?

Full-Time

Part-Time

Self-Employed

Unemployed

Retired

Other

6. What is your highest educational attainment?

Less than high school

High school graduate

Bachelor's degree

Master's or professional degree

Doctoral degree

7. Are you currently a student?

Yes

No

8. In terms of your political leanings, how would you rate yourself?

Extremely liberal

☐
☐
☐
☐
☐
☐

Extremely conservative

9. With what political party, if any, do you sympathize?

Democrat

Republican

Independent/none

Continue

Remaining time: 08:55

Page 9 Instructions
All treatment conditions

In this part of the study, you will be matched with another participant referred to as "Person A." Below are six decision tasks in which you will allocate points (representing money) between yourself and Person A. Note that **Person A does not make any decisions**. This task is unrelated to the previous parts of the study.

At the end of the study, one of your six decisions from below will be randomly selected to determine the payoffs for both you and Person A.

All values are still stated in Experimental Currency Units (ECUs). Please use the sliders below to make your preferred allocations.

Figure 1 displays six screenshots of the experiment interface, illustrating the choices and allocations for two participants, You and Person A, across different scenarios. Each screenshot shows a horizontal bar representing the total endowment (10 ECUs) and the allocation to each participant.

- Top Left:** You choose 8.5 ECUs. You receive 8.5 ECUs, and Person A receives 5 ECUs.
- Top Middle:** You choose 9.3 ECUs. You receive 9.25 ECUs, and Person A receives 3.25 ECUs.
- Top Right:** You choose 6.8 ECUs. You receive 6.75 ECUs, and Person A receives 9.25 ECUs.
- Bottom Left:** You choose 6.8 ECUs. You receive 6.75 ECUs, and Person A receives 6.75 ECUs.
- Bottom Middle:** You choose 7.5 ECUs. You receive 7.5 ECUs, and Person A receives 7.5 ECUs.
- Bottom Right:** You choose 9.3 ECUs. You receive 9.25 ECUs, and Person A receives 6.75 ECUs.

button appears in 24 seconds

Remaining time: 09:56

Page 10 Instructions

All treatment conditions

Thank you for participating in this study!

As an additional bonus, you will be assigned the role of Person A in the decisions made by another person. This bonus payment will be processed within 72 hours of completing the study.

1. What did you think this study was about?

remaining characters 300

2. Please leave any comments you may have below:

remaining characters 300

button appears in 1 seconds

Remaining time: 09:58

B.7 Experiment Interface for the BCDG

This appendix contains screenshots of the experimental interface for the BCDG.

Page 1 Instructions

All Treatments

Thank you for participating in this study. All payment amounts will be stated in Experimental Currency Units (ECUs) which are converted at a rate of 40 ECUs = \$1.

You will be paid a minimum of 50 ECUs for completing this study, but depending on the decisions you make, you can earn additional bonus payments to be provided within 72 hours after the study has been completed. You may withdraw from the study at any time. However, if you end the study before completion, you will forfeit all payment, including any bonuses earned.

Please do not leave your computer, and keep your attention on this browser tab until the study has been completed. There will be timers and attention checks. If you remain idle for an extended amount of time, your participation in the study will be terminated.

Continue

Page 2 Instructions

All Treatments

In studies like this one, we typically present people with a scenario in which one person chooses between two options that determine their own payoff and that of another person. We are going to describe one of these scenarios to you.

Here is an example to help you understand what is going on. After you answer some questions about it to confirm your understanding, we will describe the actual scenario to you.

There are two people: Person 1 and Person 2. Person 1 picked between options A and B to decide how much they and a second individual, Person 2, will receive.

Person 2 does not make any decisions.

In this example, the payoff table below shows that if Person 1 chooses B, then they get 30 ECUs and Person 2 receives 40 ECUs.

Example Payoff Table		
Person 1 Picks	Person 1 Gets	Person 2 Gets
A	10	20
B	30	40

Question 1: In this example, if Person 1 chooses A, then how many ECUs does Person 1 receive?

Question 2: In this example, if Person 1 chooses A, then how many ECUs does Person 2 receive?

Continue

Attempts left to answer the control questions: 3

Page 3 Instructions

All Treatments

At the bottom of this page, you will answer some questions to confirm your understanding.

In the actual scenario, you will act as **Person 1** and decide between how much you and Person 2 receives.

According to this payoff table, Person 1 will always receive 65 ECUs if they pick option A, and 50 ECUs if they pick option B. Person 2 will receive either 0 or 50 ECUs depending on what Person 1 picks.

Payoff Table		
Player 1 Picks	Player 1 Gets	Player 2 Gets
A	65	0
B	50	50

Question 3 : If you choose B, what will you receive?

Question 4: If you choose A, what will Player 2 receive?

Continue

Attempts left to answer the control questions: 3

Page 4 Instructions

Belief After only

Now we will ask you to act in the role of **Person 1**. You will earn additional ECUs based on your decision.

You Pick	You Get	Person 2 Gets
A	65	0
B	50	50

We will ask you to make your decision on the next page.

Continue

Page 5 Instructions

Belief After only

You are **Person 1**. Make your decision.

You Pick	You Get	Person 2 Gets
A	65	0
B	50	50

A
B

Continue

Page 6 Instructions

Belief After - Option A framing only

Person 1 Picks	Person 1 Gets	Person 2 Gets
A	65	0
B	50	50

Now we will ask you to predict how other people made the same type of decision you just made.

In another study, participants were asked to participate as Person 1. What percentage of people acting in the role of Person 1 do you believe picked option A?

Use the slider and plus or minus buttons below to make your prediction.

No one picked A		Everyone picked A
-	Percentage of people who picked A: %	+

button appears in 19 seconds

Page 6 Instructions

Belief After - Option B framing only

Person 1 Picks	Person 1 Gets	Person 2 Gets
A	65	0
B	50	50

Now we will ask you to predict how other people made the same type of decision you just made.

In another study, participants were asked to participate as Person 1. What percentage of people acting in the role of Person 1 do you believe picked option B?

Use the slider and plus or minus buttons below to make your prediction.

No one picked B		Everyone picked B
-	Percentage of people who picked B: %	+

button appears in 20 seconds

Page 4 Instructions

Belief First - Option A framing only

Person 1 Picks	Person 1 Gets	Person 2 Gets
A	65	0
B	50	50

You will be assigned the role of Person 1. But before you make your decision, we will ask you to predict the behavior of other people.

In another study, participants were asked to participate as Person 1. What percentage of people acting in the role of Person 1 do you believe picked option A?

Use the slider and plus or minus buttons below to make your prediction.

No one picked A

Everyone picked A

-

Percentage of people who picked A: %

+

button appears in 21 seconds

Page 4 Instructions

Belief First - Option B framing only

Person 1 Picks	Person 1 Gets	Person 2 Gets
A	65	0
B	50	50

You will be assigned the role of Person 1. But before you make your decision, we will ask you to predict the behavior of other people.

In another study, participants were asked to participate as Person 1. What percentage of people acting in the role of Person 1 do you believe picked option B?

Use the slider and plus or minus buttons below to make your prediction.

No one picked B

Everyone picked B

-

Percentage of people who picked B: %

+

button appears in 17 seconds

Page 5 Instructions

Belief First only

Now we will ask you to act in the role of **Person 1**. You will earn additional ECUs based on your decision.

You Pick	You Get	Person 2 Gets
A	65	0
B	50	50

We will ask you to make your decision on the next page.

Continue

Page 6 Instructions

Belief First - Option A framing only

You predicted that 72% of people picked option A.

You are **Person 1**. Make your decision.

You Pick	You Get	Person 2 Gets
A	65	0
B	50	50

A
B

Continue

Page 6 Instructions

Belief First - Option B framing only

You predicted that 78% of people picked option **B**.

You are **Person 1**. Make your decision.

You Pick	You Get	Person 2 Gets
A	65	0
B	50	50

A
B

Continue

Page 7 Instructions

All treatments

Please respond to the questions below. When you reach the bottom, click the button to move to the next page.

Payoff Table		
Person 1 Picks	Person 1 Gets	Person 2 Gets
A	65	0
B	50	50

You may reference the table above for the questions below.

How socially appropriate do you **personally** believe it is to **pick option A**?

Socially Inappropriate



Socially Appropriate

How socially appropriate do you **personally** believe it is to **pick option B**?

Socially Inappropriate



Socially Appropriate

button appears in 7 seconds

Remaining time: 09:59

Page 8 Instructions

All treatments

Please respond to the questions below. When you reach the bottom, click the button to move to the next page.

1. Briefly explain your decision when acting as Person 1.

remaining characters 300

2. Briefly explain how you made your prediction regarding the percentage of people acting as Person 1.

remaining characters 300

3. Have you ever participated in a study that asked you to do the same task as this one?

Yes
No

button appears in 6 seconds

Remaining time: 10:00

Page 9 Instructions

All Treatments

Please respond to the questions below. When you reach the bottom, click the button to move to the next page.

1. What is your assigned sex at birth?

Male
Female
Other

2. How many years old are you?

3. What is your race, ethnic, or ancestral background?

White
Black or African-American
Hispanic or Latino/Latina/Latinx
East Asian
Southeast Asian
Middle Eastern or North African
Other

4. What is your annual income before taxes?

\$0
\$20,000 or less
\$40,000 or less
\$60,000 or less
\$80,000 or less
\$100,000 or less
Greater than \$100,000

5. What is your employment status?

Full-Time
Part-Time
Self-Employed
Unemployed
Retired
Other

6. What is your highest educational attainment?

Less than high school
High school graduate
Bachelor's degree
Master's or professional degree
Doctoral degree

7. Are you currently a student?

Yes
No

8. In terms of your political leanings, how would you rate yourself?

Extremely liberal ☐ ☐ ☐ ☐ ☐ ☐ Extremely conservative

9. With what political party, if any, do you sympathize?

Democrat
Republican
Independent/none

Continue

Page 10 Instructions
All Treatments

In this part of the study, you will be matched with a third participant referred to as "Person A" who is unrelated to Person 2 from before. Below are six decision tasks in which you will allocate points (representing money) between yourself and Person A. Note that **Person A does not make any decisions**. This task is unrelated to the previous parts of the study.

At the end of the study, one of your six decisions from below will be randomly selected to determine the payoffs for both you and Person A.

All values are still stated in Experimental Currency Units (ECUs). Please use the sliders below to make your preferred allocations.

Figure 1 displays six screenshots of the experiment interface, arranged in a 2x3 grid. Each screenshot shows a choice screen where a participant (You) selects an allocation for Person A from a set of options. The interface includes a title "Choose your allocation.", a list of allocation options, and the resulting payoffs for both parties.

- Top Left:** Allocation options: 8.5, 8.5, 8.5, 8.5, 8.5, 8.5, 8.5, 8.5, 8.5. Person A receives 8.5 ECUs, and Person A receives 5 ECUs.
- Top Middle:** Allocation options: 8.5, 8.7, 8.9, 9.1, 9.3, 9.4, 9.6, 9.8, 10.0. Person A receives 9.25 ECUs, and Person A receives 3.25 ECUs.
- Top Right:** Allocation options: 5.0, 5.4, 5.9, 6.3, 6.8, 7.2, 7.6, 8.1, 8.5. Person A receives 6.75 ECUs, and Person A receives 9.25 ECUs.
- Bottom Left:** Allocation options: 5.0, 5.4, 5.9, 6.3, 6.8, 7.2, 7.6, 8.1, 8.5. Person A receives 6.75 ECUs, and Person A receives 5.75 ECUs.
- Bottom Middle:** Allocation options: 10.0, 9.4, 8.8, 8.1, 7.5, 6.9, 6.3, 5.6, 5.0. Person A receives 7.5 ECUs, and Person A receives 7.5 ECUs.
- Bottom Right:** Allocation options: 10.0, 9.8, 9.6, 9.4, 9.3, 9.1, 8.9, 8.7, 8.5. Person A receives 9.25 ECUs, and Person A receives 6.75 ECUs.

button appears in 28 seconds

Remaining time: 10:00

Page 11 Instructions

All Treatments

Thank you for participating in this study!

As an additional bonus, you will be assigned the roles of Person 2 (the recipient) and Person A in the decisions made by two different Person 1 participants. This bonus payment will be processed within 72 hours of completing the study.

1. What did you think this study was about?

remaining characters 300

2. Please leave any comments you may have below:

remaining characters 300

Continue

Remaining time: 09:56

Appendix C

C.1 Demographics

Table C.1 reports detailed demographic characteristics of participants across all treatment arms in the Moral Wiggle Room Game (MWRG). The variables include age, gender, student status, race and ethnicity, household income, education level, and political party affiliation.

Table C.1: Summary Demographics for the Moral Wiggle Room Game

	Treatment Arm				
	Belief After n=181	Belief First n=171	Full Info n=80	Self/Self n=62	Total n=494
age	40.10 (11.60)	39.97 (12.17)	41.00 (11.65)	40.85 (10.48)	40.29 (11.65)
female	0.52 (0.50)	0.47 (0.50)	0.50 (0.50)	0.49 (0.50)	0.50 (0.50)
student	0.09 (0.29)	0.12 (0.33)	0.10 (0.30)	0.09 (0.29)	0.10 (0.31)
race					
White	109 (68.6%)	113 (72.4%)	59 (81.9%)	41 (74.5%)	322 (72.9%)
Black	13 (8.2%)	16 (10.3%)	5 (6.9%)	8 (14.5%)	42 (9.5%)
Hispanic	8 (5.0%)	10 (6.4%)	2 (2.8%)	2 (3.6%)	22 (5.0%)
East Asian	12 (7.5%)	5 (3.2%)	2 (2.8%)	2 (3.6%)	21 (4.8%)
Southeast Asian	5 (3.1%)	8 (5.1%)	2 (2.8%)	1 (1.8%)	16 (3.6%)
Other Race	12 (7.5%)	4 (2.6%)	2 (2.8%)	1 (1.8%)	19 (4.3%)
income					
\$0	2 (1.3%)	2 (1.3%)	2 (2.8%)	0 (0.0%)	6 (1.4%)
less than \$20,000	19 (11.9%)	28 (17.9%)	9 (12.5%)	5 (9.1%)	61 (13.8%)
less than \$40,000	26 (16.4%)	24 (15.4%)	10 (13.9%)	15 (27.3%)	75 (17.0%)
less than \$60,000	32 (20.1%)	39 (25.0%)	13 (18.1%)	13 (23.6%)	97 (21.9%)
less than \$80,000	31 (19.5%)	28 (17.9%)	10 (13.9%)	5 (9.1%)	74 (16.7%)
less than \$100,000	16 (10.1%)	13 (8.3%)	10 (13.9%)	5 (9.1%)	44 (10.0%)
greater than \$100,000	33 (20.8%)	22 (14.1%)	18 (25.0%)	12 (21.8%)	85 (19.2%)
education					
Less than high school	1 (0.6%)	0 (0.0%)	1 (1.4%)	0 (0.0%)	2 (0.5%)
High school graduate	42 (26.4%)	60 (38.5%)	16 (22.2%)	19 (34.5%)	137 (31.0%)
Bachelors	81 (50.9%)	63 (40.4%)	37 (51.4%)	23 (41.8%)	204 (46.2%)
Masters	28 (17.6%)	30 (19.2%)	14 (19.4%)	12 (21.8%)	84 (19.0%)
Doctorate	7 (4.4%)	3 (1.9%)	4 (5.6%)	1 (1.8%)	15 (3.4%)
political_party					
Republican	33 (20.8%)	45 (28.8%)	22 (30.6%)	17 (30.9%)	117 (26.5%)
Independent	57 (35.8%)	47 (30.1%)	22 (30.6%)	12 (21.8%)	138 (31.2%)
Democrat	69 (43.4%)	64 (41.0%)	28 (38.9%)	26 (47.3%)	187 (42.3%)

Note: This table presents detailed demographic characteristics of participants across all treatment conditions. Variables include age, gender, student status, race/ethnicity, income, education, and political affiliation. Each cell shows the proportion or mean value for the group, with standard deviations in parentheses where applicable.

C.2 Analysis of Demographic Covariates and Result on *Ignorance Expectant* Dictators

This appendix examines whether specific demographic variables systematically influence the likelihood of a participant being classified as an *reveal expectant* or *ignorance expectant* dictator. To assess this, we regress the probability of being a *ignorance expectant* dictator on a range of demographic covariates, including age, gender, income, education, political affiliation, and religious background.

The results indicate that no single demographic characteristic is disproportionately associated with *ignorance expectant* dictators. In other words, *ignorance expectant* and *reveal expectant* dictators appear to be relatively balanced across demographic groups, suggesting that Result 2 is not driven by an overrepresentation of any particular subgroup. This strengthens the conclusion that differences in norm responsiveness are behavioral rather than a function of underlying demographic composition.

Table C.2 presents the full regression results.

Table C.2: Demographics on Being a *Ignorance Expectant* Versus *Reveal Expectant* Dictator

	(1)	(2)	(3)
	Ignorance Expectant	Ignorance Expectant	Ignorance Expectant
Ignorance Expectant			
Age	-0.00214 (0.00506)	-0.0128 (0.00790)	0.0104 (0.00722)
Female	0.0155 (0.104)	0.0646 (0.156)	-0.0442 (0.154)
Black	0.178 (0.165)	0.368 (0.247)	0.0806 (0.236)
Hispanic	-0.105 (0.219)	0.229 (0.335)	-0.379 (0.328)
East Asian	0.267 (0.240)	0.509 (0.346)	-0.0470 (0.376)
Southeast Asian	0.188 (0.286)	0.500 (0.390)	-0.0550 (0.485)
Other Race	-0.149 (0.316)	0.195 (0.416)	-0.438 (0.591)
less than \$20,000	0.336 (0.315)	0.324 (0.499)	0.265 (0.431)
less than \$40,000	0.272 (0.333)	0.566 (0.521)	-0.217 (0.465)
less than \$60,000	0.132 (0.335)	0.0528 (0.532)	0.111 (0.456)
less than \$80,000	-0.0887 (0.347)	0.183 (0.543)	-0.571 (0.486)
less than \$100,000	0.211 (0.363)	0.0329 (0.584)	0.130 (0.490)
greater than \$100,000	0.180 (0.347)	0.260 (0.560)	-0.0198 (0.473)
Part-Time	-0.0881 (0.157)	-0.257 (0.232)	0.0102 (0.239)
Self-Employed	-0.388** (0.186)	-0.452* (0.266)	-0.362 (0.286)
Unemployed	-0.0583 (0.167)	-0.323 (0.245)	0.0878 (0.246)
Retired	-0.00701 (0.317)	-0.196 (0.494)	0.0245 (0.465)
Other Employment	-0.470* (0.251)	-0.955** (0.416)	-0.265 (0.355)
High school graduate	-1.113 (0.940)	-0.435* (0.250)	-0.102 (0.693)
Bachelors	-0.980 (0.941)	-0.370 (0.232)	0.0315 (0.684)
Masters	-0.801 (0.948)	0 (.)	0.182 (0.697)
Student	-0.00992 (0.171)	-0.202 (0.270)	0.213 (0.244)
Independent	-0.0797 (0.140)	-0.391* (0.219)	0.194 (0.203)
Democrat	-0.119 (0.140)	-0.156 (0.206)	-0.0254 (0.208)
Judaism	-0.334 (0.418)	-0.218 (0.672)	-0.505 (0.595)
Hinduism	-0.0831 (0.743)	-0.101 (0.862)	0 (.)
Buddhism	-0.748 (0.575)	0 (.)	0.125 (0.809)
Agnosticism	0.0796 (0.141)	0.169 (0.207)	-0.0360 (0.210)
Atheism	0.205 (0.152)	0.403* (0.222)	-0.00192 (0.228)
Other Religion	0.238 (0.165)	0.663*** (0.249)	-0.175 (0.245)
Condition	All	Known	Unknown
Observations	884	429	429

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.3 Social Value Orientation (SVO) Moderation

This appendix reports results for the preregistered hypotheses on Social Value Orientation as a moderator in the MWRG. The rationale for this amendment (August 7, 2025) was that participants' willingness to avoid information might depend not only on treatment assignment but also on their underlying social preferences. In particular, those with a more prosocial orientation may naturally gravitate toward the fair option, while those with more individualistic or competitive orientations may be less sensitive to normative cues. If belief elicitation works by activating internalized fairness norms, then its effects should be strongest among participants who already value prosociality.

The preregistered hypotheses therefore specified that SVO would moderate treatment effects in the following ways. First, low-SVO participants (Individualist/Competitive) were expected to show higher levels of information avoidance in social settings (*Belief First* and *Belief After*) relative to the *Self/Self* condition, reflecting a tendency to sidestep norm-related image costs. Second, high-SVO participants (Prosocial) were predicted to exhibit little or no increase in ignorance across social settings, since their preferred allocation already aligns with fairness. Finally, it was hypothesized that belief elicitation—particularly in the *Belief After* condition—would reduce avoidance most strongly among Prosocial participants, as these individuals would be most likely to internalize the norm activation triggered by forecasting others' behavior. Table C.3 summarizes the results.

Table C.3: SVO moderation of ignorance in the Moral Wiggle Room Game

Contrast	Δ pp	95% CI	p
Prosocial			
Belief First – Belief After	–25.3	[–38.5, –12.0]	< .001
Self/Self – Belief After	–26.7	[–45.9, –7.4]	.007
Self/Self – Belief First	–2.0	[–28.7, 24.7]	.874
Individualistic			
Belief First – Belief After	+14.8	[–1.4, 29.8]	.075
Self/Self – Belief After	–35.0	[–50.9, –19.8]	< .001
Self/Self – Belief First	–20.1	[–32.8, –7.4]	.002

The preregistration further hypothesized that treatment effects would be

moderated by Social Value Orientation. Logistic regressions of ignorance on treatment by SVO type revealed significant heterogeneity. Among **Individualistic participants**, ignorance rates were substantially higher in the social settings than in *Self/Self* (*Belief After* – *Self/Self*: +35pp, $p < .001$; *Belief First* – *Self/Self*: +20pp, $p = .002$), consistent with the idea that they avoided information to sidestep norm-related image costs. By contrast, **Prosocial participants** showed little difference between *Belief First* and *Self/Self* (–2pp, n.s.), but were more ignorant in *Belief After* compared to both *Self/Self* (–27pp, $p = .007$) and *Belief First* (–25pp, $p < .001$). Thus, belief elicitation reduced avoidance most strongly among Prosocials, whereas Individualists remained relatively unresponsive. These results are directionally consistent with the preregistered moderation hypothesis, though the Prosocial increase in avoidance under *Belief After* suggests the effect of elicitation timing is more complex than initially predicted.

C.4 Robustness Check Regression on Ignorance Norms

As a robustness check for Result 1, I estimate a probit regression where the dependent variable is an indicator for ignorance (1 = did not reveal the payoff table). The key independent variable is a numeric treatment measure coded as 70, 30, or 10, reflecting the actual rates of ignorance observed in each norm condition (i.e., the percentage of prior participants who chose to remain ignorant).

The model includes demographic controls: age, gender, race, income, education, employment status, student status, political affiliation, and religion. I find a positive and statistically significant effect, indicating that higher ignorance environments increases the likelihood of information avoidance. The results confirm a monotonic decline in ignorance as norms increasingly favor information acquisition.

C.5 Sentiment and Linguistic Analysis of Subjects' Open Response

To assist in interpreting the sentiment and thematic patterns, I used OpenAI's ChatGPT (version GPT-4, accessed July 2025) to synthesize trends across groups. ChatGPT was used exclusively for descriptive synthesis and interpretation; no generated text was used verbatim in the manuscript. All underlying data were

Table C.4: Probit Regression on Ignorance Choice and Ignorance Environments

	(1)	(2)	(3)
	ignorance_choice	ignorance_choice	ignorance_choice
ignorance_choice			
Belief First (1=Before, 0=After)	-0.465*** (0.136)	-0.542*** (0.158)	-0.596*** (0.160)
Demographics Control	No	Yes	Yes
SVO Control	No	No	Yes
Observations	352	310	310

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

produced by participants as part of the experiment. No model fine-tuning or custom training was applied. Using Python’s TextBlob library, I extracted sentiment polarity (valence from -1 to $+1$) and subjectivity (scale from 0 to 1) for each explanation. Additionally, I used TF-IDF (term frequency–inverse document frequency) to identify prominent words in each group’s reasoning. The results are presented in Table C.5.

Reveal expectants expressed higher polarity (mean = 0.214) and greater subjectivity (mean = 0.503), suggesting more positive and opinion-driven explanations. Their most common terms—such as “right,” “fair,” “know,” “others,” and “information”—reflect moral language, indicating that many framed information acquisition as a normative good. One such participant wrote, “*Since it’s free and helps everyone make a better decision, I figured most would choose to reveal it to keep things fair and transparent,*” signaling a belief in shared prosocial values. This pattern aligns with the behavioral finding that *reveal expectants* were relatively insensitive to descriptive norm treatments, consistent with internally anchored moral motivation.

By contrast, *ignorance expectants* used more instrumental and self-interested language, marked by lower polarity (0.159) and subjectivity (0.456). Their dominant terms—“maximize,” “money,” “gain,” “avoid,” and “easy”—reflect a reasoning style focused on utility maximization, cost-benefit tradeoffs, and strategic simplicity. For example, one *ignorance expectant* explained, “*I thought that the majority of people would look to maximize their own earnings without caring what Person 2 received,*” emphasizing assumptions about others’ self-interest rather than moral considerations.

This linguistic framing is consistent with the behavioral pattern that *ignorance expectants* were significantly more responsive to norm information, especially in the 10% Ignorance condition. In one interpretation, *ignorance expectants* resemble

conditional cooperators—individuals who are not intrinsically moralizers but who adjust their behavior when they perceive prosocial action as common or expected. Their reasoning and behavior reflect strategic norm-following rather than principled commitment, suggesting that norm-based interventions may be particularly effective within this subgroup.

Together, these results reinforce the paper’s main conclusion. While some individuals exhibit stable moral motivation, others engage in strategic ignorance based on perceived normative expectations. Recognizing these differences is essential for designing effective nudges or norm-based interventions in ethical decision-making contexts.

Table C.5: Sentiment and Linguistic Theme Analysis by Expectant Type

Measure	Reveal Expectants	Ignorance Expectants	Interpretation
Sentiment Polarity (mean $\pm$ SD)	0.214 ± 0.232	0.159 ± 0.237	<i>Reveal expectants</i> used more positive tone
Subjectivity (mean $\pm$ SD)	0.503 ± 0.230	0.456 ± 0.253	<i>Reveal expectants</i> expressed more opinion-driven reasoning
Top Linguistic Themes (TF-IDF)	<i>right, know, information</i>	<i>fair, others, money, gain, easy, avoid</i>	Moral framing vs. instrumental reasoning
Moral Reasoning Style	Deontological Principled	Conditional Norm-dependent	Reflects differences in norm sensitivity
Motivational Profile	Internalized norms	Strategic norm-following	Matches observed behavioral patterns

C.6 Moral Universalism Trust Survey

At the end of the questionnaire, participants had the option to complete the Moral Universalism Trust Survey, which measures the extent of trust individuals place in people across different social distances (Enke et al. (2022)). A higher score reflects greater generalized trust toward strangers, whereas a lower score indicates a stronger preference for in-group trust.

The regression table below examines the relationship between a participant’s universalism score and their likelihood of revealing the payoff tables. Interestingly, universalism is negatively correlated with information avoidance behavior only in

the *Known* treatments, suggesting that a *self-referential* framing engages one's universalism traits.

Table C.6: Effect of Universalism Score on the Likelihood of Avoiding Information

	(1)	(2)	(3)	(4)
	avoided info	avoided info	avoided info	avoided info
universalism score	-0.00742* (0.00393)	-0.00977** (0.00480)	-0.00280 (0.00411)	-0.00327 (0.00499)
Condition	Known	Known	Unknown	Unknown
Controls	No	Yes	No	Yes
Observations	341	328	296	291

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The distribution of subjects' universalism trust scores does not significantly differ between the two groups. A Kolmogorov-Smirnov test indicates that the distributions are statistically indistinguishable ($D = 0.0662, p = 0.491$). Figure C.1 shows the distribution of universalism scores.

As additional collaborative evidence, Table C.2 shows that in the *Unknown* condition, political affiliation has no statistical effect, while in the *Known* condition, Republicans, relative to Democrats, were more likely to avoid information, a result closely aligned with Momsen and Ohndorf (2023a). Furthermore, these results are in line with Enke et al. (2024) who found that universalism is positively correlated with liberal beliefs. Given that universalism becomes more statistically significant once controls are added, I consider this suggestive evidence that *self-centric* framing activates innate traits.

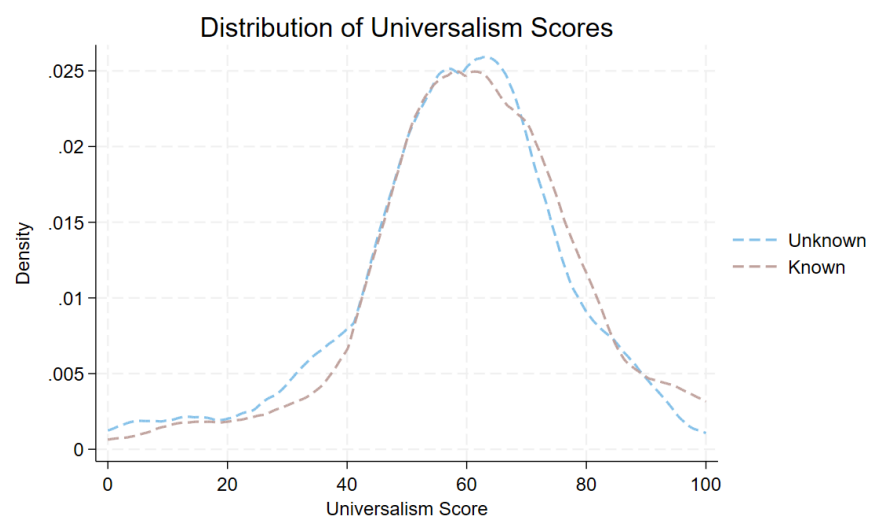


Figure C.1: Distribution of participants' Universalism Trust scores.

C.7 Conformity Score

In this appendix, I examine the Conformity Score, a psychological measure of norm-following behavior (Mehrabian and Steff (1995)). The first step is to assess whether this score has predictive power. To do this, I create a binary variable, `followed_norm`, which equals 1 if a subject follows the majority’s behavior—choosing to reveal when the majority revealed or choosing not to reveal when the majority did not reveal. The regression analysis suggests a positive correlation between the Conformity Score and norm-following behavior, but the results are not statistically significant ($p = 0.19$). Results reported in Table C.7.

Next, I investigate whether *ignorance expectant* dictators are more likely to conform to norms. To test this, I regress subjects’ Conformity Scores—used as a proxy for norm-following tendencies—against their likelihood of being classified as *ignorance expectant* dictators. The results indicate no statistically significant relationship, suggesting that neither *ignorance expectant* nor *reveal expectant* dictators are inherently more likely classified as norm followers.

Table C.7: Effect of Conformity Score on the Likelihood of Following Norms or Being a *Ignorance Expectant* Dictator

	(1)	(2)	(3)	(4)
	<code>followed_norm</code>	<code>followed_norm</code>	Ignorance Expectant	Ignorance Expectant
main				
conformity score	0.0548 (0.0421)	0.0611 (0.0464)	-0.00387 (0.0477)	-0.0104 (0.0518)
Controls	No	Yes	No	Yes
Observations	675	654	675	637

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.8 Social Appropriateness Ratings of Ignorance

This appendix examines self-reported social appropriateness ratings associated with the decision not to reveal information. Two distinct measures of social appropriateness were collected:

1. Social Appropriateness: Participants predicted the modal response of other

subjects regarding the appropriateness of not revealing. They were incentivized with an additional ECU for correctly matching the most common response.

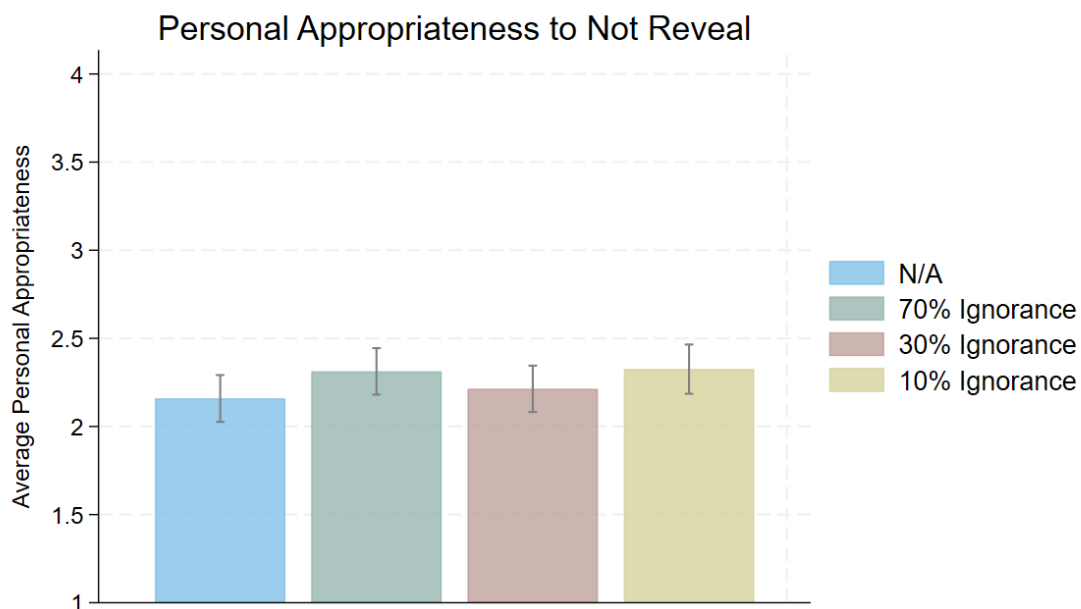
2. Personal Appropriateness: Participants provided their own personal evaluation of whether choosing not to reveal was appropriate, uninfluenced by incentive alignment.

The following pre-registered hypothesis was tested:

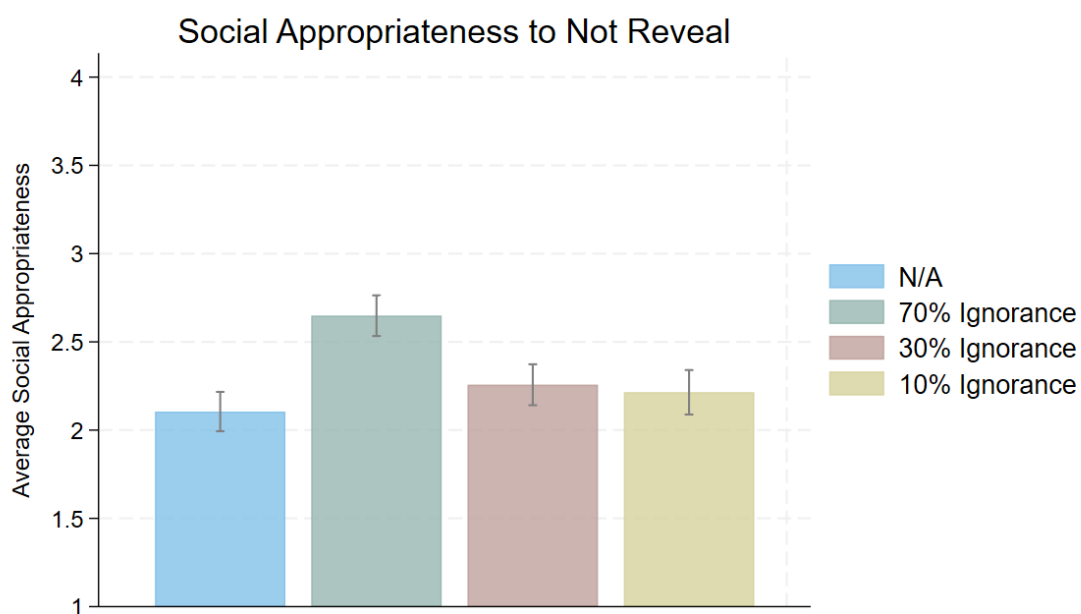
Hypothesis - Self-justifying Beliefs: Ignorant dictators will be more likely to rate a selfish ignorant action as appropriate than an informed dictator. Similarly, an informed dictator will be more likely to rate a prosocial reveal choice more favorably than an ignorant dictator.

The results indicate that personal appropriateness ratings remained statistically consistent across all treatment conditions, suggesting that individuals' moral evaluations did not depend on their information status. In contrast, perceptions of others' social appropriateness judgments varied systematically, indicating that dictators updated their beliefs about injunctive norms. Specifically, dictators in the 70% *Ignorance* condition believed that others viewed choosing not to reveal information more favorably compared to other conditions, suggesting that perceived social norms influence second-order beliefs.

Figure C.2 presents these results, highlighting the distinction between personal appropriateness judgments and incentivized social appropriateness perceptions.



(a) Subject's personal belief in the appropriateness of choosing not to reveal.



(b) Subject's incentivized prediction of how they believe others rated the appropriateness of choosing not to reveal.

Figure C.2: Perceived social appropriateness of information avoidance across norm treatments.

Notes: Bars indicate 95% confidence intervals.

C.9 Experiment Interface for the MWRG

Page 1 Instructions

All Treatments

Thank you for participating in this study. All payment amounts will be stated in Experimental Currency Units (ECUs) which are converted at a rate of 40 ECUs = \$1.

You will be paid a minimum of 50 ECUs for completing this study, but depending on the decisions you make, you can earn additional bonus payments to be provided within 72 hours after the study has been completed. You may withdraw from the study at any time. However, if you end the study before completion, you will forfeit all payment, including any bonuses earned.

Please do not leave your computer, and keep your attention on this browser tab until the study has been completed. There will be timers and attention checks. If you remain idle for an extended amount of time, your participation in the study will be terminated.

Continue

Page 2 Instructions

Belief First, Belief After, and Full Information Treatments only

In studies like this one, we typically present people with a scenario in which one person chooses between two options that determine their own payoff and that of another person. We are going to describe one of these scenarios to you.

Here is an example to help you understand what is going on. After you answer some questions about it to confirm your understanding, we will describe the actual scenario to you.

There are two people: Person 1 and Person 2. Person 1 picked between options A and B to decide how much they and a second individual, Person 2, will receive.

Person 2 does not make any decisions.

In this example, the payoff table below shows that if Person 1 chooses B, then they get 30 ECUs and Person 2 receives 40 ECUs.

Example Payoff Table		
Person 1 Picks	Person 1 Gets	Person 2 Gets
A	10	20
B	30	40

Question 1: In this example, if Person 1 chooses A, then how many ECUs does Person 1 receive?

Question 2: In this example, if Person 1 chooses A, then how many ECUs does Person 2 receive?

Continue

Attempts left to answer the control questions: 3

Page 2 Instructions

Self/Self Treatment only

In studies like this one, we typically present people with a scenario in which one person chooses between two options that determine how much they will receive. We are going to describe one of these scenarios to you.

Here is an example to help you understand what is going on. After you answer some questions about it to confirm your understanding, we will describe the actual scenario to you.

Person 1 picks between options A and B to decide how much they will receive. Their payoff is a combination of two values. The amount they get can be calculated by adding the two values together.

In this example, the payoff table below shows that if Person 1 chooses B, then they get $30 + 40$ ECUs, for a total of 70 ECUs.

Example Payoff Table		
Person 1 Picks	Person 1 Gets	Person 1 Also Gets
A	10	20
B	30	40

Question 1: In this example, if Person 1 chooses A, then how many ECUs does Person 1 receive?

Question 2: In this example, if Person 1 chooses B, then how many ECUs does Person 1 receive?

Continue

Attempts left to answer the control questions: 3

Page 3 Instructions

Belief First and *Belief After* Treatments only

At the bottom of this page, you will answer some questions to confirm your understanding.

In the actual scenario, Person 1 will see the Hidden Payoff Table which hides Person 2's payoffs from Person 1. However, Person 1 may **optionally** reveal the Hidden Payoff Table to determine if it is one of **two possible** versions of the payoff tables (shown below): Payoff Table 1 or Payoff Table 2, with an **equal chance** of it being either.

The computer will choose the actual payoff table ahead of time. Revealing the payoff table only shows which version is being used. **Person 2 will not know if Person 1 chooses to reveal** the Hidden Payoff Table.

According to these two payoff tables, Person 1 will always receive 65 ECUs if they pick option A and 50 ECUs if they pick option B. Person 2 will receive either 0 or 50 ECUs depending on the actual payoff table.

Hidden Payoff Table		
Player 1 Picks	Player 1 Gets	Player 2 Gets
A	65	???
B	50	???

Payoff Table 1		
Player 1 Picks	Player 1 Gets	Player 2 Gets
A	65	0
B	50	50

Payoff Table 2		
Player 1 Picks	Player 1 Gets	Player 2 Gets
A	65	50
B	50	0

Question 3 : If Person 1 did not reveal the Hidden Payoffs Table and chooses B, what will Person 2 receive?

Question 4: If Person 1 chooses A, what will Person 1 receive?

Continue

Attempts left to answer the control questions: 3

Page 3 Instructions

Full Info Treatment only

At the bottom of this page, you will answer some questions to confirm your understanding.

In the actual scenario, you will act as **Person 1** and decide between how much you and Person 2 receives.

According to this payoff table, Person 1 will always receive 65 ECUs if they pick option A, and 50 ECUs if they pick option B. Person 2 will receive either 0 or 50 ECUs depending on what Person 1 picks.

Payoff Table		
Player 1 Picks	Player 1 Gets	Player 2 Gets
A	65	0
B	50	50

Question 3 : If you choose B, what will you receive?

I will always get 65
I will always get 50
I will always get 0
I will get either 65 or 115
I will get either 50 or 115

Question 4: If you choose A, what will Player 2 receive?

Player 2 will always get 0
Player 2 will always get 50
Player 2 will always get 65
Player 2 will always get 10
It depends.

Continue

Attempts left to answer the control questions: 3

Page 3 Instructions

Self/Self Treatment only

At the bottom of this page, you will answer some questions to confirm your understanding.

In the actual scenario, you will act as **Person 1** with the real payoffs below. As from before, your payoff will be a sum of two values based on your decision.

Person 1 will see the Hidden Payoff Table which shows one part of their payment but hides the other part. However, Person 1 may **optionally** reveal the Hidden Payoff Table to determine if it is one of **two possible** versions of the payoff tables (shown below): Payoff Table 1 or Payoff Table 2, with an **equal chance** of it being either.

The computer will choose the actual payoff table ahead of time. Revealing the payoff table only shows which version is being used.

According to this payoff tables, Person 1 will not know exactly how much they receive unless they choose to reveal the payoff tables.

Hidden Payoff Table		
You Pick	You Get	You Also Get
A	65	???
B	50	???

Payoff Table 1		
You Pick	You Get	You Also Get
A	65	0
B	50	50

Payoff Table 2		
You Pick	You Get	You Also Get
A	65	50
B	50	0

Question 3 : If you did not reveal the Hidden Payoffs Table and choose A, what will you receive?

I will always get 65
I will always get 50
I will always get 0
I will get either 65 or 115
I will get either 50 or 100

Question 4: If you did not reveal the Hidden Payoffs Table and choose B, what will you receive?

I will always get 65
I will always get 50
I will always get 0
I will get either 65 or 115
I will get either 50 or 100

Continue

Attempts left to answer the control questions: 3

Page 4 Instructions

Belief After Treatment only

You are **Person 1**. If you choose "Reveal Payoff Table", the next page will reveal the payoffs Person 2 will receive when you choose between A and B.

You Pick	You Get	Person 2 Gets
A	65	???
B	50	???

A
B
Reveal Payoff Table

Continue

Page 4 Instructions

Belief After Treatment only

Hidden Payoff Table		
Person 1 Picks	Person 1 Gets	Person 2 Gets
A	65	???
B	50	???

Now we will ask you to predict how other people made the same type of decision you just made.

This is the Hidden Payoffs Table that participants acting in the role of Person 1 sees when making their decision. Person 2's payoffs are hidden by "???" but **Person 1 can choose to reveal the payoffs before choosing between option A and option B. Choosing to reveal is free.**

In another study, participants were asked to participate as Person 1. What percentage of people acting in the role of Person 1 do you believe revealed the payoff tables?

Use the slider and plus or minus buttons below to make your prediction.

No one revealed		Everyone revealed
-	Percentage of people who revealed: %	+

button appears in 20 seconds

Page 4 Instructions

Belief After Treatment only

Actual payoff table shown:

You Pick	You Get	Person 2 Gets
A	65	0
B	50	50

Please select one of the options, then click to continue.

A

B

Continue

Page 4 Instructions

Belief First Treatment only

Hidden Payoff Table		
Person 1 Picks	Person 1 Gets	Person 2 Gets
A	65	???
B	50	???

You will be assigned the role of Person 1. But before you make your decision, we will ask you to predict the behavior of other people.

This is the Hidden Payoffs Table that participants acting in the role of Person 1 sees when making their decision. Person 2's payoffs are hidden by "???" but **Person 1 can choose to reveal the payoffs before choosing between option A and option B. Choosing to reveal is free.**

In another study, participants were asked to participate as Person 1. What percentage of people acting in the role of Person 1 do you believe revealed the payoff tables?

Use the slider and plus or minus buttons below to make your prediction.

No one revealed

-

Percentage of people who revealed: %

+

Everyone revealed

button appears in 23 seconds

Page 4 Instructions

Belief First Treatment only

Now we will ask you to act in the role of **Person 1**. You will earn additional ECUs based on your decision. Recall that these are the two possible payoff tables, with both being equally likely.

Payoff Table 1		
You Pick	You Get	Person 2 Gets
A	65	0
B	50	50

Payoff Table 2		
You Pick	You Get	Person 2 Gets
A	65	50
B	50	0

You will not know which payoff table is being used, but you can choose to reveal which payoff table is being used. We will ask you to make your decision on the next page.

Continue

Page 4 Instructions

Belief First Treatment only

Actual payoff table shown:

You Pick	You Get	Person 2 Gets
A	65	0
B	50	50

Please select one of the options, then click to continue.

A
B

Continue

Page 4 Instructions

Full Information Treatment only

Actual payoff table shown:

You Pick	You Get	Person 2 Gets
A	65	50
B	50	0

Please select one of the options, then click to continue.

A
B

Continue

Page 4 Instructions

Self/Self Treatment only

If you choose "Reveal Payoff Table", the next page will reveal the part of your payoffs.

You Pick	You Get	You Also Get
A	65	???
B	50	???

A
B

Reveal Payoff Table

Continue

Page 4 Instructions

Self/Self Treatment only

Actual payoff table shown:

You Pick	You Get	You Also Get
A	65	0
B	50	50

Please select one of the options, then click to continue.

A
B

Continue

Page 5 Instructions

All Treatments

Please respond to the questions below. When you reach the bottom, click the button to move to the next page.

How socially appropriate do you **personally** believe it is to **reveal** the payoff table, in order to show Person 2's payoffs?

Socially Inappropriate ☐ ☐ ☐ ☐ Socially Appropriate

How socially appropriate do you **personally** believe it is to **not reveal** the payoff table, leaving Person 2's payoffs hidden?

Socially Inappropriate ☐ ☐ ☐ ☐ Socially Appropriate

Continue

Remaining time: 09:50

Page 6 Instructions

Belief First and *Belief After* treatments only

Please respond to the questions below. When you reach the bottom, click the button to move to the next page.

1. Briefly explain your decision when acting as Person 1.

remaining characters 300

2. Briefly explain how you made your prediction regarding the percentage of people acting as Person 1 who revealed the payoff tables in the other study.

remaining characters 300

3. Have you ever participated in a study that asked you to do the same task as this one?

Yes
No

button appears in 5 seconds

Remaining time: 10:00

Page 6 Instructions

Full Information and *Self/Self* treatments only

Please respond to the questions below. When you reach the bottom, click the button to move to the next page.

1. Briefly explain your decision when acting as Person 1.

remaining characters 300

3. Have you ever participated in a study that asked you to do the same task as this one?

Yes
No

button appears in 1 seconds

Remaining time: 09:53

Page 7 Instructions

All Treatments

Please respond to the questions below. When you reach the bottom, click the button to move to the next page.

1. What is your assigned sex at birth?

Male
Female
Other

2. How many years old are you?

--

3. What is your race, ethnic, or ancestral background?

White
Black or African-American
Hispanic or Latino/Latina/Latinx
East Asian
Southeast Asian
Middle Eastern or North African
Other

4. What is your annual income before taxes?

\$0
\$20,000 or less
\$40,000 or less
\$60,000 or less
\$80,000 or less
\$100,000 or less
Greater than \$100,000

5. What is your employment status?

Full-Time
Part-Time
Self-Employed
Unemployed
Retired
Other

6. What is your highest educational attainment?

Less than high school
High school graduate
Bachelor's degree
Master's or professional degree
Doctoral degree

7. Are you currently a student?

Yes
No

8. In terms of your political leanings, how would you rate yourself?

Extremely liberal ● ● ● ● ● Extremely conservative

9. With what political party, if any, do you sympathize?

Democrat
Republican
Independent/none

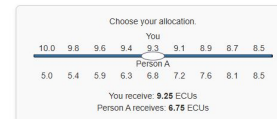
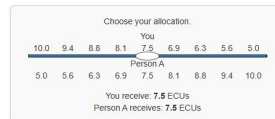
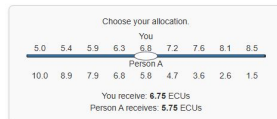
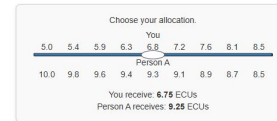
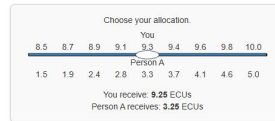
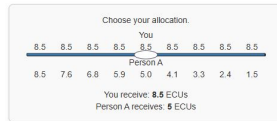
Continue

Remaining time: 08:55

In this part of the study, you will be matched with a third participant referred to as "Person A" who is unrelated to Person 2 from before. Below are six decision tasks in which you will allocate points (representing money) between yourself and Person A. Note that **Person A does not make any decisions**. This task is unrelated to the previous parts of the study.

At the end of the study, **one of your six decisions from below will be randomly selected** to determine the payoffs for both you and Person A.

All values are still stated in Experimental Currency Units (ECUs). Please use the sliders below to make your preferred allocations.



button appears in 28 seconds

Remaining time: 10:00

Page 9 Instructions

All Treatments

Thank you for participating in this study!

As an additional bonus, you will be assigned the roles of Person 2 (the recipient) and Person A in the decisions made by two different Person 1 participants. This bonus payment will be processed within 72 hours of completing the study.

1. What did you think this study was about?

remaining characters 300

2. Please leave any comments you may have below:

remaining characters 300

Continue

Remaining time: 09:56